Verh. 5. int. Kongr. Phon. Wiss., Münster 1964, pp. 1–24
(S. Karger, Basel/New York 1965).

Die Bedeutung der Sprachstruktur für die Analyse des Sprechens
Problemgeschichtliche Erörterung

Von EBERHARD ZWIRNER, Münster

1. Geschichtlicher Teil

Am Ende des 18. Jahrhunderts hat der deutsch-dänische Archäologe *Georg Zoega*[1] in seinem Werk «De origine et usu obeliscorum» (39) eine classis notarum phoneticarum vier «rätselhaften» Klassen hiëroglyphischer Zeichen gegenübergestellt. Mit dem Begriff der nota phonetica hat er das neulateinische Wort phoneticus geprägt. "To George Zoega", schrieb 1853 *M. Russell*[2], "also belongs the merit of employing (1797) the term phonetic" und 1880 wiederholte *Renouf*[3]: "That the oval rings" – nämlich auf der 1799 bei dem jetzt zerstörten Fort St. Julien der unterägyptischen Hafenstadt Rosette gefundenen Inschrifttafel (6) – "contained royal names was first pointed out by the Danish scholar Zoega, who was also the first in modern times to essert that some hieroglyphic characters were phonetic."

Ein lateinisches Wort phoneticus gab es nicht. Im Griechischen kommt *φωνητικός* als Adjektiv – und zwar nur in der Zusammensetzung *φωνητικὰ ὄργανα* – Sprachorgane – vor, ferner *φωνητικόν* als Teil der Seele[4] – also etwa Sprachvermögen; *Borinski* vermutet, daß – «nach griechischem Sprachgebrauch» – bei *φωνητική* zunächst *τέχνη* zu ergänzen sei; danach erst – als medius terminus – zugleich die damit beschäftigte *επιστήμη* oder *θεωρία*[5].

Bereits fünf Jahre nach dem Erscheinen des Werkes von *Zoega*

[1] Vgl. (38); (20).
[2] (28) S. 434.
[3] (26) p. 13; vgl. (22).
[4] *Plutarch* moral, p. 898.
[5] (4) S. 1.

stellte der schwedische Archäologe *Åkerblad* (1) ein phonetisches Alphabet der Kursive auf, die seit *Herodot* als demotische Schrift bezeichnet wird (21, 10). In ihr drücken die phonetischen, also die lautbezeichnenden Zeichen entweder *einen* – das sind die sog. alphabetischen Hiëroglyphen – oder *mehrere* Konsonanten aus (34). In den folgenden Jahrzehnten blieb der Begriff der Phonetik zunächst noch auf die Kennzeichnung von phonetischen Schriftzeichen beschränkt. «Le monument de Rosette nous présente l'application de ce système auxiliaire d'écriture que nous avons appelé phonétique, c'est-à-dire exprimant les sons, dans les noms propres des rois Alexandre, Ptolémée, etc.» schreibt *M. Champollion* 1822 in seinem offenen Brief an *Dacier*[1]. In der Auseinandersetzung mit *Young* und *Salt* (30) von 1825, welche einige Jahre danach von *Russell*[2] ins richtige Licht gesetzt worden ist, wird der Begriff der phonetischen Hiëroglyphen bereits als feststehender wissenschaftlicher Terminus gebraucht.

Von 1833 an benutzt auch *Franz Bopp* den Begriff – allerdings bereits mit einer schon stärkeren Hinwendung zur Aussprache. Er gebraucht ihn im Sinn des offenbar von ihm geprägten Terminus «euphonisch» und immer mit dem Zusatz «rein phonetisch» oder – noch öfter – «bloß phonetisch». «Ich glaube, in Kompositen wie... die Verlängerung des Endvokals des ersten Gliedes der Komposition nicht als rein phonetisch, sondern als Folge der Dualflexion ansehen zu dürfen[3]» oder – etwa 40 Jahre später – «Ich erkenne in dem Zischlaute dieser Form weder einen Zusammenhang mit dem Charakter des Futurums noch mit dem des Desiderativums, sondern einen bloßen phonetischen Zusatz».

Drei Jahre später – 1836 – spricht *Wilhelm von Humboldt*, der *Champollion* zitiert, – neben einer «intellectuellen» – von einer «phonetischen Technik», «deren sich die Sprache zur Erreichung ihrer Zwecke bedient»[4].

Im selben Jahr unterschied *Friedrich Diez* – noch einmal ungefähr in *Zoegas* Sinn – Lautbezeichnungen, die auf das phonetische Prinzip gegründet sind, von solchen, die sich auf das etymologische Prinzip stützen[5].

1837 gebraucht *Rudolf von Raumer* den neuen Begriff nicht mehr

[1] (7) S. 4.
[2] (29) S. 184–191.
[3] (3) S. 1428.
[4] (16) S. 89. Akad. Ausg. VII, 1, S. 84.
[5] (8) S. 67.

nur für geschriebene Zeichen, sondern nun zum erstenmal expressis verbis auch für die Untersuchung der Aussprache selbst. Zwar spricht auch er gelegentlich von «phonetischer Schreibweise», die er der historischen gegenüberstellt. Im übrigen aber unterscheidet er – ähnlich wie vier Jahre später *Robert Gordon Latham*[1] – zwischen graphischer, etymologischer und phonetischer Identität[2]. «Da die Umwandlung der Wörter nicht auf den geschriebenen Zeichen beruht und auf den Ähnlichkeiten derselben», heißt es in der Erstlingsschrift des Zweiundzwanzigjährigen in Anlehnung an Auffassungen *Johann Andreas Schmellers*[3], «sondern auf den gesprochenen Lauten, so müssen eigentlich mit aller Etymologie phonetische Untersuchungen Hand in Hand gehen.»

So selbstverständlich dies klingt – und vermutlich (zum Schaden der Sache) immer geklungen hat –, so schwierige Probleme verbergen sich hinter diesem «selbstverständlichen» Schluß, daß «eigentlich mit aller Etymologie phonetische Untersuchungen Hand in Hand» zu gehen hätten.

Denn so gut wie alles ist hier problematisch, und es hat eines vollen Jahrhunderts bedurft, um zu sichern, was hier unter «aller Etymologie», unter «phonetischen Untersuchungen», vor allem aber, was hier unter «Hand in Hand» zu verstehen sei.

Daher scheint es nicht unangebracht, heut – rund 130 Jahre später – einen «Kongreß für Phonetische Wissenschaften» mit einer Erörterung der hier von dem jungen *Raumer* aufgeworfenen Probleme zu eröffnen.

Leicht zu beantworten ist noch, was *Raumer* unter phonetischen Untersuchungen verstand. Er selbst hat diese Frage in einer Abhandlung von 1858 beantwortet, die den Titel trägt: «Die sprachgeschichtliche Umwandlung und die naturgeschichtliche – oder wenn man lieber will (fügt er 1863 hinzu) naturwissenschaftliche[4] Bestimmung der Laute[5].» Hier also findet sich zum erstenmal diese Gegenüberstellung, die bis zu den ersten Veröffentlichungen der

[1] (19) S. 124; vgl. (40) S. 46 f. u. S. 127.
[2] (25) S. 18 f.
[3] (35; 33).
[4] Der Begriff der «Naturwissenschaft» stammt von *Christian Wolff.* Er findet sich zuerst in seiner Schrift «Vernünftige Gedanken von Gott, der Welt und der Seele des Menschen, auch allen Dingen überhaupt» von 1719. Im Jahre 1849 hat ihm *Schiel* (in seiner Übersetzung von *John Stuart Mills* System of logic) den Begriff der «Geisteswissenschaften» (für social sciences) gegenübergestellt. Vgl. dazu (14) S. 163; sowie (9) S. 5.
[5] (25) S. 368.

Prager Phonologen, ja bis heute eine so wichtige – um nicht zu sagen: verhängnisvolle – Rolle für die Aufhellung dieses Fragenkomplexes spielt.

Was versteht *Raumer* unter «naturgeschichtlicher» oder «naturwissenschaftlicher» Bestimmung der Laute? «Die naturgeschichtliche Bestimmung der Laute», schreibt er, «hat sich zuvörderst ganz zu halten an die Laute der *Gegenwart* als das Object *unmittelbarer Beobachtung*. Der Hauptgegenstand der Beobachtung ist die Art der *Hervorbringung* der Laute[1] (an anderer Stelle: die «Bewegung der Lautwerkzeuge»). Unterschiede, welche das Ohr wahrnimmt oder wahrzunehmen glaubt, sind nicht von der Hand zu weisen. Aber in den Kreis scharfer naturgeschichtlicher Beobachtung treten sie erst dann ein, wenn es gelingt, die Verschiedenheit der *Hervorbringung* mit Bestimmtheit nachzuweisen.» Dazu als Anmerkung: «Ich stimme hier den Ansichten *Brückes* bei.» *Raumer* bezieht sich in dieser Notiz von 1863 auf die 1856 erschienenen «Grundzüge der Physiologie und Systematik der Sprachlaute» des Wiener Physiologen *Ernst Brücke* (5).

Unter «naturgeschichtlicher Bestimmung» versteht *Raumer* also sowohl das, was wir heute als sprachphysiologische, als auch das, was wir heute als auditive Methode kennen, wobei letztere bei ihm in Abhängigkeit von der ersteren verstanden wird – offenbar sowohl in der Meinung, daß der hörbare Laut die Wirkung der ihn verursachenden Bewegung der Lautwerkzeuge, als auch, daß die auditive Unterscheidung von minderer Exaktheit als die physiologische sei. Daß dies die allgemeine Meinung der Zeit war, zeigt ein Vortrag von *Hermann von Helmholtz* aus dem Jahre 1862 «über das Verhältnis der Naturwissenschaften zur Gesamtheit der Wissenschaften» (14). Eine physikalisch-akustische Bestimmung der Sprachlaute gab es in der Mitte der vorigen Jahrhunderts noch nicht.

Aber angesichts der Tatsache, daß Sprache nicht durch Beobachtung der «Lautwerkzeuge» (wie *Raumer* sie nennt), sondern durch Hören – also von Mund zu Ohr – tradiert wird, erscheint es zweifelhaft, ob der auditive Gesichtspunkt so vernachlässigt werden darf, wie es *Raumer* – und nach ihm die gesamte Experimentalphonetik bis in die dreißiger Jahre dieses Jahrhunderts – getan hat. Hinzu kommen heute selbstverständlich die physiologischen Probleme der nervösen Erregung der im Sprechen in Aktionsgemeinschaft tretenden Muskeln aus den beiden sich kreuzenden anatomisch-physiologischen Systemen: dem Magendarmtraktus – vom Mund bis zum Pharynx – und dem Atmungstraktus – von den Nasenöffnungen bis zum Larynx –, sowie die Frage ihrer zentralnervösen Steuerung, die seit der Entdeckung des motorischen Sprachzentrums durch den französischen Chirurgen und Anthropologen *Paul Broca* und des sensorischen Sprachzentrums durch *Karl Wernicke* immer stärker in den Mittelpunkt der Aphasieforschung – und damit der Hirnforschung überhaupt, der Kybernetik[1] im besonderen – getreten sind.

[1] Daß bereits *P. Nigidius* die physische Entstehung der Laute beobachtet hat, erhellt aus einer Bemerkung über die gutturale Nasalis bei Aulus Gellius (1.1) im 2. nachchristlichen Jahrhundert.

Zwar schränkt auch *Raumer* in der Tradition der phonetischen Lautzeichen *Zoegas* das Gebiet der Phonetik in der Regel auf die Erforschung der Sprachlaute ein. Aber bereits 1837 spricht er doch auch von dem «Nacheinander» der Laute in der «Melodie»[2], später auch von der Umgestaltung eines Sprachlautes «durch die Nachbarschaft eines anderen»[3], und in einem «Offenen Brief an den Herausgeber der Zeitschrift für die deutschen Mundarten» Karl Frommann von 1857 (25a) wünscht er die wirkliche Sprache verschiedener einzelner Menschen aus einer und derselben Gegend mit diplomatischer Genauigkeit kennenzulernen und will, «daß man deren Sprache so treu wie möglich zu Papier brächte»[4].

«Hätten wir einen Apparat», heißt es dort, «der das Gesprochene ebenso treu auffaßte und auf dem Papier befestigte, wie das Daguerreotyp das Gesehene, so würden dessen Leistungen dem entsprechen, was ich wünsche.» In dieser Weise will er erstens «den Satzbau des Sprechenden» festgehalten wissen, dann «die vom Sprechenden wirklich gebrauchten grammatischen Formen», und erst an dritter Stelle «würden noch die Laute des Sprechenden möglichst treu wiederzugeben sein».

Hier weitet sich zum erstenmal das Gebiet der Phonetik auf das ganze Problem der Realisierung geltender – und das heißt: in geschichtlichen, geographischen und sozialen Räumen geltender – Strukturen aus, was freilich nach ihm durch die sich seit *Rousselot* entfaltende Experimentalphonetik bald wieder – wenigstens dem Programm nach – auf die Untersuchung der Laute eingeschränkt

[1] (18) S. 31–71.
[2] (25) S. 15.
[3] (25) S. 376.
[4] (25a) S. 365.

worden ist, obwohl in den Problemen der Sprachmelodie, des Akzents, der Quantität, der Pausen immer schon auch die den Einzellaut transgredierenden suprasegmentalen Eigenschaften behandelt worden sind.

Zugleich mit seiner Forderung, das gesamte Gebiet der Realisierung geltender Strukturen ins Auge zu fassen, äußert *Raumer* den Wunsch, auch das bis dahin noch so gut wie nicht beachtete Problem der Schwankungen – und zwar sowohl im Gebrauch grammatischer Formen als das «lautliche Schwanken vieler Wortformen» zu beobachten[1]. An anderer Stelle spricht er von einer «Mannigfaltigkeit von Zwischenlauten in der lebendigen Rede» –, während die Schrift z.B. bei /t/ und /d/ nur den «weichsten und den härtesten Grad» bezeichne. (Hier kündigt sich das Problem des «Gegensatzes» an, von dem *Jespersen* spricht[2] – bzw. der «Opposition» der Prager Phonologen.) Wenn *Raumer* fortfährt, daß «alles, was dazwischen liegt, sich gefallen lassen müsse, unter /d/ oder unter /t/ eingereiht zu werden», so versteht er diese beiden Laute hier bereits als statistische Klassen – wenn auch in dem Irrtum befangen, daß dieses «Sichbegnügen» (wie er es nennt) eine Besonderheit der Schrift und zwar – gemessen an der Fluktuation der gesprochenen Sprache: ein Mangel der Schrift sei, der die «scharfe» – und das heißt wohl auch: autochthon naturgeschichtliche – Beobachtung als das eigentlich phonetische Verfahren geradezu fordere.

Wiederaufgenommen wird die Berücksichtigung dieses «lautlichen Schwankens vieler Wortformen», der «Mannigfaltigkeit von Zwischenlauten in der lebendigen Rede» und der «ganzen Reihe der Mittelstufen» zwischen den «äußersten Grenzen» von der junggrammatischen Theorie des Lautwandels. Besonders prägnant findet sich diese Frage in *Hermann Pauls* «Prinzipien der Sprachgeschichte» (23) entwickelt: «Diese Variabilität der Aussprache», heißt es darin, «die wegen der engen Grenzen, in denen sie sich bewegt, unbeachtet bleibt, enthält den Schlüssel zum Verständnis, der sonst unbegreiflichen Tatsache, daß sich allmählich eine Veränderung des Usus in bezug auf die lautliche Seite der Sprache vollzieht, ohne daß diejenigen, an welchen die Veränderung vor sich geht, die geringste Ahnung davon haben[3].»

Zum erstenmal ausgesprochen wurde diese Erscheinung übri-

[1] (25a) S. 366.
[2] (17) S. 108ff.
[3] (23) S. 55.

gens schon Jahrhunderte vorher – und zwar durch *Dante*. In seiner nach 1305 entstandenen Abhandlung «Von der Volkssprache»[1] heißt es: «Was sich langsam bewegt, können wir am wenigsten beobachten. Und je längere Zeit das Beobachten der Veränderung eines Dinges beansprucht, für um so feststehender halten wir es. Wir wundern uns also nicht, wenn die Meinungen der Menschen ... dahin gehen, daß eine und dieselbe Stadt stets mit derselben Sprache gelebt hat, da die Veränderung der Sprache innerhalb derselben Stadt nur in einem sehr langen Zeitablauf allmählich vor sich gegangen und das Leben der Menschen ihrer Natur gemäß sehr kurz ist. Wenn also in ein und demselben Volk die Sprache sich, wie gesagt, verändert im Lauf der Zeiten und auf keine Weise feststehen kann, so muß sie bei getrennt und entfernt Lebenden sich mannigfach verändern.»

Hinzuweisen ist an dieser Stelle auf Unterscheidungen noch eines anderen deutschen Sprachforschers und originellen Denkers aus dem ausgehenden neunzehnten Jahrhundert: auf das 1891 erschienene Werk «Die Sprachwissenschaft» des Professors der ostasiatischen Sprachen und der allgemeinen Sprachwissenschaft an der Universität Berlin, *Georg von der Gabelentz*, dessen Bedeutung für die moderne Linguistik wohl allzu sehr hinter dem posthum erschienenen Werk seines großen Genfer Zeitgenossen *Ferdinand de Saussure* zurückgetreten, um nicht zu sagen: vergessen worden ist.

Die Aktualität des nun fast 75 Jahre zurückliegenden Werkes liegt zunächst in der von *Gabelentz* gesehenen Verschiedenheit der damals allgemein betriebenen «*historisch-genealogischen Sprachforschung*» von der von *Gabelentz* geforderten «*einzelsprachlichen Forschung*» – eine Verschiedenheit, die erst auf Grund der *Saussureschen* Unterscheidung diachronischer und synchronischer Sprachwissenschaft durch Phonologie und Strukturalismus der letzten 30 Jahre allgemein akzeptiert worden ist. (Zu beachten ist, daß *von der Gabelentz*, wie nach ihm *de Saussure*, die einzelsprachliche bzw. die synchronische Sprachforschung *vor* der genealogisch-historischen bzw. der diachronischen abhandelt.)

Von der «lebenden Sprache» im Sinne von Einzelsprache sagt *Gabelentz*[2], daß sie «in jedem Augenblicke etwas Ganzes sei». «Man bildet sich nur zu gern ein, man wisse, warum etwas jetzt ist, wenn

[1] De vulgari eloquentia; erste Ausg. (in ital. Sprache) 1529; kritischer Text von *Bertalot* 1917; Deutsch von *Dornseiff* und *Balogh* 1925. Zitiert nach (2) S. 44.
[2] (12) 9f.

man weiß, wie es früher gewesen ist, und die einschlagenden Gesetze des Lautwandels kennt. Das ist aber nur insoweit richtig, als diese Gesetze allein die Schicksale der Wörter und Wortformen bestimmen. Weiß ich z.B., daß lateinisches /f/ im Spanischen zu /h/, /li/ vor Vokalen zu /j/ (sprich χ), die Endung der zweiten Deklination im Singular /o/ im Plural /os/ geworden ist: so ist es mir erklärlich, wie filius zu hijo werden mußte. Gesetzt nun, jedes Wort und jede Form der spanischen Sprache wäre auf diese Weise genetisch abgeleitet: wäre damit die spanische Sprache erklärt? Sicherlich nicht, denn die Sprache ist ebensowenig eine Sammlung von Wörtern und Formen, wie der organische Körper eine Sammlung von Gliedern und Organen ist. Beide sind *in jeder Phase ihres Lebens (relativ) vollkommene Systeme*, nur von sich selbst abhängig... Nicht die früheren Phasen einer Sprache erklären die lebendige Rede, sondern die jeweilig im Geiste des Volkes lebende Sprache selbst.»

«Die synchronische Sprachwissenschaft», heißt es 15 Jahre später bei *Saussure*[1], «befaßt sich mit logischen und psychologischen Verhältnissen, welche zwischen gleichzeitigen Gliedern bestehen, die ein System bilden, so wie sie von einem und demselben Kollektivbewußtsein wahrgenommen werden.»

«Die diachronische Sprachwissenschaft untersucht dagegen die Beziehungen, die zwischen aufeinanderfolgenden Gliedern obwalten, die von einem in sich gleichen Kollektivbewußtsein nicht wahrgenommen werden, und von denen die einen an die Stelle der anderen treten, ohne daß sie unter sich ein System bilden.»

Die Beziehungen zwischen dem *Gabelentzschen* «*Ganzen in jedem Augenblick*» zu den *Saussureschen* «*gleichzeitigen Gliedern, die ein System bilden*» ergeben sich von selbst. In beiden Fällen aber bedarf der Begriff dieses Ganzen bzw. dieses Systems noch einer Erläuterung.

Zunächst einmal manifestiert sich dieses «Ganze» oder dieses «System» ausschließlich in Sprechakten, die ihrerseits dem *Saussureschen* Linearitätsprinzip unterworfen sind: in der chaîne parlée. Es kann sich bei dieser Ganzheit also nur um ein funktionales Bezugssystem handeln, das zwischen allen sprachlichen Aktualisierungen der Sprachgemeinschaft besteht. – Und es tritt die Frage auf, mit welchem Recht und wie weit von den jeweils untersuchten Sprechakten auf dieses Ganze, d.h. auf die nicht untersuchbare Sprache,

[1] *Saussure* hat in der junggrammatischen Schule gelernt; er hat zwischen 1877 und 1881 in Genf, Leipzig und Berlin studiert; *Georg von der Gabelentz* war von 1878–1889 Professor in Leipzig, von 1889–1893 Professor in Berlin.

die diese Ganzheit darstellt, geschlossen werden darf. Sodann ist die Frage nach der Art der «Grenzen» dieses Systems zu stellen – und zwar wiederum nach den geographischen, sozialen im weitesten Sinn und historischen Grenzen. In dieser Fragestellung offenbart sich der Vorrang der einzel-sprachlich-synchronischen Forschung vor der genealogisch-diachronischen, die sich dem System dieser Fragen geradezu einfügt. *Von der Gabelentz* spricht daher höchst präzis auch nur von einem «relativ vollkommenen System».

«Man wird mich nicht mißverstehen», schreibt er[1], «wenn ich die Gesichtspunkte der einzelsprachlichen und der sprachgeschichtlichen Forschung einander schroff entgegensetze. Die Gleichberechtigung beider erkenne ich ja an, und ich suche zu zeigen, wie die beiden sich am Ende ineinander verweben müssen.»

Die Aktualität seines Werkes liegt ferner in der Unterscheidung der «*Sprache als Erscheinung*, als jeweiliges Ausdrucksmittel für den jeweiligen Gedanken, d.h. als Rede», der «*Sprache als einer einheitlichen Ganzheit* solcher Ausdrucksmittel für jeden beliebigen Gedanken» – etwa als «Sprache eines Volkes, einer Berufsklasse, eines Schriftstellers usw.», als «Gesamtheit derjenigen Fähigkeiten und Neigungen, welche die Form, derjenigen sachlichen Vorstellungen, welche den Stoff der Rede bestimmen». Und drittens der *Sprache als* «*Sprachvermögen*, d.h. der allen Völkern innewohnenden Gabe des Gedankenaustausches durch Sprache» – «ebenso wie das Recht und die Religion ein Gemeingut der Menschen»[2].

Man erkennt auch hier wieder sogleich die Beziehungen dieser Dreigliederung zu *Saussure*, dem sie – 15 Jahre später – auch durch die Unterscheidungen der französischen Sprache: parole, langue und langage, nahegelegt worden sind.

Und wie *Rudolf von Raumer* – und später *Hermann Paul* – betont auch *von der Gabelentz* die Bedeutung der Schwankungen der Aussprache. Ja, klarer als *Raumer* sieht er das Verhältnis der endlichen Zahl von Lauten einer Sprache zu der Fülle ihrer verschiedenartigen Realisierungen. Sie verhielten sich «wie Arten zu Individuen, wie Kreise zu Punkten... immer duldet die Sprache einen gewissen Spielraum». Und nicht mehr wie *Raumer*, hält er es für einen *Mangel* der Lautschrift, daß sie den Reichtum aller Nuancen der Rede nur in einer endlichen Zahl von Lautzeichen fassen könne, sondern er betont, daß «der Sprachforscher» bestätige, daß in der

[1] (12) S. V.
[2] (12) S. 3f.

Mannigfaltigkeit der Realisierungen doch «immer dasselbe Wort... immer richtig ausgesprochen» worden sei. Klarer ließ sich vor rund 75 Jahren kaum sagen, was uns in den letzten drei Jahrzehnten – nach Ausbildung einer physikalischen Akustik der Sprachlaute – immer stärker zum Problem geworden ist.

2. *Systematischer Teil*

Ungelöst blieben bei *Georg von der Gabelentz* ebenso wie bei *Ferdinand de Saussure* die Fragen der «allgemeinen Sprachwissenschaft» bzw. die Probleme der langage. Beide definieren die Sprache in diesem Sinn als «menschliches Sprachvermögen». Angesichts dieser Definition erhebt sich aber natürlich sogleich die Frage nach der wissenschaftlichen Methode, die es erlaubt, dieses «menschliche Sprachvermögen» zu untersuchen. Wort wie Begriff des «Vermögens» verweisen auf psychologische und biologische Zusammenhänge: das Vermögen zu sprechen im Unterschied zum Vermögen zu atmen, bzw. das Vermögen des Menschen im Unterschied zu dem Vermögen der Tiere. Das aber hieße natürlich nichts anderes, als daß die «allgemeine Sprachwissenschaft» keine autonome Wissenschaft mehr wäre, sondern mit ihrem Gegenstand langage in einem System anderer «Vermögen» als Teilgebiet einer umfassenderen Disziplin stünde, deren Methode und deren Begriffssystem bisher freilich reichlich unbestimmt geblieben sind. Eine ähnliche Kalamität ergibt sich übrigens seit *Saussure* durch die Versuche, die sprachlichen Zeichen in ein System von Zeichen im weiteren Sinn zu verweisen, ohne dabei zu berücksichtigen, daß diese Zeichen im weiteren Sinn nicht ohne sprachliche Zeichen im engeren Sinn bestimmbar sind. Natürlich gilt für jede Wissenschaft, daß ihre Urteile sprachbezogen, d.h. auf das System der Einzelsprachen bezogen sind. Hier aber, wo Zeichen zum Gegenstand werden, ist es wichtig, sie nicht mit den Zeichen des Urteils zu verwechseln. *Ungeheuer* deutet die Schwierigkeit an, die sich aus diesem Wechselverhältnis ergibt.

Prüft man, was *Gabelentz* in der «allgemeinen Sprachwissenschaft» zusammenfaßt, so sind es vier methodisch höchst heterogene Problembereiche, die er unter diesem Titel behandelt: nämlich psychologische, physiologische (also biologische), linguistische und

– last not least – erkenntnistheoretische Fragenkreise, die zu entwirren eine noch heute lohnende problemgeschichtliche Aufgabe wäre.

Wie wenig *Gabelentz* ein derartiges Abgleiten auf andere Ebenen aber eigentlich beabsichtigte, ergibt sich aus der Art, in der er diesen Begriff der Sprache (gleich langage) einführt und ihn neben die Begriffe des Rechts und der Religion stellt. Hier wird klar, daß es sich nicht um ein «Vermögen» in irgendeinem vagen psychologischen oder gar physiologischen Sinn handelt, sondern um ein System, das er von anderen – vergleichbaren – Systemen unterschieden wissen will. Das heißt, daß sich für ihn dieser Begriff der «Sprache überhaupt» auf das System jener Begriffe und Relationen bezieht, die der Linguist voraussetzen muß, um sein Objekt, die Einzelsprachen, in synchronischer und diachronischer Hinsicht ins Blickfeld und in den Griff zu bekommen.

Bei dieser Auffassung tritt die *Linguistik* mit ihrem spezifischen Ziel und ihrem System von Voraussetzungen neben die übrigen Wissenschaften, die durch ihre Ziele und Voraussetzungen definiert sind. Der Natur der Sache nach sind das zunächst die ebenfalls auf Quellen bezogenen Disziplinen der *Geschichte* und der *Literaturwissenschaft*[1], sodann die *Psychologie* mit ihren Beziehungen zur inneren und äußeren Psychophysik im Sinne *Fechners*[2], d.h. zu Physiologie (Biologie) und *Umwelt* (Soziologie und Geographie). Erst danach folgt der Bezug auf die «*Mathematik*» und die «mathematischen Wissenschaften» *Physik* und *Chemie*.

Überblickt man das halbe Jahrhundert, das seit dem ersten Internationalen phonetischen Kongreß im Jahre 1914 vergangen ist, so wird man zwei Phasen der Forschung unterscheiden können: die erste, die zu dem von *Jacob van Ginneken* – wohl in Anlehnung an den englischen Begriff der phonetics – geprägten Begriff der «*phonetischen Wissenschaften*» geführt hat; und die zweite, die die Zusammengehörigkeit dieses Systems phonetischer Wissenschaften wieder sichtbar werden läßt. Dieser Zusammenhang hat sich inzwischen als ein Gefüge gegenseitiger Abhängigkeiten erwiesen, deren Struktur nicht ohne Bezug auf das eben angedeutete System der Wissenschaften untersucht werden kann.

Die für den Fortgang der phonetischen Forschung entscheidende Erkenntnis scheint mir dabei die Einsicht in den *Vorrang der lin-*

[1] Auf deren Beziehungen zu den Bühlerschen Kategorien des Appelles und der Kundgabe kann hier nur hingewiesen werden.

[2] (11) S. 10f.

guistischen Zielsetzung und Fragestellung zu sein. Das heißt, daß das System der Sprachen auch Ziel und Aufgabe aller sog. «phonetischen Wissenschaften» beherrscht, wodurch alle praktischen, nicht auf reine Forschung abgestellten Bestrebungen: mutterprachliche Erziehung und fremdsprachlicher Unterricht, Logopädie und Phoniatrie, automatische Spracherkennung, Lehr- und Übersetzungsmaschinen – von so großer Bedeutung sie natürlich sind – einerseits klar von den rein wissenschaftlichen Richtungen unterschieden, andererseits in ihrer Abhängigkeit von dieser wissenschaftlichen, linguistisch-phonetischen Behandlung des Fragenkreises erkannt werden.

Ich beschränke mich im folgenden auf die Systematik der wissenschaftlichen Aufgaben und Methoden und auf ihre Bedeutung für die Struktur der Phonetik als Teilgebiet der Linguistik, d.h. als «Sprachwissenschaft vom Sprechen», wie *Saussure* sie nennt. In welcher Weise Linguistik und Phonetik in den 35 Jahren seit dem Hervortreten der Prager Phonologie zusammengewachsen sind, bedarf heute keines Hinweises mehr. Dieses Zusammenwachsen ist von beiden Seiten des Tales erfolgt, das die diachronische Schriftlinguistik von der naturwissenschaftlichen Experimentalphonetik der Jahrhundertwende trennte. Die verbindende Brücke zu schlagen, ist technisch durch den Schallträger: durch Edisonwalze, Schallplatte und Tonband, durch Tonfilm und Röntgentonfilm, ermöglicht worden.

Unter wissenschaftlichem Aspekt begann jener Brückenbau auf der einen Seite durch die Hinwendung von Linguisten zunächst zu schriftlosen Sprachen, vor allem den nordamerikanischen Indianersprachen, sodann zur Phonologie bzw. zur Phonematik, auf der anderen Seite durch die Erkenntnis der Abhängigkeit der quantitativ-phonetischen Forschung von jener phonologischen bzw. phonematischen Forschung und deren Ergebnissen. Um die Aufhellung jener Beziehungen bzw. jener Abhängigkeiten ging es in den seinerzeit als «Phonometrie» zusammengefaßten Bestrebungen auf akustischem, sprachphysiologischem und statistischem Gebiet.

Immer stärker sind in diesen drei Jahrzehnten die Probleme der *auditiven* Phonetik in den Vordergrund getreten – einerseits in Form der Abhängigkeit des Sprachverstehens von den linguistischen Strukturen der verstandenen Sprache, andererseits in Form der Abhängigkeit der akustischen und sprachphysiologischen Forschung von der auditiven Segmentierung.

Wo der modernen Linguistik ihre Hinwendung zu den Problemen der Realisierung linguistischer Strukturen heute noch bestritten wird, kommt dies dem Versuch gleich, der Biologie ihre Hinwendung zu Genetik und Verhaltensforschung zu bestreiten mit dem unvorstellbaren Argument, daß die Paläontologie an ihren fossilen Objekten über die Petrefaktenkunde hinaus nicht weiter als bis zur Paläobiologie vordringen könne und daß die Biologie daher ebenfalls nicht weiter vorzudringen brauche.

In dem halben Jahrhundert, das seit dem ersten Internationalen Kongreß für Phonetik vergangen ist, hat sich sowohl durch erkenntnistheoretische Erörterungen (40) als durch Deskription und Analyse eines immer größer werdenden Materials klar und nun wohl unwiderleglich gezeigt, daß auch der sogenannte «Sprechvorgang» in all seinen auditiven, physiologischen und akustischen Seiten nicht analysierbar ist, wenn solchen Untersuchungen nicht die einzelsprachlichen Strukturen zugrunde gelegt werden, ohne welche der Sprech vorgang eben kein Sprech vorgang ist. Die Ergebnisse dieser Erörterungen der letzten drei Jahrzehnte dürfen daher dahingehend zusammengefaßt werden, daß das System der phonetischen Aspekte: der auditive und der akustische, der sprachphysiologische und der heterograd-statistische als Inbegriff der Realisierung geltender Strukturen zu den Voraussetzungen der Linguistik selbst gehört. Und wenn *Rudolf von Raumer* gefordert hatte, daß «mit aller Etymologie» – wir müßten heute (und zwar durchaus im Sinne *Raumers*) sagen; daß «mit aller Linguistik» phonetische Untersuchungen «Hand in Hand» zu gehen haben, so heißt das nichts anderes. Es bleibt zu fragen, wie dieses «Hand in Hand» für die linguistische und phonetische Forschung im einzelnen zu verstehen ist.

Bei der Mannigfaltigkeit dieser Forschungsaspekte und erst recht bei der längst dem Einzelnen unübersehbar gewordenen Fülle der Forschungsergebnisse, kann dieses «Hand in Hand» nur in einem beinahe wörtlichen Sinn verstanden werden: nämlich als ein «Hand-in-Hand-Arbeiten» eines Teams von Forschern verschiedener Ausbildung, die sich entschlossen haben, sich gemeinsam *einem* Ziel: dem linguistischen Ziel der Erhellung des Systems der Sprachen und Mundarten zu unterwerfen. Dabei ist das Auswahlprinzip für ein solches Team eine Funktion der Struktur der Linguistik selbst: indem der Linguist – auch im engen und herkömmlichen Sinn – einem Objekt gegenüber auch nur von «Schrift» spricht, muß auch er bereits voraussetzen, daß es sich dabei um spezifische Symbole einer *gesprochenen* Sprache handelt. Und weiter muß er voraussetzen, daß die in geographischen, sozialen und historischen Räumen geltende Struktur dieser Sprache sich im Sprachgebrauch und d.h. im Sprechen einzelner – menschlicher – Sprecher dieser Gemeinschaft realisiert, nur als solche überhaupt eine sprachliche Struktur ist und daher im Grund auch nur *in* dieser Realisierung gefunden werden kann. Damit aber sind die auditiven, akustischen, physiologischen und heterograd-statistischen Aspekte als Bedingungen *jeder* linguistischen Forschung – keineswegs nur der phonologischen – erwiesen.

Dabei dürfen wir die physiologischen, akustischen und auditiven Aspekte zugleich als die drei Hauptphasen der phonetischen Forschung des letzten Halbjahrhunderts verstehen; der fast ausschließlich muskel-*physiologischen*, sog. «genetischen» Phase der Phonetik vor und nach der Jahrhundertwende ist seit den dreißiger Jahren die immer stärker in den Vordergrund gerückte *akustische* Phase gefolgt. Und wenn nicht alle Zeichen trügen, stehen wir heute am Beginn der dritten Phase, in der zu diesen beiden klassisch phonetischen Aspekten die noch in ihren Anfängen steckende *psychologische:* nämlich auditive *und* sprechmotorische Forschungsrichtung, tritt (27).

Aber alle diese immer enger zusammenwachsenden Forschungsrichtungen haben sich im Lauf der letzten Jahrzehnte bereits außerordentlich differenziert. Zunächst einmal ist unter erkenntnistheoretischem, vor allem aber unter kommunikations- und informationstheoretischem Einfluß die kommunikative *Einheit von Sprecher und Hörer* wesentlich stärker beachtet worden als in der alten experimentalphonetischen Ära, so daß schon aus diesem Grund zur Physiologie des Sprechvorgangs die Physiologie des Hörvorgangs getreten ist – beide in unmittelbarer Beziehung zu psychologischen Aspekten dieser zusammengehörigen «Vorgänge», die gerade unter dem Gesichtspunkt der Psychologie ihren Charakter als in der Zeit ablaufende Vorgänge einbüßen bzw. einschränken: indem der Sprecher spricht oder sprechen will – und das heißt immer: jetzt – oder der Hörer sprechen hört und versteht – und das heißt wieder: jetzt –, erweisen sich diese beiden physiologischen «Vorgänge» – im Unterschied etwa zu den physiologischen Vorgängen der Herz- und Nierentätigkeit – als «präsenzielle Vorgänge»: Sprecher und Hörer *wissen*, daß sie sprechen oder hören; die «Vorgänge» des Sprechens

und Hörens erweisen sich dadurch als präsenziell gegliederte Ereignisse von eigentümlich ganzheitlichem Charakter. Die Faktizität dieses Erlebens unter Vermeidung des Psychologismus in der Urteilstheorie verankert zu haben, war – auch gegenüber der Phänomenologie *Husserls* – die epochale Leistung meines Lehrers *Richard Hönigswald*, wie für unsere Generation *Hans Wagner* noch einmal nachgewiesen hat[1].

Das Prinzip dieser Gliederung auditiver und sprechmotorischer Erlebnisse aber ist – neben anderem – auch das Prinzip der linguistischen Gliederung des Gehörten und des Gemeinten. Gliedert sich also das Gesamtgebiet der Phonetik gemäß der erkenntnistheoretisch zu bestimmenden Struktur der Linguistik, so erweist sich die Abhängigkeit der phonetischen Forschung auf all diesen verschiedenen Gebieten: auf auditivem, physiologischem und akustischem Sektor, von der Struktur der jeweils zu untersuchenden Einzelsprache.

Das sind die Zusammenhänge, in deren Rahmen sich die alte Muskelphysiologie der Lautstellungen, später der Bewegungen der Sprachorgane erweitert hat zu einer Neurophysiologie des Sprechens (bereichert vor allem in jüngster Zeit durch die Neurophysiologie des Hörens) – eingeleitet z.T. durch die hydrodynamischen Hörtheorien von *Békésy* und *Ranke* (24), zu neuen myo- und neuroanatomischen Forschungen, zu zellular- und elektrophysiologischen sowie biochemischen Untersuchungen – zu weitreichenden Einsichten also in die biologischen Bedingungen des Sprech- und Hörvorgangs. Nur im Zusammenhang neurophysiologischer und kybernetischer Erforschungen zentralnervöser Vorgänge kann künftig wohl auch die Röntgenologie der Sprechbewegungen vorwärts getrieben werden.

Im Mittelpunkt aller phonetischen Untersuchungen steht nach dem Gesagten der dem Linearitätsprinzip *Saussures* unterworfene *phonematische Text*. Er stellt das Bindeglied dar zwischen allen Formen phonetischer Deskription auf der einen, syntaktischer, lexikalischer, morphologischer und phonematischer Strukturanalyse auf der anderen Seite.

Dabei taucht zuerst das linguistische und statistische Problem auf, mit welchem Recht von der Analyse eines endlichen Textes – also eines oder mehrerer Sprecher, die in einer geschichtlichen und

[1] (37) S. 1–22 und 93–123.

daher unwiederholbaren Situation gesprochen haben – auf *die* Sprache geschlossen werden darf, deren Struktur mit vergleichbaren Strukturen anderer Sprachen, anderer Sprachschichten, anderer Sprachepochen in Beziehung gesetzt werden soll. Zusammen mit meinem Freunde *Kurt Zwirner* glaube ich, die Lösung des statistischen Problems vorbereitet zu haben (41). Wir haben gezeigt, daß hier nicht unbegrenzte Vergrößerung des Materials des Rätsels Lösung ist, sondern linguistische Beurteilung der Texte und ihre Erweiterung zu einer Grenze, die durch homograde Statistik bestimmt werden kann. Damit rücken Fragen der Zählung sprachlicher Einheiten in den Vordergrund phonetischer Forschung, wobei hier nur angedeutet sei, daß die Zählung von Wörtern infolge der Abhängigkeit ihrer Häufung vom Gegenstand der Rede oder des Gesprächs auch bei beliebiger Vergrößerung des Materials nur recht bedingten wissenschaftlichen, d.h. hier: zu Vergleichungen geeigneten Sinn hat, während ihre praktische Bedeutung für mancherlei Anwendungen natürlich nicht bestritten werden kann und bestritten werden soll.

Hier sieht sich also die phonetische Forschung – wenn auch in negativer Hinsicht – auf Inhalte der untersuchten Texte verwiesen. Unberücksichtigt darf dabei hier der Unterschied von Bedeutung und Gegenstand bleiben. Es mag genügen, auf die alten Untersuchungen *Julius Stenzels* zur Bedeutungstheorie zu verweisen (36) und darauf, daß es ein wissenschaftstheoretisches Problem ist, das System der denkbaren *Gegenstände* – und das heißt: der bestimmbaren, also der wissenschaftlich bestimmbaren Gegenstände – als System der *Wissenschaften* zu ordnen, in denen die Linguistik selbst ihren adäquaten Platz haben muß[1].

Die Erforschung einer neuen Sprache sieht sich einem eigentümlichen Wechselverhältnis der auditiven Erfassung realisierter Strukturen und der phonematischen Kenntnis jener Strukturen gegenüber: immer wieder ist seit *Trubetzkoy* die phonetische Deskription neuer Sprachen für ihre phonematische Analyse gefordert worden. Umgekehrt aber ist die Kenntnis – und die Beherrschung – des Phoneminventars Voraussetzung für die auditive Deskription der zu untersuchenden Sprache, das heißt also zunächst für die Herstellung adäquater Abhörtexte. Zwischen beiden Textformen: dem phonematischen und dem Abhörtext steht der allophonische oder der

[1] (15) vor allem S. 28f.

Lautklassentext, der die linguistische Regel des hic et nunc Gemeinten und Gehörten darstellt. Unter dem unglücklichen Zwang der Zweiteilung in Phonologie und Phonetik war *Trubetzkoy* immer bemüht, dem Phonem den Sprach- *oder* Sprechlaut gegenüberzustellen, während die Praxis dazu zwingt, den Sprachlaut vom hörbaren Laut der gesprochenen Rede, andererseits aber auch vom Phonem zu unterscheiden[1].

Zu den noch nicht genügend beachteten Besonderheiten von Abhörtexten gehört ihr phonematisch-auditiver Unsicherheitsfaktor: wer Erfahrung im Abhören von Tonbändern hat, weiß, daß auch im Abhören geübte Kenner der betreffenden Sprache nicht von Lautzeichen zu Lautzeichen gleiche Abhörtexte gewinnen: immer gibt es Stellen, an denen verschiedene Abhörer den an einer bestimmten Stelle der gesprochenen Rede hörbaren Wert verschiedenen Lautklassen zuordnen. Dies ergibt sich aus dem Verhältnis endlicher Lautklassen zu den unendlich vielen Möglichkeiten ihrer Realisierung, die alle denkbaren Übergänge zeigen können. Solche Stellen können insofern von besonderem linguistischem Interesse werden, als an ihnen zuerst Lautwandelerscheinungen zu erwarten – und möglich – sind, die sich an den vergleichsweise «festen» Stellen des Textes kaum oder jedenfalls nicht unabsichtlich einstellen werden. Daß dabei der Unterschied zwischen phonologisch relevanten, distinktiven Oppositionen und phonologisch irrelevanten Oppositionen von wesentlicher Bedeutung sein wird, bedarf kaum der Erwähnung.

Aber auch in den Abhörtexten symbolisieren die Lautzeichen noch *Klassen*, die allein es erlauben, die Fülle der nur noch quantitativ faßbaren Variationen in einer Weise zu ordnen, die linguistischer Deskription für vergleichende Zwecke genügt. Hier erst treten Sonagramm, Oszillogramm, Umhüllende, Pitch-Kurve und Röntgentonfilm ins Blickfeld des linguistisch orientierten Phonetikers. Seine linguistische Orientierung erweist sich darin, daß er alle von diesen verschiedenartigen Dokumentationsformen gewonnenen Meßdaten über den Abhörtext dem phonematischen Text zuordnet, der sich dadurch zu der Textliste erweitert. Diese erst wird – natürlich nur grundsätzlich verstanden – für die künftige Linguistik das eigentliche Objekt der Analyse werden, von dem aus sie zu den Daten vorzudringen in der Lage ist, die die zu untersuchende Sprache charakterisieren.

[1] Vgl. dazu auch (32).

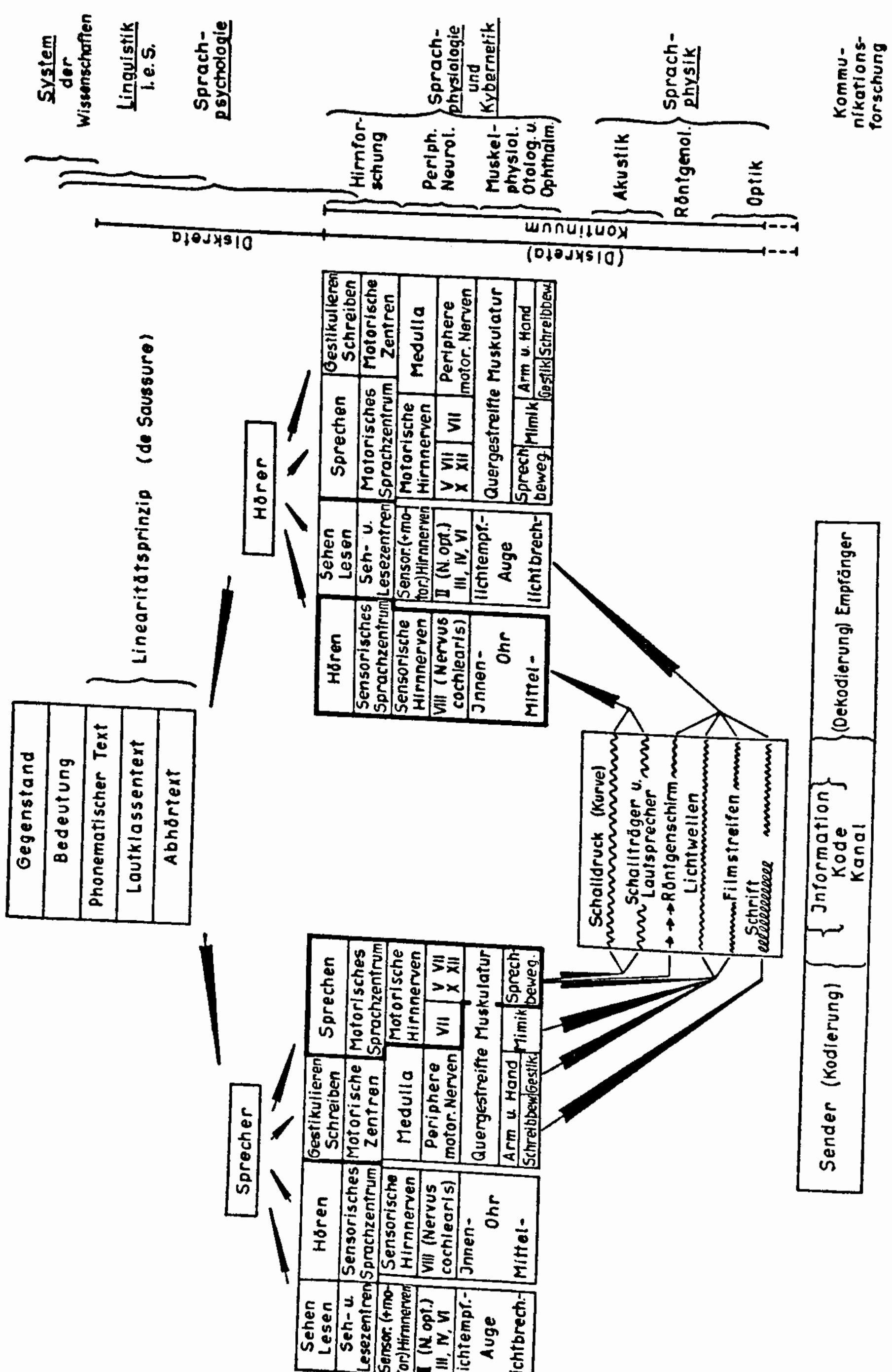


Abb. 1

Das Schema (Abb. 1) soll die Beziehungen der zum Sprechvorgang gehörenden Sachverhalte und der Formen ihrer Erforschung erläutern:

Während die beiden Außenblöcke des Feldes den Sprecher, der stets zugleich Hörer ist, und den Hörer, der stets zugleich Sprecher ist, und deren psychologisch-physiologische Funktionen darstellen, bezieht sich der mittlere Block in seinem oberen Anteil auf das Gemeinte und Verstandene, also auf das Bezeichnete und auf das Bezeichnende[1], in seinem unteren Teil auf die Schall- und Lichtwellenvorgänge, die der Lautsprache und der Gestik zugeordnet sind.

Voraussetzung exakter phonetischer Forschung ist die Fixierung jener Wellen in reproduzierbarer, graphisch darstellbarer und meßbarer Form, also durch Tonband, Film und Röntgenfilm einerseits, Schalldruck- und Bewegungskurven (13) andererseits. Erst wo die durch quantitative Bearbeitung gewonnenen Meßdaten den Abhörtexten in Form von Textlisten zugeordnet sind, ist eine erschöpfende Analyse auch der phonematischen Texte durchführbar.

Diese Komplexität des Sachverhalts «Sprache» und der Formen seiner wissenschaftlichen Erfassung verlangt die Schaffung und Institutionalisierung von Arbeitsgemeinschaften, in denen Linguistik, Logistik und mathematische Informationstheorie, Psychologie und Soziologie, Neurophysiologie und Neuropathologie, Muskel- und Sinnesphysiologie, physikalische Akustik und Röntgenologie vertreten sind. Durch solche Institute erst wird die Forderung *Rudolf von Raumers* vom Jahre 1837 erfüllt, daß «eigentlich mit aller Etymologie phonetische Untersuchungen Hand in Hand gehen» sollten.

Literatur

1. *Akerblad, J. D.:* Lettre sur l'inscription égyptienne de Rosette (Paris 1802).
2. *Arens, H.:* Sprachwissenschaft. Der Gang ihrer Entwicklung von der Antike bis zur Gegenwart. Freiburg/München: Alber (1955) = Orbis Academicus; Bd. I/6.
3. *Bopp, F.:* Vergleichende Grammatik des Sanskrit, Zend, Griechischen, Lateinischen, Litthauischen, Gothischen und Deutschen (Berlin 1833–1852).
4. *Borinski, K.:* Grundzüge des Systems der artikulierten Phonetik zur Revision der Prinzipien der Sprachwissenschaft (Stuttgart 1891).
5. *Brücke, E. Th.:* Ernst Brücke (Wien 1928).
6. *Budge, E. A. W.:* The Rosetta Stone in the British Museum (London 1929).
7. *Champollion, J. F.:* Lettre à M. Dacier... relative à l'alphabet des hiéroglyphes phonétiques employés par les Egyptiens pour inscrire sur leurs monuments les titres, les noms et les surnoms des souverains grecs et romains (Paris 1822).

[1] (31) S. 122.

8. *Diez, F.*: Grammatik der romanischen Sprachen; Teil 1 (Bonn 1836).
9. *Dilthey, W.*: Einleitung in die Geisteswissenschaften; Vol. 1, 2. Aufl. (Leipzig/Berlin 1922).
10. *Erman, A.*: Die Hiëroglyphen (Sammlung Göschen Nr. 608, 1917).
11. *Fechner, G.*: Elemente der Psychophysik; 1. Teil, 3. Aufl. (Leipzig 1907).
12. *von der Gabelentz, G.*: Die Sprachwissenschaft, ihre Aufgaben, Methoden und bisherigen Ergebnisse (1. Aufl. 1891, 2. Aufl. Leipzig 1901).
13. *Gottheiner, V.* und *Zwirner, E.*: Die Verwendung des Röntgentonfilms für die Sprachforschung. Fortschr. Röntgenstrahlen *47:* 455–462 (1933).
14. *v. Helmholtz, H.*: Über das Verhältnis der Naturwissenschaften zur Gesamtheit der Wissenschaften (1862). In: Vorträge und Reden; Vol. 1. Bd., 4. Aufl. (Braunschweig 1896).
15. *Hönigswald, R.*: Die Grundlagen der Denkpsychologie. Studien und Analysen (1. Aufl. 1920, 2. Aufl. Leipzig 1925).
16. *v. Humboldt, W.*: Über die Verschiedenheit des menschlichen Sprachbaues und ihren Einfluß auf die geistige Entwicklung des Menschengeschlechts (1836, Berlin 1935).
17. *Jespersen, O.*: Lehrbuch der Phonetik (Fonetik, Teil 1). Übers. v. *H. Davidson* (1. Aufl. 1904, 4. Aufl. Leipzig/Berlin 1926).
18. *Keidel, W.-D.*: Kybernetische Systeme des menschlichen Organismus. Arbeitsg. f. Forschung des Landes Nordrhein-Westfalen; Heft 118 (Köln/Opladen 1964).
19. *Latham, R. G.*: Facts and observations relating to the science of phonetics. In: Ldn., Edbgh. and Dbl. Philos. Mag. J. Sci. Vol. XVIII (1841).
20. *Michaelis, A.*: Georg Zoega. Allg. Deutsche Biographie. *45:* 386–402 (Leipzig 1900).
21. *Möller, G.*: Hiëratische Paläographie. 3 Bde. (Leipzig 1909–1912, 2. Aufl. 1927).
22. *Murray, J.*: New English dictionary, on historical principles. Vol. 7, 2nd (Oxford 1909).
23. *Paul, H.*: Prinzipien der Sprachgeschichte (1. Aufl. 1880, 5. Aufl. Halle 1920).
24. *Rauch, S.*: Biochemie des Hörorgans. Einführung in Methoden und Ergebnisse (Stuttgart 1964).
25. *v. Raumer, R.*: Die Aspiration und die Lautverschiebung. Eine sprachgeschichtliche Untersuchung (1. Aufl. Leipzig 1837). In: Gesammelte sprachwissenschaftliche Schriften (hrsg. v. *Heyder* und *Zimmer*) (Frankfurt/Erlangen 1863).
25a. Offener Brief an Karl Frommann. Die deutschen Mundarten. 4. Jg. (1857).
26. *Renouf, P. Le Page:* Lectures on the origin and growth of religion as illustrated by the religion of ancient Egypt (delivered 1879) (London 1880).
27. *Richter, H.*: Anleitung zur auditiv-phänomenalen Beurteilung der suprasegmentellen Eigenschaften sprachlicher Äußerungen. Phonetica *11:* 51–62 (1964).
28. *Russell, M.*: Egypt XI (1853).
29. *Russel, M.*: View of ancient and modern Egypt with an outline of its natural history (Edinburgh 1831).
30. *Salt, H.*: Essay on Dr. Young's and Champollion's phonetic system of hieroglyphics (London 1825).
31. *de Saussure, F.*: Grundfragen der allgemeinen Sprachwissenschaft (Cours de linguistique générale, 1916). Übers. v. *H. Lommel* (Berlin/Leipzig 1931).
32. *Schmitt, A.*: Über den Begriff des Lautes. In: Arch. vgl. Phonetik. *2:* 65–77, 161–176 (1938).
33. *Schröder, E.*: Johann Andreas Schmeller. Allg. Deutsche Biographie. *31:* 786–792 (Leipzig 1890).
34. *Sottas, H.* et *Driton, E.*: Introduction à l'étude des hiéroglyphes.
35. *Steinmeyer, W.*: Rudolf von Raumer. Allg. Deutsche Biographie. *27:* 423–429 (Leipzig 1888).
36. *Stenzel, J.*: Sinn, Bedeutung, Begriff, Definition. Ein Beitrag zur Frage der Sprachmelodie. In: Jb. Philol.; Vol. 1 (München 1925). Wiederabgedruckt für die wissenschaftliche Buchgesellschaft (Darmstadt 1958).

37. *Wagner, H.*: Kritische Betrachtungen zu Husserls Nachlaß. Philos. Rundschau; 1. Jg. (1953/54).
38. *Welcker, F. G.*: Zoegas Leben. Sammlung seiner Briefe und Beurtheilung seiner Werke. 2 Teile (1. Aufl. 1819) Neudruck (Halle 1912/13) = Klassiker der Archäologie, Bd. 2.
39. *Zoega, G.*: De origine et usu obeliscorum ad Pium IV. Pont. max. (Rom 1797).
40. *Zwirner, E.* und *Zwirner, K.*: Grundfragen der Phonometrie (Berlin 1936); 2. Aufl. in «Phonometrie», Vol. 1 (Karger, Basel/New York 1964).
41. *Zwirner, E.* und *Zwirner, K.*: Lauthäufigkeit und Zufallsgesetz. In: Forschungen und Fortschritte; 11. Jg. (1935). Neudruck in «Phonometrie», Vol. 2 (Karger, Basel/New York 1964).

Adresse des Autors: Prof. Dr. Dr. E. Zwirner, Institut für Phonetik der Universität, Meister-Ekkehart-Straße 7, *5 Köln* (Westdeutschland).

Discussion

Eli Fischer-Jørgensen (Kopenhagen): Aus den Ausführungen greife ich die Bedeutung der phonematischen Struktur für die Analyse der phonetischen Manifestation heraus.

Die Weise, in der dieses Problem vorgelegt wurde, ist z. T. durch die Methoden bedingt, die Herr *Zwirner* in seinen eigenen Untersuchungen verwendet, indem er 1. z. T. unbeschriebene Sprachen (deutsche Dialekte), 2. frei gesprochene Erzählungen als Texte verwendet, 3. sein Interesse besonders auf strukturell schwer faßbare Erscheinungen konzentriert (Akzent, Intonation usw.).

Die Wechselwirkung zwischen auditiver Auffassung und Kenntnis der phonematischen Struktur, die Herr *Zwirner* erwähnt, ist für die Phase der Erlernung neuer Sprachen kennzeichnend. Wenn man von dem phonematischen System mehr weiß, hört man anders. Diese Phase muß der instrumentellen Untersuchung vorausgehen. Es ist zu empfehlen, mit größeren, statistisch zu bearbeitenden Untersuchungen zu warten, bis man die Struktur der Sprache kennt. Man braucht zwar nicht das endgültige Phonemsystem aufgestellt zu haben, muß aber die Oppositionen und die Kombinationsmöglichkeiten kennen, weil man bei der statistischen Bearbeitung funktionell verschiedene Größen auseinanderhalten muß (z. B. bei Dauermessungen von Vokalen solche Vokale weglassen muß, die an der Quantitätsopposition nicht teilnehmen) und weil die wichtigste Aufgabe einer phonetischen Beschreibung darin bestehen muß, die phonetische Manifestation der Oppositionen festzustellen.

Man tut sicher gut daran, bei solchen Untersuchungen mit den von Herrn *Zwirner* verachteten isolierten Wörtern und kleinen Sätzchen anzufangen, weil die Bedingungen im zusammenhängenden Text so kompliziert sind, daß das Material inhomogen wird und die verschiedenen Faktoren sich nicht entwirren lassen.

Man kann deshalb wohl sagen, daß es gewagt ist, mit teilweise unbeschriebenen Sprachen und zusammenhängenden frei gesprochenen Texten anzufangen. Aber wer wagt, gewinnt. Und man muß gestehen, daß Herr *Zwirner* mit seinen Quantitätsuntersuchungen an deutschen Dialekten bemerkenswerte Erfolge erzielt hat. Nun stellt allerdings die Vokalquantität ein relativ einfaches Problem dar. Bei einer Untersuchung der Vokalqualität müßte man z. B. wissen, ob e und ε in den verschiedenen Dialekten zwei verschiedene Phoneme wären (was sich an einem kurzen zusammenhängenden Text nicht ganz leicht entscheiden läßt).

Es gibt aber auch einzelne Probleme, die man ohne Kenntnis der betreffenden Phonemsysteme untersuchen könnte, z. B. die durchschnittliche Größe der Tonintervalle.

Es handelt sich hier um Erscheinungen, die innerhalb einer Sprachgemeinschaft keine Funktion haben, die aber von einer Sprachgemeinschaft zur andern in charakteristischer Weise verschieden sein können.

Das Problem der Transkription ist, wenn man mit isolierten Wörtern arbeitet, kaum relevant; wenn man aber zu zusammenhängenden Texten übergeht (und das tun wir wohl alle früher oder später), wird dieses Problem aktuell. Herr *Zwirner* unterscheidet hier zwischen phonematischem Text, allophonischem Lautklassentext und Abhörtext. Aber die Prinzipien, die dem Lautklassentext und dem Abhörtext unterliegen, sind mir nicht klar. Es ist evident, daß man das Band abhören muß, nicht nur bei freiem Gespräch, sondern auch bei vorgelesenen Texten. Denn erstens kann sich ja der Lesende versprechen, zweitens muß man wissen, welches Phonem er, in den Fällen, wo die Wahl frei ist (z. B. «werden» mit [eː] oder [ɛ]), gewählt hat, drittens müssen die Assimilationen und Weglassungen festgestellt werden. Man muß auch durch vorläufiges Abhören feststellen, ob die Phoneme in verschiedenen Stellungen verschieden manifestiert werden und demnach bei statistischer Verarbeitung getrennt behandelt werden sollten. Es ist z. B. möglich, daß t und d im Anlaut durch Aspiration, im Inlaut durch Stimmhaftigkeit und Dauer unterschieden werden und das t im Inlaut unaspiriert bleibt. Aber die verschiedenen Stellungen (z. B. Anlaut, Inlaut usw.) sind ja linguistisch definiert und brauchen daher nicht im Text angegeben zu werden; von diesen sprachlich definierten Klassen sollte man ausgehen, nicht so, wie es in den ersten Phonometrischen Textlisten geschah, von durch die Abhörer festgestellten Klassen (z. B. aspiriertes und unaspiriertes t).

Die Untersuchung der phonetischen Manifestation der Oppositionen ist, wie gesagt, die wichtigste Aufgabe, aber auch der Unterschied zwischen kombinatorischen Varianten ist von Interesse; man findet hier oft Unterschiede, die von Sprache zu Sprache ähnlich sind; es muß auch erlaubt sein, bei der Untersuchung direkt auf allgemeine Züge hinzuzielen. Wir wollen doch nicht nur Einzelsprachen untersuchen und vergleichen, sondern auch das für viele oder vielleicht alle Sprachen Gemeinsame finden. Herr *Zwirner* scheint diese Zielsetzung nicht anerkennen zu wollen, vielleicht weil er denkt, daß man dafür alle existierenden (und noch nicht existierenden) Sprachen untersuchen müßte. Aber eine Induktion aus unendlich vielen Sprachen ist ja nicht notwendig. Wenn man einige Sprachen von verschiedener Struktur untersucht hat und in allen dasselbe Phänomen gefunden hat (z. B. a stärker als i), und wenn sich dies nicht aus der spezifischen Struktur der Sprachen, sondern im Gegenteil durch allgemeine physiologische oder physikalische oder kommunikationsmäßige Verhältnisse erklären läßt, dann kann man doch als eine wahrscheinliche Hypothese auf allgemeine Tendenzen schließen. Zwar muß man auch in diesen Fällen die spezifische Struktur der untersuchten Sprachen berücksichtigen, aber das Ziel ist ein anderes.

Antwort *Zwirner:* Den Unterschied zwischen *Lautklassentext* und *Abhörtext* halte ich deshalb für größer, als es Fräulein *Fischer-Jørgensen* tut, weil ich der Meinung bin, daß die Natur der Abhörtexte noch nicht genügend geklärt ist, da der Begriff des «Abhörens» mehrdeutig ist: vom Erraten und Ergänzen sinnvoller Zusammenhänge bis zum Hörerlebnis isolierter und dadurch sinnlos gewordener Lautgruppen oder Lautphasen bedarf es noch umfangreicher wahrnehmungs-psychologischer Untersuchungen. Die Lösung soll dadurch wenigstens vorbereitet werden, daß wir von vornherein neben dem allophonischen Text einen Abhörtext berücksichtigen.

Über die Frage, ob man an isolierten Wörtern oder kleinen Sätzchen arbeiten darf, gehen die Auffassungen von Fräulein *Fischer-Jørgensen* und mir wenig auseinander. Ich «verachte» die experimentelle Arbeit mit isolierten Wörtern oder kleinen Sätzen, so wie es Fräulein *Fischer-Jørgensen* darstellt, keineswegs. Was ich ablehnte, war die Meinung der Experimentalphonetik, daß mit Untersuchungen an isolierten Elementen die phonetischen Fragen eines Universalalphabets gelöst werden könnten. Kritisiert

habe ich den Verzicht der ehemaligen Experimentalphonetik, an zusammenhängenden Texten zu arbeiten, obwohl dies doch schon damals an Schallplatten möglich gewesen wäre.

Nicht ganz sicher kann ich die Frage beantworten, ob ich mit Fräulein *Fischer-Jørgensen* auch im Hinblick auf ihre Forderung einig bin, daß man Strukturen finden sollte, die allen Sprachen gemeinsam sind. Daß es solche Strukturen gibt, leugne ich nicht. Ich unterscheide jedoch zwei Typen dieser allen Sprachen gemeinsamen Strukturen. Zu den linguistisch-phonetischen gehört es, daß man viele Sprachen untersucht, um dann mit aller Vorsicht sagen zu können, daß das, was an vielen verschiedenartig gebauten Sprachen gefunden ist, vielleicht für alle Sprachen gilt.

Eine andere Form der Allgemeingültigkeit ist die, die Fräulein *Fischer-Jørgensen* in ihrer Diskussionsbemerkung erwähnt. Sobald sich herausstellt, daß bestimmte Phänomene deshalb bei zwei oder mehr Sprachen gefunden werden, weil sie physiologisch oder physikalisch oder kommunikationstheoretisch bedingt sind, also außerlinguistisch erklärt werden müssen, halte ich die Feststellung solcher Gemeinsamkeit selbstverständlich für möglich. Nur handelt es sich dann eben nicht um linguistische bzw. linguistisch-phonetische Erscheinungen, sondern um physiologische, physikalische oder andere Voraussetzungen, mit denen der Linguist arbeiten muß, die ihm aber durch andere Disziplinen gegeben werden, bei deren Erforschung es also gar nicht um das System der Sprachen geht, sondern um physikalische oder biologische Systeme oder dergleichen.

G. Herdan (Bristol): I should like to stress the importance of Mr. *Zwirner's* contribution to the subject of statistical linguistics. He was among the first who, together with *K. Zwirner*, showed phoneme occurrence in a given language to be a suitable field for the application of statistical methods. Among other things, he arrived at the conclusion that in order to get an idea of phoneme frequency it was not necessary to accumulate samples of prodigious proportions, and still be in doubt whether they were sufficient, but that an unbiassed random sample of moderate size was enough for obtaining reliable estimates of the basic or population probabilities.

P. Meriggi (Pavia): Zu den einleitenden Worten von Herrn *Zwirner* möchte ich mir eine kleine Randbemerkung nicht streng phonetischer, sondern allgemeiner Natur erlauben. Es ist zu begrüßen, daß er den Ursprung des Terminus «System» auf den heute fast völlig vergessenen *Georg von der Gabelentz* zurückgeführt hat (obwohl ich der ketzerischen Ansicht bin, daß die Sprache kein System darstellt oder doch nur mit wesentlichen Einschränkungen, eine Ansicht, die ich hier natürlich nicht begründen kann).

Wichtiger ist wohl, zu bemerken, daß die Deutschredenden es gar nicht nötig haben, die Termini «langue» und «parole» beizubehalten, als ob sie unübersetzbar wären. Wenn sie ihre alten, ihre größten Sprachforscher lesen würden, was sie anscheinend wohl nicht mehr tun, würden sie finden, daß man mit der üblichen Zuweisung an *de Saussure* der für die heutige Sprachwissenschaft wesentlichen Scheidung von «langue» und «parole» ihrem älteren Urheber *H. Paul* Unrecht antut, der nicht nur zwischen «Sprachusus» (langue) und «gewöhnlicher Sprechtätigkeit» (eben «parole») deutlich unterschieden, sondern sogar ihre kausale Verkettung entdeckt hat, indem er die Ursachen der Sprach(usus)veränderungen in der «gewöhnlichen Sprechtätigkeit» erblickte. Ein grundlegender Gedanke, den man auch bei *de Saussure* vielleicht voraussetzen, aber nicht ausdrücklich geäußert finden kann.

Antwort *Zwirner:* Ich darf darauf hinweisen, daß *Georg von der Gabelentz* von «*relativ*» geschlossenen Systemen gesprochen hat. Vor mehr als 70 Jahren hat er bereits Zweifel an der faktischen Geschlossenheit sprachlicher Systeme geäußert. Geschlossene Systeme sind überhaupt keine empirischen Größen, sondern Modelle, mit deren Hilfe empirische Systeme in vergleichbarer Form beschrieben werden können. Dankbar bin ich für den Hinweis auf *Hermann Paul. Kurt Zwirner* und ich hatten zwar die Gauß-Verteilung aller von uns untersuchten Eigenschaften der Sprache bereits gefunden, als wir *Hermann Paul* zu studieren begannen. Aber dann haben wir immer wieder auf die «Prinzipien» von *Paul* hingewiesen und darauf, daß er – nächst *Rudolf von Raumer* – wie kein zweiter der damaligen Generation auf den Zufallscharakter der Realisierung sprachlicher Strukturen hingewiesen hat. Ich würde sogar sagen, daß die Phonometrie in vieler Beziehung eine Fortsetzung junggrammatischer Bestrebungen im Rahmen der strukturellen Linguistik ist. Ich billige keineswegs die Herabsetzung der Junggrammatiker durch Vertreter der neueren Sprachphilosophie.

E. Buyssens (Bruxelles): Il a été fait allusion à la paternité de la notion de système. Il est exacte que *von der Gabelentz* a parlé de système avant que paraisse le Cours de Linguistique de *Saussure:* mais bien avant cela dans sa thème doctorale «Mémoire sur le système primitif des voyelles dans les langues indo-européennes» *Saussure* a utilisé le mot système pour l'ensemble des relations existants entre les voyelles. En particulier, il a identifié des voyelles grâce aux degrés d'apophonie.

Antwort *Zwirner:* Für den Hinweis auf die frühe Schrift von *Saussure* bin ich dankbar. Ich werde dieser Frage nachgehen, prüfen, was in der frühen Schrift von *Saussure* im Hinblick auf das uns interessierende Problem bereits enthalten war, und ob *Georg von der Gabelentz* diese Schrift von *Saussure* gekannt hat. Er zitiert sie nicht.

Proc. 5th int. Congr. phon. Sci., Münster 1964, pp. 25–42
(S. Karger, Basel/New York 1965).

From the University of Pennsylvania, Philadelphia

Phonetic Reconstruction

By Henry M. Hoenigswald

1. "Phonetic reconstruction" is a natural collocation of terms, inasmuch as the task of reconstructing a past stage of language seems to have greater affinity to the phonological than to the grammatico-lexical level of structure. Despite its origins, "comparative grammar" has largely come to be, over the generations, historical and comparative *phonology*[1] ("etymology" in earlier parlance), and the reason for this is not just – as is often thought – an all too human preference for the tidier and less exacting work in the more manageable field. The hope that a merely greater, but still analogous, effort is all that is needed to reconstruct the extra-phonological phases of language more satisfactorily may well be false, because what is reflected here is a true hierarchy. Its nature is perhaps best understood when we consider what difference there is between determining the phonemic shape of a given stem or affix in an ancestor language on the one hand, and translating a grammatical construction or a dictionary item into a reconstructed ancestor on the other hand. The first can be done even in the face of contradiction, i.e. of other than only one-to-one correspondence, among the descendant languages; the other cannot, unless recourse to quite different types of reasoning is had[2]. One of the safeguards of sound etymologizing lies in our customary reliance on identity or at least plausibility of meaning, and thus on a higher level than that (namely phonological shape) on which results are sought. As we go

[1] In the meaning of *Lautlehre, phonétique* (in the wider sense). The equivalent to *Phonologie, phonologie* used here is "phonemics".

[2] For somewhat closer, but very restricted, morphological and semantic analogies to phonemic "comparative" reasoning, see my *Language Change and Linguistic Reconstruction* (Chicago, 1960), p. 70f.

beyond phonology with our demands for results, no analogous upward appeal is available.

2. It is useless to speak about reconstruction without first gaining a picture of the change processes to which language is subject; and this, in turn, cannot be done without reference to the role of sound in language structure quite apart from a consideration of change processes. For instance, we must remember that linguistic judgment rests on the hearer's deciding, under relevant conditions[3], whether two given speech utterances are the same or different. No segmentation is as yet introduced; we do not ask the subject to distinguish sentences, phrases, words, syllables, phones, or features; we merely want his reaction to pairs of texts (however short we may like to keep these texts in the interest of simplicity). Segmentation comes later[4]. We may find it convenient to blame the difference between utterances that are judged different on one part of these utterances rather than on the whole or on some other part. Thus in the two texts, *These are beads* and *These are beets* we may wish to regard the reported difference as lodged in the final "consonants" rather than in some larger stretch such as, perhaps, the one taking in the preceding "vowels". This is to say that, although the principle of segmentation may be considered necessary, there is nothing necessary about a given way of segmenting. We may agree to use such a way of segmenting, without claiming uniqueness for it, so long as "different" utterances are not thrown together. Moreover, we are always free to take a Saussurean view of our segments, where any one of them is defined by its pattern of occurrence with all the others. We may further choose to identify certain segments, especially short ones, in the tradition of articulatory phonetics or on the basis of some other quasi-measurement procedure. The choice has usually been guided by hopes for fruitful generalization: sound types, distinctive features, or compatability rules concerning either of these may emerge as universally recognizable or may at least serve, by their incidence or absence, to delimit typological areas on the map. However that may be, we may picture our language as a large corpus of texts, recorded in a "phonemic" notation which utilizes, in the familiar fashion, some physical segmentation of the flow of speech. Besides, we know other things about these texts: we

[3] As in a "pair test"; see *Chomsky, N.:* Syntactic Structures, p. 96 (s'Gravenhage, 1957).

[4] In the logical sequence (though not necessarily in actual procedure).

have (though by no means necessarily in this order) located the morph boundaries within[5], recognized morphemes and constructions, and formulated certain relations which exist among the texts in the corpus and among all "possible" texts. We have a phonology, a grammar, and a dictionary; and we can, in principle, name the elements together with which a given element occurs in the texts[6].

3. Statements about linguistic change, both conventional and unconventional, typically take the form "*A*[*B*. . > *M*[*N*. .", i.e. "*A* in the environment *B*. . is replaced by *M* in the environment *N*. .", where *A*, *B*. . are elements[7] in the texts of one language, and *M*, *N*. . elements in the texts of another language whose descent from the former has been established[8]. The two languages are termed earlier stage and later stage respectively. The implication is that many texts of the later stage may be identified as equivalents of texts of the earlier stage, on grounds furnished by a theory of translation. (At the same time other earlier-stage texts are lost without replacement, and other later-stage texts come into existence without replacing a prototype – especially texts containing morphemes with obsolete and with new meanings respectively.) Thus NHG *zwei* replaces *zween* (aside from also replacing *zwo* and *zwei*); NE *meat* replaces *flesh* in some contexts but not in, say, *flesh*[*-wound;* Germanic *p* replaces IE *p* after *s* but not in most other environments; in Greek, an older *s* is in certain environments replaced by *s*, in others by 0 (nothing); and so on. These replacement processes are classified as *sound changes* if and only if the environment which needs to be accorded to *A* in order to make the statement valid does not combine with *A* itself to produce a stretch such as to be always coextensive with a morph or morph sequence[9]. The elements which most appropriately fill the positions of *A*, *B*. ., *M*, and *N*. . are of course phonemes, phonemic components, distinctive features, and the like. We shall hereafter concentrate on phonemes, and represent a sound change by writing *a*[*b*. . > *m*[*n*. ., where *a*, *b*. . are the phonemes (in

[5] Morphs are smallest meaningful phoneme stretches, not necessarily continuous. Morphemes are classes formed from morphs and defined by the pattern in which they contrast with each other or (which may amount to the same) by their grammatical or lexical meanings.

[6] The formulation must in many cases be in terms of a generative grammar.

[7] Including "0", or absence of any element. "Environments" must be mutually exclusive in order for such a notation as ours not to lead to any contradictions.

[8] On descent, see footnote 25.

[9] "Sound change operates without exceptions" is, of course, a definitional statement.

some valid phonemicization) of the earlier stage, and *m*, *n*.. are the phonemes of the later stage[10].

It should be noted that *a* and *m* are defined separately, each with purely synchronic reference to all the other phonemes with which it contrasts. Thus, IE *p* is "*p*" inasmuch as it contrasts with IE *t*, *bh*, *gʷ*, *e*.., while Germanic *p* is "*p*" as distinct from Germanic *t*, *f*, *h*... The replacement relation between the two is an independent finding. This does not mean, however, that additional relations may not exist between the replaced and the replacing phoneme, nor that certain further observations cannot be made about the replacement process.

One possible kind of observation concerns the connection of a replacement with other replacements. Some replacements are *one-to-one* in the sense that there is only one source, namely *a*[*b*.., for *m*[*n*... All instances of Arabic *f* come from Proto-Semitic *p*; so do all instances of Hebrew *p*. All instances of Germanic *p* in position after *s* replace an IE *p*. Other replacements are parties to *mergers* of original contrasts, where both *a*[*b*.. and *c*[*b*.. go to *m*[*n*... An Italian *ę* replaces both Latin *ĭ* and *ē*. A Germanic *t* when preceded by *s* and followed by *r* may go back to an IE *t* preceded by *s* and followed by *r*, but also to "0 preceded by *s* and followed by *r*". With this dichotomy between one-to-one processes and mergers there intersects another, according as *a* goes to *m* in all environments (*a*[*b*..> *m*[*n*..; *a*[*d*..> *m*[*n*.. or > *m*[*o*..) or only in some environments (*a*[*b*..> *m*[*n*..; but *a*[*d*..> *p*[*n*.. or > *p*[*o*..). To illustrate such a *conditioned* sound change, or *split:* English *θ* > later *θ* in one set of environments but > *ð* in another, with the difference between the environments themselves partly lost (hence *wreath: wreathe*); IE *p* after *s* > Germanic *p*, but > *f* and > *b* in two other environment classes; IE *s* > Greek 0 between vowels, but > *s* in certain other environments. The second dichotomy is, however, far less incisive than the first since it often depends on the notation chosen. If, say, *sb* were written instead of IE *sp* (a possible, though for various reasons less elegant alternative[11]) there would be no split, since IE *b* in general > Germanic *p*.

---

[10] The phones or properties of /*b*../ and /*n*../ may occur before, simultaneously with, and after the phones of /*a*/ and /*m*/ (in contact or discontinuously) in the speech stream.

[11] Chiefly because of the parallel "*sd*" which is a better notation for *Brugmann's* "*zd*" in items like *nisdos* 'nest'.

We may also wish to take a more systemic view of sound changes. There may be grounds for asserting that a descriptive parallelism exists, in part or in toto, between the two phonemic systems; *a*, *b*, *c*.. together with their privileges of mutual occurrence on the one hand, and *m*, *n*, *o*.., likewise, on the other hand. In this sense we may in particular make some assertion such as that *a* is homologous to *m*, or perhaps even (by some extension of whatever principle underlies here) that *a* is homologous to the union of *m* and *n* or that the union of *a* and *b* is homologous to *m*. It then becomes a matter of interest to state the extent to which phonemes are replaced by their homologs. Suppose, for instance, that the sequences *ab*, *ac* and *cb* occur but *bc*, *ba*, *ca* do not; that *nm*, *om* and *on* occur but *mn*, *mo*, *no* do not; and suppose moreover that the notion of parallelism from system to system is in this case based, by agreement, on such clustering behavior. This makes *a*, *b*, *c* homologous with *o*, *m*, *n* respectively. The question is now whether or not it is also true that *a* > *o*, *b* > *m*, *c* > *n*. The stop subsystems of IE and of Germanic may well be judged parallel, in which case IE *b*, *p*, and *bh* and Germanic *b*, *p*, and *f*, in this order, would be plausible homologs; yet *bh* is not replaced by *f*.

It has become increasingly clear that the systemic properties which are here considered have typological importance, in the sense that they tend to be characteristic of languages, related or unrelated, spoken in parts of the same (often large) area during a given (often long) period. This means that throughout such periods of stability the processes of sound change may be expected to keep moving within a framework of more or less readily identifiable structure points. It also means that when important alterations do appear in the structural complexion of a language, these alterations are likely to have a specific, area-wide direction of some sort. In the absence of detailed knowledge of the machinery that is at work here, it has proved most fruitful to treat language as if it were subject to typological pressure toward a goal, be it in the nature of preserving a prevalent, stable type or of working along with some readily recognizable trend toward a new type. But it must be stressed again that these pressures had better be thought of as specific factors in history, in spite of their broad and slow way of operating; there is no reason to acknowledge as circumscribing agents only such allegedly "universal" features as the general symmetry of linguistic (particularly, phonemic) systems (possibly modified by anatomic asymmetries), the finite character of possible distinctive features, certain constants having to do with functional load and with redundancy, and generally the minimal requirement that a language remain "a language"[12].

After this digression, let us return to the classification of our artificially isolated individual replacements. We have considered their possible concatenation through merger and split, and we have considered their relation to structural stability and instability. There remains a third criterion: that of the phonetic properties of our replacements. This criterion looms so large in the existing literature that we need not concern ourselves with it in detail. The textbooks of general linguistics and the introductory chapters to our great historical and comparative grammars are full of enumerations of varieties of phonetic change: assimilation, simplification of clusters, loss, emergence of glides, diphthongization, dissimilation, metathesis, and many others. What interests us more deeply is the relationship between this criterion and the other two points of view introduced above. It is for instance a fact that the examples of loss and cluster simplification are frequently also examples of merger with "0" (as when *hr–* is replaced by *r–* in most Germanic languages and thereby produces homonymy with earlier *r–*, i.e. "0*r*–"). But this is not necessarily true: suppose that a language in which all utterances end in consonant + vowel drops its final vowels ($..CV\#$ $> ..C\#$). This will constitute a case of $C]V[\# > 0$, but not a merger, since there is no previous "0" between C and pause. On the other hand, it contributes to, and may be the principal tool of, a striking change in the structure of the language. Or consider the frequently used and even more frequently implied notion of a sound "remaining unchanged". As a mere physical description this notion is apt to be meaningless and unverifiable, but this is irrelevant in any case, since statements of this kind usually turn out not to be intended as mere physical description. At the very least, the intention may be to describe the phonic substance of an *m* as *more* similar to that of the *a* which it replaces than to that of the *b*'s and *c*'s replaced by *n*'s and *o*'s, under criteria of similarity taken either from an universalistic framework like traditional articulatory phonetics, or from an array of distinctive features. But the intention may also be typological and refer to homologs in a stable area. Only on grounds like these might

[12] *Martinet, A.:* Eléments de linguistique générale, p. 215ff. (Paris 1960).

one justify saying that IE *p* "remains" *p* in Germanic after *s*, but "changes to" *f* or *b* in other environments – a distinction which cannot meaningfully refer to phonic identity or near-identity in any absolute sense. It is of course common practice to regard certain types of phonetic change (such as assimilation) as more plausible to postulate than certain other types (say, unvoicing of consonants between vowels). But there are enough counter-instances to such things to suggest that the ease or difficulty with which the many physical varieties of change operate is itself somewhat subject to typological restriction in space and time, hence less universal than is thought. We sum up some of our argument by recommending that in weighing the merits of a given reconstruction these two factors be taken into account: the leeway that exists for language structure in terms of the area and the period insofar as these are known; and the available replacement machinery, also in terms of areal plausibility.

4. There are three contexts in which we might speak of phonetic reconstruction. First, we may be asked to make pronouncements about the phonetic properties of a language extant in the form of written records. Secondly, our task might be that of reconstructing such properties from a known later state of the 'same' language. The third task is the most elaborate of all, namely that of retrieving detail from material which is linked with the language in question by the hypothesis of common descent. Like many distinctions, this threefold division is somewhat impure; in particular, it does not neatly partition the processes of history to which the three procedures are applied, since each procedure may be used in conjunction with the other two. Still, as procedures, they are, to an extent, separate and remain separable.

The simplest observations of sound change, and consequently the simplest reconstructions of sound are made upon written records. In the favorable case a descendant form of the language is known, preferably from scientific present-day observation. Since this is not the place to discuss the problems of decipherment, we may limit ourselves to phonemic, and especially to alphabetic scripts, that is, to those scripts which are designed to render minimum sound. The way in which this is done is rather well known and requires no great elaboration. Alphabets are, ideally, (morpho-)phonemic writings, in which there is a letter to each phoneme as well as a phoneme to each letter. Once a script is recognized as alphabetic, two kinds of question arise: In what respects does it fall short of the ideal of phonemic consistency?; and: What can we know in the way of phonetic fact that goes beyond the matter of phonemic consistency? On the whole, the two questions are approached in quite separate ways.

A script may fail to render certain contrasts in any way. This is often fatal, especially where whole phonemic subsystems are affected. The classical instance are the intonations and many of the stress and "juncture" traits of practically all languages known from alphabetically written records but not fully analyzed from spoken texts; the near-neglect of these features is of course part of the alphabetic tradition. But even where the area of neglect is more capricious it is difficult or impossible to infer the existence of particular contrasts that are never represented in writing. It takes additional information to suspect or discover it: it is of course well known how a knowledge of alliteration, riming, quantitative prosody, etc., can help reveal the presence of unwritten distinctions. Vowel length could be reconstructed for Latin, and correctly assigned to a large number of particular Latin words, in this way, even if this reconstruction could not, as is actually the case, be checked against the same result obtained in other ways. If we suspect that a contrast has been left unexpressed, a knowledge of the history of the script itself may tell us that, short of invention, no means of expression existed, thus making the neglect plausible. The opposite kind of inconsistency is easier to see. Both promiscuous and regulated uses of more than one letter for a phoneme can obviously be detected; if the mutual distribution of Q and C in Latin were not in turn complicated by the two (phonemically different?) uses of V for either *u* or *v* we should have no trouble deducing that QV represents a phonemic *cv*[13]. It is perhaps unnecessary to remind ourselves that this does not mean that the writers recognized, and intended to write, positional variants of phonemes (in this case, the rounded variety of *c* before *v*). In all known instances the complementary use of letters is no more than an analog to the way in which allophones complement each other; its origin lies in the vagaries of writing itself, and it is often an episode in the gradual and mechanical reduction of a superabundance of letters over phonemes[14]. Only by accident may it be expected to correspond to some major allophonic division within a phoneme. Deviations from phonemic writing which attach consistantly to dictionary items may be given away by even small numbers of informal and semi-literate writings where the irregularity tends to be removed.

[13] There is no good reason to see a "unit phoneme" in "qu", if importance were accorded to this very secondary question. See *Transactions of the American Philological Association 80:* 274[18] (1949).

[14] As in the case of the Greek qoppa.

A knowledge of descendant states of the same language introduces considerations which will occupy us later. It contributes most to the interpretation of written records where a finding of existing contrast needs to be confirmed, or where the absence of such a finding leaves us in much ignorance. If metrical [and other] evidence were not available to establish the unwritten contrast between short and long vowels in Latin, it would nevertheless have to be inferred from the contrast which most "descendants" [e.g. Italian] exhibit in the relevant places.

Within the limitations indicated, then, the internal evidence of alphabetically written texts furnishes information on the number of phonemic contrasts, and on the incidence and distribution of contrast. It provides us with the means for a phonemic reconstruction in one very narrow sense of the word. To give body to such a skeleton of phonemic structure points, other evidence and additional considerations are needed. Some of the evidence comes to us along with alphabet lore itself; we know of the use of the same alphabet (alphabetic writing is, after all, historically connected throughout) for other languages, and if part of that knowledge is granted, all kinds of presumptions concerning the phonetic content of the phonemic entities that are represented may be obtained. Familiarity with Greek alphabetic practice would go a long way toward establishing facts about Latin phonemics *and* phonetics even if Latin had to be deciphered without any other outside aid. It is further well known that decipherers make crucial use of notions which they regard as self-evident and which are indeed either linguistic universals or at least solid characteristics of the typology which happens to be involved. Here belong, in particular, certain assumptions on syllable structure, on the relative frequency of phonemes with vocalic and with consonantal phones, on the special properties of utterance-initial and utterance-final phoneme sequences, on the expected frequency of "word" boundaries and its relation to signs (including spaces and punctuations) suspected or known to be word or phrase dividers, and so on. There is, moreover, the evidence (if its availability be granted) of borrowings from and into other known languages with its many intricate phonetic and phonemic aspects[15].

[15] *Sturtevant, E. H.:* Pronunciation of Greek and Latin[2] (Philadelphia 1940), passim.

5. We are now ready to turn to internal reconstruction – a subject which has been treated extensively in recent years[16]. Internal reconstruction is based on the grammatical alternations between phonemes (on the morphophonemics) of a language state. It aims at the recovery of the processes which have resulted in the alternations. In a manner quite similar to what is true of the interpretation of alphabetically written records, it yields primarily the phonemic shapes, or approximations to the phonemic shapes, of morphs and morph sequences as they must have existed before the operation of the processes in question. Only incidentally and secondarily does it provide more strictly phonetic information. In addition, it has other limitations.

The principles of internal reconstruction may be derived rather simply from the theory of sound change as outlined earlier, as well as from a handful of simple assumptions governing morphs and morphemes. There is a superficial, partly valid and partly deceptive analogy between the allomorphic structure of morphemes and the allophonic structure of phonemes. Thus, there are positional allomorphs which are non-contrastive (e.g. NHG *gut: bes–* [in *besser*, *best*]; NHG */vált/ : /véld–/* [*Wald:Wäld(er)*]) in the sense that neither of the two[17] forms in partnership ever occurs surrounded by the same morphs. Sets of allomorphs are of two kinds: suppletive or morphophonemic. The latter are distinguished from the former by the fact that the phonemic differences between the co-allomorphs recur in other sets of allomorphs; /a/ alternates with /e/ in *arg: Ärger*, *alt: Eltern*, /t/ with /d/ in *Feld: Felder*, *lud: laden*, whereas no comparable statement could be made about *gut: best*. It may be taken as axiomatic that suppletion is the effect of some grammatical process (often against a background of syncretism where formerly contrasting morphs develop a complementary and hence allomorphic distribution[18]) or else the last remnant of a sound change which then, in its very isolation, is already at the point at which it ceases to be useful as a datum from which to reconstruct[19]. Morphophonemically related allomorphs, on the other hand, are the typical remnants of sound change.

[16] *Chafe:* Language *35:* 477–495 (1959); my note in Language *36:* 191–192 (1960); and *Kuryłowicz* in the Proceedings of the Ninth International Congress of Linguists (ed. *Horace G. Lunt*), p. 9–36, give the most recent treatment.
[17] Or more.
[18] *LCLR*, p. 48f.
[19] Ibd., p. 49.

In order to recover the change process it is necessary to examine the terms of the phoneme alternation as it recurs in the relevant sets of allomorphs. Some alternations are *irregular*, in the sense that the environments which determine the choice of one allomorph over the other (or rather: of the allomorphs with one of the two alternating phonemes rather than those with other) require the naming of specific morphs. The alternation of NHG /a/ with /e/, where it occurs at all, can only be stated in terms of the grammar of inflections and derivations. Other alternations between phonemes are *regular* in the sense that where they occur at all, their distribution can be described in terms of (often very small) phonemic stretches in the environment which are not coextensive with morphs. Thus, where a Latin *r* does alternate with *s*, we find *r* between vowels and *s* otherwise[20] (*gerō: gestus*, *corporis: corpus*); in *Wald: Wälder* and in all the other instances of *t: d*, *d* occurs before vowels, *t* before pause, etc.[20]. But between the Latin and the German example there is a further significant difference. The phonemic system of Latin allows both *s* and *r* in either kind of environment, intervocalic and otherwise (*gerō nisi mīsī; gestus certus*), whereas the rules of German phonemics exclude *d* precisely from those surroundings in which the allomorphs with *t* appear instead of those with *d*. In other words the alternation *t: d* is not only regular, that is, conditioned; we may say that the phonemic system imposes it (provided that we designate *d* as the "basic" partner). The alternation is a *compulsory* one. The play of *r* and *s* is not compulsory; it is, so far as it is stated here, merely regular.

There are good reasons why it is not always feasible to classify an alternation unambiguously along these lines; one of the limitations of our work lies here[21]. But under reasonably favorable conditions the classification can be valuable. This is so because compulsory alternations are the typical results of certain sound changes; because certain additional sound changes will transform a compulsory into a non-compulsory alternation; and finally because it is also possible to name the type of sound change which produces irregular alternations.

What is fundamental in all these cases is the simple circumstance that a split will lead to an alternation if some of the instances

[20] Simplified.
[21] On this and on the effects of "analogic change", see *LCLR*, pp. 99–111. In connection with the mainly negative implication of the term "compulsory", see *Wells*, Language *25:* 99–116 (1949).

of $a[b..$ and $a[d..$ on p. 28 above involve a given morph containing a so that a morph boundary passes between a and its significant environment, whichever it happens to be. If the split satisfies the description $a[b.. > m[n..; a[d.. > p[n..$ (that is, if b has merged with d), the result will be an irregular alternation between m and p; the voiceless and the voiced "th" in *wreath* and *wreathe* alternate irregularly inasmuch as the statement can only be given in grammatical terms ("noun:verb", or "underlying noun:verb derived with zero-suffix"); this has come to pass because the differences in the original conditioning environment (the presence vs. absence of a following vowel in the verb ending) have been merged into "0". If, contrariwise, the split is of the type $a[b.. > m[n; a[d.. > p[o..$, m will alternate with p in such a way that m occurs in the environment n, and p in the environment o – that is, in regular form. That this regular alternation is moreover likely to be compulsory follows from a corollary of the split process. There must be, outside the area in which the alternations arise, instances of m in some environment in which p also occurs; or else m and p are not in contrast, hence not separate phonemes. This will in particular happen if there are other sources for one of the two, say for m, in all or in part of the environment o. At this stage, then, we shall have m rather than p in those environments in which m is permitted but p is not (namely, in the environment n), while the converse is not true.

For the reason sketched here it is possible to infer that some p's in the environment o (itself reconstructable as d) replace a, while a in environment b has gone to m. If other replacement processes create instances of m in environment o as well, they will there contrast with p and thus contribute to eliminating the compulsory character of the alternation created by the earlier process but they will not thereby make it useless for the purpose of recovering it. Thus, pre-Greek s between vowels goes to 0 (thereby merging with other ø's [e.g. from y]); between vowel and word-end it becomes s, a phoneme which for a while does not occur between vowels. As a result, *génos* 'family' alternates with *gene0–* (in, say, *genéōn* g. pl.) in a regular and compulsory manner which permits us to infer that a phonemic entity of some sort split into s and 0 under statable conditions. When later *tw*, which occurs between vowels, is replaced by s, the earlier alternations become in part non-compulsory but lose nothing of their indicative value[22]. The principle, in brief, must be

[22] See footnote 21.

to give chronological precedence to alternations from phoneme to phoneme over the later bonds which only hold the constituent phones of a phoneme together[23].

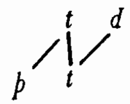
It is a weakness of internal reconstruction that it is silent on pure (unconditioned) merger processes and that it has sometimes contributed, by its selectivity, to giving the wrong impression that all language becomes more complicated morphophonemically as time goes on[24]. It shares a further weakness with our procedures for the interpretation of written records: namely that of providing, in principle, proto-forms in their phonemic makeup but without physical detail; it gives us information about the incidence of contrast but none about the nature of the contrasting phones. We have been careful to stress this at the cost of adhering to a complex, abstract formulation in which no unwarranted identification between the phonemes of the extant stage and those of the inferred stage is surreptitiously introduced. This is necessary because, as we know from our earlier discussion, a study of the relation between the two must be independent of the replacement formula.

Suppose that we have two stages with a trivially simple replacement relation obtaining between the two: all $a > m$, all $b > n$, ... It is normal to expect the physical characteristics of a and of m to be in some sense "the same", and likewise for all other pairs. It is, however, also clear that this is not necessarily the case; there may be a "shift". Thus, in the popular simplified version of the Germanic consonant shift, Germanic *þ, t, d* replace IE *t, d, dh* one-to-one, and yet with a difference which has importance both as a collection of isolated physical fact and also with reference to a more or less stably continuing typological framework. It is an open question to what extent such a displacement could be inferred simply from a later stage. On the other hand it is also doubtful that major shifts ever go on in complete purity; just because some of the properties of the framework within which the shift occurs are best thought of as deep-seated areal (or general) characteristics, the shifting of, say, a stop (in all its environments) to a spirant location in the framework may lead to a situation where all the occurrences of the stop cannot be accommodated precisely because some of them are characteristic of stops rather of spirants under the prevailing areal

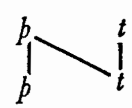
[23] But involve (at least prior to analogical action) different lexical material in each partition.

[24] *Chafe*, p. 495.

type. Therefore, we may use the occasionally observable (and internally reconstructable) minor conditioned sound changes on the fringes of seemingly simple replacement processes as phonetic indications. In fact, the Germanic change does not just shift *t* to *þ* and *d* to *t;* it also lets *st* go to *st*.



In this situation it is possible to argue in reverse, with a great deal of typological concreteness, that the "*a*" which split up into *þ* and *t* had [t]-like rather than [θ]-like phones, because it is less likely, from what we know about the languages of the family, area, and period, that [sθ] existed or if it existed should have become [st], than that *t* should have been spirantized except when it followed *s*. Another possibility, namely that "*a*" had always had two clearly different positional variants, *t* after *s* and *þ* otherwise, runs into another typological objection: the other phoneme, viz. the one which ends up as Germanic *t* is also known to have occurred after *s*, in contrast with "*a*", as in the words NHG *Nest, Ast*, from IE *sd.* Such is the nature of the argument against a phonetic interpretation like this:



It seems, in general, that it is fruitful to distinguish those split-cum-merger processes in which the splitting and the merging phoneme are homologous (as they are in the case of IE and Germanic *t*) from those which are not (as in the case of the unvoicing of German *d* to *t*, in *Wälder*, above).

The irregular alternations pose a different phonetic problem. These alternations exist between sounds which may have remained substantially unchanged but which used to be allophonically different until the alternative conditioning factors in the environment became merged (*a*[*b*.. > *m*[*n*..; *a*[*d*.. > *p*[*n*..). We may wish to know where these factors were located in the flow of speech and what their physical characteristics were. Satisfactory answers to these questions are often possible, since there seems to be great uniformity in the assimilatory (and otherwise conditioning) action from segment to segment. In the light of our accumulated experience it is reasonable to suspect that the voicing of *wreathe* has to do with an intervocalic environment, or that Germanic umlaut points to front vowels in the final part of the word.

There are situations in which internal reconstruction is the only available avenue to the past. This is true when all we possess is either one synchronic corpus or a reconstructed ancestor language in the antecedents of which we are interested. Considering the limitations of the method the results will be uneven, and more rewarding for languages with elaborate morphophonemic alternations and with histories of extensive conditioned sound than for languages of a different build. Fuller use of internal reconstruction may be made where it supplements the comparative method.

6. The "comparative" method of reconstruction utilizes the circumstance that when separate and different replacements operate on one and the same language (thereby breaking up an original community of speech), the recurrent phoneme correspondences which result among the descendant languages must in a majority of cases behave like allophones of a phoneme in being complementarily distributed and "similar". Suppose that $a > m$, $c > n$ in one descendant, (I), while $a > t$ as well as $c > t$ in the other descendant language, (II)[25]. The effect will be that two correspondences are recorded, *m*/*t* and *n*/*t*. These are "similar" in sharing their language II component (namely, *t*) but they contrast in the sense that they can be preceded and followed, in the morphs in which they occur, by the same correspondences. Thus, Low German *d* corresponds to both High German *t* and *d*, but both correspondences occur in initial antevocalic position (as in the words for *du* and *tun*), thus testifying to two separate entities in the common (West-Germanic) ancestor and to the merger of these in LG. This amounts to saying that no condition can be named under which an alternatively reconstructed single entity would have split up in HG. On the other hand, a LG *t* corresponds both to HG "*z*" *(zehn)* and to HG *t (treu, stehen, Stroh)* so that the two correspondences are once again "similar". They are, however, not in contrast, since *t*/*t* occurs after *s*/*š* or before *r*/*r* or under both conditions combined, where *t*/"*z*" is missing. Consequently, a single phonemic source may be posited, along with a conditioned sound change in HG. This method of

[25] Note that, contrary to appearance, the notion of descent is derived from the comparative method itself: if two related languages, III and IV are subjected to the comparative method, and the result is "identical" with III, III is itself the ancestor of IV (*LCLR*, p. 145f.).

inference is far more powerful than internal reconstruction: it is not dependent on the special nature of the morphological structure of the language, nor does it break down before the task of retrieving a merger process. Its two principal drawbacks are these: it will naturally miss the independent duplication of a merger (or of one portion of a merger) on the part of both descendant languages ($a > m$, $c > m$ in I, and $a > t$, $c > t$ in II; or similarly for conditioned changes); and there will be difficulties when the same correspondence arises in two unconnected ways. (IE *p* > Germanic *f* but > Germanic *b* after unaccented vowel, IE *bh* > Germanic *b;* IE *p* > Greek *p*, IE *bh* > Greek *ph* but > Greek *p* if an aspirate begins the following syllable. The result is that the distribution of the correspondence Germanic *b*/Greek *p* is not, as a whole, complementary with that of either *f*/*p* or of *b*/*ph*.) Both these difficulties are diminished as more than two languages, or pairs formed from a collection of more than two languages, are examined.

Once again, the immediate result is what some scholars feel a disembodied system of contrasts endowed with but little substance, and once again we must look for the considerations, concealed or explicit, which lend concreteness to a proto-language[26]. In part these considerations have already been taken up in connection with our other methods. The same subtle balance between a belief in universals and a recognition of areal type plays a role in assessing plausibility. Where requirements are made explicit they have occasionally been overly severe. Authors who insist that the ancestor structure must resemble that of the descendant forget that we are often faced with a history in which all or most descendants are also members of an area and thus subject to parallel pressures away from the type represented by the proto-language. If they have responded to these pressures with somewhat varied mechanisms, we are lucky, because this is what keeps the original structure within our reach. The simplest examples are those in which the areal trend is toward a smaller phonemic system or subsystem. Almost all the Semitic languages have reduced their sibilant contrasts. The fact that they have done so in different ways has brought it about that the number of contrasting correspondences for the lot exceeds the number of sibilant phonemes in each separate descendant. The requisite reconstruction is then different, in this small respect, from the type

[26] Language *35:* 410 (1959); *Allen*, TPS 52–108 (1953); *Pike, K. L.:* Axioms and procedures (Revised Edition, Glendale 1957), passim.

which prevails in the family, and, being different, has been suspect to some critics. The suspicion is allayed when South Arabic, situated at the margin of the Semitic world, turns up with the same aberrantly high number of sibilants. Similar experiences are frequent enough.

The old question whether asterisked proto-words are "only formulae for observed correspondences" – an alternative which has never been taken quite seriously – or have some "reality" turns, not perhaps entirely but still to a large extent, on the phonic content of the reconstructed phonemes. Having a broader basis, the "comparative" approach is a little better off than the more elementary forms of guessing at the linguistic past. Rightly or wrongly there is an inevitable inclination to believe that a phonetic consensus among the daughter languages establishes a phone as ancestral. On this basis, nobody would doubt that proto-Romance had a voiceless fortis unaspirated bilabial stop for its **p*, and what we know about Latin from other sources confirms this. In cases of discrepancy there is sometimes a tendency, other things (like intrinsic phonetic plausibility) being equal, to abide simply by some kind of majority count, where what should also be considered is not only the size but the relative geographic position of the aberrant phenomenon. This presupposes, of course, that the descendants have not already been shown, by the comparative method itself, to be related through some kind of definite sub-ancestry. If this *is* the case, an innovation, even just a suspected phonetic (sub-phonemic) one which would not by itself contribute much to the task of establishing sub-ancestry in the first place, must of course be counted as having occurred only once (in the sub-ancestor), regardless of how many descendants from it are extant.

The breadth of a "comparative" foundation may support a fair knowledge of positional variety within the phonemes or distinctive components in the proto-language. After all, the lines along which the daughter languages let the phonemes break up under conditioned sound change must correspond to earlier allophonic groupings within the phonemes. They may of course be of different age, and it would be poor method to project all of them back into the ideally uniform ancestor. Yet some of them must be ancient. IE aspirates, if followed by another aspirate in the next syllable, lost their aspiration both in Greek and in Indic (i.e. in those two languages in which aspiration is an active distinctive feature). While each descendant accomplished this by an entirely separate merger, it is easy to believe that aspirates had untypical allophones when occurring before a syllable with another aspirate. Another extremely instructive example is furnished by "Sievers' Law" (including the doctrine of the so-called schwa secundum) as it now appears to some Indo-europeanists. Its essence is that an automatic, non-phonemic element of syllabicity would crop up in sequences of more than two consonants after every two consonants (*ere epre*, but /*erpre*/[erp$_{\text{v}}$re]). These predictable, hence non-distinctive supports had a way of merging with otherwise existing, non-predictable phonemes in the individual IE languages (though not with the same ones in all each languages). Thus, *erpre* [erp$_{\text{v}}$re] appears as *arpura* in Sanskrit as does an IE *erpwre;* in Attic Greek *erpre* merges with IE *erpHere* (or something of the sort) into *erpare.* A correspondance *u*/*a* is established; as it is found to occur predictably in consonant settings it must be classified as a non-phonemic feature in the ancestor.

7. To return to the topic of this congress, we must admit that reconstruction does more for "The Phoneme" than for "Its Realization" as it existed in the dim past. But nobody will say, these days, that it is possible to keep the two apart. Even in a historical context we cannot possibly deny our constructions and reconstructions their physical, phonetic substance.

Author's address: Henry M. Hoenigswald, 23 Bennett Hall, University of Pennsylvania, *Philadelphia 4*, Pa. (USA).

Discussion

Martinet (Paris): Le texte de la communication de M. *Hœnigswald* pourrait faire croire que la conception de changements phonétiques dirigés vers un but (*goal*) fait partie des principes explicatifs qu'il m'attribue à juste titre. Je voudrais rappeler que la conception de l'économie des changements phonétiques que je préconise n'implique aucune téléologie, mais une succession de causes et d'effets.

Proc. 5th int. Congr. phon. Sci., Münster 1964, pp. 43–60
(S. Karger, Basel/New York 1965).

Code Theory and 'Discrete Mathematics' in Phonology

By Henning Spang-Hanssen, Copenhagen

Language is public property. This is true of language as an instrument of thought and communication and of language as a source of emotional, e.g. aesthetic impressions. However, it seems to hold true of language even as an object of research. Apparently there is no end to the list of disciplines and sciences which lay claim to language as a genuine part of their field of study. Among other things language has been appointed a set of social habits, a kind of individual behavior, a logical – though perverted – structure (a calculus), a hierarchy of mathematical relations, and a – rather irrational – code. Accordingly, linguistics has been considered part of anthropology, psychology, logic, mathematics, and information theory or communication theory.

Whether language *is* a behaviour or a code, etc., and whether linguistics *is* part of logic or mathematics, etc., are to a considerable extent analytical questions, to be answered differently according to adopted definitions of the terms in question, including the word *is*. Thus the various statements concerning the nature of language may all be true at the same time, on condition that they are individually interpreted as "Language *also* is ...". And in fact it is very unlikely that all empirical linguistic phenomena can adequately be studied within the frames of any one existing discipline or science. With regard to phenomena of language expression this fact has been duly recognized in the naming of the series of congresses, of which this forms the fifth, as Congresses of Phonetic *Sciences*.

It is definitely to be hoped that still new disciplines by laying claim to language will contribute to the study of language. On the other hand, it sometimes happens that a new claim for some time attracts attention and meets with approval to a degree that is hardly motivated by the actual gain of knowledge likely to be achieved from this new point of view. To keep up with fashion is a factor of some importance even in linguistic research. In particular it may often reasonably be asked whether a new claim or approach is anything but a new terminology. This question, though sceptical, may not be depreciatory, as the development of linguistic terminology forms an important part of the development of linguistics and of the relations between linguistics and other fields of knowledge. But in order to further the study of linguistic phenomena new claims and statements concerning the nature of language – or of linguistics – must at any rate describe what is less known ("language" or the like) by what is better known, or expound what is complex and less intelligible by what is simple and more easy to understand. In this respect the above statements, as usually presented, leave much to be desired.

"*Linguistics is a discrete mathematics*"

Our present concern is with the claim that language is (also, or partly) a code, and the claim that linguistics is (also, or partly) a discrete (discontinuous) mathematics [cf., for instance, *Martin Joos*' paper in JASA *22:* 701–708 (1950)]. These statements are related to some extent, cf. later; at the outset, however, they will be discussed separately.

As regards the relationships between linguistics and discrete mathematics I may refer to the discussion I have given in a paper read to The 9th Internat. Congress of Linguists (Preprints Cambridge, Mass, pp. 133–138, 1962; Proceedings 724–730). My main points are the following:

1. Qualitative linguistics (as distinct from research including statistical or other quantitative aspects of language) may be called mathematical, in the sense of *axiomatic* (i.e. making use of some explicit model containing axioms or postulates). But as various axiomatic models since long have been developed within linguistics proper, independently of mathematics as a particular science, an extension of the term 'mathematical', so as to cover what is usually called structural linguistics, only seems to blur a useful distinction (between mathematical and axiomatic) by putting a "mathematical cross" on every axiomatic model in any field of study. Thus in this respect no clarity is achieved by describing linguistics as a (discrete) mathematics.

2. The statement "Linguistics is a discrete mathematics" may be read in a *normative* sense, claiming that linguistic models ought to be models developed within mathematics. The part of discrete mathematics most often suggested to deal with linguistic matters – at least in the field of phonetics and phonology (or phonemics) – is *set theory* (Mengenlehre, teorija množestv, théorie des ensembles). In fact this discipline is often regarded as part of logic (symbolic logic) or as (part of) mathematical logic or logical algebra.

Set-theoretical models in phonology

Detailed phonological (phonemic) studies applying set-theoretical models are those of *Harary* and *Paper* [Language *33:* 143–169 (1957)] and of *Sigurd* and *Gårding* [Studia Linguistica, pp. 8–34 (1955)]. These papers lack nothing in clarity, and they form important contributions to the particular linguistic matters dealt with. But at the same time they clearly illustrate the limitations of set theory. In my above-mentioned paper I have pointed to certain inadequacies of set-theoretical models in relation to linguistic matters:

a) The notion of *order* is a fundamental in set theory and in the adjoining theory of relations. Thus a fundamental distinction is made between symmetric relations (i.e. sets of combinations in which for any pair of elements x and y both the combination xy and the combination yx are found) and asymmetric relations (in which at least one combination xy has no counterpart yx). But since linguistic conditions of combinability are in important cases neutral as regards order (cf. the combinatory difference between vowels and consonants, or relationships such as concord and government), models introducing order as a necessary notion may have the effect of a strait-jacket. No doubt set theory can be modified in this respect, but not without consequences to other parts of the theory (in particular to the relational property called transitivity), and in any case this modification forms an illustration of a necessary adaptation to linguistic conditions.

b) More serious inadequacies of set-theoretical models are due to their *synthetic* nature. In set theory a set is defined by its elements, and a set of combinations by the collection of *given* pairs of elements. Accordingly, a set-theoretical model may well serve the purpose of describing empirical phoneme combinations by their constituent phonemes, but it is not particularly suited for classificatory analysis of phonemes in terms of combinatory possibilities (combinability). "It is our aim to provide a technique for describing and quantifying phonemic interaction, thereby sharpening the concept of distribution. In none of this is there any claim that analysis will be aided, although, to be sure, we know of no other adequate technique for handling distribution as rigorously or as effective." [*Harary* and *Paper*, Language *33:* 145 (1957).]

Whether distribution may be handled by other techniques, is partly a question of what is understood by *distribution*. By reference to my report on this subject to The 8th Internat. Congress of Linguists (Oslo 1957) it may be said that when distribution is regarded as a basis of classificatory phonemic analysis, other models, in particular those based on the glossematic concept of relation (or function), seem to be more adequate than set-theoretical models. It is interesting – and promising indeed – that the glossematic typology of relations may be mapped on the set-theoretical properties of relations, and vice versa. As an illustration one may imagine an inventory of "syllables" or "word expressions" like that given in the diagram.

	b	d	f	g	h	a	e	i	o
b						ba	be		
d						da			do
f							fe	fi	fo
g							ge		go
h								hi	
a									
e									
i									
o									

In set-theoretical terms this inventory forms an intransitive relation: Two elements (e.g., *a* and *e*) that are found in combination with one and the same element (*b*) are never found in mutual combination. In glossematic terms this inventory is analyzable by means of the relation (function) called solidarity, in that two solidary classes (categories) of constituent parts are recognizable: A member of the class *b, d, f, g, h* never occurs without being accompanied by a member of the class *a, e, i, o* – and vice versa.

In this particular example it seems a pure matter of taste which model to prefer, but in cases of partly transitive sets (e.g., if even the combination *ae* has occurred) the latter model is superior as regards classificatory power (cf. my "Probability and Structural Classification", Copenhagen 1959), and moreover it permits handling the question of accidental gaps versus excluded combinations.

c) With regard to this and other questions of chance set-theoretical models are unsuitable – in accordance with their nature of being defined by *given* sets of combinations. This fact forms a serious restriction to the application of such models in phonology and, considering the development of linguistic analysis towards problems of grammaticalness and generative grammar, even in other linguistic areas. Models of quite a different nature are needed to deal with the prognostic, extrapolating, and generalizing aspects of linguistics, including even the characterization of language as such. "Woher aber kommt dieses *Geltenmüssen* für nicht untersuchte Sprachen: noch nicht untersuchte oder prinzipiell nicht untersuchbare (weil sie nämlich noch gar nicht existieren)? Ein solches 'Muß' ist ein Fremdkörper in der von Bopp, Rask, Grimm begründeten Sprachwissenschaft..." (*Eberhard Zwirner*[19], p. 136).

The considerations relevant to such questions are not to be found in mathematics of the type meant when linguistics is spoken about as a discrete mathematics. With regard to these questions interest is focussed on the calculus of probability and mathematical statistics.

Summing up it may be said that with regard to linguistics set theory is definitely more than a new terminology, but on the other hand the descriptive power of such models is limited. Until now their chief merit is to have thrown light on certain pecularities of *linguistic* problems.

This conclusion seems to be in accordance with the aim of other applications of set theory (or symbolic logic) to phonological problems. *J. Cantineau* [Word *11:* 1–9 (1955)] has compared the set-theoretical notions of relation with *Trubetzkoy's* system of phonological oppositions (Lit. 16, pp. 68–99). *G. Ungeheuer* [Studia Linguistica, pp. 69–97 (1959)] applies logical algebra to "1. die formalen Prinzipien einer binären Klassifikation von Schallereignissen" (in terms of distinctive features); "2. die formalen Prinzipien der analytischen Transkription, die auf dieser Klassifikation aufbaut".

In these papers logical algebra is applied to models already set up within linguistics, in order to formalize and to develop more consistently the bases of classification and reduction (cf. also *W. Plath's* comments – in Lit. 11, p. 39f. – on *Ungeheuer's* paper). No doubt the conceptual basis of phonological description will benefit by such attempts (quite recent contributions are found in "Problemy strukturnoj lingvistiki", Moskva 1963). However, important basic problems are beyond their compass; among other things the description of phonemes by distinctive features gives rise to the problem of where to end the analysis. If the ultimate elements of the expression structure of a (spoken) language are a set of distinctive features, and if structure is defined as (a hierarchy of) rules of combination (or more precisely: of combinability), it will be necessary to account for the empirical fact that the number of phonemes is – usually, and possibly always – smaller than the number of imaginable combinations of features. This can be done *either* by stating rules governing the combinability of features, *or* by describing missing combinations as accidental gaps.

In the latter case the empirical inventory of phonemes cannot be regarded as exhaustive, i.e. as a structural fact about the language in question. If one is unwilling to accept this consequence, i.e. if one insists on regarding phonemes as structural units, the distinctive features cannot be preserved as structural elements of the language in question; they may, however, be regarded as elements belonging to a certain structure of expression *manifestation*, and as such they may be relevant to more than one empirical language.

Which one of the alternative ways of description to be chosen, i.e. whether to carry on the structural analysis of a language expression below the level of phonemes or not, will mainly depend on the actual possibility (for the language in question) of ascertaining rules governing the restricted combinability of distinctive features. This is a problem of the type accidental gap versus excluded occurrence, and – as mentioned earlier – set theory does not provide tools for solving such questions.

The problem just discussed is reflected in the following quotation from *Jakobson* and *Halle*[4] (p. 217): "... this code includes all the distinctive features to be manipulated, all their admissible combinations into bundles of concurrent features termed 'phonemes', and all the rules of concatenating phonemes into sequences...". It will be seen that no mention is here made of *rules* of concatenating or combining the distinctive features into phonemes.

Another point to notice in this quotation is the use of 'code' to mean linguistic inventories and rules. This brings us to the discussion concerning language and code.

What is a code?

All of us should like to know what language is, but if we look for an answer in the statement "Language is a code" we shall be disappointed. For opinions differ as much about the meaning of 'code' as about the meaning of 'language'. Since this paper deals with code in relation to language (in particular: to language expression), an attempt at a general survey of code theories shall not be made. With regard to linguistic applications it is hardly possible to distinguish between communication theory and information theory; in fact, only a few experts seem to manage this distinction (cf. the similar situation of logic and mathematics sharing set theory).

Code is a basic concept in information and communication theory. It might be regarded as an indefinable, and in fact no definition of 'code' is found in *W. Meyer-Eppler's* detailed and comprehensive "Grundlagen und Anwendungen der Informationstheorie" (1959). However, in other works various types of definition or characterization of 'code' are met with:

I. "Quand on parle de code, nous pensons souvent à des secrets ou à des intrigues internationales, mais dans ce livre nous emploierons ce mot dans un sens beaucoup plus général. Tout système de symboles qui, par convention préalable, est destiné à représenter et transmettre l'information entre la source et le point de destination sera appelé un code. Ainsi, en ce sens, la langue française est un code et la langue allemande un autre." *G. A. Miller* (p. 14 of the French edition of "Language and Communication"). According to this kind of definition the statement "Language is a code" becomes trivial: it holds true because code has been defined so as to include language. But with regard to linguistic research this conception of code amounts to the non-trivial hypothesis that language and (other) symbol systems having a communicative function may profitably be studied together. Or in a normative version: Language and (other) communicative systems of symbols ought to be studied together.

II. A similar, yet different conception of code is connected with the view that language should be studied together with (other) symbol systems not because of a common communicative function but because of common formal features. In other words: Language and certain (probably not all) symbol systems are assumed to be of the same internal nature, and 'code' may be used as a designation of any system of that nature. Opinions differ, however, as to the nature of this nature: Which formal features or characteristics are to form the basis of such a notion 'code'?

a) The characteristic of a code may be sought in the existence of rules governing the combination of elements; thus *combinability* (sočetaemost') is the basic feature. Cf. by *E. V. Padučeva*[10] (pp. 114, 115 transl.): "The resemblance of language to a code is above all based on the fact that a description of the combinability of elements plays an important part both in technical codes and language... The description of a language from the point of view of the combinability of its units will be labelled 'description of a language as code'." As she also remarks, this conception of code basically coincides with a description in terms of distribution. Thus, as regards the qualitative aspects of distribution, reference may be made to the earlier discussion on mathematical (set-theoretical) models applied to phonology; the more so, as in information theory, etc. the description of combinatory conditions of codes (linguistic and non-linguistic) usually is given in such mathematical terms.

In information theory, however, the qualitative description is supplemented by quantitative, in particular statistical points of view, and therefore the above-mentioned conception of code includes the hypothesis that the *quantitative* aspects of combinatory conditions in language may profitably be studied together with combinatory conditions found in (other) symbol systems. Cf. later.

b) In discussions concerning the nature of 'le signe linguistique' attention is often focussed on *arbitrariness of signs* as a characteristic of language and of certain other sign systems. *Saussure*, having suggested "une science qui étudie la vie des signes au sein de la vie sociale ... nous la nommerons sémiologie"[14] (p. 33), describes the main subject of this science as "l'ensemble des systèmes fondés sur l'arbitraire du signe"[14] (p. 100).

Besides language (la langue) he mentions among other examples the system of military signals as belonging to the field of semiology. Even though the designation 'code' was not used by *Saussure*, certain aspects of modern code theory (or theories) have a striking resemblance to his idea – which is not tantamount to saying that code theory has been influenced by *Saussure's* suggestion. Among linguists *Saussure's* idea of a semiology has not found wide

acceptance; the designation has survived, in particular in works by *Buyssens* and by *Hjelmslev*, but there it covers other, and divergent, conceptions as to the basis on which language ought to be studied together with other sign or symbol systems (a more detailed discussion is found, e.g., in my "Recent Theories...[15]).

c) Since all the technical codes in which modern code theory originates are systems of arbitrary signs or symbols, arbitrariness does not form a particularly interesting formal feature from the point of view of code theory, and the claim for language does not in particular seem to be motivated in the arbitrariness of language signs. On the contrary, language (together with certain sign systems) may be found terminologically distinguished from code just by reference to difference of *origin*, in some way related to the question of arbitrariness; "... we distinguish sharply between *language*, which is developed organically over long periods of time, and *codes*, which are invented for some specific purpose and follow explicit rules". *Colin Cherry*[1] (p. 7).

However, this basis of distinction does not seem important to the question of whether language may profitably be studied together with other sign systems. For an organically developed sign system, e.g. the decimal cipher system, may well be studied together with "invented" systems, e.g. the binary or the tetradic cipher system.

Speaking of a binary code it shall be mentioned that the present paper does not deal with the much-debated question whether linguistic description – or if one prefers: the linguistic metalanguage or metatext – is profitably constructed in terms of some particular code, e.g. a binary code. The discussion in this paper is confined to the question of to what extent the study of language phenomena can be furthered by comparing language with other systems of symbols or signs.

III. According to still another conception a code may be defined as a communicative system manifested in a particular way (i.e. in a particular expression substance, or by a particular use of expression substance). This definition or characterization covers among other things flag codes and secret codes expressed by ciphers or by letters occurring with unusual values; moreover, certain ways of symbolization introduced for scientific, technical, or commercial purposes.

By this kind of definition nothing is said about the structure of the system manifested; hence a code (in this sense) may be identical as to structure with some sign system manifested in a usual or "normal" way, for instance a language. And since by definition the manifestations of such codes deviate from those of normal languages, the study of language expression is not likely to be furthered by the study of codes in this sense – exception made for the very fact that one and the same structure may be manifested in different ways.

A variant of this conception of code is to characterize a code as being *secondary* to some other communicative system of signs or symbols. This kind of definition is also discussed by *E. V. Padučeva* (op. cit., p. 115). But apart from facts of historical origin the question of what is secondary and what is primary does not basically apply to systems but to *texts* (messages) or to a particular *process* of communication. For instance, a secret code or a technical code will be secondary to the original message (in plain text) with regard to the process of encoding, but the reverse holds true with regard to the process of decoding (deciphering).

Here, and in general, the linguistic distinction between text (utterance, message) and system (language) proves essential to the application of code theory to linguistic matters. In the literature on information or communication theory the word code is not rarely used indiscriminately in both ways, thereby causing confusion about the nature of coding as a link in a process of communication. A text may be converted into another version, but not into a system; thus a message cannot be converted into a code in the sense of a system.

IV. The confusion becomes even greater since in the relevant literature 'code' is often *defined as a transformation*, 'eine Zuordnung', or the like, i.e. as a system for converting messages, whereas the *examples* of codes given by the same authors point to the conception of a code as a system of symbols or signs. "... a code is an agreed transformation, usually one to one and reversible, by which messages may be converted from one set of the signs to another. Morse code, semaphore, and the deaf-and-dumb code represent typical examples." (*Colin Cherry*, op. cit., p. 7.) "Unter einem Code versteht man eine Zuordnung zwischen zwei Listen von Zeichen oder Zeichenserien; ... Ein ... Code ist das indoarabische Zahlensystem, das allen möglichen Zahlen eine Serie zuordnet, die aus den 10 Ziffern von 0 bis 9 entnommen ist; ... Das normale Alphabet ist ein Code mit einer Liste von (z.B.) $k = 26$ Zeichen." (*Heinz Zemanek*[18], p. 30.)

Here a clear distinction between code, coding (encoding, decoding, recoding), and the result of coding would be useful. At any rate rules for converting from one system to another must not be identified with the systems; the system (the rules) used for converting a decimal number (say 9) into a binary number (viz. 1001) is not identical with the system of decimal numbers, nor with that of binary numbers; rules for transcribing written Russian into a sequence of Latin characters do not form part of the Russian alphabet (or graphemic structure), nor of the Latin alphabet.

Language can hardly be regarded as a code in the sense of a system of rules connecting one system of signs with another system of signs. It is worthy of mention, however, that *L. Zabrocki*[17] (pp. 64 to 73) interprets the structural hierarchy of any language as a code, i.e. as a coding process proceeding from 'Lautgefüge' to 'Wort', from 'Wortgefüge' to 'Satz', etc. He expressly points to the duality of his concept 'code': "Der *Sprachkode* ist im Grunde genommen ein Transponierungsprozeß. Er enthält zugleich die Gesetze der Transponierung" (p. 73). It is difficult to see, however, whether *Zabrocki's* theory of language as a code is compatible with the conceptions of code developed by information theory etc.

Processes in which two linguistic systems of signs take part are sometimes talked about as being of the same nature as the transformations effected in technical coding processes. "According to our definition, transforming a printed message into Morse code, transliterating from the Cyrillic to the Roman alphabet, enciphering for cryptographic purposes, and replacing decimal numerals by binary numerals belong to one family with translating *Macbeth* into German ..." *A. G. Oettinger*[9] (p. 104). Since, however, the crucial point of machine translation is whether the relation between texts in different languages can actually be substituted by a set of correspondences of the one-to-one type characteristic of technical coding, it seems premature to regard translation on a par with obvious coding processes like transliteration. Cf. also *A. A. Reformatskij's* criticism[12] (pp. 208–215) of the term 'recoding' (perekodirovanie) used of the relation between normal and tactile language as well as of e.g., transliteration.

Phonological applications of code theories

Summing up this discussion on the ways in which code is conceived with regard to language, it may be said that for the time being the notion of 'code' is too complex or too vague to serve as a basis of a uniform study of linguistic matters. Accordingly, papers presented as applications of code theory to language may have little in common – except for the very term "code".

In certain papers code theory seems to be little more than a new terminology applied to considerations built on combinatorics – a discipline known through centuries, and in a variety of sciences (classical probability, genetics, etc., even linguistics) – and to reflections on efficiency, e.g. in linguistic change, previously discussed by linguists (among others by *Otto Jespersen*). It should be noticed, moreover, that information theory etc. usually narrows down the problems of linguistic efficiency to the question of how texts (messages) are efficiently *communicated*. This narrowing down is quite natural – and in fact necessary – from the point of view of the technical applications (telecommunication) in which information theory originates.

Due to its technical perspectives information theory is *normative* and *evaluating:* Code systems are studied with the aim of achieving the greatest possible efficiency, such as saving time or equipment, and as avoiding disturbances. This kind of approach is extremely important, but it does not exhaust the study and the description of language expression, not even with regard to questions of efficiency. Conditions of *acquiring* a code or a language (in childhood or by conscious learning) form a different basis of evaluating expression systems; and basically different from the question of transmitting given information is also the question of permitting new information to be formed and expressed.

The latter question is, among other things, bound up with the possibilities of introducing new word expressions, e.g. admissible but hitherto not exploited syllables, on the basis of a given inventory of phonemes. The possibility of creating new words, and in general of creating new combinations of elements and units, seems to be a characteristic of natural languages. From this point of view the conditions of language expression prove to be more complex than can adequately be dealt with in notions such as 'rationelle Sprache' in the sense suggested by *W. Fucks*. "Wir betrachten eine Sprache (im allgemeinsten Sinn des Wortes), in der zusammenhängende Symbolaggregate (Komplexionen) von maximal n Elementen vorkommen. Diese Komplexionen sollen *Wörter* genannt werden, ohne daß diese Bezeichnung mehr als eine formale Ähnlichkeit mit den Wörtern einer Nationalsprache ausdrücken soll. Ein erstes Charakteristikum einer solchen Sprache ist ihr *Wortvorrat* oder *Vokabular*, d.h. die Zahl der aus ζ verschiedenen Symbolen durch Bildung von bis zu n-stelligen Komplexionen zu gewinnenden verschiedenen Wörtern. Eine Sprache, bei der diese Wortbildungsmöglichkeit restlos ausgenutzt wird, heiße nach *Fucks* eine *rationelle Sprache*." *W. Meyer-Eppler*[7] (p. 86).

Due regard should be payed to the reservations found in this quotation, but on the other hand it is hard to see how code theory may at all contribute to the study of language structure, if a formal similarity (eine formale Ähnlichkeit) is not considered a sufficient basis. At any rate the similarity has been deemed close enough to justify an adoption of the designations Wort and Sprache.

But in the first place it applies that what according to the above definition is rational, may in another respect appear irrational; this fact is duly recognized elsewhere in information theory, stressing the importance of redundancy as a safeguard against mistakes. Secondly, it should be noted that the failure of a natural language to fulfil the conditions for being 'eine rationelle Sprache' may be due to factors of two different kinds: It may be due to properties of the structure (in that certain combinations – 'Komplexionen' – are structurally excluded), or it may be due to conditions of *usage* only (in that not all admissible combinations have empirically occurred as word expressions).

It is true that in a number of applications of code theory this kind of difference is taken into account; various papers of this nature form important contributions to the qualitative and quantitative description of phonemic and graphemic conditions in various languages. But it may well be asked, whether the linguistic perspectives of these contributions exceed the implications of earlier approaches to phonology (or phonemics) and to phonological statistics; in the latter field one may in particular think of papers by *V. Mathesius*[6] and by other Czech phonologists, cf. the survey by *J. Krámský*[5] in Phonetica.

In various respects the descriptive power of code theory with regard to phonology is obviously hampered by pecularities of technical codes. In discussing the question of functional load (exploitation, Belastung) in phonemics, *J. Rischel* [in Statistical Methods in Linguistics *1:* 13–23 (1962)] points to certain difficulties of a description in purely sequential terms, among other things to "the well-known fact that certain contrasts between phonemes are 'neutralized' under specific conditions, cp. the neutralization of the opposition aspirated stop: unaspirated stop in final position in Danish" (p. 15). "In all contributions to linguistic methodology which build on the basic notions of information theory, it seems to be implied that linguistic sequences are built up of members from one inventory" (p. 16).

Rischel proposes "to introduce the concept 'neutralization' in the purely sequential aspect of language as a Markoff-process. This would involve that surely not only the probabilities but the code *inventory* itself varies throughout sequences: *after certain sequences the difference between two elements is neutralized*" (pp. 15–16). It would take us too far to discuss this contribution to code theory, but anyhow *Rischel's* remarks are noticeable by their suggesting, in fact, that in this field codes may profitably be studied together with language, and not the other way round.

An important fact about most – or all – technical codes is their synthetic nature: The code system consists of a *given* number of elements, and of *given* (explicit) rules for combining them into (potential) messages. Owing to this fact discrete signalling systems and their functioning have been studied by information theory without it being necessary to tackle problems of how to find out the (or a) system that corresponds to – "underlies" – a given message. When dealing with the phonemic or graphemic aspect of linguistic messages, code theory in practice draws upon the results of a phonological (phonemic) or graphemic analysis, carried out on a linguistic basis (cf. the situation previously discussed for set-theoretical descriptions of language expression). "Auf welche Weise man die Nachrichtenobjekte im Informationsvolumen" (e.g., the phonemes of an utterance) "erkennen kann, bleibe zunächst offen. Die hierzu geeigneten *Analysiermethoden* bilden eines der schwierigsten Probleme bei der praktischen Anwendung informationstheoretischer Methoden. Wir setzen voraus, die Analyse sei bereits durchgeführt...." *W. Meyer-Eppler*, op. cit., p. 58.

Thus it is an open question whether code theory can furnish a new and more general basis of setting up elements like phonemes. So far there seems to be no theory overbridging the division which *Saussure* introduced at once in the semiology he had suggested himself: "La langue présente donc ce caractère étrange et frappant de ne pas offrir d'entités perceptibles de prime abord, sans qu'on puisse douter cependant qu'elles existent et que c'est leur jeu qui la constitue. C'est là sans doute un trait qui la distingue de toutes les autres institutions sémiologiques" (Cours p. 149).

In the case of continuous signals, including speech in its *physical* aspect, the situation is different. Since the notion of code – whatever particular definition adopted – is always bound up with discrete (discontinuous) signals, code theory becomes relevant to the speech continuum only where attempts are made at quantizing speech into recurrent elements. The background of telecommunication research in this field lies in the importance of compressing speech into signals occupying less channel capacity; in addition to ways of compression by which the speech signal remains continuous, various ways of discrete (parametric) compression, making use of vocoders of different constructions, are being developed (cf. for instance the recent survey by *M. A. Sapožkov*[13]).

It is beyond the scope of the present paper to discuss whether human perception of speech involves neurological processes similar to such technical coding processes – a view advocated in cybernetics. But it is highly relevant to phonology whether it is possible by technical means to quantize speech into code elements corresponding to phonemes. In her report to The 8th Internat. Congress of Linguists (Oslo 1957), *Eli Fischer-Jørgensen*[2] discusses the possibility of obtaining 'phonemes from curves' (by means of "phoneme detectors") but answers in the negative. It is, indeed, unlikely that a machine of human ability in recognizing phonemes will ever be constructed. But it is a matter of human ability in constructing machines, *in what degree* a mechanical segmentation of speech into linguistically relevant elements is possible; among other things, the elements obtained need not be of the same extension as phonemes.

Considering the economical interests attached to this possibility – closely related to conversion of speech into writing – a good deal of effort will probably be devoted to such tasks, and phonology and phonetics may from this research learn something about the nature of phonological analysis, in the same way as research on machine translation has thrown light on certain hitherto unnoticed presuppositions of man-oriented grammar. Anyhow, these perspectives are only by-products of code theory, and they are connected with the conception of language as a code in an indirect way only.

The conclusions arrived at do not raise the expectations from applying code theory to questions of phonology (in the wider sense).

In other linguistic fields the prospects of code theory may be others; but conclusive judgments in the negative occur in the literature: "Two opinions current in MT writings on language are that language is a code and that the code is fundamentally binary. Both these views are, from the standpoint of a communication engineer, tenable and useful. From the linguistic standpoint, however, these views are both questionable and unhelpful; and they have hampered MT work because they misrepresent the functioning of language both in its internal relations and in its relations to non-language" *M. A. K. Halliday*[3] (p. 146).

Anyhow, the question of whether language is a code has important theoretical implications, and even answers in the negative may be useful; for – on condition that a clear definition of code is given – we shall learn something about the nature of language by finding out to what extent language is *not* a code.

References

1. *Cherry, C.:* On human communication (Wiley, and Chapman & Hall, New York/London 1957).
2. *Fischer-Jørgensen, E.:* What can the new techniques of acoustic phonetics contribute to linguistics? Proc. the VIIIth int. Congr. Ling. Oslo 1957, pp. 433–480.
3. *Halliday, M. A. K.:* Linguistics and machine translation. Z. Phon. Sprachw. Kom. *15:* 145–158 (1962).
4. *Jakobson, R.* and *Halle, M.:* Phonology in relation to phonetics, in *L. Kaiser's* Manual of phonetics (Amsterdam 1957).
5. *Krámský, J.:* Quantitative phonemics in the last decade. Phonetica *8:* 166–185 (1962).
6. *Mathesius, V.:* La structure phonologique du lexique du tchèque moderne. TCLP *1:* 67–84 (1929). – Zum Problem der Belastungs- und Kombinationsfähigkeit der Phoneme. TCLP *4:* 148–152 (1931).
7. *Meyer-Eppler, W.:* Grundlagen und Anwendung der Informationstheorie (Springer, Berlin 1959).
8. *Miller, G. A.:* Language and communication (McGraw-Hill, New York 1951; French edition 1956).
9. *Oettinger, A. G.:* Automatic language translation (Cambridge, Mass. 1960).
10. *Akhmanova, O.S.* (red.): O točnykh metodakh issledovanija jazyka (Moskva 1961).
11. Trends in European and American linguistics (Utrecht 1961).
12. *Mološnaja, T. N.* (red.): Issledovanija po strukturnoj tipologii (Moskva 1963).
13. *Sapozkov, M. A.:* Rečevoj signal v kibernetike i svjazi (Moskva 1963).
14. *Saussure, F. M.:* Cours de linguistique générale, 3e éd. (Lausanne/Paris 1931).
15. *Spang-Hanssen, H.:* Recent theories on the nature of the language sign. TCLC *9* (1954).
16. *Trubetzkoy, N. S.:* Grundzüge der Phonologie. TCLP *7:* 68–99 (1939).
17. *Zabrocki, L.:* Sprachkode. Z. Phon., Sprachwiss. Kom. *14:* 64–73 (1961).
18. *Zemanek, H.:* Elementare Informationstheorie (Oldenbourg, Wien/München 1959).
19. *Zwirner, E.:* System der Sprache und System der Wissenschaften. Indog. Forsch. *68:* 133–148 (1963).

Author's address: Dr. H. Spang-Hanssen, Moltkesvej 40, *Copenhagen* (Denmark).

Discussion

Herdan (Bristol): I do not know who ever described the application of mathematics to language in the words used by the lecturer, namely "Language is a code" and "Language is discrete mathematics". Rightly understood, no such identification is intended by the introduction of mathematical methods in linguistics. Such application has a sound empirical basis. It was simply found that language has certain aspects in common with artificial codes, and since we know the structure of such codes very well, they being of our own making, it is quite natural that we should make use of this when it comes to describing language. This is all it comes to. There is no identification implied in this between natural languages and artificial codes.

Mr. *Spang-Hanssen* said that he did not see the connection between *de Saussure's* teaching and code theory, or that there was no historical connection between the two. He evidently does not take into due consideration the *Saussure's* axiom of the independence of sound and meaning. If this is understood in all its implications it leads to the conception of language being in certain aspects, and specially so on the phonemic and alphabetic levels, very similar to artifical coding systems. In particular, Information Theory which works with the conception of language as a code is only possible if the frequency distribution of phonemes and of alphabetic symbols remains sensibly stable regardless of the content of sufficiently long messages or texts.

As to the statement that "Language is discrete mathematics", if it was ever made like this, it ought not to have been. Although we make extensive use of continuous mathematics in engineering, nobody has ever said that engineering is continuous mathematics. To say so would only provoke the obvious objection that engineering is engineering and mathematics is mathematics. Similarly, it would not make sense to say that language is a discrete mathematics. Both engineering and language are fields in which mathematics can be applied profitably. They are thus fields of applied mathematics, not the mathematics themselves.

Bés (Buenos Aires): Je voudrais souligner l'observation faite par M. *Spang-Hanssen*, laquelle s'appuie, à son tour, sur les remarques de M. *Rischel*. Il est bien connu que pour certaines formulations phonologiques on emploie non seulement les unités qui précèdent à un élément donné mais aussi celles qui le suivent dans la chaîne. En espagnol et dans beaucoup d'autres langues, entre voyelle et /p/, la seule nasale qu'on trouve c'est [m]; donc [m] a une information plus faible que si elle était en opposition avec les autres nasales, c'est à dire, quand elle est suivie par une voyelle. Si on exprimait ce fait par le seul recours aux éléments précédents, la formulation qui en résulterait ne serait pas acceptable du point de vue phonologique à cause de sa complexité. On peut donc tirer la conséquence que non seulement il y a des différences d'inventaire dans les différents points de la chaîne mais que celles ci, à son tour, sont parfois déterminées, au moins partiellement, par les phonèmes qui suivent. Cela pose de sérieux problèmes à l'application du procès de Markoff.

Fry (London): I would put in a plea for using the expression "artificial code" as Mr. *Herdan* has done in contrast with "natural languages".

This seems to me much clearer than the distinction between codes in which the rules are "explicit" or "implicit", as Mr. *Spang-Hanssen* has characterized them. In an artificial code the rules are formulated before the code is used, and thus we know what the rules are from the start; in natural languages we do not start by knowing the rules and we have to try to discover them. But unless a code is in use, whether it be an artificial code or a natural language, the rules are always implicit.

Tillmann (Bonn): Die Frage, ob Sprache ein Code sei oder nicht, kann offenbleiben, wenn man darauf hinweist, daß Sprache immer in einer mittelbaren, sagen wir: codierten Form zugänglich ist. Selbst wenn man es aus wissenschafts- und erkenntnistheoretischen Überlegungen unterlassen muß, der Beziehung zwischen Sprache-an-sich und deren Transformation, d.h. Codierung, und ihrer Beschreibung weiter nachzugehen, so kann meines Erachtens doch die Auffassung der beschriebenen Sprache als Code manche Scheinprobleme auflösen helfen, z.B. das der mono- oder biphonematischen Wertung von Diphthongen und Affrikaten. Ja, man könnte sogar von der «codetheory» Argumente für die Möglichkeit der an sich relativ arbiträren Entscheidung für das eine oder das andere erwarten.

Proc. 5th int. Congr. phon. Sci., Münster 1964, pp. 61–81
(S. Karger, Basel/New York 1965).

Suitable and Unsuitable Mathematical Models in Language Statistics, and their Consequences

By GUSTAV HERDAN, Bristol

Originally the word 'mathematical model' meant a three-dimensional geometric structure in wood or cardboard representing visually the relation between three variables; later, it came to be used in the sense of hypothesis or theory, by which to explain an observed relationship between variables. More precisely, and more ambitiously, it is the name for the differential equation set up on the basis of an hypothesis about the behaviour of two or more variables, which by integration would lead to the empirically established relation between those variables. There is nothing to be said against the use of the term 'model', provided one is clear about the sense in which it is to be understood. But from the explanation given above, it is difficult to see what advantage it is to call an hypothesis, or a differential equation, or a theory, a 'model'. In the original sense, a model was very effective in making us visualize an observed mathematical relationship; in its metaphorical sense, it does nothing of the kind, but may even obscure the matter if it makes us forget that 'model' stands here for hypothesis or theory. If so, it tends to make us construct such models independently of one another, and regardless of the wider implications of the observed relation. This could be a great disadvantage, considering that it is the virtue of a theory to reveal hitherto unrecognized relationships between often widely different parts of a field of knowledge. If, therefore, in this address, I use the term 'model', it should be understood according to the above interpretation as hypothesis or, better, theory.

I.

1. For mathematical models to be of real value it is necessary that (1) the relation of events of which the mathematical structure

Table V

Macauley: (1) the vocabulary "expected" on the basis of the *Bunyan* distributions, table II, cols. 2 and 3, and (2) the observed distribution from table II, cols. 4 and 5 (summed).

1 Initial letter of noun	2 Expected vocabulary	3 Observed vocabulary	1 Initial letter letter of noun	2 Expected vocabulary	3 Observed vocabulary
A	207.1	245	O	69.0	70
B	169.8	156	P	339.6	329
C	370.0	377	Q	13.3	14
D	256.2	235	R	227.4	188
E	118.8	162	S	358.2	372
F	155.6	145	T	164.4	172
G	95.6	100	U	15.8	21
H	129.9	107	V	84.1	63
I	139.3	169	W	97.7	86
J	40.9	23	X	–	–
K	17.1	19	Y	4.8	8
L	108.8	117	Z	2.1	4
M	197.6	205			
N	53.8	50	Total	3436.9	3437

difference between the *Bunyan* and *Macaulay* distribution of vocabulary simply in terms of the respective proportions of OE and LR words. By simply varying the proportions in the *Bunyan* vocabulary, we can account for $r^2 = 0.9727$, or roughly for 97 % of the variance of the *Macaulay* distribution.

Regarding the general linguistic aspect of language contact, we have learnt from the above analysis that the relative distributions of nouns in literary texts of English according to the initial letter for each, the Teutonic and the Romance component of vocabulary, can be regarded as random samples of the corresponding probability distribution in the language. As random samples, they are independent of text length and content; moreover the distribution of each component, OE and LR, is independent of, and without influence upon, the other. It is these findings which provide the theoretical basis for drawing the inference of different proportions of the OE and LR components in the vocabularies of two English writers from an observed difference in the alphabetic distribution of their vocabularies.

2a. An obvious and interesting question is whether the stability of the alphabetic distribution of nouns of a given component, say the LR component, is the manifestation of the form of that type of

distribution in the parent language, Latin in our case. If this could be established it would throw quite unexpected light upon the *mechanism of 'borrowing'* and the *mixing* of languages. Unexpected, because the current view in this matter among linguists is that change and not stability is the effect of such language contact.

Table VI gives the answer to this question. Col. 2 gives the mean *Bunyan-Macaulay* ranking of the alphabet for nouns of LR origin: col. 3 gives the ranking for Mediaeval Latin nouns (2454 altogether) from the 'De Imitatio Christi', together with samples

Table VI

Comparison of the ranking of initials (1) in nouns of Latin-Romance origin in *Bunyan* and *Macaulay*, col. 2; (2) in Latin nouns occurring in samples from the Imitation, *à Kempis* and *Gerson*, col. 3.

1 Rank	2 *Bunyan* and *Macaulay*, RL component	3 *à Kempis* and *Gerson*, Mediaeval Latin	1 Rank	2 *Bunyan* and *Macaulay*, RL component	3 *à Kempis* and *Gerson*, Mediaeval Latin
1	C	C	14	G	O
2	P	P	15	L	N
3	S	S	16	O	H
4	A	A	17	H	G
5	D	I	18	N	B
6	R	D	19	J	J
7	M	M	20	U	U
8	I	F	21	Q	Q
9	E	R	22	W	Z
10	T	T	23	Z	K
11	F	L	24	K	W
12	V	V	25	X	X
13	B	E	26	Y	Y

from miscellaneous works by *Thomas à Kempis* and theological writings of *Gerson*. There is evidently good general agreement between the two, in spite of some differences in rank for certain letters. As a summarizing measure for the agreement, we use again the rank correlation coefficient which for our two series results as 0.960, which means a highly significant correlation. In fact, it is not appreciably different from the ranking correlation coefficient between the two LR series from *Bunyan* and *Macaulay*, which results as 0.986.

The high correlation between the alphabetical distribution of the LR component in both writers, and the corresponding distribution from works written in Mediaeval Latin provides the explanation for the stability discussed above. It is the *tenacity of the functional burdening* of particular sounds (represented by letters in our illustration) when used as noun initials which has made the same probability distribution persist from mediaeval times through the subsequent development of English.

A new and important item of knowledge we derive from the investigation in this respect is that the borrowing of Mediaeval Latin by Old English and the initimate mixing of the two main components of English has left the original alphabetical distribution of nouns in Mediaeval Latin unaltered. It follows that *such alphabetical distributions are to a very high and significant extent independent from one another, in spite of the intimate mixing of the components.*

2b. A point of methodological interest to which I should like to draw attention is that the investigation shows how what started as research of an apparently literary nature only, namely as stylo-statistics, could lead to *highly relevant linguistic results.* This, I submit, is one of the most valuable concomitants of literary or stylo-statistics, to use its conventional name. That this feature should have escaped the attention of *W. Plath,* the author of the chapter on 'Mathematical Linguistics' in "Trends in European and American Linguistics 1930 to 1960" (Utrecht 1961) is most deplorable and must be regarded as a severe fault of his presentation of the subject, in particular where it deals with 'Statistical Linguistics'. According to *Plath,* there has been no important development of the statistical study of language since *Yule,* apart from my new derivation and interpretation of the predominant parameter of vocabulary distribution, *Yule's* Characteristic K, as the coefficient of variation of the mean.

Mr. *Plath* has here missed the essential point that whereas *Yule* was only concerned with the statistical study of vocabulary for purposes of characterizing *individual* style in an objective manner, I have been mainly concerned with extending the analysis of word frequency, and of frequency of linguistic forms in general (phonemes, letters, morphemes, syllables, etc.) to language as such, which led to *statistical linguistics* as the quantitative interpretation of *de Saussure's* 'langue-parole' dichotomy, and with it to a new branch of linguistics as a science in the sense in which *L. Bloomfield* uses the term. That this point should have been missed so completely in a publication which is primarily meant to acquaint linguists with the development of this branch of knowledge is highly regrettable.

3. The question now arises whether the calculus of statistical conditions in language in the form of a multinomial law also applies on the vocabulary level. It is quite conceivable that this formula being only part of a general calculus of linguistic observations, may not be the appropriate one on the higher levels of language.

As it turns out, the difference between the comparatively small number of phonemes in a language (say between 20 and 50) and the very large number of vocabulary items (of the order of magnitude of 50,000) is such as to make it practically impossible to apply the multinomial law if the variable is the occurrence probability of the individual vocabulary items. This would mean a series of as many p_i as there are vocabulary items in a comprehensive dictionary of the language.

However, the multinomial law becomes applicable if instead we let the p_i denote not the probabilities of particular vocabulary items – against which there might also be objections of a theoretical nature –, but the probability of a vocabulary item belonging to one part of the text, to two parts, ... to all parts into which the original text, or complex of texts, has been equally divided. This means that instead of with vocabulary occurrence frequency, we work with what is known as *vocabulary partitioning,* or *vocabulary connectivity,* a most useful and important characteristic of the vocabulary structure in a language.

In place of the p_i in the statistical universe of phonemes, we have here the probabilities of vocabulary connectivity according to the Random Partitioning Function (R.P.F.)*. Their series serves as a yardstick against which to judge an observed vocabulary connectivity with a view to arriving at a decision whether the differences between theory and observation are compatible with regarding the observed series as a random sample of the universe, represented by the theoretical series*.

3a. To examine, the vocabulary connectivity in say four samples from *Macaulay's* Essay on *Bacon,* we construct by the random partitioning function the chance model of vocabulary connectivity as consisting of all possible combinations in the group, in order to compare the observed connectivity in the members with what would obtain by chance. This enables us also to decide whether a particular member differed significantly from the chance model, and hereby, from the

* *G. Herdan:* Type-Token Mathematics, esp. chapter 3, section 6.3 and chapter XVIII A (Mouton & Co., The Hague 1960).

rest, and if so, in what points, i.e. with regard to which particular classes of combination. A significant χ^2 would mean that the member in question was significantly different from the other members – all or some – in the group, as regards vocabulary and/or occurrence frequency.

The calculation of the required probabilities from first principles soon becomes very troublesome. I have therefore prepared a table from which they can be read off directly for combinations up to the order 30, and occurrence frequencies up to 100 (Type-Token Mathematics, pp. 341–412).

The four cell frequencies for the chance model resulted as 277.51, 61.60, 71.58, 273.66: total 684.35. For the χ^2-test they are adjusted to the observed total 681.50, as in the second row of table VII. This provides our yardstick of chance with which we compare the observed class frequencies from the essay on *Bacon*, as shown in the first row of the following table.

Table VII

Vocabulary connectivity (averaged) in four samples from essay on *Bacon**.

	(ABCD)	(ABCδ)	(AB$\gamma\delta$)	(A$\beta\gamma\delta$)	Total
Observed	271.00	59.25	67.50	283.75	681.50
Calculated	276.40	61.35	71.29	272.57	681.61

$\chi^2 = 0.839$

Since χ^2 remains far below the value required for significance (7.815) for the given degrees of freedom, we conclude that the vocabulary connectivity between one sample (averaged) and the rest is, by and large, that expected on random partitioning within the same universe.

In a comparison 7 samples from works by *Bunyan*, the calculated cell frequencies were 449.17, 75.64, 76.02, 259.53: total 860.36. For the χ^2-test they were proportionally adjusted to the observed total 718.58, as shown in the second row of the table VIII.

Contrary to the previous comparison, the value of χ^2 substantially exceeds that required for significance, and we conclude that the vocabulary connectivity between the samples (averaged) and the rest

* The presence of a vocabulary item in the four samples is denoted by the capital Latin letters A, B, C, D, and their absence by the Greek letters α, β, γ, δ.

Table VIII

Vocabulary connectivity (averaged) in four works by *Bunyan.*

	(ABCD)	(ABCδ)	(AB$\gamma\delta$)	(A$\beta\gamma\delta$)	Total
Observed	259.00	72.75	74.33	312.50	718.58
Calculated	375.15	63.18	63.49	216.76	718.76

$\chi^2 = 81.527$

differs significantly from what is expected by random partitioning. The reason is evidently that we are here dealing with samples from three different works by *Bunyan*, which statistically must also be regarded as different universes.

4. As if to disprove the view held by some linguists that literary statistics could not contribute to our knowledge of language, or – as the magic formula goes – were "linguistically not relevant", the Italian Semiticist, *P. Franzaroli*, has adapted the method of vocabulary connectivity to provide a method for the investigation of philological or, as we say, linguistic phenomena. His problem was the classification of six semitic languages: Babylonian, Ugaritic, Hebrew, Syrian, Arabic and Ge'ez (abbreviated: Ba, Ug, Heb, Sy, Ar, and Ge).

For the investigation he selected certain phonological and morphological characteristics, altogether 217 isoglosses, as a representative sample of such features, and recorded for each of the six languages whether a particular characteristic was present or not. The following is a sample from the fundamental table listing the results of the investigation.

Table IX

	Ba	Ug	Heb	Sy	Ar	Ge
1. *p conservato*	+	+	+	+	–	–
2. $p > f$	–	–	–	–	+	+
3. $\underline{t}$ *conservato*	–	+	–	–	+	–
4. $\underline{t} > s$ *(s in Ge'ez)*	+	–	+	–	–	+
5. $\underline{t} > t$	–	–	–	+	–	–

He then chose the characteristic of vocabulary connectivity – which here becomes that of connectivity in phonological and morphological features, though, in order to avoid the longer form, we shall continue to use the term vocabulary connectivity – and recorded the numbers of characteristic features which each language had in common with 1, 2, ... 5 other languages (table X).

Table X

Languages	1	2	3	4	5	Total
Ba	33	11	18	15	8	85
Ug	3	18	17	17	18	73
Heb	9	22	22	20	16	89
Sy	12	14	16	21	17	80
Ar	12	28	14	20	16	90
Ge	14	21	12	11	15	73
Average	13.8	19.0	16.5	17.3	15.0	

For comparing one language with the rest, we require again a chance model of connectivity. We can, however, not directly use the numerical table of the random partitioning function because the basic data are here given in a different form from how they were provided in the illustration under 3. There the basic information was the frequency distribution of vocabulary according to the occurrence number of the vocabulary items. These numbers could vary from 1 to anything. Here every one of the 217 phonological or morphological items can only be present or absent, and in symbols have thus the occurrence numbers 1 or 0, in each language. However, the chance model can here be obtained directly from the column sums of table X by dividing the column sums by the number of rows, that is, as the average vocabulary connectivity.

Illustration

For the purpose of comparison, the theoretical figures are reduced to the observed totals, e.g. for Babylonian:

Table XI

	1	2	3	4	5	Total
Ba	33	11	18	15	8	85
Chance model*	14.3	19.8	17.2	18.0	15.6	84.9

$$\chi^2 = 32.59$$

* Last row of table X adjusted to a total of 85.

which for the given degrees of freedom means that a χ^2 like this or greater could on pure chance occur less often than 1 in 100. We, therefore, conclude that Babylonian is exceptional in its vocabulary connectivity with the other languages.

* * *

For the sake of completeness, and to avoid confusion, it should be remembered that just as our model is not the whole of the Calculus of Statistical Conditions in language, so that Calculus is not the whole of what I have called the Calculus of Linguistic Observations, which in addition to statistics comprises Combinatorics and certain branches of Geometry.

References

Franzaroli, P.: Prospettive di Metodo Statistico nello classificazione delle lingue Semitiche. Atti della Accademia dei Lincei, Serie Ottava., XVI, fasc. 1–12, Roma 1961.

Herdan, G.: Type-token mathematics (Mouton & Co., 's-Gravenhage 1960).

Herdan, G.: The calculus of linguistic observations (Mouton & Co., 's-Gravenhage 1962).

Mandelbrot, B.: An informational theory of the statistical structure of language. *W. Jackson,* Communication Theory (Butterworth, London 1953).

Plath, W.: Mathematical linguistics. In Trends in European and American Linguistics 1930–1960 (Spectrum Publ., Utrecht 1961).

Yule, G. U.: The statistical study of literary vocabulary (Cambridge University Press, Cambridge 1944).

Zipf, G. K.: Selected studies of the principle of relative frequency in language (Harvard University Press, 1932).

Zwirner, E. und *Zwirner, K.:* Die Häufigkeit von Buchstaben- und Lautkombinationen. F. u. F. *12:* 286–287 (1936).

Autor's address: Gustav Herdan, University of Bristol, *Bristol* (Great Britain).

C.R. 5e Congr. int. Sci. phon., Münster 1964, pp. 82–104
(S. Karger, Basel/New York 1965).

Les problèmes de la phonétique évolutive

Par André Martinet

Pour comprendre comment se posent, aujourd'hui, les différents problèmes de la phonétique évolutive, il est moins important d'être familiarisé avec les pratiques de la linguistique structurale contemporaine que d'avoir pris conscience de ce qu'elle représentait, à ses débuts, par rapport à ce qui avait précédé. La phonologie d'où, par filiation ou par réaction, sont sortis la plupart des mouvements structuralistes, est, avant tout, l'affirmation et la démonstration que – pour dire les choses en termes un peu naïfs, mais clairs – les sons font partie de la langue au même titre que le sens. Sans doute *Saussure* avait-il auparavant définitivement établi que le signifié n'est une réalité linguistique que parce qu'il correspond à un signifiant qui appartient à la langue au même titre que le signifié. Mais l'articulation du signifiant en segments phoniques successifs restait, pour *Saussure* et les saussuriens, un aspect purement marginal de l'organisation de la langue. La phonétique demeurait, pour eux, ce qu'elle avait été pour les générations de penseurs qui les avaient précédés, une science auxiliaire de la linguistique. La phonologie a montré que les segments phoniques successifs dont se compose le signifiant sont des unités linguistiques, autres que le signe, certes, puisqu'elles sont distinctives et non significatives, mais qui existent et qui fonctionnent dans les mêmes conditions que les signes. On est aujourd'hui pleinement conscient du fait qu'il existe une syntaxe des phonèmes comme il y a une syntaxe des signes, même si l'on préfère en pratique réserver ce terme aux rapports entre les unités douées de sens. Aux rapports associatifs de *Saussure*, restes d'une psychologie dépassée, on a substitué des rapports paradigmatiques, ceux que l'on relève entre les unités susceptibles d'apparaître dans les mêmes contextes. Ces rapports paradigmatiques existent aussi bien entre les unités distinctives, les phonèmes, qu'entre les unités significatives, les monèmes, et c'est, en fait, l'exemple des classes

paradigmatiques de phonèmes qui a donné le courage d'innover, sur le plan des signes, par rapport à l'enseignement saussurien.

Pour bien apercevoir toutes les implications et toutes les conséquences de ces parallélismes d'un type d'unités à l'autre, il n'est pas inutile de recourir à des analyses et à des formulations plus concrètes et plus explicites.

Soit un locuteur et un auditeur. Si la communication s'établit entre eux, c'est qu'ils parlent et comprennent la même langue. Le locuteur pourrait devenir auditeur et vice versa. La langue qu'ils pratiquent est le français. Le locuteur commence avec /l.../ prononcé [lə]. L'auditeur peut, dès cet instant, faire l'hypothèse que si l'expérience à transmettre avait été quelque peu différente, le locuteur aurait dit /œ̃/ (article indéfini) au lieu de [lə] (article défini): devant le contexte qui va suivre, le locuteur avait le *choix* entre /œ̃/ et [lə], et son choix a été *dicté* par ce qu'il avait à dire. Ce qu'il avait à dire est précisément ce que l'auditeur doit comprendre. Il y a choix du locuteur, mais un choix qui n'intéresse l'auditeur que du fait de ce qui le conditionne. Si son hypothèse relativement à [lə] est correcte (s'il s'agit bien de l'article), l'auditeur sait que ce qui va suivre représente un nouveau choix du locuteur, choix toujours strictement conditionné par l'expérience qu'il veut transmettre. Les circonstances dans lesquelles a lieu l'échange linguistique rendent plus vraisemblables certains choix que d'autres. Si la scène se passe dans la boutique d'un chapelier, le choix de /ʃapo/ *(chapeau)* est plus probable que celui de /fyzi/ *(fusil)* ou de /ʃamo/ *(chameau)*, mais ceux-ci ne sont nullement exclus. En fait n'importe quel substantif (de genre masculin et commençant par une consonne) peut figurer après [lə] initial. C'est même à cette possibilité qu'on reconnaît en français un substantif (appartenant aux sous-classes définies ci-dessus).

Mais revenons à nos deux protagonistes. Le locuteur a effectivement prononcé /ʃapo/. Son choix du signe *chapeau* impliquait celui des phonèmes /ʃ a p o/ dans cet ordre. Mais, bien que dicté par le choix préalable du signe *chapeau*, le choix de chacun des phonèmes constitutifs du signifiant n'en reste pas moins un choix, tout comme reste un choix celui de *chapeau*, bien qu'il soit dicté par ce qu'il y a à dire. Dire que *chameau* est distinct de *chapeau* du fait de la présence du phonème /m/ là où *chapeau* présente le phonème /p/, c'est admettre que le /ʃa...o/ de *chameau* et celui de *chapeau* sont linguistiquement identiques, c'est-à-dire ne diffèrent que du fait des contextes dans lesquels ils apparaissent, /m/ et /p/ faisant partie de ces contextes. Sur le plan du comportement du locuteur, ceci implique que la production du signifiant /ʃapo/ ne représente pas une habitude motrice unique, mais qu'elle résulte de la succession de quatre habitudes motrices distinctes correspondant à chacun des phonèmes du mot. Le locuteur, une fois qu'il a prononcé /ʃa.../ peut prononcer /...p.../ ou /...m.../ ou tel autre phonème du français; son désir de transmettre un message correspondant à son expérience du moment lui fera choisir /...p.../ au lieu de /...m.../ ou tel autre phonème que diverses circonstances pourraient l'inciter à prononcer. On peut supposer, par exemple, qu'il a eu récemment à prononcer, avec une particulière fréquence, des mots où /...a.../ était suivi de /...m.../, de telle sorte que /...m.../ viendra plus naturellement après /...a.../ que toute autre consonne de la langue. Ou encore, la vue, accidentelle, d'un tableau représentant un chameau peut le pousser à dire /ʃam.../ même si les besoins de la communication réclament en l'occurrence une autre consonne que /...m.../. Pour prononcer /ʃapo/, il lui faudra surmonter la tentation de faire suivre /...a.../ de /...m.../. Peu importe que tout ceci soit conscient ou inconscient. Il lui aura fallu maintenir le voile du palais relevé contre la paroi postérieure du pharynx et interrompre les vibrations glottales, alors que pour /m/ le voile se serait abaissé et la voix aurait persisté. L'étude des lapsus montre que les tentations du type de celles que nous supposons ici, sont souvent si fortes qu'on y cède. Les contrepèteries, comme celles qui ont rendu célèbre le «don» *Spooner* d'Oxford (*half-warmed fish* pour *half-formed wish*, etc.), les allitérations, les assonances, les rimes, les devinettes que mentionne *Troubetzkoy*[1], sont autant de preuves que l'analyse des signifiants en phonèmes successifs n'est pas un simple procédé descriptif inventé par les phonologues, mais correspond à quelque chose d'observable dans le comportement linguistique de l'homme. A chaque point du discours, il y a donc *choix d'un phonème* parmi tous ceux qui pourraient paraître dans le même contexte si le message à transmettre était autre.

La façon dont un phonème se réalise dépend d'un certain nombre de facteurs comme le contexte phonique, la conformation des organes de celui qui parle et l'état de son humeur. Elle ne dépend jamais, en principe, du sens du mot ou du monème dans lequel

[1] Arch. f. vergl. Phonetik *1:* 129–153 (1937).

le phonème figure. Ceci est une preuve supplémentaire de la réalité du phonème et du caractère fondamental de la seconde articulation du langage humain, celle selon laquelle les signifiants s'articulent en unités distinctives successives. Si le signifiant [ʃapo] correspondait à une habitude motrice unique comportant, par hasard, une suite de traits qu'on pourrait *approximativement* identifier à certains types phoniques notés [ʃ], [a], [p], [o], rien ne pourrait empêcher ce signifiant d'évoluer, par exemple pour mieux s'adapter, selon la fantaisie des usagers, à l'expression du signifié «chapeau». Les seules limitations qu'on pourrait imaginer pour cette évolution, résulteraient de la nécessité de distinguer le signe *chapeau* des autres signes de la langue. Comme il n'y a pas en français de signe à signifiant *[sapo], ou *[ʃapu], ou *[ʃepo], rien ne s'opposerait à ce que [ʃapo] évolue vers l'une quelconque de ces formes, et bien au-delà. Mais, bien entendu, [ʃa], signifiant de *chat*, ne pourrait évoluer vers [sa], puisque cela entraînerait une confusion avec [sa] signifiant de *ça;* [ʃapo] pourrait passer à [ʃapu], mais [po] *(pot)* devrait rester distinct de [pu] *(pou)*, etc. En d'autres termes, l'évolution phonique se ferait au hasard, au gré de l'humeur des locuteurs, et l'incessante variation des signifiants censée, dans l'esprit des usagers, viser à une meilleure adaptation de la forme au sens aurait pour corollaire qu'à une certaine forme correspondrait un certain sens et qu'à l'infinité des nuances phoniques correspondrait une infinité de nuances sémantiques qui aboutirait à diluer le lexique en un enchevêtrement de nébuleuses. Bien entendu, on ne constate, dans les langues, rien de semblable: /ʃapo/ ne pourrait évoluer vers /sapo/ que si tous les /ʃ/ initiaux devant /a/ évoluaient vers /s/, aussi bien dans /ʃa/ *(chat)*, où l'évolution amènerait à identifier *chat* et *ça*, que dans /ʃapo/ où elle n'aboutirait à aucune confusion. Toutes les réalisations d'un même phonème sont solidaires les unes des autres. Ceci ne veut pas dire que le résultat acoustique et l'évolution à venir seront partout les mêmes, car la pression du contexte phonique pourra, à la longue, entraîner des déviations considérables. Mais il n'y a pas de solidarité entre le sens du mot et la forme du signifiant. Tout ceci n'est, bien entendu, qu'une version synchronique de l'enseignement néo-grammairien relatif à la régularité des changements phonétiques. Les déformations expressives, gémination, allongements et autres, pour fréquentes qu'elles soient, sont des exceptions dont le caractère marginal ne fait que mieux ressortir l'indépendance que confère à la forme le caractère discret des unités distinctives.

Il ne peut y avoir choix qu'entre des éléments distincts, et la nécessité de choisir implique la préservation des distinctions. Or, réaliser une distinction demande nécessairement un effort, que ceci se place sur le plan des monèmes ou sur celui des phonèmes: on se fatiguerait moins à employer toujours *chapeau* dès qu'il s'agit d'un couvre-chef qu'à essayer de distinguer entre la casquette, le béret, le chapeau proprement dit, et, parmi les chapeaux, le melon, le haut-de-forme, le canotier, etc. Mais les besoins traditionnels de la communauté française réclament qu'on distingue absolument entre les couvre-chef à bord, les couvre-chef à visière et les couvre-chef sans bord ni visière. Sur le plan des phonèmes, on réaliserait une considérable économie d'énergie si, dans l'articulation des consonnes françaises, on laissait au contexte phonique le soin de décider si la glotte doit vibrer et le voile s'abaisser ou se relever; mais il en résulterait la confusion des trois phonèmes /p/, /b/ et /m/ entraînant celle de *chapeau*, *chabot* et *chameau* et de centaines d'autres triades ou d'autres paires que distingue la langue. Il y a conflit permanent entre la tendance de l'individu à restreindre sa dépense d'énergie et les besoins de la communauté qui réclament le maintien de distinctions jugées nécessaires par l'ensemble des usagers de la langue. C'est ce conflit, que résume la théorie du moindre effort qu'on désigne également comme le principe d'économie.

On objecte parfois à la théorie du moindre effort le fait bien établi de la dépense gratuite de surplus d'énergie, dans le jeu par exemple. Dans le cas du langage, on constate à tout moment son emploi à des fins non communicatives, dans le monologue, par exemple, ou dans certains dialogues qui ne sont que des monologues déguisés. Mais lorsque le langage n'est qu'un jeu, il n'est joué de façon satisfaisante par le joueur que si celui-ci se conforme aux règles qui sont celles du langage communicatif, et la tricherie porte en elle-même sa sanction. En tout cas, si celui qui joue au langage déviait, consciemment ou inconsciemment, des règles établies, il se verrait contraint de s'y conformer dès qu'il lui faudrait se faire comprendre d'autrui. C'est pourquoi l'économie du langage est bien pour l'essentiel réglée par le moindre effort, c'est-à-dire, il faut le rappeler, l'équilibre entre l'inertie naturelle et la satisfaction des besoins.

Parmi les facteurs d'inertie, il convient de mettre en valeur ceux qui s'exercent sur l'axe du discours, d'une unité à l'autre d'un même énoncé. Sur le plan des unités significatives, on sait comment le sens d'un monème ou d'un mot est précisé et limité par le contexte sémantique où il figure et avec quelle fréquence des contextes favorisés entraînent la fixation de certains glissements sémantiques: l'anglais *bead* est ainsi passé du sens de «prière» à celui de «grain de collier ou de bracelet». Sur le plan des unités distinctives, c'est tout le chapitre des changements dits «conditionnés» qui est en cause. Comme nous l'avons vu ci-dessus, le sens d'un mot ne saurait, en principe, influencer la réalisation des phonèmes qui composent sa face signifiante; mais l'habitude motrice qui s'identifie à un phonème particulier sera nécessairement infléchie à son début et à sa fin, par l'habitude motrice qui la précède et par celle qui la suit; l'adaptation du phonème à son contexte est inéluctable. Il en résulte qu'un phonème n'est jamais attesté que sous la forme de variantes contextuelles, ou allophones, dont partent certains linguistes pour définir le phonème comme un groupe d'allophones ou une famille de sons concrets. Il n'y a guère de limites aux modifications que le contexte peut déterminer chez un segment phonique: dans l'espagnol populaire *ocupao*, qui dérive, en dernière analyse, du latin *occupatum*, l'assimilation de *t* à son contexte, par voisement tout d'abord (cf. la graphie *ocupado*), par relâchement ultérieur de l'occlusion (cf. la prononciation soignée [ɔkupaðo]), a conduit à l'élimination pure et simple du segment. Mais il est clair que tel n'est pas le sort de tout [t] intervocalique: celui du russe *živĕte* subsiste depuis plus de quatre ou cinq mille ans sans aucune trace d'adaptation au contexte, et, en espagnol même, le [t] de *matar*, qui est intervocalique depuis bien des siècles, est toujours sourd et occlusif. Il est donc clair que l'inertie, face aux pressions du contexte, peut être compensée par autre chose qui contribue au maintien des contrastes de la chaîne par le maintien des oppositions du système.

C'est là où en était restée la phonétique évolutive traditionnelle, celle qui dénonçait son impuissance en désignant comme des changements «spontanés» ceux pour lesquels elle ne trouvait pas de justifications dans son arsenal, comme si un changement pouvait ne pas avoir de cause. Pour poser correctement le problème, il faut dégager tout ce que ne saurait expliquer la pression des contextes particuliers et qui comprend, outre le conditionnement des changements qui affectent toutes les variantes contextuelles d'un phonème donné, par exemple le passage de *ū* latin à [y] français, les facteurs qui permettent à la pression du contexte de se donner libre cours (esp. *ocupao*) et ceux qui empêchent cette pression de s'exercer.

En face de la carence de la phonétique évolutive traditionnelle dans tous ces cas, les linguistes ont réagi de trois façons différentes, selon leur tempérament, leur éducation, ou l'école à laquelle ils appartenaient.

Les uns ont invoqué des facteurs non linguistiques, facteurs raciaux ou facteurs géographiques divers[2]. Même sous leurs formes les plus récentes et les plus élaborées, les hypothèses de ce type restent un ensemble de vues de l'esprit, parfois séduisantes, souvent comiques, mais toujours plus amusantes que convaincantes.

D'autres ont préféré écarter les problèmes en cause comme définitivement ou temporairement insolubles. Ce sont ceux pour qui compte surtout l'apparence de la rigueur: puisqu'en tout état de cause, nous ne pouvons pas tout expliquer, n'expliquons rien et contentons-nous de présenter les faits dans leur succession tels que nous les constatons. C'est un genre de linguistique à laquelle on doit étendre l'épithète de «descriptive» si malencontreusement réservée aujourd'hui aux travaux des synchronistes. En face d'un problème comme celui de l'évolution des occlusives intervocaliques en roman occidental, le comportement du descriptiviste consiste à relever et à dater les faits de graphie qui paraissent impliquer une modification de l'articulation des intervocaliques. Comme, toutefois, considérer les intervocaliques comme un ensemble de faits susceptibles de recevoir un traitement analogue laisse transparaître quelque préjugé explicatif (la position intervocalique serait partiellement responsable de l'évolution qu'on va constater), les descriptivistes les plus convaincus traitent à part de chacune des consonnes du latin sans jamais comparer les phénomènes qu'on constate dans une position déterminée[3]. Un des résultats les plus sûrs de cette méthode est d'écarter des études de linguistique diachronique ceux pour qui la compréhension des phénomènes est la récompense d'heures d'études longues et austères.

D'autres enfin ont invoqué les influences qu'exercent les langues les unes sur les autres. L'hypothèse la plus connue est celle du substrat qui, ni au premier abord, ni à plus ample examen, ne mérite le mépris dans lequel la tiennent certains linguistes contemporains[4].

---

[2] Par ex. *Ginneken, J. van:* TCLP *8:* 233–261 (1939); *Koppelmann, H. L.:* Ursachen des Lautwandels (Leyde 1939).

[3] Sans doute par souci d'éviter toute source de controverse dans un manuel élémentaire, chez *Bourciez, E.:* Phonétique française, 177–179, 198–200, 203–205, 227–230 (Paris 1937).

[4] Cf. le point de vue nuancé, mais plutôt hostile de *Jungemann, F. H.:* La teoría del sustrato y los dialectos hispano-romances y gascones, 17–27 (Madrid 1955).

On peut parfaitement écarter l'explication substratiste pour le *u* du français ou la *gorgia* toscane après examen des dossiers respectifs, sans pour cela rejeter définitivement le substrat comme principe d'explication. La seule attitude scientifique en la matière est de vérifier, dans le monde d'aujourd'hui, dans les sociétés accessibles à l'observation, ce qui se passe réellement lorsque deux langues sont en contact. Les réponses que fournira l'observation seront valables, non seulement dans le cas assez particulier du substrat, mais dans toutes les situations bilingues ou plurilingues.

Le danger auquel on s'expose constamment lorsqu'on estime avoir découvert quelque nouveau principe d'explication c'est, bien entendu, de vouloir en faire un principe universel. C'est à cette tentation qu'ont cédé beaucoup de ceux qui, ayant étudié dans le détail la façon dont se propagent certains changements phonétiques, que ce soit à l'intérieur d'une communauté homogène ou de langue à langue par le chenal de bilingues, ont voulu voir, dans l'imitation d'usages, de dialectes ou d'idiomes différents un moyen d'expliquer tous les changements phonétiques.

Il y a deux façons de concevoir l'action d'une langue sur une autre en matière de changements phonétiques. On peut d'abord supposer l'existence dans une langue A d'un processus évolutif: telle voyelle est en train de se diphtonguer, [e:] par exemple devient [ej]; des bilingues, parfaits ou imparfaits, là n'est pas la question, reçoivent des unilingues de langue A cette habitude de diphtonguer le [e:] et ils la pratiquent dans leur autre langue, B, aussi bien que dans A; des unilingues de langue B imitent à leur tour les bilingues, et c'est ainsi que la diphtongaison de [e:] passe d'une langue à une autre. Ce qui a été emprunté, dans ce cas, c'est un processus ou, comme le disent certains, une tendance.

On peut d'autre part envisager que de dialecte à dialecte ou d'une langue à une autre langue, lorsque, par parenté génétique ou emprunts massifs, existe un vocabulaire commun, la forme de certains mots sera modifiée pour l'identifier à celle des mots correspondants du dialecte voisin. Soit un dialecte B qui présente /ka/, /kã/, /kapo/, /kato/ là où le dialecte A offre /ʃa/, /ʃã/, /ʃapo/, /ʃato/. Il se peut que B emprunte à A les formes /ʃapo/ et /ʃato/ et les emploie régulièrement au lieu de /kapo/ et /kato/, alors qu'il conserve /ka/ et /kã/ en face des /ʃa/ et /ʃã/ de A. Un peu plus tard, ces derniers finiront aussi par s'imposer en B aux dépens de /ka/ et /kã/. Les linguistes examinant le dialecte B quelques siècles plus tard, seront légitimement tentés, s'ils n'ont pas de données ponctuant les différents temps du phénomène, d'y retrouver un changement «régulier» d'un /k/ primitif en /ʃ/, alors qu'en fait il y a eu remplacement de certaines formes du parler par les formes d'un autre dans des conditions qui rappellent celles de l'emprunt lexical.

L'une et l'autre modalités d'action sont largement attestées et ont été assez bien décrites[5]. Il est souvent possible d'identifier la seconde du fait de la persistance de mots qui n'ont pas «fait» le changement (dans le cas présenté ci-dessus, ceux qui ont gardé /ka/ au lieu du /ʃa/ attendu) soit parce qu'ils n'avaient pas de correspondants dans le dialecte qui a fourni les formes nouvelles, soit parce qu'ils étaient d'une nature sémantique telle qu'ils n'avaient guère de chance d'être employés au cours de relations interdialectales (mots très familiers, obscènes ou bas). Il n'est donc pas question d'écarter les contacts de langue, c'est-à-dire le bilinguisme et le plurilinguisme, lorsqu'il s'agit de rendre compte de l'évolution de la phonie des langues. Mais on ne saurait en faire l'unique principe d'explication.

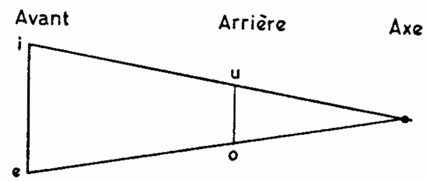
Ce que la phonétique évolutive traditionnelle n'a jamais fait entrer en ligne de compte, c'est l'action que peut avoir, sur la nature articulatoire des phonèmes, la nécessité de les maintenir distincts les uns des autres. On aperçoit assez bien les raisons de ce désintérêt: seul le changement pouvait retenir l'attention du spécialiste de phonétique évolutive; s'il s'agissait pour les phonèmes de rester distincts, la solution la plus simple semblait être dans l'immobilité, le maintien du *statu quo;* il était impossible d'imaginer la préservation des distinctions comme le moteur initial d'un changement. Or, ce qu'on recherchait, un peu naïvement nous semble-t-il aujourd'hui, c'était ce moteur initial, la cause première et unique des changements particuliers, voire même des changements phonétiques en général. On avait, bien entendu, souvent constaté des changements en chaîne[6], /u/ passant à /y/, /o/ à /u/, /ɔ à o/ par exemple, et ceci aurait pu suggérer qu'un déplacement particulier, disons celui de /o/ à /u/, n'avait pour cause que la nécessité, pour les locuteurs, de distinguer les mots et les formes présentant le /o/ traditionnel de ceux qui présentaient un /ɔ/, au moment où le /ɔ/ envahissait le domaine du /o/. Mais comme on voulait tout expliquer d'un seul coup, on pouvait penser que le principe qui rendrait compte, un jour, d'un des chaînons permettrait de les expliquer tous, et ce principe ne pouvait être le besoin de préserver les distinctions puisque le besoin semblait satisfait avant que commence le déplacement en chaîne. En attendant la découverte de ce principe, on tentait de coller, sur l'ensemble du phénomène, une étiquette comme «fermeture», «ouverture», ou «palatalisation» qui donnait un peu l'illusion qu'on avait compris, même lorsque, comme ici, on hésitait à loger à la même enseigne l'avancée de /u/ vers /y/ et la montée de /o/ vers /u/, encore que, comme l'observation contemporaine l'a montré, il n'y ait, en réalité, pas de virage brutal sur le chemin qui mène de /ɔ/ à /y/.

[5] Références chez *Martinet, A.:* RomPh. *6:* 5–13 (1952), et *Weinreich, Uriel:* Languages in contact, 14–28 (New York 1953).

[6] Un bel exemple chez *Rogers, F. M.:* Insular Portuguese Pronunciation. Hispanic Review *16:* 1–32.

La nécessité de préserver les oppositions phonologiques ne peut s'imposer comme d'une importance fondamentale en phonétique évolutive que lorsqu'on a acquis la conviction que la phonie de toute langue est, à tout instant, en voie d'évolution, parce que l'équilibre entre l'inertie et les besoins est toujours précaire et instable. L'étude des grandes langues de civilisation qui jouent d'autant mieux leur rôle de liaison qu'elles sont plus uniformes et plus stables, nous fait souvent prendre l'idéal de stabilité pour une réalité. Mais c'est là une grave illusion qui s'évanouit dès qu'on étudie, sans préoccupations normatives et sans préjugés esthétiques, les usages linguistiques observables. Un déplacement en chaîne, comme celui que nous avons discuté ci-dessus, n'est qu'un moment d'une évolution ininterrompue, moment privilégié, si l'on veut, dans ce sens qu'il représente une réorganisation impliquée dans le système phonologique, mais freinée et stoppée, pendant un temps, par divers facteurs, comme l'imitation d'un parler directeur moins évolué, une langue commune à fortes traditions, par exemple.

Le maintien des distinctions phonologiques implique, d'une part, ce qu'on a appelé la différenciation maxima et, d'autre part, parmi les phonèmes appartenant à une même zone articulatoire continue, comme les voyelles, ce qu'on désigne métaphoriquement comme l'équidistance entre les unités distinctives. Les phonèmes d'une langue seront aussi différents qu'il est possible de l'être sans que les désavantages divers résultant de cette différenciation (articulation délicate, résultats acoustiques peu satisfaisants) l'emportent sur l'avantage résultant, pour la communication, de la différenciation: un /i/ se distinguera au mieux des autres voyelles du système en se fermant au maximum vers l'avant, mais, en tant que support de syllabe, il ne pourra pas passer à [j]. L'équidistance signifie que, dans une langue qui possède cinq phonèmes vocaliques, ces phonèmes seront articulés de telle façon qu'ils soient acoustiquement également distincts les uns des autres; l'équidistance est celle qu'on constaterait sur un diagramme qui viserait à représenter les relations acoustiques entre les phonèmes. Le principe d'équidistance se heurte très vite à des résistances de types divers qui en limitent l'application. Il y a des résistances dues à la variété et à l'asymétrie des organes qui réduisent beaucoup les zones continues où pourrait se manifester l'équidistance: même dans le champ vocalique, privilégié à cet égard, l'économie articulatoire évidente que représente le choix du même angle d'ouverture du maxillaire pour les voyelles



d'avant et celles d'arrière doit entraîner un écartement plus considérable à l'avant et une «distance» acoustique plus grande entre [e] et [i] qu'entre [o] et [u]. Le choix du même angle d'ouverture n'est pas une vue de l'esprit, comme le montre l'étude des faits de diphtongaison. La tendance à l'équidistance se manifeste par la fréquence des systèmes où la série d'avant a plus d'unités que la série d'arrière, c'est-à-dire où un type d'économie (tendance à l'équidistance) l'emporte sur l'autre (identité des angles d'ouverture). Il y a donc là, pour tout système phonologique, une source possible de déséquilibre. D'autres entorses à l'équidistance pourront résulter du fait que certaines oppositions sont, en pratique, plus importantes que d'autres, comme nous le verrons plus loin.

Supposons tout d'abord que les distinctions phonologiques soient toutes également utiles au fonctionnement de la langue et que la préservation de chacune d'entre elles soit également désirable. La solution idéale serait évidemment le maintien du *statu quo* une fois qu'un équilibre satisfaisant a été trouvé. Mais nous venons de voir que, même si les pressions de langue à langue ou de dialecte à dialecte étaient exclues, même si les besoins de la communication étaient supposés immuables, il resterait, dans l'asymétrie même des organes de la parole, une source d'instabilité. L'expérience montre

que certaines langues conservent de larges pans de leur structure phonologique sans y rien changer pendant des siècles, mais aussi que toutes celles qui restent des moyens normaux de communication orale présentent immanquablement une phonie en voie d'évolution sur quelque point.

Ces changements en progrès ne tendent pas nécessairement à éliminer certaines oppositions et à en préparer d'autres. Il peut y avoir simplement modification de la nature articulatoire et acoustique d'un ou de plusieurs phonèmes, dans toutes les positions où ils apparaissent ou dans certaines positions seulement. De telles modifications peuvent être dues à l'imitation de locuteurs d'une autre langue; par exemple, le remplacement de [r] par [ʀ] ou [ʁ] dans une grande partie de l'Europe. Bien entendu, il peut y avoir, dans ce cas, non point adoption d'un son nouveau, mais imitation d'un processus menant à ce son nouveau, une diphtongaison en cours d'établissement par exemple.

Ces modifications pourront être aussi un temps particulier d'une réorganisation en cours; à la suite d'un déplacement en chaîne de /ɔ/ à /y/, l'articulation d'un /a/ peut reculer vers l'espace laissé vacant par la «montée» de /ɔ/ si l'équidistance le demande. Une telle réorganisation peut s'étendre sur des siècles ou des millénaires; celle qui a consisté, pour le roman d'Espagne, à remplacer les géminées héritées du latin par des articulations de types divers, de quantité non distinctive, sans les confondre (sauf dans le cas de *-m-*/*-mm-*) avec les simples correspondantes, a dû commencer il y a plus de mille ans et ne sera terminée que lorsque /r̄/ et /r/ (dans *cerro* et *cero*) se seront, dans tous les usages de la langue, qualitativement différenciés[7]. A l'origine de réorganisations de cette espèce, qui comprennent ce qu'on appelle les mutations comme celle qui a abouti au consonantisme du germanique ancien, il peut y avoir un contact de langue (substrat); mais on peut aussi supposer un conditionnement interne impliquant parfois des répercussions sur le plan phonologique d'évolutions sur d'autres plans de la langue, par exemple, une évolution du système morpho-syntaxique entraînant une modification du système accentuel et, par contre-coup, celle du système phonématique.

On peut enfin imaginer, pour de telles modifications, d'autres conditionnements; une mode, par exemple, qui favorisera telle dé-

[7] Cf. *Martinet, A.:* Economie des changements phonétiques, 273–288 (Berne 1955).

viation accidentelle; mais le cas diffère en fait assez peu de celui, déjà considéré, de l'emprunt à un autre idiome.

Les modifications du type que nous venons de considérer ne changent pas le nombre des phonèmes de la langue ou, plus exactement, s'il s'agit de modifications limitées à certains contextes, le nombre d'oppositions phonologiques disponibles dans une position déterminée. On ne peut pas dire cependant que le système ne soit pas affecté, puisque les rapports des phonèmes dans le système peuvent fort bien être tout autres, à l'issue du processus: lorsqu'en haut-allemand, un /d/ ancien devient /t/ (anglais *do*, allemand *tun*), il acquiert un trait, la sourdité, qu'il a désormais en commun avec /p/ et /k/ et perd celui de sonorité qu'il partageait précédemment avec /b/ et /g/; c'est une réorganisation du système qui a permis à l'ancien /d/ d'occuper la place d'un ancien /t/ passé à /ts/ ou /ss/ dorsoalvéolaire; les phonèmes ne se sont pas confondus dans le système, ils ne se sont pas télescopés dans la chaîne parlée, mais ils sont désormais dans des rapports nouveaux dont on devra tenir compte si l'on veut comprendre l'évolution à venir.

Il y a d'autres changements phonétiques, parmi les plus fréquemment attestés, qui peuvent aboutir à modifier le nombre des phonèmes, aussi bien dans la chaîne que dans le système, sans cependant que les latitudes distinctives de la langue en soient affectées, c'est-à-dire sans que, de ce fait, un segment quelconque du discours puisse jamais en venir à se prononcer de la même façon qu'un autre qui était au départ différent. Il s'agit du transfert d'un segment à un autre d'un trait distinctif. Soit une langue qui présente trois phonèmes vocaliques /a/, /i/ et /u/; dans certaines conditions, ailleurs que sous l'accent par exemple, la voyelle perd son timbre propre et le transfère sur la consonne qui précède; /–ta/, /–ti–/ et /–tu–/ deviennent respectivement, dans ce cas, /–tə–/, /–t'ə–/ et /–tʷə–/. La langue voit, de ce fait, le nombre de ses phonèmes consonantiques multiplié par trois; tout ce qui y était distinct au départ demeure distinct à l'arrivée; mais elle n'acquiert pas, au cours du procès, des possibilités nouvelles de distinguer une forme d'une autre. Soit encore une langue où, à la finale de syllabe, le choix du type de consonne nasale est toujours déterminé par le contexte (c'est la situation en espagnol, par exemple); une modification se produit qui consiste à anticiper l'abaissement du voile du palais, caractéristique de la nasale, de façon à le faire coïncider avec la voyelle précédente; /–anta/ passera donc à /–ãta/; toutes les consonnes nasales implo-

sives disparaîtront, ce qui diminuera considérablement le nombre des phonèmes successifs de la chaîne, mais multipliera par deux le nombre des phonèmes du système vocalique, puisqu'à chaque phonème oral correspondra désormais un phonème nasal. Cependant les latitudes distinctives de la langue n'auront pas changé. Cette fois-ci, le système est modifié non seulement qualitativement, mais quantitativement, et, bien entendu, ces modifications seront décisives pour l'évolution qui suivra.

Contre le point de vue de ceux qui mettent en relief l'importance, pour l'explication de l'évolution phonétique, de la préservation des distinctions, on fait souvent valoir l'existence, voire la fréquence, des confusions de phonèmes. Puisque, argue-t-on, les phonèmes sont là pour assurer des distinctions, s'il est prouvé qu'ils peuvent se confondre, n'en peut-on conclure que l'évolution phonétique est aveugle, ou, en d'autres termes, qu'elle se produit sans égard à la fonction des unités distinctives? Cet argument serait décisif s'il était vrai, de tous les points de vue, qu'une opposition phonologique qui ne sert à distinguer qu'une seule paire de mots est à mettre sur le même plan que celle qui assure à elle seule la distinction de centaines de quasi-homonymes. Du point de vue de la description synchronique qui doit aboutir à un système graphique marquant tout ce qui peut différencier un mot d'un autre, une forme d'une autre, ce principe est parfaitement justifié. Tant qu'un Français peut, s'il le veut et même s'il ne le fait pas normalement, distinguer dans le discours entre un *mètre* /metr/ et un *maître* /mētr/, il est du devoir de celui qui décrit la phonologie du français général de signaler l'opposition d'un /e/ bref à un /ē/ long en syllabe fermée et de prévoir des notations adéquates. Mais lorsqu'on considère le même problème sous un angle évolutif, il est indispensable de distinguer, d'une part, les oppositions phonologiques largement utilisées, /p/ ∾ /b/ en français, par exemple, d'autre part, parmi celles qui servent peu, les oppositions rarement mises à profit, mais qui sont stables parce qu'elles se fondent sur la présence, ou l'absence, d'un trait distinctif largement utilisé par ailleurs (/θ/ ∾ /ð/, /ʃ/ ∾ /ʒ/ en anglais), et celles qui, servant peu et étant d'un type isolé dans le système, sont, en fait, en voie d'élimination (/e/ ∾ /ē/ en français).

Lorsqu'on observe, dans une langue contemporaine, l'élimination d'une opposition phonologique, on constate qu'elle se produit lorsque la confusion des deux phonèmes ne peut plus affecter très sérieusement la compréhension de ce qui est dit. Il ne pourrait y avoir d'exceptions que dans le cas où l'élimination se réalise par imitation de certains traits, statiques ou dynamiques, d'une langue de prestige: une opposition /r/ ∾ /r̄/, analogue à celle du castillan, qui rend quelques services dans certains dialectes de France, s'y maintient mal sous la pression du français général qui a éliminé cette opposition depuis longtemps[8]; le processus constaté, qui n'est pas simple (/r/ ∾ /r̄/ > /r/ ∾ /ʀ·/ > /r/ ∾ /ʀ/ > /ʀ/), doit offrir la possibilité de se protéger, par des innovations lexicales, des conflits homonymiques qui pourraient en résulter.

Il y a trop de cas où se vérifie l'hypothèse que la survie d'une opposition dépend, pour une part, de ce qu'on appelle son rendement fonctionnel (*funktionelle Belastung*, *functional yield or load*) pour qu'on puisse l'écarter, même si l'on ne s'est pas mis d'accord sur la meilleure façon d'évaluer le rendement d'une opposition. On ne saurait en fait se prononcer sur ce qui est décisif en la matière: la fréquence des cas où une négligence dans la réalisation correcte de l'opposition entraînerait réellement l'incompréhension (elle a les *cheveux blonds* /blõ/, elle a les *cheveux blancs* /blã/) ou la fréquence générale des deux phonèmes dans les mêmes contextes indépendamment des conflits réels aboutissant, de la part de l'auditeur, à une demande d'explication. Sur le plan de la méthode et sans se prononcer sur le fond de l'affaire, on recommandera le procédé le plus simple, c'est-à-dire un relevé de fréquence dans les textes, à condition de traiter à part des différents contextes.

La vérification de l'hypothèse relative à l'influence du rendement fonctionnel sur le sort de l'opposition ne peut se faire que sur des langues dont on peut observer directement le fonctionnement. On peut l'appliquer à des évolutions historiques dans la mesure où l'on est convaincu de sa validité. Mais on ne peut faire valoir contre elle des exemples empruntés à des états de langue disparus pour lesquels il est difficile de réunir la documentation nécessaire à toute vérification sérieuse et où l'on est le plus souvent en peine pour identifier les étapes successives du phénomène; les quatre étapes du déroulement de l'élimination de /r/ ∾ /r̄/ indiquées ci-dessus ont été constatées au même moment chez des gens dont l'âge s'échelonnait de 70 à 40 ans. Une observation directe, mais moins attentive, aurait pu faire croire que /r/ et /r̄/ s'étaient confondus directement

[8] Cf. *Martinet, A.:* La description phonologique, 64–67 (Genève/Paris 1956).

en /ʀ/, ce qui aurait exclu tout rapprochement avec les processus de remplacement de /r/ ∾ /r̄/ par /r/ ∾ /ʀ/ constatés de l'Amérique du Sud à la Suède centrale[9]. Dans tous les cas où une opposition phonologique se neutralise dans une situation donnée ou est totalement éliminée, ainsi que dans celui où un phonème disparaît par amuissement, on doit toujours envisager la possibilité que le trait distinctif éliminé (ou l'un des traits s'il s'agit d'un amuissement) ait été transféré sur un voisin dans les conditions exposées ci-dessus. Ce trait, le trait nasal par exemple, peut se fixer plus ou moins définitivement (français *Martin*, *catalan*) ou être assez vite éliminé (catalan *Marty*, *català*). La combinaison du trait distinctif en cause avec ceux du phonème auquel il s'ajoute, ne donne pas nécessairement un produit d'excellente qualité, ni du point de vue de l'articulation, ni de celui de la perception: les voyelles nasales dans la production desquelles tout l'air qui passe par les fosses nasales est perdu pour l'identification de l'articulation buccale, sont des combinaisons peu stables, soit que leur timbre spécifique tende à se modifier (comme dans le français /æ̃/ de *vin* venu de [i] nasal), soit que la nasalisation y disparaisse comme elle a dû le faire en catalan. Il n'est nullement invraisemblable que l'élimination de la nasalité dans une opposition /i/ ∾ /ĩ/ soit précédée d'une période où la différence, souvent insuffisante dans la pratique, entre l'orale et la nasale ait amené les locuteurs à prendre des précautions en éliminant toute quasi-homonymie dangereuse, comme les Gascons ont évité le danger de l'homonymie de *gat* «chat» avec **gat* «coq» en remplaçant ce dernier par *hazan* ou *bigey* avant même, sans doute, que l'ancien **gall* ait effectivement abouti à **gat*. Le transfert latéral d'un trait pertinent ne serait donc souvent qu'une solution provisoire.

En face d'une élimination partiellement déterminée par le faible emploi d'une opposition, il faut peut-être envisager des cas d'amuissement où une excessive fréquence d'un phonème aurait entraîné son affaiblissement dans certaines positions, celles où un affaiblissement est le plus vraisemblable. On pense au cas du *s* implosif de l'espagnol d'Andalousie et de certaines contrées d'Amérique. La faiblesse des implosives est presque un trait général du parler humain, bien qu'elle se manifeste de façon très variable selon les langues. Elle est particulièrement nette en espagnol contemporain, ce que nous constaterons ici sans chercher à l'expliquer. Or le *s* est,

[9] Cf. *Malmberg, B.:* Studia Linguistica *15:* 100–101 (1961).

dans cette langue, d'une toute particulière fréquence à l'implosion, notamment du fait de son utilisation comme marque de pluriel. On pourrait donc supposer que la faible information entraînée par la haute fréquence a abouti à un affaiblissement en un [h] qui, à son tour, peut s'amuir. Comme toutefois le pluriel *las mesas* ne saurait se confondre avec *la mesa*, la disparition de [h] s'accompagne de différenciation du timbre des voyelles lorsqu'elles étaient suivies de /s/ > [h][10]. Ce qui renforce l'hypothèse que la fréquence du phonème /s/ a pu avoir son mot à dire, c'est que les dialectes d'Andalousie, ceux qui ont le plus profondément influencé les formes transatlantiques de l'Espagnol, ont une fréquence de /s/ bien supérieure à celle des autres usages de la Péninsule, car les deux phonèmes /s/ et /θ/ s'y sont confondus (dans certaines zones sous la forme de /θ/, par nature plus susceptible encore de s'affaiblir que [s])[11]. Il n'y a pas conflit entre l'hypothèse relative au rôle du rendement fonctionnel et celle selon laquelle un phonème d'une extraordinaire fréquence pourrait s'affaiblir; l'opposition /h/ ∾ zéro est probablement tout à fait satisfaisante du point de vue distinctif, tant qu'elle se maintient. Ce qui peut créer une situation délicate c'est le fait que l'évolution [s] > [h] est irréversible, que la seule évolution connue pour [h] est l'amuissement, et que les oppositions de timbres vocaliques qui, en s'accentuant, pourraient fournir une solution permanente au problème des distinctions morphologiques et lexicales à préserver, tendent probablement à s'éliminer du fait du prestige du castillan et de son vocalisme élémentaire. Mais il n'y a pas de téléologie dans le fonctionnement de la langue. Ceux qui ont laissé passer [s] à [h] n'avaient évidemment aucun moyen de s'imaginer le mauvais service qu'ils rendaient à leurs descendants.

L'apparition de nouveaux phonèmes dans une langue ne se produit sous l'influence directe de nouveaux besoins distinctifs à satisfaire que dans le cas d'emprunts de mots comportant des articulations inconnues à la langue emprunteuse: /ŋ/ dans le français *meeting*, /ã/ et /ʒ/ dans l'allemand *rangieren*, etc. Ces emprunts sont, on s'en doute, facilités quand les traits distinctifs qui assurent l'identité du nouveau phonème préexistaient dans le système: l'articulation de [b] est relativement facile pour ceux qui connaissent une opposition /p/ ∾ /t/ et une opposition /t/ ∾ /d/. Emprunts mis à part, les nouveaux phonèmes résultent nécessairement de variations

[10] Cf. *Navarro Tomás, T.:* TCLP *8:* 184–186 (1939).
[11] Cf. *Zamora Vicente, A.:* Dialectología española, 236–244 (Madrid 1960).

contextuelles qui, de façon ou d'autre, se trouvent, à un moment donné, ne plus être entièrement déterminées par le contexte qui les a fait naître: si, dans le contexte /–ati/, /t/ se prononce [t'] (et, en conséquence, /–ati–/ se prononce [–at'i]), son statut phonologique n'a pas changé; mais si, sur ces entrefaites, /–i/ final tombe régulièrement et si le /t/ de /–ati/ continue à se prononcer [t'], /–ati/ deviendra /–at'/ et la langue aura acquis un nouveau phonème /t'/. Un processus de ce type aboutit à réduire le nombre des phonèmes successifs du discours et à multiplier le nombre des phonèmes du système. Il représente, pour ceux qui le réalisent, une économie, puisqu'il consiste pour ceux-ci à ne plus articuler certains segments du discours. Ils n'ont pas à apprendre à articuler et à distinguer de nouveaux phonèmes, puisque les articulations en question leur étaient naturelles. Mais la situation est tout autre pour leurs descendants qui, eux, doivent continuer à apprendre, au cours de leur première enfance, à distinguer entre /i/ et les autres voyelles comme l'ont fait leurs parents, mais qui, en plus, vont avoir à s'efforcer d'articuler de façon distincte un /t/ et un /t'/, un /p/ et un /p'/, etc.

Il y aura nécessairement une limite à l'accumulation des distinctions phonématiques qui résultent, au cours des siècles, des processus de transfert des traits distinctifs. Ces processus seront alors freinés ou stoppés, au moins pour un temps, en attendant que les moins utiles parmi les distinctions phonologiques existantes aient été éliminées: à l'aube du XVIII^e siècle, il semble que le français de Paris ait présenté vingt, et peut-être vingt-quatre phonèmes vocaliques différents[12]; les jeunes Parisiens d'aujourd'hui n'en utilisent guère plus de treize. Les oppositions qui ont été éliminées ne l'ont pas toujours été sous la forme qu'elles avaient en 1700; chacune a eu son histoire particulière, bien que, pour la plupart d'entre elles, cette histoire entre dans le vaste chapitre de l'élimination de la quantité.

De tout ceci, on retiendra surtout que ce qui est économique pour celui qui réalise le changement représente fréquemment une complication pour les générations à venir: il y aura toujours des économies à réaliser en transférant des traits distinctifs sur les phonèmes voisins de la chaîne et, ultérieurement, d'autres économies en éliminant du système les oppositions de faible rendement.

Il est clair toutefois que ces conflits internes qui, à eux seuls per-

[12] Voir *Martinet, A.:* BSL *43:* 15–21 (1946).

mettent de comprendre la permanente instabilité des systèmes phonologiques, ne sont pas seuls en cause; il reste indispensable d'attirer longuement l'attention sur eux, puisque l'enseignement traditionnel n'en faisait pas mention. Mais ceux-là mêmes qui insistent pour qu'on leur accorde l'attention qu'ils méritent seraient les premiers à protester si l'on devait négliger les facteurs qui, de l'extérieur, peuvent contribuer à déséquilibrer le système phonologique, qu'il s'agisse de l'influence de systèmes concurrents ou des répercussions de l'évolution des besoins communicatifs de la communauté, eux-mêmes sous la dépendance directe de l'évolution de la société. Il n'entre pas dans le cadre du présent examen de préciser comment l'évolution de la société influence la première articulation du langage, c'est-à-dire la façon dont les locuteurs analysent leur expérience en unités significatives successives. Cette influence, évidente en matière de lexique, n'est pas niable en ce qui concerne les traits grammaticaux, ne serait-ce que parce que unités grammaticales et unités lexicales se complètent pour couvrir une même domaine, et que la grammaticalisation d'un domaine sémantique, comme le temps, a nécessairement des répercussions sur la fréquence et l'inventaire des unités lexicales du même domaine. C'est essentiellement par le chenal de traits prosodiques comme l'accent et de faits d'expressivité que l'évolution des besoins communicatifs se répercute jusque dans le système phonologique. Il convient, en la matière, de se défaire du préjugé qui voyait, dans l'accent, une cause première, un phénomène inexplicable, éclatant comme l'orage dans un ciel d'été. Sans doute y a-t-il des cas où un certain type accentuel a dû être emprunté à une autre langue et où l'influence du modèle a été assez forte pour s'imposer *aux dépens* de la langue imitatrice. Mais, dans chaque cas, on retiendra, tout d'abord, l'hypothèse que c'est l'évolution même de la langue et des besoins de la communauté qui la parle qui a entraîné une réorganisation du système accentuel. On dira, de façon un peu sommaire, mais assez juste, que la syllabe qui a reçu l'accent est celle qu'il convenait, pour le succès de la communication, de mettre en valeur.

Il n'est pas indispensable de rappeler ici les changements phonétiques qui peuvent coïncider avec une réorganisation du système accentuel. Mais on attirera l'attention sur l'influence des procédés expressifs, comme l'allongement ou la gémination. Ces procédés, véritables modes, dont l'apparition dans telle ou telle langue est favorisée par la structure même du système phonologique, peuvent

prendre une extension considérable. Si leurs effets se fixent, c'est-à-dire en viennent à caractériser de façon permanente certains éléments du vocabulaire, ils aboutissent à déséquilibrer le système en bouleversant la fréquence respective des unités distinctives. Les effets de tels bouleversements peuvent se répercuter à travers des millénaires.

La formation philologique qui reste celle de beaucoup de linguistes prépare mal à concevoir le fonctionnement de la causalité interne des systèmes phonologiques. Il faut, pour le comprendre, observer les échanges linguistiques tels qu'ils ont lieu en fait autour de nous, et, partant de là, essayer de s'imaginer la façon dont les locuteurs d'une époque révolue ont résolu les problèmes que posait alors la compréhension mutuelle. Soit un déplacement en chaîne, celui par exemple que l'on constate dans l'ancien roman de l'ouest et qui amène /-d-/ à /-ð-/, /-t-/ à /-d-/, /-tt-/ à /-t-/. L'observation semble indiquer que l'affaiblissement, par spirantisation ou voisement des intervocaliques, ne se produit que là où existent des géminées de fréquence comparable aux simples. L'information fournie par les unes est donc analogue à celle qui est fournie par les autres, et l'on peut s'attendre à ce que les locuteurs tendent à affaiblir l'articulation des géminées jusqu'au moment où le rapport de l'énergie dépensée à l'information fournie sera analogue à celui qui existe dans le cas des simples. Tout ceci semblerait indiquer que ce sont les géminées qui ont amorcé le processus en s'affaiblissant, ce qui a déterminé, de proche en proche, le voisement de la simple sourde et la spirantisation de la sonore. Certains objectent que les géminées ne peuvent avoir été les premières à se déplacer, puisque l'on constate, par l'examen des documents, que la simplification des géminées est ultérieure au voisement et à la spirantisation des simples. Mais c'est oublier que la préservation des distinctions réclame que les géminées ne soient définitivement simplifiées que lorsque les /-t-/ seront parfaitement voisés chez tous les locuteurs et en toute circonstance, ce qui implique qu'au préalable tous les /-d-/ se seront dûment spirantisés. Il suffit qu'une classe de la société ou que quelques cantons manifestent un attachement à la tradition pour que le processus soit freiné, voire même stoppé. Il ne faudrait pas qu'à la phonétique «sur le papier» d'une époque révolue succédât une phonologie qui ignore les situations sociolinguistiques réelles. La *gorgia* toscane[13] désigne l'affaiblissement en spirantes des occlusives et des affriquées intervocaliques d'où il résulte que *la casa* se prononce [laxasa] et *la cena* [la ʃena]. Lorsque son histoire aura été parfaitement restituée, la *gorgia* sera l'illustration parfaite d'une mutation déterminée, au départ, par une tendance des géminées à s'affaiblir et qui a abouti à spirantiser (et non à «aspirer» comme on s'obstine à l'écrire) la consonne simple correspondante là où elle est en opposition avec la géminée. Cependant, continuellement battue en brèche par l'influence de la graphie et celle des usages romains, elle n'a jamais pu aboutir à la simplification des géminées, puisqu'il y avait toujours des locuteurs pour qui /-k-/ intervocalique restait [k] et qui auraient mal interprété un /-kk-/ trop affaibli. Puisque le /-k-/ simple pouvait se réaliser aussi bien comme [k] que comme [x], la géminée correspondante devait conserver une articulation prolongée. Il est, bien entendu, de moins en moins question de simplifier la géminée puisque la prononciation [x] pour /-k-/ est en sérieuse régression.

Quelle que soit l'évolution phonétique qu'on étudie, que l'on soupçonne au départ l'action d'une autre langue, l'imitation d'un processus, le remplacement mot par mot d'une articulation par une autre jusqu'à élimination complète de la première, la pression des besoins lexicaux ou grammaticaux, une modification quelconque de la fréquence de certaines catégories phonologiques, on ne saurait oublier que tout se tient dans une langue et qu'aucun changement ne s'y produit dans le vide. Les hypothèses qu'il faudra s'efforcer de vérifier dans chaque cas devront toujours se fonder, non sur d'autres hypothèses, même si ces dernières ont pour elles le support d'une longue tradition, mais sur une observation attentive du comportement linguistique des êtres humains.

Adresse de l'auteur: M. André Martinet, 3, Place de la Gare, *Sceaux*, Seine (France).

Discussion

Pilch (Freiburg): In aller Bescheidenheit möchte ich mir zwei kritische Bemerkungen erlauben zu Fragen, die m. E. innerhalb von *Martinets* Theorie weiterer Ausarbeitung bedürfen:

1. Die Termini *équidistance*, *angle maxillaire* u. dgl. sind offenbar artikulatorische bzw. akustische Metaphern für auditive Realitäten im Sinne der «imitation label technique». Nur so kann es etwa zu verstehen sein, daß mehr Platz für vordere als für hintere Vokale vorhanden sei. An der tatsächlichen Artikulation ist der gesamte Artikulationskanal beteiligt, nicht nur die Vorderzunge.

[13] Dans son exposé du problème, *Weinrich, H.:* Phonologische Studien zur romanischen Sprachgeschichte, 105–174 (Münster 1958), a malheureusement pris au sérieux le terme traditionnel d'«aspiration» et s'est laissé induire en erreur par une conception simpliste des mutations phonologiques.

2. Schwache funktionelle Belastung als solche halte ich – selbst bei fehlender Integration – nicht für eine ausreichende Erklärung für die Aufgabe einer Opposition (Zusammenfall zweier Phoneme). Nur wenige Oppositionen sind funktionell hoch belastet (gerechnet nach der Zahl der Minimalpaare). Die hohe Belastung anlautender englischer /p/ und /b/ halte ich für eine Seltenheit. Dtsch. /p/ und /b/ sind nur gering belastet, ebenso die distinktive Konsonantenlänge des Finnischen. Sprecher, die diese Unterschiede nicht beachten, können (unter günstigen Verständigungsbedingungen – Fehlen von Nebengeräusch) durchaus noch verstanden werden. Gewiß könnten nicht etwa alle Phoneme zu einem einzigen zusammenfallen. Einige tatsächlich vorhandene Oppositionen könnten es aber durchaus, ohne wesentliche Verständigungsschwierigkeiten zu verursachen. Vielleicht wäre dem Problem mit informationstheoretischen Methoden beizukommen.

Kiparsky (Helsinki): Question: l'opposition (e:) : (ɛ:) en français moderne est-elle en train de disparaître vu qu'on ne distingue plus entre *serai* et *serais* et même (d'après *Lewi-Strauss*, à Bloomington, Ind. au 1952) on confond *été* et *étais*, *épée* et *épais?*

A M. *Pilch:*

Die Opposition /k/ : /k k/ und ähnlich ist im Finnischen nach wie vor sehr wichtig.

Vachek (Praha): Mr. *Martinet's* synthetic presentation of the problems of historical phonology must be duly appreciated, especially by members of the Prague group who have followed analogous trends of argument since the late twenties of this century. It might be questioned, however, whether economy, though undoubtedly important, is really the prime motive of the development of the phonic and grammatical patterns of language. A large number of language facts are demonstrably non-economic (see, e.g., gender phenomena in the nouns denoting inanimate beings in Slavic languages), but rather redundant. This redundancy is, of course, motivated by the function of language to serve as means of communication even in unfavourable circumstances. It appears that non-economic factors are allowed full play in the development of language if they do not jeopardize the communicative functioning of language. If such jeopardization should occur, the factor of economy is most likely to intervene. In other words, economy does not seem to act as an initiative but rather as a controlling factor of language development.

Fourquet (Paris): Economie deckt zwei Begriffe. «Wirtschaftlichkeit» und «Ersparnis». Die Zweideutigkeit des französischen Ausdrucks ist zu bedauern.

Réponse *Martinet* à *Pilch:* Les hypothèses fonctionnelles que sont en fait la différenciation maxima et l'équidistance des phonèmes gagneront à être vérifiées par des examens établissant sur quels plans elles se réalisent effectivement. L'asymétrie des organes que suggère l'observation des articulations se vérifie dans des examens acoustiques un peu poussés (cf. le diagramme dans «Acoustic Phonetics» de *Martin Joos*).

Les arguments fixés du rendement fonctionnel valent surtout en ce qui concerne l'élimination des oppositions. Il y a des examples classiques d'opposition de très faible rendement qui se maintiennent parce que bien «intégrées» dans le système.

Le faible rendement des oppositions d'occlusives en allemand et leur haut rendement en anglais expliquent la fréquence des confusions des deux séries d'occlusives dans les usages allemand et le strict maintien de leur distinction en anglais.

A *Kiparski: Lévi-Strauss* se presse un peu trop de sonner le glas de l'opposition *é/è* en français: elle est bien vivante dans les usages parisiens et, en général, dans la France non méridionale. Dans le Midi, elle ne s'est jamais imposée. La distribution des deux phonèmes (là ou ils sont distincts, c'est-à-dire à la finale absolue) peut varier, certes, d'une région à une autre de la France non méridionale, d'un usage à un autre, mais *épée* et *épais* restent, au Midi près, universellement distincts. L'absence de distinction assez générale entre *j'irai* et *j'irais*, je *prendrai* et je *prendrais*, etc., est un cas particulier.

A *Vachek:* Ce que l'on range sous la rubrique «Economie linguistique» dépend naturellement de la conception qu'on se fait de cette économie. Définie, selon moi, comme l'équilibre entre l'inertie et les besoins changeants de la communication, elle n'exclut absolument pas la redondance qu'on sait être indispensable au bon fonctionnement du langage et à l'apprentissage du langage par l'enfant. On comprend pourquoi on insiste surtout sur les usages prosaïques du langage lorsqu'on traite de l'économie, mais les problèmes de la communication littéraire et poétique qui sont un peu spéciaux n'en figurent pas moins dans un traitement général de l'économie linguistique. Les situations bilingues entrent très naturellement dans le même cadre.

Hammarström (Uppsala): Il est sans doute opportun de souligner que l'hypothèse de l'importance du rendement fonctionnel fait partie d'un complexe d'hypothèses (et M. *Martinet* est le premier à le savoir) et qu'un changement déterminé peut se produire même si le rendement d'une opposition est élevé. Ainsi à Stockholm la différence /ɛ:/ – /e:/ disparaît.

Proc. 5th int. Congr. phon. Sci., Münster 1964, pp. 105–119
(S. Karger, Basel/New York 1965).

University of Michigan

On the Grammar of Intonation

By KENNETH L. PIKE

Since the publication of my first structural studies of English intonation[1] various theoretical principles have developed in connection with my work with grammatical and phonological data. This paper suggests how these principles might illuminate, in retrospect, some of the empirical problems reported earlier. The over-all approach embodying these concepts I shall call tagmemic[2] theory.

1. *Particle, Wave, and Field*

Perhaps the assumption most crucial to the tagmemic approach is that language structure – and all of life's behavior – is far too complex to be seen completely through any one simple model. Three perspectives must be utilized. The same data at different moments are best seen as *particle*, and as *wave*, and as *field*.

Nor is one perspective independent of another. The insights gained are neither separable from each other, nor simply additive. When one approach comes into focus, the other two must necessarily comprise its background. No logical progression can lead from one – isolated – to the others. They complementarily reside simultaneously in permanent relation. If one is taken as starting point, the others are somewhere, in some sense, present as undefined terms or as unstated assumptions.

---

[1] See Lit. 7, Vol. 1: Pronunciation; revised and incorporated in my *Intonation of American English* (Ann Arbor, 1945). Page references in this paper refer to the latter volume, unless otherwise specified.

[2] Named after one of the units of the theory, the tagmeme, for which see my *Language in Relation to a Unified Theory of the Structure of Human Behavior* (see Lit. 9).

1.1 *Intonation as Particle*

The particle perspective leads one to perceive intonation segments *as if* they were static bits. Through it one finds phonemes of intonation – if his assumptions leave room for any phonemes at all. From this implicit perspective come our four levels of contrastive pitch (§ 2.1), which we shall here symbolize as E(xtra-high), H(igh), M(id), L(ow)[3].

The particle perspective yields lexical and grammatical segments, as well as phonemes. Symbolization of a primary contour such as °M-H represents a lexical particle, a morpheme. Its formal character comprises pitch movement from mid to high over the marked segment, beginning with the primary stress represented by the degree sign. Its semantic components include 'incompleteness' and 'sequence'.

The re-casting of the intonation morphemes into classes of forms filling specific grammatical roles – the presentation of grammatical tagmemes each with its relevant functional slot and appropriate class of fillers in an intonation construction or set of constructions – constitutes one priority task for grammatical analysis. The "stem" tagmeme in intonational "words" (cf. § 3) may be one such tagmeme, one particle in intonation grammar.

1.2. *Intonation as Wave*

The wave perspective leads us to see units of intonation as *dynamic*, flowing morphemic contours, or as relevant phonemic points in such contours. Peaks of the waves comprise nuclei[4] of the contours; ends of waves, or troughs between them, comprise margins of the contour units. In L-°H-L-M, for example, the nucleus is marked by the degree sign before the stressed syllable (27–28); premarginal and postmarginal components occur at L- and -M. Change points – the intonation phonemes – occur at L-, °H-, -L-, and -M; a sequence of a dozen syllables might still have only these change points, with pitches of the other syllables fitting into the curve with indeterminate levels.

[3] To correspond to our numbers 1, 2, 3, 4 – high to low – of 1945. We avoid the numbers here to lessen confusion with the works of authors who represent these levels by the symbols 4, 3, 2, 1. A glide or step from one phonemic level to another is shown as a combination of symbols, such as M-L for mid to low.

[4] See Lit. 10.

Merging of two contours leads to a syllable in double function which occur as the end margin of first and beginning margin of second contour. Note the indeterminacy – the double function – of -M- in M-°H-M-°H-L, for *The 'book of 'stories* (37, 67) from slow M-°H-M M-°H-L (where an intonation break between -M and M- makes the juncture determinate).

1.3. *Intonation as Field*

A field perspective leads us to look for dimensions[5] of contrast in networks made by the intersecting of categories of form and of categories of meaning; for system resulting from intersecting hierarchies[6] of lexicon, phonology, and grammar; for style output resulting from intersecting dynamic and intonational factors; for the intersection of voice quality[7] with the emic structures of segmental and suprasegmental characteristics. Such a task cannot be attempted in a short paper, but one suggestion may be given for each of these kinds of data.

Intersecting dimensions: Matrix 1 is constructed of primary contours from a subset of primary contours which have no internal change point. Their contrastive beginning points comprise vectors of one dimension and their contrastive end points the other. Interesting semantic relations[8] show up as shared meanings of blocks of cells in the matrix. The diagonal matric elements are all level contours. These often have meanings of 'unification' and 'implication' (61, 64). All lower diagonal elements are rising contours, with a class meaning of 'incompleteness' (51–60). Upper diagonal elements – all falling – carry the class meaning of 'attention'. The upper row and lefthand column, with E, carry 'intensity' (47) or 'surprise' (49) as class meaning – but this may be weakened or idiomatically specialized in °H-E and °M-E to 'politeness' (51, 59). The lower row (with initial L) has class meaning of 'deliberateness' (54–55) possibly with weakening on the °L-L (cf. 62–63).

Intersecting hierarchies: The place where the nuclear, stressed syllable of a primary contour may fall is in general determined not

[5] See Lit. 11.

[6] See Lit. 9, Vol. 3, and Lit. 12.

[7] See (99–103), and Lit. 9, Vol. 3, §§ 13.4, 13.81.

[8] The semantic components are drawn from the 1945 volume since we are interested in studying the old data in the light of the newer outlook. If this task can be finished, new and profitable research tasks should then appear.

	-E	-H	-M	-L
°E-	°E-E	°E-H	°E-M	°E-L
°H-	°H-E	°H-H	°H-M	°H-L
°M-	°M-E	°M-H	°M-M	°M-L
°L-	°L-E	°L-H	°L-M	°L-L

Matrix 1. Primary contours. Internal blocks represent semantic groups, for which see text.

by the system of pitch contrasts but by the element of the lexical hierarchy (27, 77, 84, 118). Such placement of stress may be changed only by intersection with a further superimposed dimension of contrastive placement of stress (84, 124) (as in *I said 'divert not 'revert*, rather than normal *re'vert*, and *di'vert*). Breaks between intonation contours often reinforce breaks between grammatical constructions rather than interrupting them, illustrating intersection of intonation with the segmental grammatical hierarchy. Reinforcement occurs in such an instance as:

The manager sold him the book
M- °H -M °H- -M M- °H-L

in which each phrase included in the clause begins and ends with an intonation contour. Clash between grammar phrasing and intonation is exploited when the last syllable of a line of a poem is lengthened when the line is without grammatical relevance (11); or when the expected pause or juncture between sentences is delayed until after the first word of the second, to prevent an interlocutor from 'getting a word in edgewise'; or when a juncture occurs after a preposition or determiner but before its accompanying noun, so that attention falls most strongly on the noun (as in some advertising announcements).

Intersecting dynamics: We have already mentioned (§ 1.2) that two contours can fuse, sharing a syllable. Here we add that a speed change forcing such smearing can itself be viewed as an intersecting dynamic factor. When such fusion is in analytical focus, a wave perspective becomes dominant in relation to the manifestation of the units in sequence.

When a style with syllable timing (35, 71, 109) leads to syllables of approximately even length, it contrasts with a more normal American English style with stress-group timing (35, 109) in which primary contours (with or without precontours) are of approximately even length (unless a double nucleus is present in the contour). Style dynamics intersect with normal intonational structure.

To some extent special emphasis (85–86) can be viewed as a style change. Greater differences, however, are brought in where formal or deliberate or casual attitudes are reflected in assignment of primary contours or their internal manifestation. When two sentences are deliberately pronounced so as to be simultaneously present in a single homophonous utterance, puns may develop which require specific intonations for the coexistent actualizations (45–46).

Voice quality differences – harshness, and so on (99–104) – add further intersecting dimensions to the total field perspective of speech.

2. *Unit*

Closely related to the assumption about particle, wave, and field is one about the nature of units. Any unit – whether an event, a thing, or a concept – is assumed to be well-known if and only if the analyst knows[9] 1. its contrastive-identificational features, 2. its range of variability with concomitant physical manifesting components, and 3. its distribution (a) as a member of class, (b) in slots of an appropriate set of environments, (c) with relation to a network of intersecting vectors of an emic system (i.e., an emic matrix).

2.1. *Contrast*

A unit itself as thus defined becomes a particle perceived (or a construct conceptualized) by an observer. An observer component implies coordinates specified from within the system for an emic view, but from outside the system for an etic view. No "thing-in-itself", studied apart from an observer relationship, is treated by the theory.

It is an observer relationship that requires that a unit be

[9] For application of the criteria to constructions, see Dimensions... (Lit. 11).

contrastive. A unit is not perceived for what it is, until it is perceived for what it is not. This theory allows a component to be used as identificational in contexts where the contrast does not occur, after it has been established as emic through use of contrastive contexts. This specification of the theory allows the equating of units from one context to some others even when contrast is neutralized.

Four contrastive levels of intonation were postulated in my analysis. Fewer would not leave room for the difference of function within any one general height of speech, and within one style. More levels added no new contrasts within such a style. (But general voice height could change a style as a whole – 76–77 – with all levels changing with it to higher, lower, spread, and so on.)

Primary contours are seen as units of a different – lexical – hierarchy of the intonation system. These contrast with each other in form (e.g., °H-L versus °M-H) and in meaning (normal attention, 44, versus incomplete sequence, 51). Tagmemes of intonation contrast in their structural role (primary contour tagmeme as the stem of intonation "word", versus precontour as intonation "prefix"; see §§ 1.1, 2.3).

Since the contrastive levels as phonemes and the contrastive contours as morphemes were treated above (§ 1.1) as particles, it becomes evident that theoretical particle and theoretical unit overlap as constructs in the theory. Particles and units result from alternative perspectives to be exploited when changing purpose makes one or the other the more advantageous. The advantage, in turn, will be determined by desirability of further relating the elements in view to an immediate context which will profit more by relation to a discussion in terms of related wave and field, or one in terms of correlated variation and distribution.

2.2 *Variability*

Just as contrastive impact is in part closely integrated with perspective as particle, so the range of variability of a unit is closely integrated with a wave perspective. Units which merge or fuse at their borders must be identified, with the help of an observer component, as waves of physical activity against an emic background. The peak of the wave becomes the nucleus; segmentation may be indeterminate or arbitrary – see above, § 1.2, and fn. 4. Since the merging into varied environments (via anticipation or delayed decay of contrastive movements in sequence) leads to a

difference in manifestation for the unit in each of its environments, range of variability is correlated with the wave character of units.

The theory requires a physical component for each unit. Neurological activity may serve as the physical component of concepts.

Yet the overlapping of phoneme wave units (from the perspective of segmented particles) may lead either to a redundancy of contrastive features or to alternate analyses of the essential contrasts. Alternate analyses of phoneme segmentation may arise when a total unit-as-wave, including its anticipatory and decay components, is analyzed as unit-as-mere-physical-segment. In this instance, one or more anticipatory components of a first wave which are simultaneous with the peak of a second wave are treated as contrastive components of the first rather than being abstracted and treated as part of the second. In such analysis the underlying assumptions may include a particle view which is much too crude, with too-sharp physical, linear segmentation without adequate wave and field components in relation to observer. An adequate view must treat a particle as itself an emic abstraction, allowing for etic smear and overlap via wave form: under some conditions, transition characteristics of a vocoid segment must be attributed to the following contoid which it in part anticipates.

Assumptions concerning the nature of definition must be revised in order to leave room, within a well-defined unit, for range of variability, wave overlap, redundancy, and indeterminacy of borders or of class membership. Specification of a unit, by this theory, requires statement of variability of a unit. The theory parts company with any treatment of definition which is satisfied whenever the definition includes just enough data to specify membership within a class and differentiation of the members of the class. These latter elements are important but insufficient for specifying enough characteristics of units to allow a member of the community to act adequately within the roles of that community. Adequate theory must lead to possibility of adequate "accentless" action. It must be generative of adequate bits – e.g., sounds and sentences – and the much larger behavior elements within which these take their appropriate place.

Specification of the variability of units is highly relevant to intonation studies. Wave characteristics of a contour such as °H-L, coupled with its application to elements ranging from one to several syllables in length, result in variant manifestations both of contour and of its included intonation phonemes. A one-syllable phrase manifests the contour as a down glide (24), a two-syllable phrase, as a down step (24). A multiple-syllable phrase appears either as a slow series of descending steps or as a series which varies from somewhat delayed – or somewhat earlier – drop (74), and so on (with indeterminacy as to the point where early drop or rise becomes contrastive).

Sound-wave frequency comprises the physical component of intonation. Relevancy, however, involves relativity to the observer (with speaker being a special instance of observer). General height of pitch (76–77) and spread of interval (76) differ according to general style or personal characteristics of the speaker, and affect the height manifestation of the intonation phonemes. These kinds of differences, also, enter the specification of the range of variation of an intonation unit.

2.3. *Distribution*

The distributional component of a well-defined unit is closely correlated with the notion of a perspective of language as field, or system. The system in which units are embedded may be considered both in a static subperspective, or in a dynamic one.

As elements of a static field the units might conceivably be defined in terms of forces of repulsion and of attraction. Emic contrast – opposition – involves psychological, observer forces which keep units apart. Negative definition[10] places units as points in such a field of forces.

Positive forces of psychological attraction would, on the contrary, draw etic elements into a single emic unit[11] as allo-units.

[10] See *Eugene A. Nida's* penetrating concept of the tokens of one morpheme as sharing a 'common semantic distinctiveness' from other morphemes (see Lit. 6). Compare *Leonard Bloomfield's* earlier negative definition, with contrastive morphemes bearing 'no partial phonetic-semantic resemblance' (Lit. 1).

[11] Which leads to *Nida's* earlier (p. 6) – and conceptually easier – use of a more traditional rule-of-thumb for morphemes as 'minimal meaningful units'. Compare *Bloomfield's* 'smallest meaningful [lexical] unit' (p. 264).

Perhaps the "positive" definition seems the simpler if one wishes purely a particle perspective. The essential nature of the negative definition becomes easier to grasp from the perspective of field. Both negative and positive are, however, necessary for well-defined particles, and both are necessary before field can be fully treated. Positive and negative perspectives become complementary in description, as do particle, wave, and field.

The allo-unit variants would be distortions of the emic units under diverse conditions of the field. The requisite conditions, in turn, would include specific orderings of the units in various sequences, where they affect one another.

Patterns of attraction may be represented as lists of emes accompanied by their respective allos. Patterns of repulsion show up as static charts in which contrastive vectors of dimensions represent a structural network.

A dynamic perspective of a unit's distribution leads to a treatment of units moving through a field. This concept is seen most easily in reference to the special case of a sequence of tones relevant to each syllable of a tone language. Let us suppose that there are four phonemic levels of tone. One tone occurs on each syllable. Each syllable may be viewed as a "fence" across the tone "field" which an utterance must pass through by way of one of the four "gates" (one of the four tones) which occur at possible choice sites – at any one syllable place – unless in some particular fence there are fewer, where contrast is "neutralized". The pitch melody picks its way through these gates[12].

English intonation units have distributional relations which can be viewed both statically and dynamically. Statically viewed, intonation contours are members of classes of contours. The primary contours comprise a class contrasting with precontours (29–30) and with composite units named as total contours (30). In a static view, also, one sees the units as appropriately filling slots in intonation constructions – as a total contour may be a construction made up of a precontour plus primary contour (e.g., M- plus °H-M for *It's a 'table*). The static view of distribution includes, furthermore, the place that a contour fills in an intonation matrix (see § 1.3).

Dynamically, on the other hand, the English intonation melody, like tone, may be treated as moving through a contrastive field of four alternative levels. At the same time, however, it moves through a set of grammatical and lexical intonational forces. The interlocking of segmental grammar with contrastive pitch, in which interpretation of segmental grammar-constituents or junctures is

[12] See figure 3 in my Operational Phonemics in Relation to Linguistic Relativity (Lit. 13). See, also, figure 5 for simultaneous multiple pathways through a complex of subsystems.

A related approach to this kind of problem is seen in *E. Colin Cherry, Roman Jakobson's* 'Distinctive Features' as the Normal Co-ordinates of a Language (Lit. 3, pp. 60–64).

affected by intonation constituents or junctures, points in this direction[13]. Similarly, the intonation system interlocks with the stress system in that primary contours of the intonation system begin with a stressed syllable. A stressed syllable is shared[14] by a unit of the rhythmic system, as a nuclear point of such a unit.

3. *Levels of Intonational Grammar*

The interlocking of intonation structure with the segmental grammar, however, should not be allowed to obscure the fact that *the intonation system has concursively a grammar of its own.* Intonation grammar involves its levels of construction, with tagmemic components in the sequence.

Morphemes (cf. 177, concerning *Harris*) of intonation, which comprise the classes of forms which fill the tagmemic slots, have contrastive forms and contrastive meanings. The formal component of an intonation morpheme is its phoneme sequence – the intonation of contrastive levels at key points in a contour. The semantic component of the intonation morphemes and of morpheme classes derive from the intersection of contrastive meanings of the semantic field (§ 1.3).

A primary contour such as °H-L typifies the "stem" of an intonation word; the precontour M-, an intonation "prefix"; the total contour M-°H-L, the "word". The stems are free, not requiring a prefix; the intonation prefixes are bound, occurring only with an intonation stem (or in a hesitation form, 32–33, 40). Intonation "suffixes" occur as resumed contours (72, 39, 41) which echo the end of a primary contour, without an additional primary stress, and without added semantic components; note *in the house* in *The class studies in the house (but...).* Postcontours are, similarly,
M- °H- -L-M L- L-M

suffixes occurring after a slight pause (cf. 65, 74, 40).

Do it your way then, he said.
M- °H- -L / -L

Compound intonation words occur when two (or more) primary stresses occur in a single rhythm unit with no intervening junctural drawl or junctural decrescendo:

[13] I have further illustrated this phase of the intonational problem in The Hierarchical... (Lit. 12).

[14] See reference to article in fn. 6.

a 'big 'table; a 'big 'black 'bug
M-°H °H-L M- °H °H °H-L

(61–62, 78, 39). Within the style and dialect recorded here (normal to me) the first stressed syllable, in my continuing perception of it, is level in pitch and stress with the next syllable. Many American scholars, however, fail to hear any such level stresses, interpreting them all as sequences of secondary-primary, or primary-secondary; when they repeat these sequences aloud to me, they fail to satisfy me, as native speaker, but make one or the other too loud and high. (I also have a few such compounds in my speech on single segmental morphemes such as *'sar'dine*, or *'um'brella* or in *'fif'teen 'men*, 77, 62.) For compound pronunciation, the first stressed syllable must be pronounced short; when drawled it breaks the expression into two contours (and other scholars then are more likely to hear the two stresses as level, with juncture between). In relation to the dynamics of rhythm units, the level stresses are treated as a double nucleus[15].

Levels higher than that of the intonation word are still obscure. It is tempting to try to relate them to pause groups. This often fails, however, since a pause may come even within an intonation word (e.g., before an intonation suffix; 33, 40–41, 74–76). Presumably, therefore, intonation phrases, intonation clauses, intonation sentences, and intonation utterance-response and discourse levels (if some or all of these emerge in analysis) must turn up as some kind of functional interrelation of intonation words in sequence. Miscellaneous instances abound, but their over-all pattern is not yet clear.

Note, for example, the unity of the taunting chant (35, 71)

Susie is a tattle tale;
°H- -H °M- -E °H- -H °M- -M

or of a descending stress series with more than four levels (70) – perhaps to be treated as an idiom, since it does not otherwise seem to fit the system; or the singsong °M-H °M-L (72); or the double rise (73–75); recurrent level contours in threat (cf. 87) or in repeated fall-rise contours (cf. 87); of alternative proposals in sequence (47); of higher unities in mathematical oral bracketing (62).

When the sequence arrangements of these contours and others

[15] See fn. 4.

have been more fully understood, it should be possible to describe the tagmemes (with their manifesting classes of contours) and the larger constructions into which they fit.

In sum, I suggest that it is time for intonational structure to be restudied in view of newly-available theoretical perspectives. These should lead to discoveries and descriptions substantially beyond our earlier structural insights.

In addition to conceptual tools through complementarity of particle-wave-field and of unit as well-known through contrast-variation-distribution, search needs to be made to see how these approaches could profit from recent empirical work of scholars such as *R. Kingdon*[16] or *W. Jassem*[17]. The transformationalist[18] approach, furthermore, might help to illuminate some of our former problems concerning the relation of intonation to potential and partially suppressed stressed (11, 87, 189[23]), and the relation of intonation to length (96-98) or juncture (30–33, 37, 40–41) or to stress in lexical and grammatical forms (78–88, 188[118]).

The largest gap in theory, so far as I can see it now, is the need

[16] See Lit. 5. *Kingdon* by his symbolism more effectively keeps a close tie between stress placement and the nucleus of a primary intonation contour than do writers such as *George L. Trager* and *Henry Lee Smith* (Lit. 15). *Kingdon* also (68–99) discusses numerous instances of "tune combinations" which should be studied for suggestions as to structural groupings which could be used to illustrate intonational clauses or other levels than that of intonation word.

[17] See Lit. 4. *Jassem* (40, 45) has some criticism of my handling of rhythm units which would in part be met by a handling of intonation in hierarchical levels as we have suggested here. His compound tonal units (54) give suggestions which would contribute toward a level higher than intonation word, but his treatment does not go further; his texts, marked for intonation (63–82) are not analyzed for possibilities of still higher structural groupings. His nuclear tunes (57–58, 60) are closer to the model of *Palmer's* work than to my primary contours. He also treats (46–47) free and conditioned variation.

The work of *Dwight L. Bolinger* is less useful here, since his assumptions about the nature of intonational structure are further removed from ours, but his comments about pitch and stress (e.g., in A Theory of Pitch Accent in English, Lit. 2) cannot be ignored in further work.

[18] For an initial attempt to add an intonational component to generative and transformational formulas see *Robert P. Stockwell* (Lit. 14). Unfortunately, for our purposes, he has an a priori assumption that ties intonation breaks somewhat too closely to grammatical ones: 'the validity of transformations which do not thus neatly predict intonational breaks may be seriously questioned' (365). This fails to leave sufficient freedom for an adequate concursive intonational grammar in which intonationally well-placed junctures sometimes occur at segmentally awkward spots; and interrogatives occur with a large variety of contours (not exclusively a rise, only, for yes – no questions – see *Stockwell* 366, versus my extensive data in 1, 53–54, 163–168, 176[40]).

for adequate explanation of the reason why several native observers reach quite rapidly a high degree of consistent agreement on the contrastive phonetic content of a segmental sequence (even though their phonemic interpretation of these data may differ) but find it much more difficult to agree on or be consistent in hearing the contrastive intonational data. Even in hearing the contrastive pitches of a tone language several foreign observers – once agreed on an analysis – may reach greater agreement than on the details of their own intonation. Presumably some theoretical insight or technology should provide a model or method which one day will allow intonation agreement as well.

References

1. *Bloomfield, L.:* Language (New York 1933).
2. *Bolinger, D. L.:* A theory of pitch accent in English. Word *14:* 109–149 (1958).
3. *Cherry, C. F.: Roman Jakobson's* 'Distinctive features' as the normal co-ordinates of a language, in For *Roman Jakobson* (Mouton & Co., Gravenhage 1956), pp. 60–64.
4. *Jassem, W.:* Intonation of conversational English (Warsaw 1952).
5. *Kingdon, R.:* The groundwork of English intonation (London 1958).
6. *Nida, A.:* Morphology[2] 7 (Ann Arbor 1949).
7. *Pike, K. L.:* An intensive course in English for Latin-American students. Vol. *1:* Pronunciation (Ann Arbor 1942).
8. *Pike, K. L.:* Intonation of American English (Ann Arbor 1945).
9. *Pike, K. L.:* Language in relation to a unified theory of the structure of human behavior. Vol. *1* (1954), 2 (1955), 3 (1960) (Glendale – Santa Ana, Cal.).
10. *Pike, K. L.:* Practical phonetics of rhythm waves. Phonetica *8:* 9–30 (1962).
11. *Pike, K. L.:* Dimensions of grammatical constructions. Language *38:* 221–244 (1962).
12. *Pike, K. L.:* The hierarchical and social matrix of suprasegmentals, in For Professor *Doroszewski* (Warszawa Phon 1962).
13. *Pike, K. L.:* Operational phonemics in relation to linguistic relativity. JASA *24:* 618–625 (1952).
14. *Stockwell, R. P.:* The place of intonation in a generative grammar of English. Language *36:* 360–367 (1960).
15. *Trager, G. L.* and *Smith, H. L.:* An outline of English structure. Studies in Linguistics. Occasional papers 3 (Norman 1951).

Author's address: Professor Kenneth L. Pike, Linguistics Department, University of Michigan, *Ann Arbor, Michigan* (USA).

Discussion

Herdan (Bristol): I shall not discuss Mr. *Pike's* conception of the grammar of intonation, but I should like to say something about the three basic concepts he mentioned: particle, wave and field, because this will enable me to show how structuralism and mathematical linguistics are connected. In order to show how these are connected with the three levels of language, I shall show a diagram from a recent paper by *Longacre* which very aptly sums up the ideas and aims of structuralism. He calls it the 9-box schema.

	Particle	String	Field
Phonology	1	2	3
Grammar	4	5	6
Lexicon	7	8	9

When I looked at this, I realized that the structuralists call particle, string and field what the mathematicians call point, line, and relation between point and line in the plane respectively I believe that it is generally accepted among linguists that the combinatorial method which has been so fully applied on the phonemic level, does not work equally well on the other two. What is here meant by the combinatorial method is that morphemes can be conceived as random combinations of the basic phonemes as units, at least in first approximation, which must be supplemented by what we know about certain preferences between phonemes, that is, by digram, trigram, etc. analysis. But as I said, when structuralists tried to extend this to the combination of words, they were so far not very successful. In doing so they naturally tried to apply the combinatorics which had proved helpful on the first level. Now this is where they went wrong. The branch of combinatorics which is suitable for 26 alphabetic units of 40 phonemes, is not the right one for using with a vocabulary of perhaps 50,000 vocabulary items, as we have for English for example. Even as a first approximation, it would here be wrong to assume complete randomness of combinations, since words are not used randomly, but selected according to meaning, and this must here be taken into account right from the start. Such methods are available now, and I should like to demonstrate their usefulness for the linguist, but the time for my intervention is not enough for that. The same applies on the grammar level. There too, the branch of combinatorics must be exactly tailored, so to speak, to the peculiarity of the situation. Such a branch is also available.

I am quite aware of the possibility that some linguists will say that all this was perhaps not really useful to the linguist. It might interest the mathematician to express the relation in question in general terms as those of points and lines in a plane, but the linguist wants to go into detail. Well, I can only assure you, not having sufficient time for demonstration, that the new concepts are more than just new names put to old ideas. They are productive of new methods of research, and these methods are such that they will enable you to go into such details as you require.

In conclusion, I should like to say this. It was the non-mathematical linguist, as structuralist, who first used the terms particle, string, field, which, in this connection, I can show stand for what the mathematician calls point, line, and the point-line relation in the plane. From what I said it follows that if the linguist really means what he says when speaking about particle, wave, and field – and, clearly, if not he should stop using these terms –, there must come the time when combinatorial mathematics can no longer be avoided.

Buyssens (Bruxelles): Je voudrais d'abord dire à M. *Pike* que je trouve son exposé un des plus beaux que j'aie entendu depuis longtemps. D'autre part, je voudrais faire une remarque. J'attache une très grande importance à la terminologie que nous utilisons pour communiquer entre linguistes. M. *Pike* a employé les termes prefixes et suffixes pour désigner autre chose que ce que l'on désigne par là en grammaire; je trouve cela regrettable. Je préfèrais infiniment qu'il trouve des termes nouveaux pour ces notions nouvelles.

Kiparsky (Helsinki): Has Mr. *Pike* succeeded in establishing general patterns of intonation for American English only or do common patterns exist for several languages?

Jassem (Poznań): One may agree or disagree with Mr. *Pike's* paralells between the various levels of structural analyses, but there is no doubt that they are stimulating and

instructive. In this connection I should like to ask whether we would agree that on the level

1. of intonation there is only one hierarchic class, viz. the pitch phoneme to correspond to two hierarchic classes on the segmental level, viz. phonemes and distinctive feature. I suggest that while a segmental phoneme is, or may be, phonetically multidimensional, and thus form a bundle of distinctive features, a pitch phoneme is unidimensional because there is only one parameter to deal with, so that no need for the establishment of distinctive features arises. Or, alternatively, could we consider intervals, relations between pitches, quasi absolute pitch ranges, etc. as distinctive features of pitch phonemes?

2. By applying some of *B. Bloch's* postulates (see Set of Postulates, etc.) it is possible, I believe, to establish structural pitch constrats purely distributionally. Very roughly, different pitches and pitch configurations that are found in the same contexts (which have all contexts in common) are free variants. Different pitches and pitch configurations that never occur in the same contexts are contextual variants. Different pitches and pitch configurations that have some, but not all, contexts in common, are contractive. An attempt to apply these principles to the analysis of Polish intonation has been made by *M. Steffen-Batogowa* in Poznań. The study is now ready for publication.

Proc. 5th int. Congr. phon. Sci., Münster 1964, pp. 120–141
(S. Karger, Basel/New York 1965).

From the Speech Transmission Laboratory, Royal Institute of Technology (KTH)
Stockholm

Formants and Cavities

By Gunnar Fant

1. Introduction

The scene is a phonetics laboratory. The investigator has just turned off his Sona-Graph, taken off the paper, and attempts to read the spectrogram. What does he see and what does he make out of it? Obviously this depends on how well he is acquainted with the language of Visible Speech[17,9,10] and all these details in electronics, acoustics, physiology, linguistics, and psychology which provide the theoretical substance of general phonetics.

In order to establish a language of Visible Speech it would be sufficient to correlate linguistic elements with spectrographic patterns. In this task one inevitably returns to questions like, what is the origin of all these details that we have made visible and what is their auditory significance?

What I am especially interested in is the theory of speech production. With some knowledge of the underlying acoustics processes the spectrogram relates a story of the articulatory and phonatory history of the utterance. Those looking for an efficient rationale for decoding speech messages out of the acoustic structure will find useful analogies and concepts in speech production. This level of specification, although not as easily accessible, is less obscured by details than is the spectrographic pattern.

Speech production once the major object of classical phonetics now enjoys a healthy renaissance and there are many sub-levels to be considered, such as the motor command or neurological level, the dynamics of muscular movements, and the dynamics of cavity size variations. With those ambitions in mind my topic here today

is rather restricted, since I am dealing with the static aspects alone. What are the formant-cavity relations in speech?

2. The Concept of Formant

First of all let us see what is meant by the terms formants and cavities. A negativistic starting point would be to declare that the whole topic is antiquated and of academic interest only since people differ so much in their concepts of formants and experts declare that any part of the spectrum of a sound is dependent on all parts of the air filled interior of the vocal tract. What is a cavity and what is not in the continuous structures involved?

The real trouble might start already when taking spectrograms since under unfavorable conditions of high fundamental pitch the formants will be poorly defined in a harmonic spectrum and the so-called broad-band filter of the Sona-Graph will portray harmonics and not formants in these instances, see Figure 1.

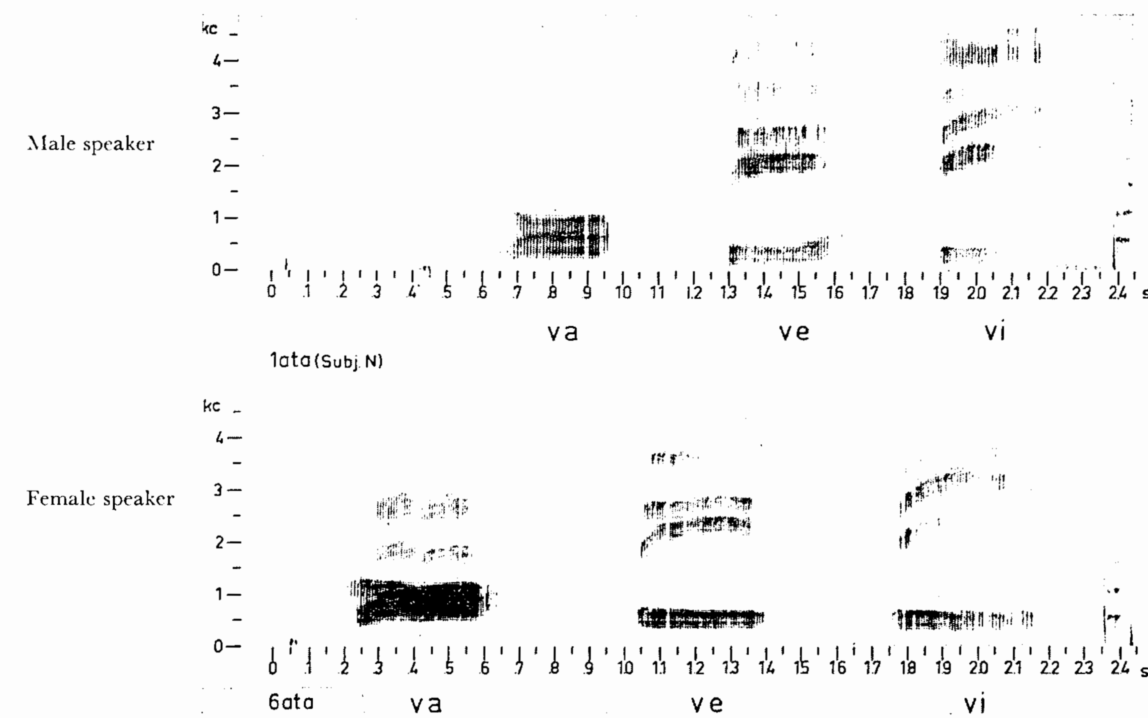


Fig. 1. Broad filter (B = 300 c/s). Sona-Graphic analysis.
A. Male voice
B. Female voice

A vowel formant of number n has the attributes of frequency F_n in c/s, bandwidth B_n in c/s[7], and amplitude level[12] L_n in dB. Of these the frequency and the bandwidth is defined as the corresponding quantities of the associated vocal tract resonance in mathematical language pole. A formant is not identical with any specific harmonic and the formant frequency is therefore not identical with the frequency of a specific harmonic. For practical purposes, dealing with male voices, we can simply and accurately measure the formant frequency as the center of the formant band of a wide-band spectrogram.

Sometimes the spectrum is obscured by two or more formants constituting a single cluster with a single peak, and a spectrum matching procedure is needed in order to decide how many vocal resonances were present and their characteristics. This view may be criticized by those preferring auditory criteria since they can claim that a formant should be defined as an auditory unit.

A single spectrum maximum should accordingly be described as a single stimulus. This point of view is indeed fruitful to follow in theories of speech perception. When studying speech structure on the acoustic level with reference to articulation, however, we need to identify formants as separate individuals even if two resonances incidently merge in a single peak. It is the demand for continuity of physical parameters intended as correlates of continuous articulatory motions and continuous perceived qualities that calls for this consistency.

To a first approximation, neglecting spectrum and voice fundamental frequency F_0, vowel qualities may be expressed in terms of formant frequencies alone, a specification in terms of three formants providing a reasonable accuracy. This type of rationale [8,16] is primarily developed for the study of non-nasalized vowels. With nasalization present it is generally possible to find out from continuity considerations what is a formant of the true vowel system and what is not. The frequencies of these formants, the so-called F-pattern, may also be traced without too great trouble in liquids and in semivowels and in voiced fricatives.

When dealing with nasal murmur sounds there are obvious difficulties and a new set of formants (constituting an N-pattern) is adopted. The established view[7,13] on the number of poles and zeros in nasal murmur segments is fit for a revision. Recent experiments performed by *O. Fujimura*[14] at the Speech Transmission Laboratory

have revealed several more poles and zeros than anticipated in our previous studies.

In stops and in fricatives the F-pattern is often identifiable at least in vocalic boundary regions. The formant structure of the higher part of the spectrum, e.g. the region above 4000 c/s in fricatives could of course also be described in terms of the F-pattern. In this frequency region it is more practical to make use of larger units in terms of prominent parts of the spectrum without reference to the detailed peak structure. Spectrum matching techniques and synthesis experiments are helpful in arriving at valid approximations.

3. Vocal Cavity Structures

The procedure for investigating cavity formant relations should start with a study of what these cavities of the speech organs look like. Secondly the description must be simplified in order to arrive at a model that can be subjected to calculations. Finally certain systematic variations in the geometry of the model should be introduced in the calculations in order to find out which aspects of the structures determine a specific pattern of formant frequencies. Out of this general study there can be derived statements such as to what extent a specific formant frequency under specified articulatory or acoustic conditions will be dependent on a specific cavity or structure or any other characteristic of the compound resonator system.

What I have to say here is essentially a summary of some of the findings reported in ref.[7]*. In addition, I have samples of more recent physiological and acoustical data, especially of nasals and quite recently gained evidence of the significance of cavity wall vibrations as a factor determining formant frequencies.

The first step in any calculation is to estimate the cross-sectional area perpendicular to a central place coordinate in the vocal tract, running from the lips to the glottis. Typical examples of a tracing from a sideview of the subject and a corresponding "area function" in continuous and quantized form is shown in Figure 2. In this example the calculations were made from an area function divided in 21 successive parts of varying length and cross-sectional area.

The calculations performed on the X-ray material of Russian speech analyzed in ref.[7] were more successful than anticipated in

* Some of the material of ref.[7] is reviewed in detail in ref.[19].

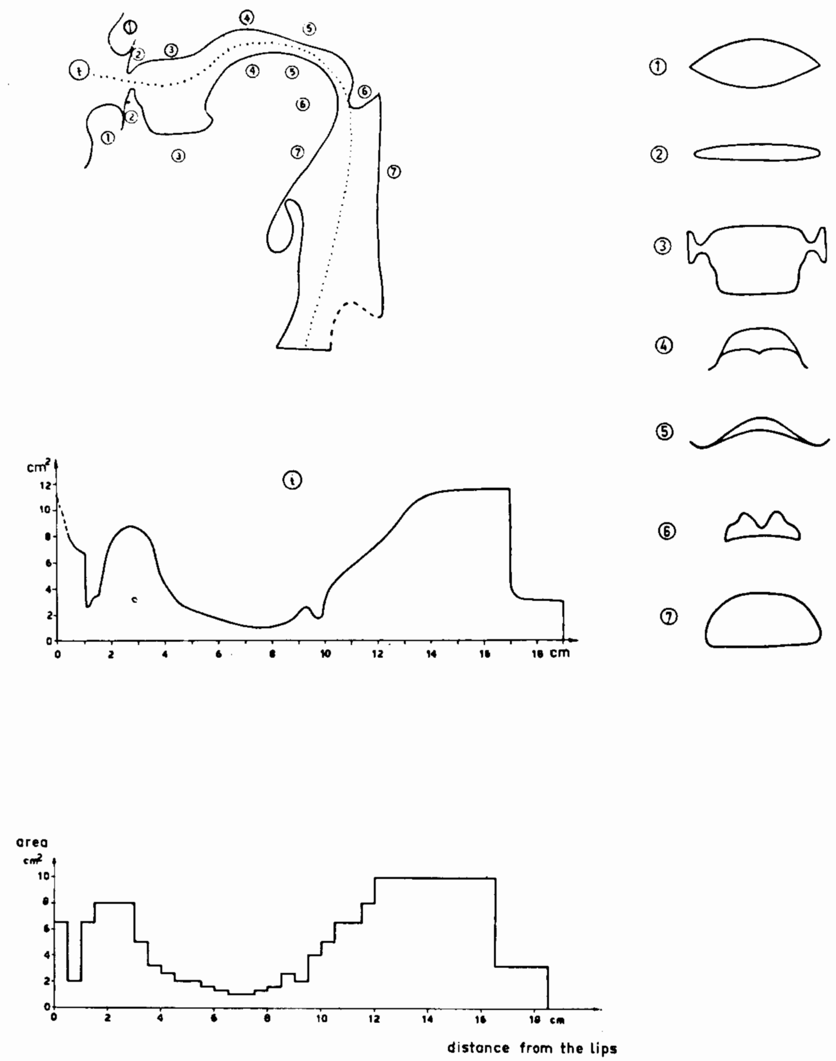


Fig. 2. The procedure for arriving at a stepwise approximation of the vocal tract area function from the X-ray data. Cross-section figures are made for a number of planes perpendicular to the central line from the glottis to the lips. A continuous outline of the area function is drawn and finally the area data within successive 0.5 cm sections are quantized according to the requirements of either the numerical calculations or those performed with LEA. (See ref.[7].)

view of the limited insight available of the lateral dimensions, i.e. from the right to left side of the subject.

The fact that the calculations came out very neatly does not guarantee that the physiological interpretations were correct in all

details although the major assumptions with respect to lateral dimensions appear to have been valid.

One obvious difficulty was to gain any insight in the structure of the nasal pathways. The calculations in ref.[7] were carried out on a nasal area function, Figure 3, derived from anatomical data* fitted to observable overall dimensions of the Russian subject.

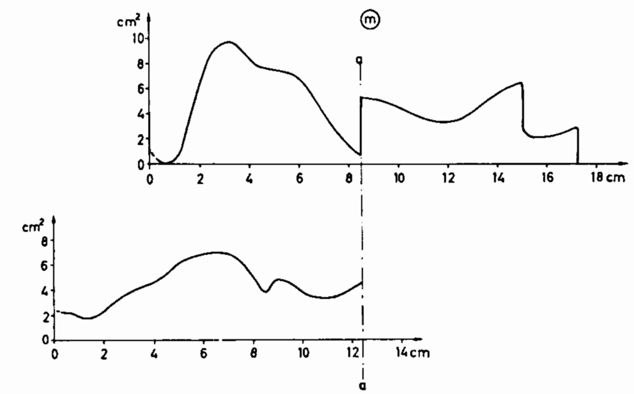


Fig. 3. Area function of complete vocal tract including nasal system (above) for the nasal consonant [m]. After *Fant*, ref.[7].

The representation of a nasal cavity system in terms of an area function alone as in Figure 3 is an oversimplification and disguises several interesting features which have acoustic relevance, as will be discussed in section 4. The primary data from perpendicular cuts of a casting of the nasal pathways, as determined by *Bjuggren*, are shown in Figure 4. The two surfaces on each side of a cut are displayed with an order scale numbered in cm along a central coordinate from the nostrils to the connection with the epipharynx. The left and right passages meet at about 8 cm from the nostrils. The symmetry is not perfect and much larger differences between the left and the right passage occur frequently. The complicated outline with an upper medium and lower channel and a very large circumference which is up to 4 times as large as that of a circular section of the same area is typical of the central parts of the nasal cavity system.

One way of visualizing cross-sections of the vocal tract is by means of so-called "laminographic" or "tomographic" X-ray

* Post-mortem anatomical study performed by Dr. *Gunnar Bjuggren*, Sabbatsbergs Sjukhus, Stockholm.

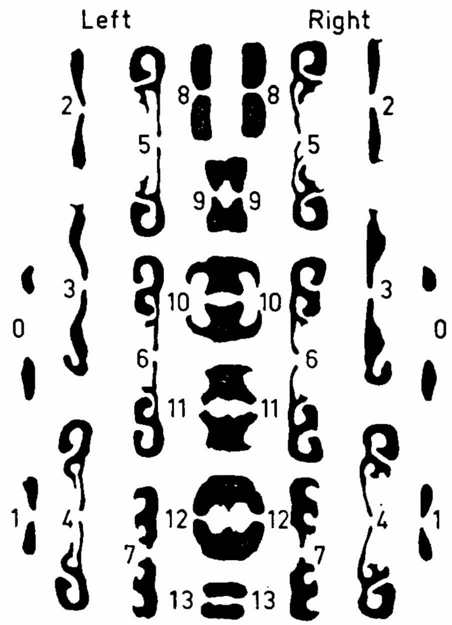


Fig. 4. Cuts through a casting of nasal cavities. Numbers pertain to distance in cm from the nostril end. Data by *G. Bjuggren.*

photos. Figure 5 is an ordinary sideview of a Swedish male subject articulating the vowel [a]. Figure 6 is a frontal tomographic view through the larynx and pharynx of the same subject articulating the vowel [i]. Figure 7 shows tomographic cuts through the middle of the mouth cavity of the same subject articulating [a] and [u]. Figure 8 pertains to cuts in a horizontal plane through the middle of the pharynx of the vowels [*a*], [u], and [i].

In these pictures made by the Swedish radiologist *P. Edholm** are several interesting facts to observe. The extreme differences in pharynx cross-sectional area comparing [i] and [a] are not only a matter of the distance between the back of the tongue and the posterior pharynx wall. The narrow pharynx passage of the vowel [*a*] is also accentuated by the short distance in the left-right direction. This general trend quoted by *Chiba* and *Kajiyama*[4] was anticipated in my work, ref.[7].

Here I would like to insert the comment that I consider not only the Swedish long vowel [a:] but also the short [a] to be a back

* Of Karolinska Sjukhuset, Stockholm. The tomograph equipment was designed by Professor *Lindblom.*

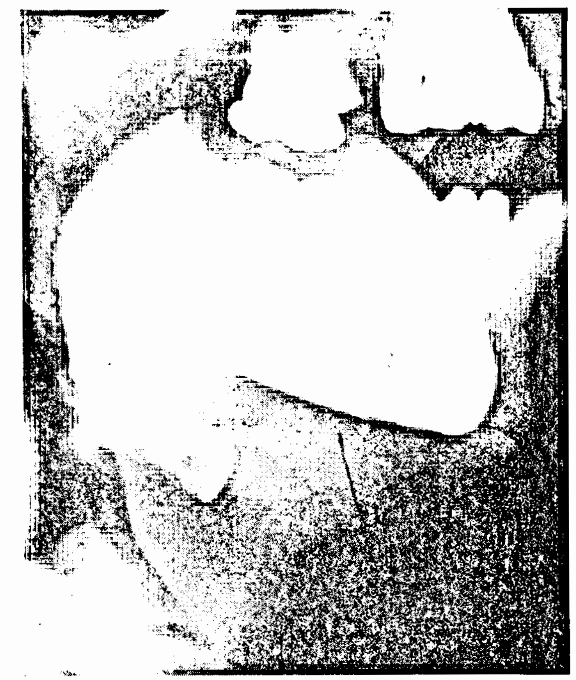


Fig. 5. Tomogram of male subject articulating the vowel [*a*]. (Photo by *P. Edholm.*)

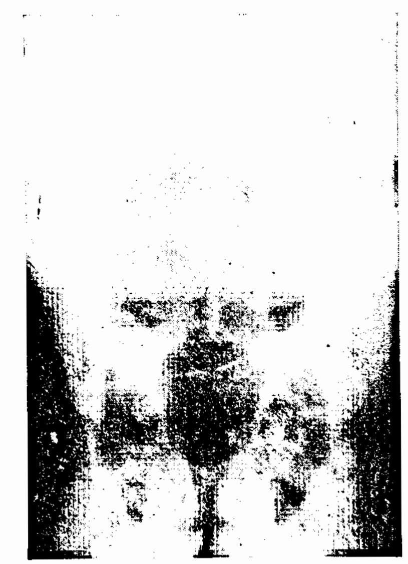


Fig. 6. Frontal tomographic cut through the larynx and pharynx of the same subject as in Fig. 5 articulating the vowel [i].

vowel and that I find several reasons for extending this articulatory classification to other languages since the main distinctive feature of the vowel [a] compared with front vowels is the relative narrow pharynx. The highest point of the tongue, being the holy reference of classical articulatory phonetics, has no acoustic relevance in this system and is fit for revision itself. The narrowing of the pharynx on the other hand is a necessary requirement for making an [a].

The vertical cuts through the mouth of Figure 7 reveal the plastic narrowing of the tongue when elevated as in the vowel [u] leaving appreciable columns of air on both sides.

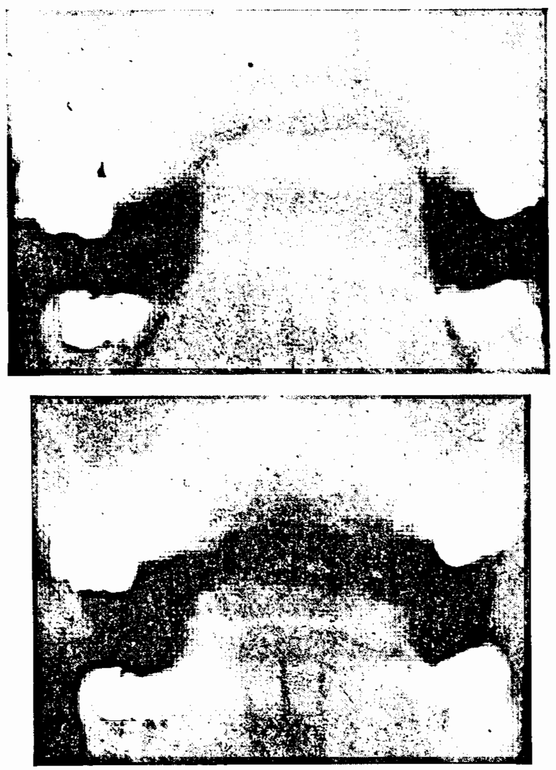


Fig. 7. Tomographic cuts through the middle of the mouth cavity of vowels [ɑ] and [u].

The frontal view of the pharynx in Figure 6 cuts through the larynx tube displaying the small Sinus Morgagni just above the vocal cords and the Sinus Piriformis side pockets on both sides of the larynx tube. These have been incorporated in some of the models submitted to calculations in ref.[7].

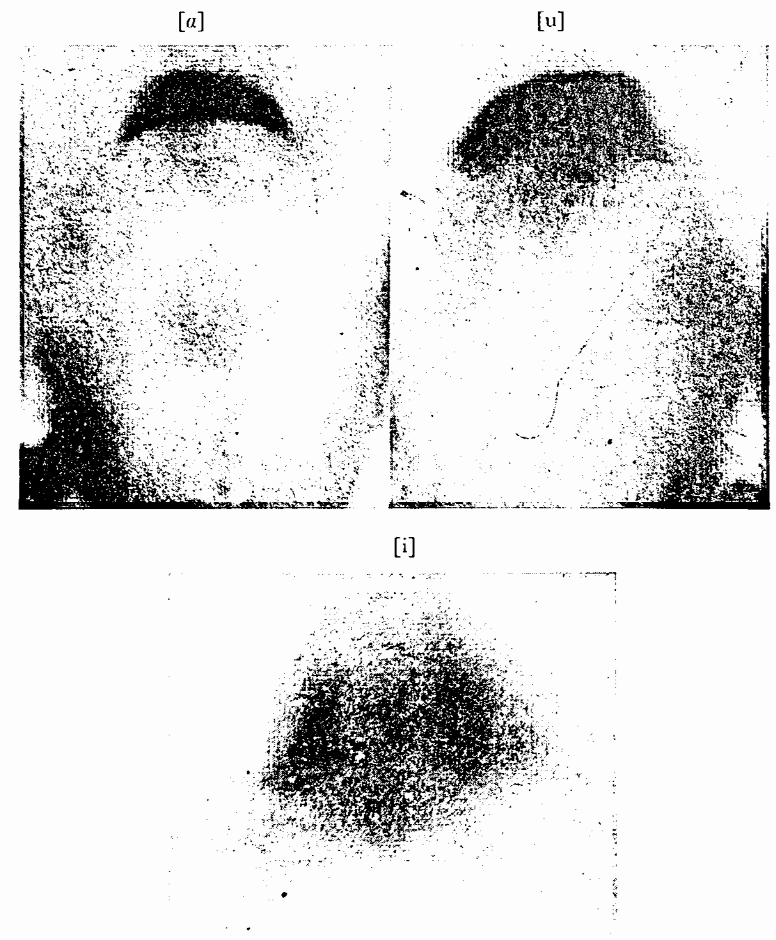


Fig. 8. Tomographic cuts horizontally through the middle of the pharynx of [ɑ], [u], and [i].

Figure 9 finally shows tracings of sideviews of articulations of Swedish vowels. Complete spectrum sections of standardized vowels selected in specific quanta of F_1 and F_2 are shown in Figure 10. These two figures have been published before[9] but are quoted here since they effectively correlate articulatory and acoustic data in terms of both formant frequency patterns and typical formant amplitude levels.

4. *Vocal Tract Models*

Chiba and *Kajiyama*[4] should be credited for being the first to effectively show the transmission line structure of the vocal tract.

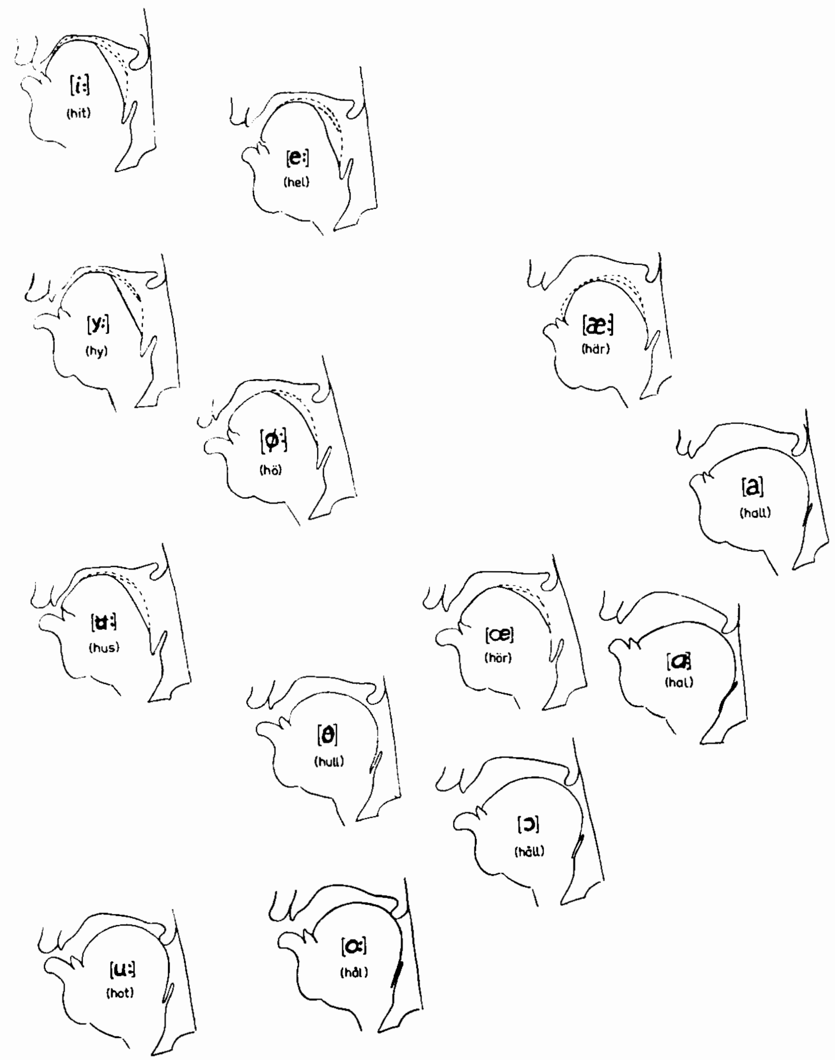


Fig. 9. X-ray tracings of Swedish vowels arranged to conform with an F_2 versus F_1 vowel diagram as in Fig. 10.

Figure 11 taken from their book[4] illustrates the single tube representation of a neutral sound. Pressure and velocity of the sound wave inside the tube have a phase difference of 90 degrees. At the frequency of a formant there is always a pressure minimum and velocity maximum at the lips and a pressure maximum and velocity minimum at the glottis. At the frequency of the first formant the standing wave pattern occupies $^1/_4$ of a full wavelength and $^3/_4$ and $^5/_4$ wave-

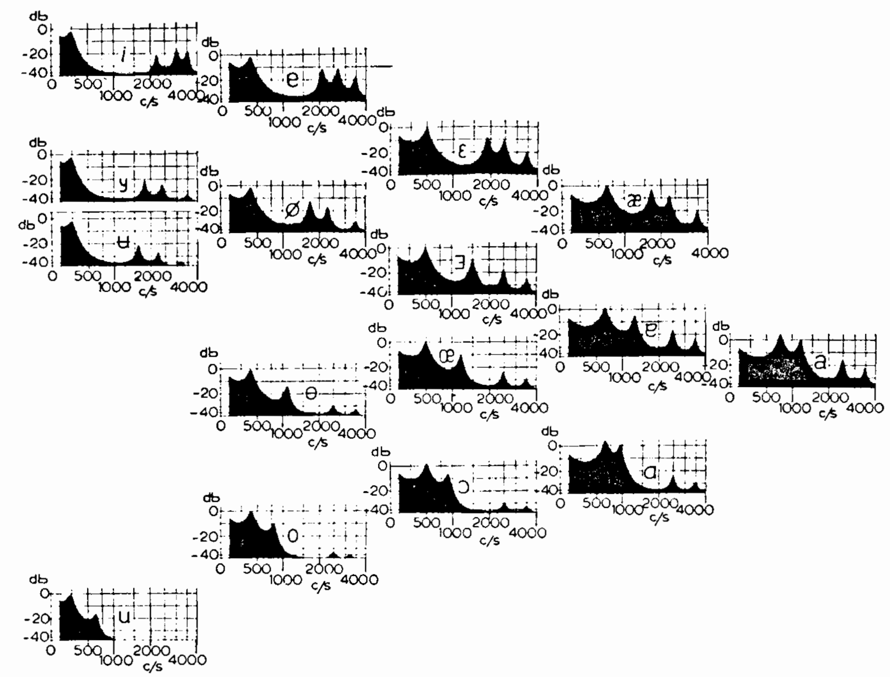


Fig. 10. Spectra of synthetic vowels assuming a standard voice source. The vowels are selected in terms of combinations of quantal values of F_1 and F_2 and F_3.

lengths at F_2 and F_3 respectively. The length of the standard tube, 17.5 cm end-corrections included, corresponds to a F-pattern of 500, 1500, 2500 c/s, etc. The most simple rule governing the relations between vocal tract configurations and F-patterns is that if a homogeneous tube is constricted at a place where one of its formants has a velocity minimum, there will follow an increase of the formant frequency whereas a constriction at a place of a velocity maximum results in a decrease of the formant frequency. These rules are related to the fact that the standing wave carries largely kinetic energy at a velocity maximum and largely potential energy at a velocity minimum, i.e. at a pressure maximum. In terms of circuit elements the region of velocity maximum may be replaced by an inductance and the region of pressure maximum by a capacitance as far as the particular formant is concerned*.

A simple but very effective approximation of the vocal tract is by means of two connected tubes of different lengths and cross-

* Calculations based on the Webster horn equation have been successfully carried out by *Ungeheuer*[25].

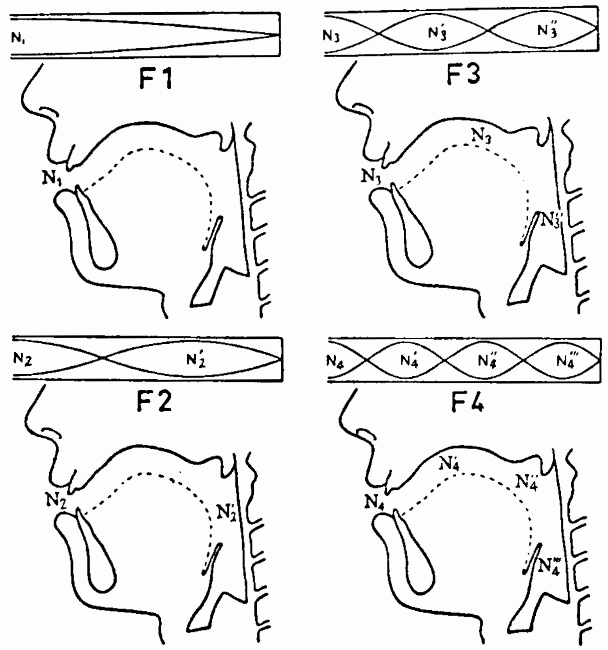


Fig. 11. Distribution of volume velocity at the frequencies of each of the first four resonances of an ideal neutral articulation in which the vocal tract simulates a tube of constant cross-sectional area (after *Chiba* and *Kajiyama*, ref.[4]).

sectional area as exemplified by Figure 12 which originates from the very early years of my work on the acoustics of speech. Of special significance is the two-tube approximation of [a] and [i] which are polar opposites in terms of articulatory features.

The double Helmholtz resonator shown in Figure 13 has played a very important role in the history of acoustic phonetics and I still use it for some approximate calculations. When the double resonator model is represented by lumped circuit elements, i.e. by the two volumes and the two necks, circuit theory will account for two formants only. As a matter of fact F_1 and F_2 of the Russian vowel [ï*] shown in Figure 2 may be calculated with a reasonable accuracy from this model. Formant frequency F_1 of most front vowels can be calculated with a reasonable degree of accuracy from a single Helmholtz resonator model of the entire vocal tract, the front part of which is the neck and the back part of which acts as the volume.

In general the vocal tract behaves like a continuously inhomo-

* Nonpalatalized allophone of [i].

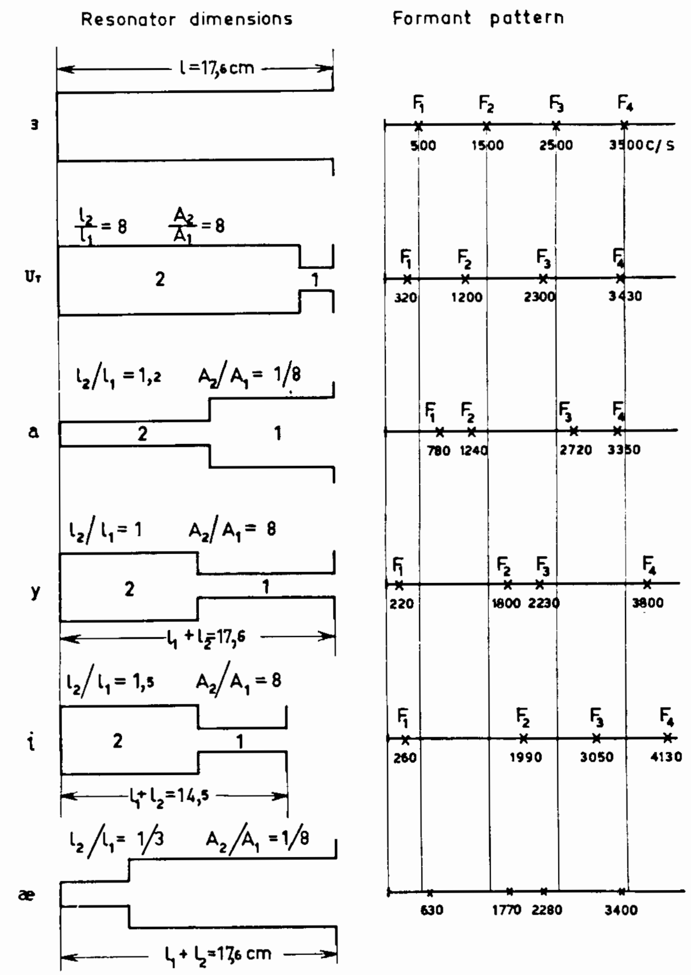


Fig. 12. Twin-tube resonators and associated F-patterns approximating a few vowels, see ref.[7]).

geneous transmission line. In order to simplify calculations *Stevens* and *House* designed a three-parameter model, whereby the complete articulation is specified by (1): the degree of opening at the lips, (2): center coordinate of the major tongue constriction, and (3): the degree of opening of the constriction.

The nomograms of Figure 15 show the relation between the place of tongue constriction and formant frequencies under conditions of a constant and small tongue constriction area and a set of varying degrees of lip-opening. This kind of nomogram is very useful for

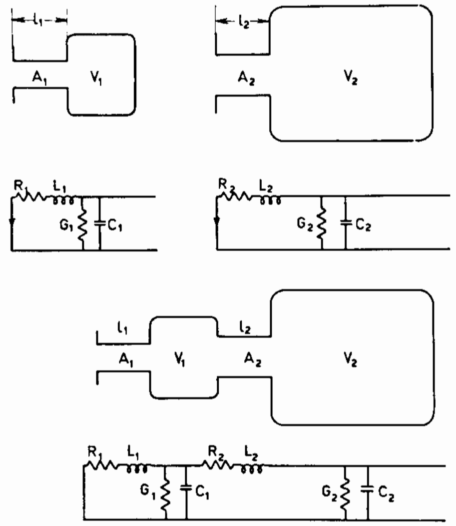


Fig. 13. Single and double Helmholtz resonators and equivalent electrical networks, see ref. [7].

translating various articulatory positions to an F-pattern. The low position of F_2 close to F_1 typical of back vowels is apparent as is the high F_2 closer to F_3 in front vowels. The place of maximum F_2 is more posterior than that of maximum F_3, the latter position representing prepalatal articulations.

5. *Formant-Cavity Affiliations*

There are several means of studying cavity formant relations. One is with reference to Figure 15. As the tongue constriction is moved from an extreme position at the glottis end of the model to the other extreme at the lips, it is to be seen how F_1 first shows a slight

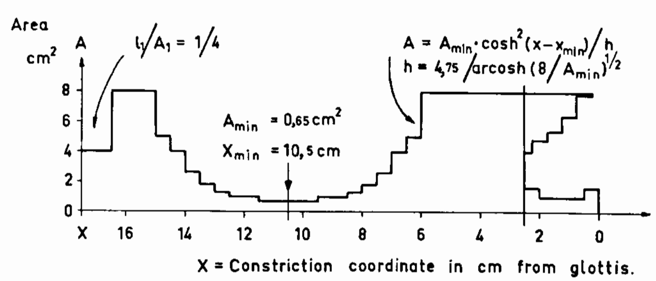


Fig. 14. Three-parameter vocal tract model based on a horn-shaped tongue section. A larynx tube, as well as the sinus piriformis cavities surrounding the larynx tube, have been incorporated as fixed cavities, see ref. [7].

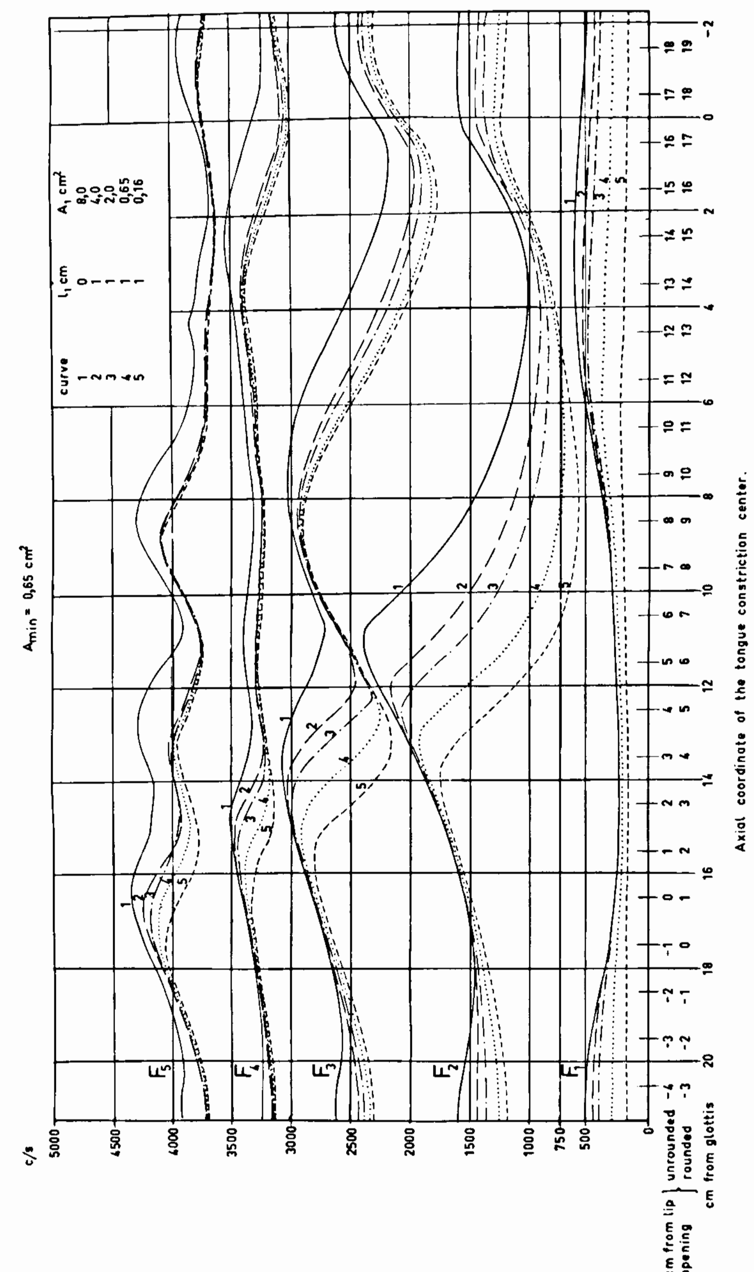


Fig. 15. Nomograms of the first five formant frequencies as a function of varying place of the major tongue constriction assuming a very narrow tongue passage and with various degrees of superimposed lip-rounding in the model of Fig. 14. (See *Fant*, ref. [7].)

tendency of rise and then falls. Here F_1 shifts from a mixed front and back cavity affiliation to being more affected by the back cavity*. In this forward movement F_2 starts with a fall and then rises as it gets more affiliated with the front cavity than with the back cavity.

However, at an advanced place of articulation in front of the place of maximum F_2 the second formant will be more dependent on the back cavity, now as a half-wavelength resonance. In this advanced location F_3 is the fundamental resonance of the front cavity. Further back at a medio-palatal to velar region F_3 is associated with a standing wave resonance of the back cavity. In back-vowels F_3 is a standing wave resonance of the front cavity.

The effect of superimposed lip-rounding on formant cavity affiliations is to shift the critical coordinates in an anterior direction, as can be seen from Figure 15. On the other hand if a lip-rounding affects one formant more than others this is a sign of an especially prominent affiliation of this formant with the front cavity. This situation is typical of the prepalatal [i] the F_3 of which is very sensitive to lip-rounding.

Quantitative measures of the degree of dependency of any formant frequency on any small variation of the vocal tract area function were given in ref.[7]. One of the several variations introduced was to make a small increase in the cross-sectional area at the coordinates of maximum area in the front and the back cavity. The tabulation is quated here, * denotes negative values.

Vowel	$\frac{\Delta F_1}{F_1} \frac{C_1}{\Delta C_1}$	$\frac{\Delta F_2}{F_2} \frac{C_1}{\Delta C_1}$	$\frac{\Delta F_1}{F_1} \frac{C_2}{\Delta C_2}$	$\frac{\Delta F_2}{F_2} \frac{C_2}{\Delta C_2}$	$\frac{\Delta F_3}{F_3} \frac{C_1}{\Delta C_1}$	$\frac{\Delta F_3}{F_3} \frac{C_2}{\Delta C_2}$
[ɑ]	0.07	0.19	0.23	0.11	0.18	0
[o]	0.05	0.33	0.37	0.22	0.25	0
[u]	0.18	0.28	0.20	0.15	0.06	0
[ï]	0.02*	0.39	0.49	0.02*	0.04	0.45
[i]	0.01*	0	0.53	0.39*	0.04	0
[e]	0.08*	0.01*	0.42	0.26*	0.01	0.23

Observe e.g. the lack of association of F_2 of [i] with the front cavity volume C_1. The same effect is observed when merely shortening the front cavity of [i] by removing one 0.5 cm section, see ref.[7], p. 120,

* This discussion pertains to the role of cavities alone. In half-open and close front vowels the tongue constriction is of equal importance as one of the basic elements of a Helmholtz resonator.

Table 2.33–4. Whilst discussing the vowel [i] it should be appreciated that what is said above pertains to the Russian and the Scandinavian [i]-vowels which are prepalatal whereas the [i] is articulated more towards the medio-palatal region in English. Also the age and sex of the speaker should be considered. Women generally have shorter necks than men, i.e. their ratio of mouth length to pharynx length is greater than in males which determines an increased affiliation of their F_2 with the front cavity.

The fourth or fifth formants of a male voice have a tendency to be strongly affected by the larynx tube. Even without such a tube there would fall a vocal tract resonance not too far away but the essential feature is that the addition of a larynx tube to a model, everything else being constant, is to place one more formant in the 3500 c/s region thus boosting this part of the spectrum. Those who have a short larynx tube or insufficient closure at the arytenoids will lack this formant. This is typical of female voices.

The effect of the Sinus Morgagni pockets is to cause a slight shift down in some of the formant frequencies and to set a sharp upper frequency limited at about 4500 c/s in the vowel spectrum, see ref.[7], p. 102, and ref.[3].

The search for formant cavity associations once approached with unsufficient theoretical tools and an overoptimistic attitude[20,21,15] was miscredited by the impact of transmission line theory in vocal tract analysis. This subject may now be treated with better theoretical tools, although our knowledge of the physiological facts has not advanced to the same extent. In fact, we still may have something to learn of the "old school". *Sovijärvi's*[20,21] ensemble of seven variable and eleven fixed formants of the speech spectrum was in part a fiction but some of these speculations deserve a renewed interest now when *Fujimura* has shown[14] that there are at least twice as many formants in a nasal murmur sound than we had anticipated from spectrum matching work[13] and the averaged area function of the nasal tract I used in my calculations[7]. A sample of *Fujimura's* observations[14] pertaining to transmission measurements with sound injected from a vibrator externally at the throat and a pick up of the sound at the nostrils is shown in Figure 16. Observe, for instance, the split first formant as in nasalized sounds which together with the several other spectral peaks can be ascribed to the asymmetry of the particular nose, the left and right passages differing in their dimensions.

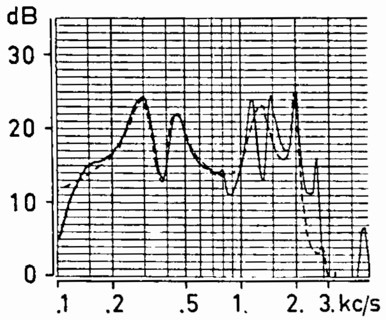


Fig. 16. Sine wave sweep frequency characteristics of the cavity system of a palatal nasal consonant co-articulated with a neutral vowel (solid line). Vibrator externally at the throat and pickup at the nostrils. A spectrum matching comprising some of the lower resonances is included (broken line). In an accurate match something like 6 poles and 3 zeros would be needed in the frequency range below 2000 c/s. (After *O. Fujimura.*)

Finally a statement on the role of the cavity walls. It was initially suggested by *van den Berg*[1,2] and I elaborated the theme in my book[7], that in sound with a very low F_1 there is a substantial energy loss from sound transmission through the walls of the vocal tract, wherever these walls are thin enough to allow vibration to be set up. In ref.[7] I calculated a limiting frequency of 150 c/s for the air inside the vocal tract resonating with the mass element of the walls all outlets being closed. Recently I discovered that this shunting mechanism is responsible for the typical nasal quality of the speech of a diver submitted to a high overpressure.

Recent experiments performed together with Dr. *B. Sonesson** have shown that the most typical characteristics of the speech of a person submitted to a high overpressure is the rise in F_1, the immobilization of F_1, and the inability of the subject to produce sounds with an F_1 lower than a certain limit set by the overpressure. What happens is that an increase in overpressure does not affect the velocity of sound, c, but it affects the density ρ of the air proportionally and thus the impedance level $\rho c/A$ whereas the impedance of the cavity walls is not affected by the pressure increase and its relative load will thus increase.

The lowest possible frequency of F_1 is simply the square root of the atmospheric pressure in ATA times a constant which is the limiting F_1 of normal speech and of the order of 200 c/s. With a pressure of 6 ATA the low limit of F_1 is thus of the order of 500 c/s.

* Will be reported on briefly in STL-QPSR 2/1964.

The typical effect is that close and half-open vowels, as well as voiced occlusives, fricatives, and semivowels and nasals, all fall within the same very narrow range of F_1 around 500–600 c/s. To the first approximation

$$F_1 = \sqrt{F_{10}^2 + F_{1k}^2}$$

where F_{10} is the F_1 without consideration to the cavity wall load and F_{1k} the limiting frequency at complete vocal tract closure.

Examples of speech at 1 and 6 ATA pressure is shown in Figure 17. This effect facilitates the calculation of the limiting F_1 in normal speech and enforces the existence of this limit to be considered when calculating the filtering properties of the vocal tract. If this effect is not taken into account calculations of F_1 will thus tend to give too low values, a fact that is supported by experience. But for this effect it would be difficult to explain the very stable base-band formant frequency at about 200 c/s of voiced consonants. It remains to study to what extent the low F_1 limit varies with speaker category.

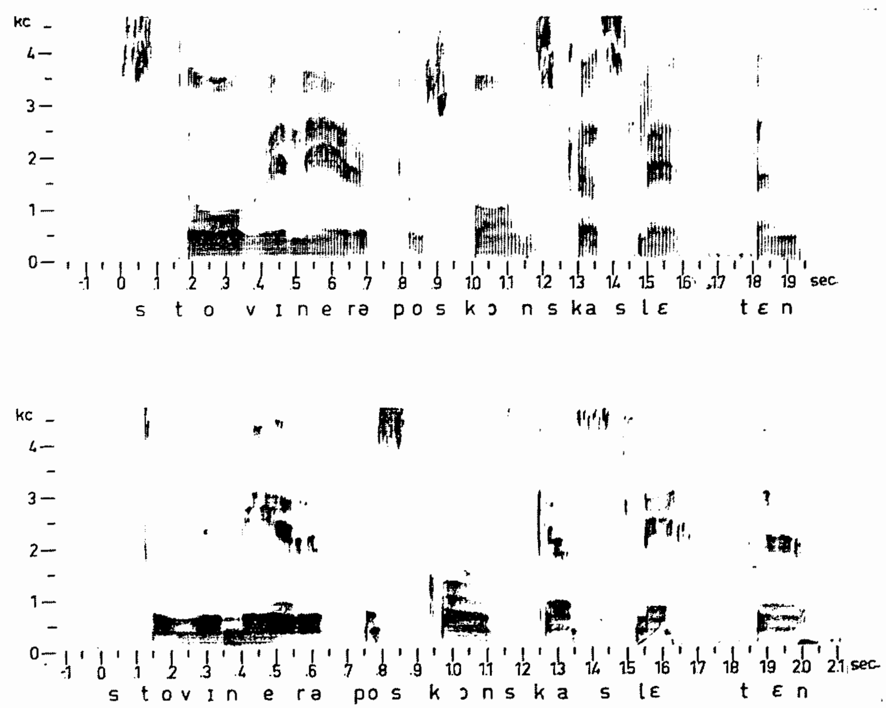


Fig. 17. The main effect on speech of increasing the air pressure in the room, where the subject is seated, is to increase the lowest limit of F_1.

References

1. *van den Berg, Jw.*: Physica van de stemvorming met toepassingen; Diss. Groningen ('s-Gravenhage 1953).
2. *van den Berg, Jw.*: Calculations on a model of the vocal tract for vowel /i/ (meat) and on the larynx. J. acoust. Soc. Amer. *27:* 332–338 (1955).
3. *van den Berg, Jw.*: On the role of the laryngeal ventricle in voice production. Folia phoniat. *7:* 57–69 (1955).
4. *Chiba, T.* and *Kajiyama, M.*: The vowel – its nature and structure (Tokyo 1941).
5. *Delattre, P.*: The physiological interpretation of sound spectrograms. PMLA, *LXVI:* 864–875 (1951).
6. *Dunn, H. K.*: The calculation of vowel resonances, and an electrical vocal tract, J. acoust. Soc. Amer. *22:* 740–753 (1950).
7. *Fant, G.*: Acoustic theory of speech production ('s-Gravenhage 1960).
8. *Fant, G.*: Acoustic analysis and synthesis of speech with applications to Swedish. Ericsson Technics *15:* 3–108 (1959).
9. *Fant, G.*: The acoustic of speech. Proc. Third int. Congr. Acoustics, Stuttgart 1959; pp. 188–201 (Amsterdam 1961).
10. *Fant, G.*: Descriptive analysis of the acoustic aspects of speech. Logos *5:* 3–17 (1962).
11. *Fant, G.*: Sound spectrography. Proc. Fourth int. Congr. Phonetic Sciences, Helsinki 1961; pp. 14–33 ('s-Gravenhage 1962).
12. *Fant, G.; Fintoft, K.; Liljencrants, J.; Lindblom, B.* and *Mártony, J.*: Formant-amplitude measurements. J. acoust. Soc. Amer. *35:* 1753–1761 (1963).
13. *Fujimura, O.*: Analysis of nasal consonants. J. acoust. Soc. Amer. *34:* 1865–1875 (1962).
14. *Fujimura, O.* and *Lindqvist, J.*: The sinewave response on the vocal tract. Speech Transmission Lab., quart. Progr. and Status Report No. 1, pp. 5–10 (KTH, Stockholm 1964).
15. *Hála, B.*: Nature acoustique des voyelles. Acta Univ. Carol. Med. (Praha 1956).
16. *Peterson, G. E.* and *Barney, H. L.*: Control methods used in a study of vowels. J. acoust. Soc. Amer. *24:* 175–184 (1952).
17. *Potter, R. K.; Kopp, A. G.* and *Green, H. C.*: Visible speech (New York 1947).
18. *Rayleigh, L.*: The theory of sound. Vols. I and II (New York 1945). (First Amer. Ed.)
19. *Sapozhkov, M. A.*: The voice signal in cybernetics and communications (Moscow 1963). (This book provides an inventory of speech research with a large number of references.)
20. *Sovijärvi, A.*: Die gehaltenen, geflüsterten und gesungenen Vokale und Nasale der finnischen Sprache. Ann. Acad. Sci. fenn. B XLIV (Helsinki 1938).
21. *Sovijärvi, A.*: Die wechselnden und festen Formanten der Vokale, erklärt durch Spektrogramme und Röntgenogramme der finnischen Vokale. Proc Third int. Congr. Phonetic Sciences, Ghent 1938; pp. 407–420 (Helsinki 1938).
22. *Stevens, K. N.* and *House, A. S.*: Development of a quantitative description of vowel articulation. J. acoust. Soc. Amer. *27:* 484–493 (1955).
23. *Stevens, K. N.* and *House, A. S.*: Studies of formant transitions using a vocal tract analog. J. acoust. Soc. Amer. *28:* 578–585 (1956).
24. *Trendelenburg, F.*: Einführung in die Akustik (Berlin 1950). (Zweite Auflage.)
25. *Ungeheuer, G.*: Elemente einer akustischen Theorie der Vokalartikulation (Berlin 1962).

Author's address: Dr. G. Fant, Speech Transmission Laboratory, Royal Institute of Technology (KTH), *Stockholm 70* (Sweden).

Discussion

H. M. Truby (San Jose, Calif.) (*Truby's* summary of his own critique): In spite of the overwhelming effects of the past 55 min of slides, I feel I must utilize some part of the few remaining discussion moments to point up certain pertinencies. The title selected by Docent *Fant*, *Formants and Cavities*, would lead members of the Congress to assume that something of the *relationship* of formants and cavities was to be discussed, as might well be the case after all the years of research applied to the establishment of such implied correlates. However, we have not been so blessed.

I hold before you a picture of a little girl. This picture tells a complex story to me which no one else in the hall could possibly be in a position to appreciate. This picture is a single instant in time, yet when *I* look at this picture an entire dynamic tale unfolds. The little girl of the picture is my daughter, and the picture acts as a multiple stimulus, the nature of which you are all aware of by this point in my analogy, for I've seen her happy and sad and at work and at play and asleep, and so on. Now, when we see slide after slide of such single-instants-in-time, it is only as products of our own individual experiences that we are able to evaluate their static natures in dynamic terms. Certainly *formants* do not reveal themselves entirely – and ofttimes not even identifiably – on sound spectrograms or similar displays, from the standpoint of their per*cep*tion, and *cavities* cannot be defined in terms of cross-sectional areas unless they happen to be highly idealized cavities, such as are *never* found in the human or any other animal form... and rarely in nature at all, except as a man-made artifact. How then can it be proposed that a discussion of a few arbitrarily selected instants in time extracted from visual displays of acoustic continua will reveal anything of the dynamic nature of the speech signal? We know as little now, apparently, of the relationship of formants to cavities and vice versa as was reported during the Oslo Congress of Linguists in 1957, which situation I cticicized expressly as *A First Feature of Indivisibility* in my own Oslo paper, *Visible and Indivisible Speech.* I say, 'apparently', since in the twelve minutes actually allotted by the Committee for the pre-discussion phase of this paper, there *would* have been time to report on the dynamic analysis aspects *now* available to speech research. Cineradiography, spectral timesampling, overall intensity dynamics, fundamental frequency flow – all these and other media for analysis are discussable at this point of the history of the phonetic sciences. Having been exposed to the implications of 'hocus pocus linguistics', in the face of what we are asked to infer from this array of time and space *cross-sections* just presented us, I should like to propose the adjunctive term 'hocus pocus phonetics'. In short, calculations in static terms of dynamic – yea, kinematic – activities have little significance for the phonetician. The high-light of my long-time colleague's presentation was the single slide of the F_1 through F_5 nomogram which, if not dynamic, was at least spatial, and which is featured in his doctoral dissertation.

I cannot hope that my remarks will not be misconstrued in some quarters and my point overlooked, but my principle intention is to point out that the present state of the art *does* permit a dramatic account of the acoustic and physiologic correlates of speech. It's there for the telling...

Epilogue

The above Discussion was, of course, motivated by the *delivered* version of *Fant's* paper. The *written* version, carefully edited by the author after the Congress, has been altered in several places significantly relevant to certain points of my verbal critique, e.g., last line, p. 120 – first line, p. 121; lines 3–8, p. 121; etc. Even though, in this instance, my own remarks have lost none of their pertinence, presented papers should not, in my opinion, be mere rough drafts of *Proceedings* reports.

Proc. 5th int. Congr. phon. Sci., Münster 1964, pp. 142–171
(S. Karger, Basel/New York 1965).

From the Haskins Laboratories, New York City

Instrumental Methods for Research in Phonetics

By Franklin S. Cooper

Instruments have long had a part in the development of phonetics. Their use has been both praised and derogated, depending in part on individual preferences as to whether phonetics should become a science or remain an art. The trend is clearly toward quantitative description and experimental method; hence, the purpose of this report is to review briefly the instrumental developments of the past few years.

A total review would surely exceed your patience and my abilities. Fortunately, the task can be reduced: (1) We can take only passing note of several areas of instrumentation that are of concern primarily to the communications engineer. This is hazardous, however, since these same tools may so easily be adapted to research on phonetic problems. (2) We can rely on existing reviews that cover most of the important developments until quite recently, in particular, the excellent and extensive reviews by *Eli Fischer-Jørgensen*[1] and *C. G. M. Fant*[2] presented at Oslo in 1958 and the reviews by *Fant*[3] and *Cooper*[4] given in Helsinki in 1961. What remains is to update the existing reviews and then perhaps to take note of some new trends in phonetic research and the related developments in instrumentation. There are hazards, to be sure, in presuming to point out significant new trends while they are yet trending: one's biases will surely be revealed, and quite possibly one's follies.

I. Instruments and Methods

Sound Recorders and Oscillographs

The use of magnetic tape for recording and reproducing the sounds of speech is so common and convenient that one tends to

forget his almost total dependence on recordings for every kind of phonetic research. Indeed, some types of studies can be done with little more than tape recordings, a razor blade, and some splicing tape.

Research in phonetics is well served by conventional magnetic tape recorders; even so, some of the many variants on the familiar recorder may prove useful for special purposes. Thus, multitrack recorders have the obvious advantage of allowing two or more distinct signals to be stored and recovered without loss of synchronism. Frequency modulated recordings preserve the audio waveform to zero frequency and are almost indispensable for such specialized tasks as recovering the glottal waveform by inverse filtering. Digital recording has advantages that may outweigh its cost when the experiments require an extremely good signal-to-noise ratio or a high degree of reproduceability on successive playings. A rather simple variant of the conventional magnetic recorder uses cards instead of tape on reels. This permits sentence-long utterances to be recorded, annotated, and filed away with other flat records; thus, a sound spectrogram and the speech it represents can be kept together for easy reference to either pattern or sound. Psychoacoustic tests that require the pairwise comparison of brief sounds are much simplified, as are comparisons of informants' utterances. Card-type recorders seem to have been used less than one might expect, perhaps because they have not been commercially available, at least in the United States, until the last year or two[5].

The display of sound waveforms by the oscillograph is, like sound recording, so commonplace that it is taken for granted. Although the waveform has proved to be less useful in phonetic research than spectrum information derived from it, there are many purposes for which the oscillogram is highly useful. Recent developments in instrumentation and recording media make it quite easy to record speech waveforms; several channels of information containing frequencies up to 5,000 cps or more can be registered on papers that require no processing except brief exposure to a fluorescent light[6].

Spectrograms and Spectrographs

Although the speech waveform contains all the acoustic information about a speech sound, it has long been known that the spectrum of the sound provides a more useful display. This was evident already from Fourier analyses of vocalic sounds that had been made even before this century. Nevertheless, the publication by the staff of the Bell Telephone Laboratories[7] of a series of articles and a book on sound spectrograms – Visible Speech, as it was called – caused much excitement among experimental phoneticians. This happened about 1945 to 1947, so that some of you may be a little surprised to find a much earlier spectrogram in the literature, showing the now-familiar dynamic patterns of speech in the same time-frequency-intensity display that is used for visible speech. This spectrogram appeared just over thirty years ago in a paper by *John C. Steinberg*[8] entitled "Application of sound measuring instruments to the study of phonetic problems" – a timely topic to this day.

Why should phoneticians have waited more than a decade to respond? Only a part of the reason is to be found in the intervening political events. Other reasons are, no doubt, that the hand-drawn patterns of *Steinberg's* spectrogram lacked the wealth of detail and artistic appearance of the machine-generated variety, and that his product also lacked the promotion given to visible speech, first by the Bell Telephone Laboratories, and then by *Martin Joos*[9] in his remarkable little book directed specifically to linguists. But we may suspect that, despite these reasons, the sheer labor of generating the patterns would have discouraged their use in research. *Steinberg's* spectrogram was in fact prepared from separate Fourier analyses of each vocal cord period, done by hand at something like two hours of work for each period. This single spectrogram, then, represents about a month of full-time effort devoted to measurement and calculation; today's spectrograph does the same job – better – in a few minutes. The moral is clear; a method, to be useful in phonetic research, must have both conceptual power and instrumental convenience, or, put in another way, new techniques that open old doorways may be almost as valuable as new insights.

The sound spectrograph, of the type manufactured by the Kay Electric Company under the trade name Sona-graph[10], based on the original Bell Telephone Laboratories' design, is now to be found in a number of laboratories. Although the instrument has its frailties, as most of its users will attest, it has proved extremely useful and is, indeed, the central piece of research instrumentation in most phonetics laboratories. This wide acceptance of a single instrument has had an important consequence in standardizing the pictorial aspects of the sound spectrogram; it is a further tribute to the original design

that even the improvements announced within the past year do not change the basic presentation, but only the mechanisms and circuitry. The new transistorized model of the Sonagraph has features of improved flexibility, ease of calibration and speed of analysis. Another new spectrograph, reflecting a thorough redesign to achieve an improved research instrument, has been made by Prestigiacomo and co-workers at the Bell Telephone Laboratories. This spectrograph[11] retains one of the best features of an earlier model, namely, a rotating magnetic head scanner that allows spectrograms to be made directly from original tape recordings; in addition, the tape system of the new spectrograph can serve as a conventional recorder. Modular electronics permit easy modification of signal conditioning circuits, filters, frequency scales, and other display characteristics. Even more important for routine use are the rugged mechanical design, rapid analysis, and the provisions for automatic calibration and for advancing the input tape. Another research instrument with many supplementary features has been designed and built by *Peterson*[12] at the University of Michigan, primarily for use in his own laboratory.

Despite the wide acceptance of the "standard" spectrograph for research in phonetics, a number of special-purpose instruments have been built. The principal objective has been to obtain real-time spectrograms of long samples of speech, or more generally to obtain running analyses whether or not the information is displayed as a conventional spectrogram. Banks of filters have long been used for this purpose and are now to be found in several laboratories; only two recent additions will be mentioned here. The 51-channel analyzer at the Royal Institute of Technology[13] is an outstanding example; it constitutes, in fact, the central unit in a very versatile research facility with capabilities that include the display and registration of real-time spectrograms and the production of digital tapes for computer use in further analyses of spectrum information. At Columbia University, *Harris*[14] has constructed a real-time audio-frequency spectrum analyzer consisting of a bank of 54 Gaussian filters. The outputs are scanned for presentation as running spectrograms on a radarscope, or converted to digital form for computer processing. Gaussian filters were employed to obtain good response to transient aspects of the speech signal.

The use of filter banks to generate real-time spectrograms has the disadvantage of poor frequency resolution or a requirement for very many filters. Another solution is to use methods that recirculate a short sample of speech at very high speed so that the requisite number of analyses can be completed during the duration of the sample, i.e., in real time. One such device, using digital delay lines and an analyzing filter has been described by *Gill*[15]; another that recirculates the analog signal in a quartz delay line for correlation analysis has been described by *Weiss* and *Harris*[16]. Equipment of this kind, though complex and costly, can provide high resolution spectrograms in quantity.

The nature of the display itself, when it is intended primarily for the human observer, has also been the subject of recent research. *Prestigiacomo*[17], following up the earlier work by his colleagues at the Bell Telephone Laboratories, has published contour spectrograms that seek to retain information about relative intensities that is often lost in the usual spectrogram because of the restricted grey scale of facsimile paper. The contour lines of these spectrograms mark off regions that differ by 6 db in marking level, and a rough identification of each such region is provided by the darkness of shading between the contour lines. Patterns of this kind have received rather wide publicity because of the claims made by *Kersta*[18] about their usefulness as voice prints for speaker identification.

Several variants of the conventional spectrographic display have been employed by *Wood* and *Hewitt*[19] at the General Electric Company in Schenectady. Their real-time analyzer is basically a filter bank which is scanned rapidly "by a unique process that interpolates the filter outputs to create continuous cross-sections of amplitude vs. frequency". The cross-sections are recorded from a cathode-ray tube onto film; in one version, the signal representing spectrum level is used not only to brighten the recording spot but also to deflect it very slightly along the time dimension. This results in some enhancement of the intensity variations across formants. In another version, the film is marked only at peaks of the spectral cross-section; this serves to delineate the course of the formants, as in the much earlier resonagrams described by *Huggins*[20].

C.P. Smith[21] at the Air Force Cambridge Research Laboratories has used a digital print-out of time- and frequency-quantized spectrograms to record numerical data without losing the essential pattern aspects of the spectrogram. The spectrum information in digital form can, of course, be used directly for quantitative experimental studies of speech.

Other Methods of Speech Analysis

There has been renewed interest in analyzing voiced segments on a period-by-period basis, employing variants of the inverse filtering method introduced by *R. L. Miller*[22] of the Bell Telephone Laboratories. The method is basically one of removing or cancelling the lower formants, one by one, from a high-fidelity recording o the wavetrain for a single glottal period. The residue then represents the glottal pulse itself, while the adjustments required to cancel the formants give formant frequencies and bandwidths. Inverse filtering is inherently a method of adjustment, though efforts have been made to automate it. *Lawrence*[23] at the Signals Research and Development Establishment, Christchurch, has referred to his equipment for cycle-by-cycle analysis as a "speech microscope". He has been concerned principally with the extraction of formant information and with adaptation of the method to the tracking of formants on a continuous basis. *Holmes*[24] at the Joint Speech Research Unit, and other workers as well, have used the method to study the details of the volume velocity waveform at the glottis.

The interest in events at the glottis includes, of course, the repetition rate or voice pitch, as well as the details of single glottal cycles. An exciting new method for obtaining the short-time average pitch has been described by *Noll*[25] of the Bell Telephone Laboratories. The method employs a spectrum analyzer designed to operate in real time and to produce high resolution spectra without using either heterodyning methods or bandpass filter banks – a device of considerable interest in its own right. The logarithms of consecutive amplitude spectra from this analyzer serve as the input to a second similar spectrum analyzer, the output of which is the "cepstrum", or power spectrum of the logarithmic spectrum. The cepstrum shows the period of the glottal pulses as peaks that are clearly separated from the peaks due to formants; thus, variations in voice pitch observed at analysis intervals of 15 msec. may be followed throughout a voiced stretch. The method has been tested by computer simulation with initial results that seem extremely promising. The term "cepstrum technique" has been applied to this method of vocal pitch detection.

The extraction of pitch information is a classic problem in the development of bandwidth compression equipment for voice communication. A very substantial amount of engineering effort has been devoted to it recently, with some success; however, none of the resulting devices really meets the need for a simple and thoroughly reliable pitch meter of the kind that could be so useful in phonetic research. One line of engineering development has, in fact, been to evade the problem by transmitting a part of the original speech signal (roughly the first formant region) to the receiving end of a communications link for use in regenerating the speech message. Automatic formant tracking is another active area of research in voice communication; here too, advances have been made. Research on the recognition of connected speech and on speaker identification must, of course, start with information about formant patterns and excitation functions and will probably need to make use of all available linguistic information as well. This is an area in which phonetic and engineering interests overlap, but it is also an area that has shown few striking gains. Quite possibly this is because no adequate rationale is yet available, with the possible exception of a method of analysis-by-synthesis which will be described in a later section. (The discussion in this brief paragraph has merely touched on some of the areas in which extensive engineering and instrumental developments have taken place. Fortunately for the present purpose, this field was reviewed in a Speech Communication Seminar that was organized by Dr. *Fant* and held at the Royal Institute of Technology, Stockholm, in August 1962[26]. The conference was aimed primarily at the field of communications although a number of the papers had a direct bearing on research in phonetics.)

Synthetic Speech and Speech Synthesizers

The great virtue of the sound spectrogram is that it displays the acoustic information as visual patterns that are correlated with the perceived sounds of speech and with the changing configurations of the vocal tract. The latter relationships are the subject of Dr. *Fant's* paper on "Formants and Cavities"; the correspondences between spectrographic pattern and linguistic unit are our present concern.

This concern stems from the need to identify the important aspects of spectrograms if we are to employ them in studying speech sounds and yet escape the fate of the investigator who, as *Fant*[3] has warned, "too easily drowns in a sea of details of unknown significance if he attempts to make use of all observable data". Fortunately, there is a direct way to examine the relationship between pattern and linguistic unit and, in so doing, to determine the significance of various aspects of the spectrogram. The method uses the spectrographic pattern, often in simplified or derived form, to control the synthesis of sounds that are then presented to listeners for perceptual evaluation, exactly as if they were listening to "real speech". The method, as applied to research in phonetics, was pioneered by my colleagues at Haskins Laboratories[27]; it is now used to good effect in several laboratories to isolate the acoustic cues for speech perception, and to pin-point those spectrographic features that deserve the most careful attention. There will be opportunity at this Congress to hear about work based on these methods in the review paper by *Delattre*, and in a number of other papers as well.

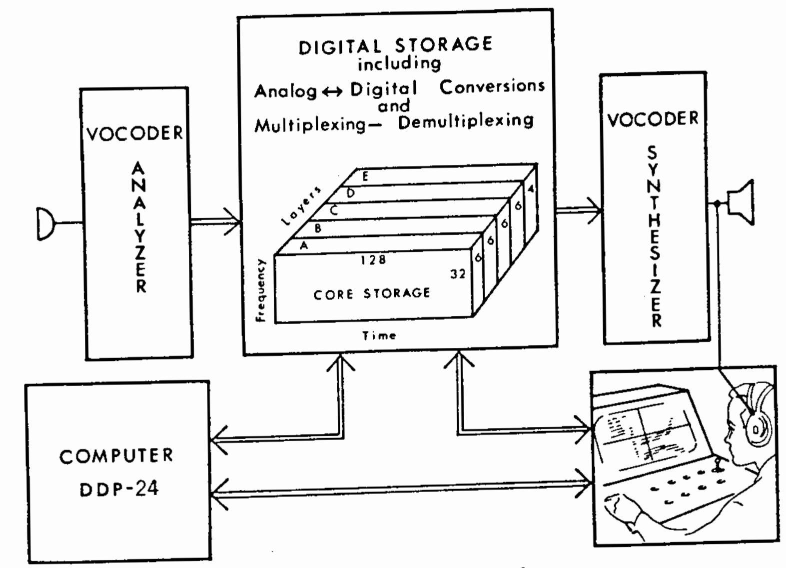
The situation with respect to speech synthesizers differs from that for sound spectrographs: synthesizers are less common, being found only in the larger laboratories, and there is as yet no "standard" instrument. This is due only in part to the complexity and cost of synthesizers. Different users have different needs and so have built synthesizers with widely differing characteristics. The relationships between intended use and machine design were discussed in a report[4] at the Helsinki Congress, with special emphasis on the control characteristics. One of the major considerations in the choice of a synthesizer is the convenience, and hence the overall effectiveness, with which the synthesis can be controlled. Conceptual convenience, as in the use of the spectrographic pattern itself for the control of synthesis, is no less important than mechanical simplicity. The overall complexity of the synthesizer and the difficulties of adequate control are functions also of the required felicity of the synthetic speech. High fidelity is indeed obtainable, but at a price; there is, in short, good reason to use care in selecting the synthesizer to match the problem.

The nature of the problem has been changing subtly as information has accumulated about the major acoustic cues. Much remains to be done at this level and with highly schematized patterns; nevertheless, some of the problems call for syntheses that are substantially closer to normal speech. Thus, in studies of stress and intonation, one may wish to impose various pitch lines on patterns that are otherwise as normal as possible; likewise, investigations of the relative importance of individual cues when more than one serves the same phonemic distinction may require controlled changes in the total spectrographic pattern derived from actual speech.

In studies for which the synthetic speech must approximate real speech, good use can be made of synthesizers of the vocal tract analog type, such as DAVO, or of the formant type that retain some of the inherent constraints of the vocal tract, such as OVE II or the series-connected PAT[3,4]. Many of you have heard synthetic speech from OVE II that was difficult to distinguish from the real speech on which the synthesis was based. This realism was obtained, however, only by painstaking attention to the accuracy of the control signals and some adjustments by ear as well. Indeed, the virtue of these devices in modeling human speech production carries with it the penalty of interactions among the control signals. These make it more or less difficult (more for DAVO and less for OVE II) to introduce point-by-point changes in the spectrum, when one is seeking to test hypotheses about acoustic cues.

Considerations of this kind – and, I must admit, a strong bias for the use of spectrographic patterns in the control of such experiments – have led to the development of a new synthesizer at Haskins Laboratories that combines realtime spectrum analysis and display with a facility for performing "microsurgery" directly on the spectrogram, and of course for hearing the sounds of both the original and modified patterns. Because digital methods are used for storing the spectrum information, and because the primary purpose is to permit manipulation of the spectrum, we refer to the device as a Digital Spectrum Manipulator.

The diagram in Fig. 1 shows the mode of operation: essentially, a vocoder analyzer and synthesizer are used to provide the spectrum and to synthesize speech from it. Between analyzer and synthesizer, there is a magnetic core memory and the conversion equipment needed to get vocoder information into and out of it. The arrangements for addressing the core memory make it look to the outside world as if it were organized in layers, like a cake, with each layer in the form of a digital spectrogram very much like the ones employed by *Smith*[21]. Each spectrogram is about two-and-one-half seconds long; the operator can look at any half-second portion of it on a cathode-ray tube that presents a flickerless spectrographic pattern. He can use a "joy stick" control to move the pattern backward and forward at slow or normal rates or to move an indicator to any frequency channel. The particular data cell marked by this indicator and a vertical line in the center of the screen is available to him and can be changed instantaneously. Two other features deserve



Digital Spectrum Manipulator

Fig. 1. Digital Spectrum Manipulator: a device that allows the experimenter to generate and examine real-time spectrograms on a cathode-ray tube, make changes in the spectrum at any point in time or frequency, and see and hear both the modified and original versions.

mention: one is that several different versions of the same pattern can be stored in the several layers of the memory for easy comparison with each other; another is that more complex changes than point-by-point modification of the spectrum can be done by a small digital computer that communicates directly with the data store and the operator.

The principal disadvantages of the system, as it now stands, are the frequency resolution and voice quality limitations inherent in a 19-channel vocoder and a corresponding graininess, or checkerboard pattern, in the display. We expect that both these limitations can be considerably eased by using 32- or 64-channel filter banks for analysis and by employing either filters or formant generators for synthesis. The advantages of the system lie not only in the fact that one works with the full spectrum of a real utterance, but also in the ease with which speech can be entered from a microphone or tape recorder and then manipulated, with immediate access to both the visual pattern and the corresponding sound of either the original or the modified utterance. The system, though designed for speech, is readily usable for other research in which multi-channel analog information is put into a computer and manipulated under the direct scrutiny and control of the experimenter. Some of the work on physiological measures of articulation, to be mentioned in a following section, is of this kind.

A device somewhat like the Digital Spectrum Manipulator was used by *Smith*[21] to generate the digital spectrograms that have already been mentioned; it, too, makes use of digital storage for the automatic processing of spectrum information about real speech. Although the design had, as its primary objective, the evaluation of a bandwidth compression system based on coded spectral cross-sections, there were provisions for manipulating and modifying the spectrum in numerous ways so that the device could serve as a general-purpose research tool. Analysis and synthesis are performed by a vocoder, with the digital spectra stored in a drum memory. The principal outputs are from a high-speed printer and a loudspeaker. A substantial part of the equipment is devoted to matching spectral sections and compiling various kinds of statistical data, some of which may be quite relevant to problems in phonetics.

The very promising development by *Stevens*, *Dennis* and their colleagues[28] at the Massachusetts Institute of Technology of synthesizers comprising dynamic analogs of the vocal tract is being continued actively. One phase of their work, following the addition of a nasal analog (DANA) to the original device (DAVO), was the introduction of digital-to-analog conversion equipment to permit control of the synthesizers by the TX-O computer. This step was aimed at the greatly increased flexibility of control by computer as compared with that provided by the original circuitry. Another phase is the construction of a new synthesizer with circuits designed for improved stability and direct digital control. Attention is being given also to computer programs for controlling the synthesis in terms that are natural and convenient for the experimenter; this involves an intimate interplay of constraints imposed by the computer, the synthesizer, and the articulatory process.

Speech stretchers are special variants of synthesizers, but of more than passing interest for research in phonetics. One type, using tape recordings and rotating magnetic heads to repeat brief segments of the recording, has been available commercially for several years; such devices yield good speech when the time expansion ratio is small. The operation of other types of speech stretchers involves both automatic extraction of speech parameters and storage of this infor-

mation for later read out in slow time. Digital devices with storage, such as those already mentioned, can serve as stretchers although analog methods are quite adequate and usually simpler. Such a speech stretcher has been built at Haskins Laboratories using a channel vocoder and a multi-track magnetic tape recorder to store and re-play the spectrum information that usually flows directly from analyzer to synthesizer. Figure 2 indicates the nature of the

Channel Vocoder

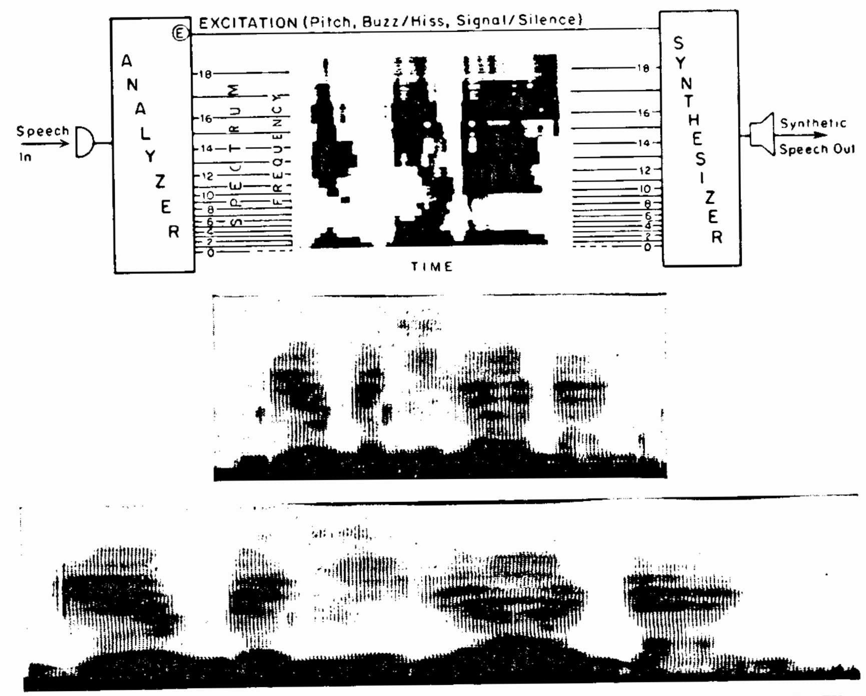


Fig. 2. One type of speech stretcher employs a vocoder for analysis and synthesis. The slowly-varying information that flows from analyzer to synthesizer, shown in essentially spectographic terms in the upper half of the figure, can be recorded on multi-track magnetic tape at one tape speed and then played into the synthesizer at a lower speed. The resulting speech will have the same spectral variations, except for the slower rate, and the same pitch as the original speech. A sample of original and stretched speech is shown in spectrographic form in the lower half of the figure.

information that is stretched, or compressed, and an example of the same speech at the normal rate and at half that rate. The quality of the stretched speech is unchanged by stretching, though it has the faults usual to vocoded speech. Stretching by factors of two-, four-, or eight-to-one, or compression by two-to-one, is done quite easily by changing motor speeds on the tape recorder. Other factors, ranging up to ten-to-one are possible, and have been used in making slow-motion X-ray movies of speech articulation[29].

II. Current Trends in Research on Phonetics

The developments mentioned in the preceding sections on instrumentation for speech analysis and speech synthesis represented mainly further progress along well established directions. Some of the devices that were described were new enough to have escaped mention at the Helsinki Congress in 1961, but not very many. There was increased emphasis on automated analysis and on synthesis applied to the complete speech spectrum. There was, in short, progress without important change of direction.

Let us turn now to some trends in phonetic research that represent significant differences in objective or major changes in research method. Here, too, we shall look in vain for abrupt departures but, over the past three to five years, the shift in direction is unmistakable.

The Use of Computers as Research Tools

Just as the mere automation of Fourier analysis by the sound spectrograph opened the way, in a practical sense, for most of the experimental advances of the past two decades, so the dexterity of the general-purpose digital computer seems about to open new paths for research. There is this difference: the spectrograph was a single-purpose instrument; the computer is extremely versatile, especially when coupled to analog devices at input and output. A survey of the applications of computers to speech, primarily for engineering purposes, would be beyond the scope of this paper; however, a sampling may serve to indicate some of the potential – and actual – uses of computers in research on phonetics.

Statistical analyses of data about speech sounds is, of course, an obvious use for the computer's extensive memory and its sorting capabilities. Such a study was reported, at an early phase, by *Denes* at the Helsinki Congress; details have since been published[30]. The initial objective of this study was to examine the frequency of occurrence of pairs of phonems in a large sample of running text as a basis for applying linguistic constraints to the operation of a phonetic typewriter. Quite aside from this practical objective, the frequencies of phoneme digrams and minimal pairs reveal some very interesting characteristics of the English language and point, in particular, to the heavy functional load carried by certain phonemic distinctions. Comparable uses of computers with higher-order linguistic units have been especially numerous: the fields of information retrieval and of mechanical translation are largely computer-based; indeed, some of you may already have made plans to attend next year an International Conference on Computational Linguistics, a label that includes, according to the call for papers, "all uses of computers to manipulate natural or artificial languages".

Direct manipulation of the speech waveform by computers equipped with suitable analog-to-digital conversion devices and data buffers has been employed for a wide variety of tasks involving both analysis and synthesis. Such applications put heavy demands on the computer facility, in part because the speech waveform requires so much memory capacity per second of speech, and in part because the real-time inputs and outputs involve both precise timing and high data rates. The use of computer methods for making decisions about pitch periods in the waveform of voiced sounds has been described by *Gold*[31] of the Lincoln Laboratories, MIT. Several criteria are employed, related to those that might be used by a human observer in hand-marking the pitch periods; the computed period is a composite measure that is more reliable than the result from any single test. *Mathews*, *Miller* and *David*[32] at the Bell Telephone Laboratories have carried out Fourier analyses of individual pitch periods, obtaining measures of both the transfer function of the vocal tract and the glottal excitation function. Pitch-synchronous analysis has the virtue that results are directly comparable for a wide range of voices without regard to their pitch. *Pinson*[33], also at the Bell Telephone Laboratories, has performed comparable analyses by fitting a set of damped sinusoids to the speech waveform. By using only the portion of the vocal cord cycle during which the cords are closed, he obtained the transfer function for the vocal tract uncontaminated by information about glottal excitation. Computer methods for calculation of the volume velocity at the glottis is being reported by *Flanagan* at this Congress.

When the speech spectrum rather than the speech waveform is to be manipulated, the usual procedure is to use a bank of filters or some form of real-time spectrum analyzer ahead of the computer rather than to compute the analysis, primarily because such computations are time consuming and therefore expensive. A recent example of spectrum manipulation by computer is the work of *Harris* and *Weiss*[34] of Columbia University and the Federal Scientific Corporation who used a high-resolution, real-time spectrum analyzer of the correlation type and an IBM 7090 computer. Their programs yield information about the spectrum power, the amplitudes and frequencies of the first three formants, and a short-time average of the voice pitch derived from harmonic spacings. The speech spectrum has been used by a number of workers as a basis for speech recognition; in particular, the *Forgies*[35], working with the TX-2 computer at Lincoln Laboratories, have made extensive studies of the automatic recognition of individual phonemes. *Sandra Pruzansky*[36], at the Bell Telephone Laboratories, has used the method of matching spectral patterns to achieve automatic talker recognition; this was an extension of an earlier experiment by *Denes* and *Mathews*[37] on the recognition of spoken digits by pattern matching.

The simulation of devices rather than their construction is another field in which computers can be especially useful. An example is provided by *Golden*[38], working with the IBM 7090 computer at the Bell Telephone Laboratories, in simulating a voice-excited vocoder. Although it required nearly 200 times as long to compute the speech from the simulated vocoder as would have been required by an actual device, the simulation was not only of high quality, but also allowed a number of design changes to be tested quickly without the cost of building and replacing equipment. Another example, already cited, is the cepstrum pitch analyzer; here, too, construction of the actual device would have been laborious and costly.

The use of computers to control the operation of other devices may require about the same peripheral equipment that is employed in simulation. Nevertheless, there is a significant difference and some of the most exciting prospects appear to lie in this area of application to phonetic research. The reason is simple enough: by divorcing the control function from the signal processing operation, it becomes easier for the human experimenter to intervene directly in the experiment. He can use a computer as a lively and tireless assistant, as my colleagues and I plan to do with our Digital Speech Manipulator, or he can guide the computer's eager efforts by supplying an element of judgment and direction. One example of the latter type is the use by *Denes*[39] of a digital computer in recovering the glottal pulse waveform by inverse filtering. The technique is essentially

that used by *Miller* and others[22–24], but without the requirement for special equipment. Another application, likewise not strictly dependent on a computer but making good use of it, is the analysis-by-synthesis procedure that will be discussed in a later section.

The production of speech directly from words or phoneme strings is another area in which computers have an advantage over analog methods in speed and facility. Various methods have been described[40], for generating speech by rules from dyads, from diphones, or from stored data in various other forms. The first use of a general-purpose digital computer in generating speech by rule from a phonemic transcription was made by *Gerstman* and *Kelly*[41]; many of you have heard the famous soliloquy from Hamlet as rendered by the IBM 7090.

These are a few of the ways computers have been used in processing speech information. The advantages have been stressed rather than the disadvantages, though there are some of the latter. Computers are expensive, both in money and in programming effort; also, for speech use, they are likely to require more analog accessories than might be supposed. Finally, they may prove to be surprisingly slow in overall performance despite the enormous speed with which their internal transactions are conducted; speech programs that run in real time are, as of today, the exception rather than the rule.

The trend to computers is nevertheless real and significant; indeed, an awareness of computer capabilities is becoming a minimal requirement for following research in experimental phonetics. The reasons for this trend were explained by *Denes*[42] in his paper at Helsinki on "The use of computers for research in phonetics". He points out that, although extensive facilities are required, "these facilities are available without the need for any electronic design on the part of the experimenter: that has been taken care of by the designer of the computer. All the phonetician has to do is to write down a series of instructions that embody the logical sequence of steps that he wants to have carried out: the computer will do the rest." In short, somebody else has done the work, insofar as equipment is concerned. *Denes* points out also the advantage of programming over construction in allowing the experimenter to keep his attention more closely focussed on the logic of the experiment. He concludes, as I would, by saying: "...there is every indication, therefore, that the availability of computers is about to produce a profound change in the way in which the experimental phonetician approaches his problems, in the range of experiments open to him and in the kind of training he requires to enable him to carry out his work." I would add only that the course of research over the last three years bears out this prediction.

Emphasis on the Changing Configurations of the Vocal Tract

A concern with the shape of the vocal tract is certainly not new in phonetics; nevertheless, the resurgence of interest in this topic is one of the significant trends in current research. What is new is not so much the area of interest as the reasons for that interest. One of these is the adoption as a working hypothesis of the idea that speech perception is somehow closely linked to articulation. This is entirely consistent with the view that sound spectrograms are so useful precisely because of an underlying relationship between the articulatory sequences they display and the linguistic units of the utterance. A second reason is the development of a powerful new method for studying speech.

The method of analysis-by-synthesis as developed by *Stevens* and his co-workers at MIT[43] was used initially as a means for locating the formant frequencies of slowly-changing vocalic sounds by matching the spectrum of the actual speech against a computed spectrum. The computation makes use of variable parameters with constraints set by a "resonance" model of speech generation. Fig. 3 illustrates the analysis-by-synthesis procedure: the incoming speech information (from a bank of filters) is held in temporary storage and presented to the comparator as the input spectra. Computed spectra are also presented to the comparator; these are formed in the spectrum generator by some strategy that will minimize the error score, i.e., the difference between the two spectra. This strategy may be carried out either by a computer or by a human observer working from a cathode-ray display. The principal limitations of the method are the adequacy of the model for speech production (incorporated into the program of the spectrum generator) and the strategy for minimizing errors, though the latter affects principally the speed and level of automation that can be obtained.

The acoustical theory[44] implicit in the operation of the spectrum generator treats the spectrum of the vocal tract output (in decibels) as the sum of a source spectrum, a transfer function, and a radiation

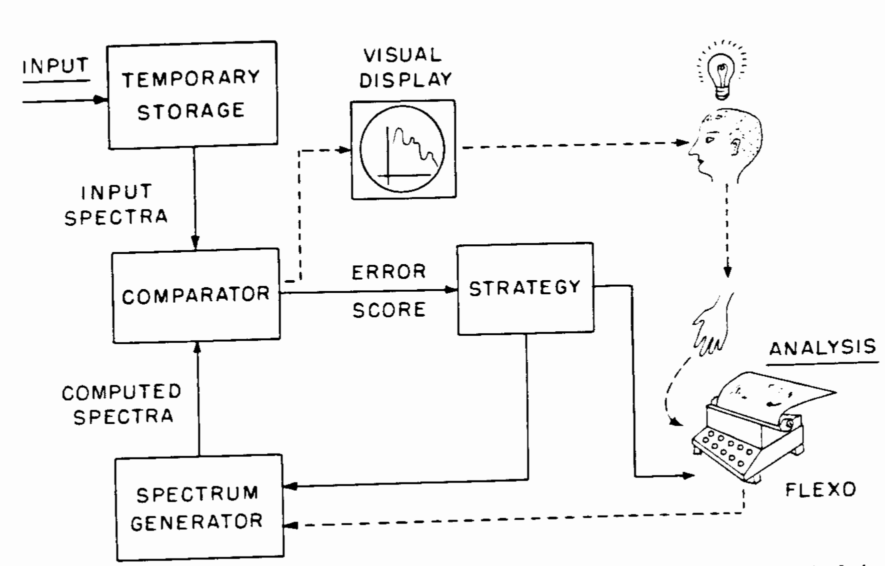


Fig. 3. Schema for the method of analysis-by-synthesis [Reproduced from *Speech Analysis and Synthesis Final Report* (Ref.[28]) with the kind permission of Prof. *K. N. Stevens*].

characteristic. Most of the spectrum variation, within a given class of speech sounds, is assignable to the transfer function which, in turn, is determined by the articulatory configuration and source location. In the case of non-nasal vowel and vowel-like sounds, the transfer function comprises only poles, or resonances; for nasal and fricative sounds, the transfer function will have zeros, or anti-resonances, as well as poles. Thus, the analysis-by-synthesis technique based on the resonance model for speech generation yields a description of the speech spectra in terms of the poles and zeros of the vocal tract transfer function (and also of generalized source and radiation characteristics). The description is, in effect, an analytic version of the sound spectrogram; ongoing articulation is represented by movement of poles and zeros over the complex plane in the one case and by formant transitions and spectral shifts in the other.

The method can be extended, however, to deal with relationships between spoken sound and vocal tract configuration, if use is made of an "articulatory" model for the generation of the computed spectra. The box marked strategy in the figure, or the human operator, must now provide information to the spectrum generator in terms of cross-sectional areas of the vocal tract at successive points along its length. The computation then proceeds in two stages. First, the poles and zeros of the transfer function are computed from the shape of the acoustic tube, and second, the computed spectra are generated from these pole-zero specifications. Both strategy and computation are now more complex, but an important advantage has been gained, particularly for the consonant sounds with their more complicated spectra. The articulatory configurations must change smoothly and continuously in proceeding from vowel to consonant, whereas the pole-zero configurations need not show comparable continuity. Also, it seems reasonable to expect that the constraint on the shape parameters, due to physical constraints on the organs of speech, might be simpler than constraints that operate at the acoustic level.

It may be feasible to extend the general method of analysis-by-synthesis still further to include a spectrum generator that makes use of rules governing the generation of speech sounds in accordance with a set of phonetic parameters "which are considered to be the result of a set of neuromuscular instructions or excitations that operate on the mechanical system consisting of lips, tongue, velum, and so forth. It is hypothesized that, for a given speech sound, these excitations assume a set of steady values, and jump discontinuously from one set of values to another as the sequence of speech sounds is generated. It is assumed that a unique set of excitations is associated with each speech sound. Relations between the excitations and the resulting articulatory motions are undoubtedly not linear, but it is convenient to characterize the mechanical properties of the articulatory structures in an approximate manner by a set of system functions[45]." Thus, it may eventually be possible to arrange the strategy box so that its signals to the spectrum generator (and to the outside world) are in the form of segmental phonemes. The assumption implicit in this extrapolation from current capabilities is that a simple relationship exists between the shape of the tract, described in phonetic parameters, and the linguistic units of the message. Whether or not this expectation is realized, the power of the analysis-by-synthesis procedure more than justifies the resurgence of interest in vocal tract configurations and their relation to the speech that is produced.

The major instrumentation used thus far in analysis-by-synthesis studies has been a filter bank, a general-purpose digital computer and its peripheral cathode-ray display that permits the operator to observe and control the spectrum matching operation.

An obvious further requirement in studying articulatory configurations is, of course, to obtain as direct information as possible about events in the human vocal tract during speech. X-ray motion pictures are a principal source of such information[46], though high-speed photography of the lips and of the vocal cords is also useful[47]. X-ray techniques were discussed at length in a review paper by the *Subtelneys* at Helsinki, and their synchronization with speech and spectrogram were described by *Truby*[48] at the same Congress. There have been extensions in X-ray technology within the past few years, primarily in the direction of higher frame rates and the use of pulsed emission, both of which improve time resolution. Viewing equipment that permits flickerless projection at very low frame rates is extremely useful, though not new. The combination of high-speed photography and speech stretching brings to slow-motion projection the added dimension of synchronous sound.

We may note, at this point, not only a trend toward studies of articulation, but also one away from an overriding concern with relationships between acoustic signal *per se* and linguistic unit. Undeniably, the acoustic signal must be heard to be understood, and it is useful to characterize the sounds of speech by the acoustic cues that ensure their recognition; nevertheless, the research trends discussed here and in the following sections do reflect a shift in emphasis from acoustic to articulatory aspects of the speech event.

Emphasis on the Perceptual Nature of Linguistic Units

The trend toward greater attention to the act of speaking is due in considerable part to gains in our understanding of the perceptual nature of linguistic units (at the level of phonemes or distinguishing features). The very special qualities of language, most notably its exceptional efficiency as an acoustic signalling system, are of course intimately bound up with a categorizing process that reshapes and regroups incoming sound into discrete units of the same kind as those that control the production of speech. Clearly this categorizing process and the perceptual nature of the categorical units themselves are of primary interest in understanding the relationship of sound to language.

The nature of this relationship has been studied intensively. One line of investigation, although it concerned itself initially with acoustic cues for perception, was led by the experimental evidence to a motor theory of speech perception. The involvement of motor reference in the perception of speech sounds, and the evidence for it, was discussed in some detail by *Liberman* in 1957 and again in 1962[49]; according to this theory, the perceptual conversion of incoming sound stuff to linguistic units can, and often does, make use of the greater distinctiveness of the articulatory gestures that produce the sounds than exists in the sounds *per se*. What concerns us here, however, is not so much the theory, or the evidence for and against it, as the questions it poses about the nature of linguistic units and the rationale it provides for certain kinds of experiments.

Much of the research on the perceptual nature of linguistic units is concerned with their distinctiveness and with the differences in distinctiveness of different classes of units. It has been found, for example, that the stop consonants are perceived in a highly categorical manner, which is roughly equivalent to saying that even when the sound stimulus differs substantially from the "norm" a listener will still hear the same phoneme and may, indeed, be unable even to detect the deviation. The situation with vowels is quite different: the listener readily detects small differences in the acoustic signal but his identifications of deviant stimuli are less sure and are strongly influenced by context[50]. It may be noted in passing that this marked difference in the perceptual nature of vowels and consonants could have been predicted from the theory. The line of investigation indicated here is currently a very active one and it seems likely to lead to additional insights into some of the fundamental problems of phonology[51]. As to techniques and instrumentation, this kind of research differs from much that has been discussed thus far in not being instrument-bound; rather, a wide variety of techniques is employed. The principal instrumental requirements are speech synthesizers for the production of controlled test stimuli and devices for measuring reaction time and muscle activity.

Emphasis on Neuromuscular Aspects of Speech Gestures

The hypothesis that speech perception is somehow closely related to articulation has served to motivate research on neuromuscular aspects of speech gestures, just as it has prompted increased attention to the changing configurations of the vocal tract. Are these essentially the same approach, or is there a significant difference? Certainly they are related, since muscle contractions and cavity shapes stand to each other as cause to effect; but muscle

contractions, being the prior events, may stand in the more direct relation to linguistic units. Thus, the sequence of events in production, which presumably starts with the intended message organized in appropriate linguistic units, proceeds by successive transformations of the message into motor command signals on the neural pathways to the articulators, then into contractions of selected sets of muscles in suitable time patterns, then into the articulatory configurations caused by these contractions, and finally into the stream of sounds that we call speech. Each of these conversions blurs still further the relationship with the initial linguistic units of the message. The encoding of motor commands into shapes and movements is an especially complex transformation because of the constraints inherent in the bone-and-muscle mechanism of the vocal tract; since the motor commands operate ahead of these complications, they escape this kind of recoding. Hence, of all the speech events – acoustic signal, articulatory shape, or neuromotor command – about which we can reasonably expect to collect information, the neural commands to the articulators would seem to provide the simplest relationship to the intended and perceived units of the message.

The experimental objective, therefore, is to get a description of speech in terms of motor commands or, since muscle contraction follows so directly from neural signal, in terms of patterns of muscle contractions and their relation to the phonemes (or distinguishing features) of the message. The total pattern of contractions may, however, contain much that is irrelevant to the phonemic distinction so that the task, put more sharply, is to seek for those component parts of the total gesture, the "characteristic gestures", that relate most directly to linguistic units[52].

The experimental procedure depends primarily on electromyography, i.e., on the fact that muscle contraction is accompanied by electrical signals that can be collected by putting electrodes on the surface of the face or tongue, or inserting small needles into the muscles. Although electromyography gives the most direct information about muscle activity, other techniques are valuable whenever reliable inferences can be drawn from the data they provide: pressure measurements both above and below the glottis, using the procedures developed by *Ladefoged*[53] give valuable information about the timing of occlusions and releases; throat microphones and equipment for transillumination[54] both yield valuable data about activity at the glottis; X-rays, particularly in slow motion with synchronous sound, can be helpful in suggesting electrode placements; and sound spectrograms can serve as a guide in planning experiments and as a check on inferences drawn from other data.

The instrumentation for research of this kind is mostly conventional though some of it must be designed for the specific task. It may be useful to describe the facilities developed at the Haskins Laboratories in starting this field of research[55, 52]. One of the first, and most troublesome, problems was to devise an electrode system that could be used inside the mouth and on the tongue, as well as on the face. A solution was found in the use of the small suction cup electrodes shown in Fig. 4. The electrode is hemispherical, about

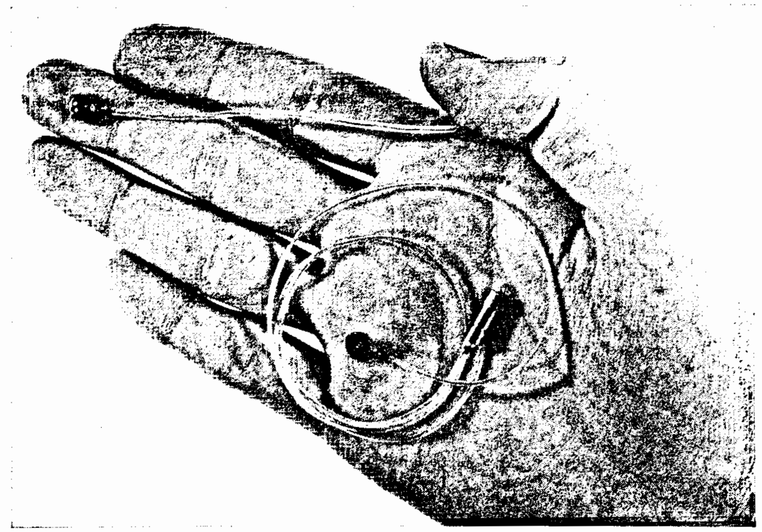


Fig. 4. Suction cup electrodes used in studies of muscle contraction by surface electromyography.

6 mm in diameter, and connected to a brass plug by a flexible plastic tube about 50 cm long. A thin steel wire threaded through the tube provides electrical connection to the plug, which can be inserted into a manifold that applies vacuum to all the electrodes as well as electrical connection from each electrode to its own preamplifier. The preamplifiers are commercial units designed for low noise, high common-mode rejection and high amplification. The system, shown diagrammatically in Fig. 5, includes eight electromyographic channels and eight other channels for varied uses. Each channel has additional amplification and facilities for monitoring signal levels

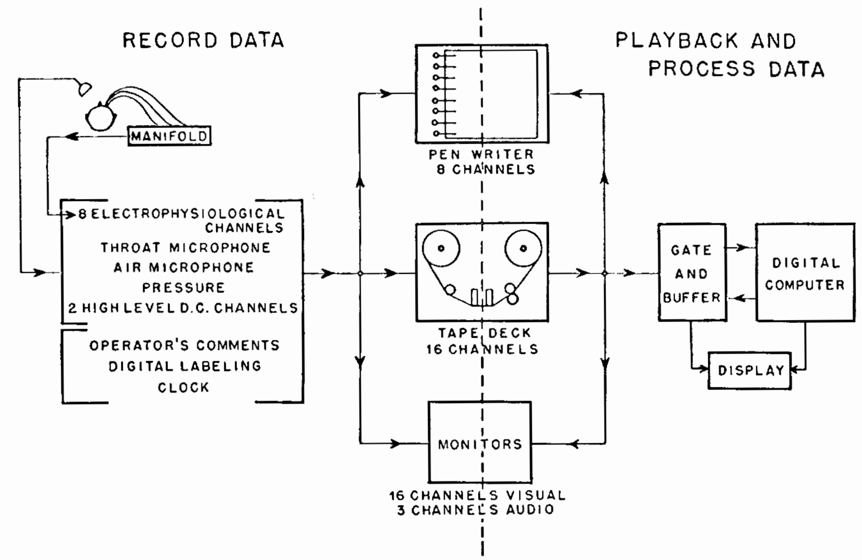


Fig. 5. Diagram of equipment and procedures for physiological studies of speech articulation. Data collection is indicated on the left of the central dashed line and data processing on the right.

during an experiment. At present, the non-myographic channels are used for a conventional microphone signal, a throat microphone, two pressure transducers, two digital signals for timing and identification of events, a pre-recorded sequence of utterances used to guide the subject's productions, and a microphone for comments by the experimenter. All sixteen channels are recorded on 1 inch magnetic tape as the experiment proceeds; parallel use is made of an 8-channel pen recorder, primarily to monitor the signals from electrodes and throat microphone. After the experimental session has been concluded, the monitoring amplifiers and pen recorders can be used to transcribe the experimental data and the event codes from the magnetic tape. Thus far, data reduction has all been done by hand methods – a slow and laborious procedure. Arrangements are almost complete to shift this burden to a small computer which will read the data directly from the magnetic tape and, with the aid of event codes and clock signals, sort and process it. The experimenter can control the operation by prior examination of the pen traces or by observing a cathode-ray display of the data as it arrives at the computer.

The trend toward physiological problems and methods in research on phonetics is not, of course, limited to work aimed primarily at a motor command description of speech production and perception. *Ladefoged*[56] and his co-workers at the University of California are following up the work reported at Helsinki and elsewhere on subglottal activity during speech with an extensive program aimed at describing and quantifying not only subglottal events in speech production but activities of the entire vocal tract. X-ray methods have been used to obtain parameters of characteristic lip positions in American English vowels, and analog models of the vocal tract are under construction. *Philip Lieberman*[57], at the Air Force Cambridge Research Laboratory, has combined techniques for high-speed photography of the glottis, waveform recording on the high-speed film, and computer-aided processing of the pictures to conduct several studies of vocal cord function. *André Malecot*[58], at the University of Pennsylvania, has developed some very ingenious equipment with which to test the ability of subjects to estimate intraoral pressures such as they themselves might generate in producing stop consonants. This bears, of course, on the question of whether or not direct awareness of such pressure differences as exist between fortis and lenis productions could provide a speaker with usable feedback information about his productions. The apparatus transmits pressure pulses of controlled magnitude and duration to the subject's mouth or to pressure plates at lips or palate. Mention has been made already of methods for transilluminating the glottis[54]; in the main, these require an open-mouth position, but it seems likely that extensions of this technique to running speech will soon be possible. A recent article by *Kozhevnikov et al.*[59] on instrumentation for phonetic research includes a description of an artificial palate with a number of electrodes that permit a running characterization of the patterns of tongue contact.

A substantial part of the physiological research on speech is, of course, medically oriented, especially to the problems of cleft-palate speech[60]. Several X-ray installations designed primarily for speech research are now available in the United States; in Sweden there is the excellent facility at the Wenner-Gren Foundation originally adapted to speech research by *Truby* and used by him for studies of the vocalization of infants.

Summary

In summary, this discussion of instrumental methods for research in phonetics has attempted, first of all, to report the principal new

instruments and techniques that have appeared since the Helsinki Congress three years ago. There were not many, though the tempo of development in related fields such as the bandwidth compression of voice signals suggests that some new tools for phonetic research may be expected soon.

The discussion then turned – or was bent – to the broader topic of the directions in which phonetic research is moving. It seems to this reviewer that some changes have occurred within the last few years and that they and the reasons for them are significant. One such trend is toward a major role for computers in many phases of research on phonetics, not because computers generate exciting new ideas, but simply because they are versatile, fast, and ready to be used on problems that were heretofore unapproachable.

The other trends for which significance was claimed depend less on new technology than on ideas about the nature of phonological units. Because these units depend on perceptual operations for their very existence as categorical entities, the linkage of perception to speaking or to listening is a crucial one. The idea that perception may be closely tied to the act of speaking represents something of a departure from the views that guided phonetic research during the 1950's. It carries with it the implication that there is a better chance of finding one-to-one relationships between the things one measures and the atomic units of language when working with articulation than when working with the acoustic signal. But measures of articulation can be either in terms of the shape of the vocal tract or of the neuromotor commands that control it. The latter are upstream with respect to the complex transformation that relates command to shape, and so may provide the simpler description – though quite possibly the ideal of a one-to-one correspondence may not be realizable short of the central nervous system.

So much for motivations. The consequence, in some laboratories at least, has been a shift in emphasis toward studies of the articulatory process by various means: notably, by the method of analysis-by-synthesis; by the use of X-ray movies; by the methods of experimental psychology applied to the perceptual nature of consonants and vowels; and by electromyographic and other physiological measures of speech gestures.

It is yet too early to exhibit the results of these lines of research and thereby to demonstrate their significance but, by the same token, your reviewer is not yet subject to the full penalties for false prophecy.

References

1. *Fischer-Jørgensen, Eli:* What can the new techniques of acoustic phonetics contribute to linguistics? Proc. 8th Int. Congr. Linguists, pp. 433–499 (Oslo 1958).
2. *Fant, C. G. M.:* Modern instruments and methods for acoustic studies of speech. Proc. 8th Int. Congr. Linguists, pp. 282–362 (Oslo 1958).
3. *Fant, C. G. M.:* Sound spectrography. Proc. 4th Int. Congr. Phonetic Sciences, Helsinki 1961, pp. 14–33 (The Hague 1962).
4. *Cooper, F. S.:* Speech synthesizers. Proc. 4th Int. Congr. Phonetic Sciences, Helsinki 1961, pp. 3–13 (The Hague 1962).
5. Language Master: Audio Instruction Device, mfg. by Bell and Howell, 2100 McCormick Rd, Chicago 45, Illinois, USA.
6. The *Visicorder:* mfg. by Heiland Division, Minneapolis-Honeywell Regulator Co., Denver 10, Col., is typical of several such recorders made in the USA; Other manufacturers include Century Electronics & Instruments, Tulsa, Okla., Consolidated Electrodynamics, Pasadena, Calif., and Midwestern Instruments, Tulsa, Okla.
7. *Potter, R. K.:* Visible patterns of sound. Science *102:* 465 (1945). – Six articles by *Potter, R. K. et al.* in J. acoust. Soc. Amer. *18:* 1–89 (1946).
   Potter, R. K.; Kopp, G. A. and *Green, Harriet C.:* Visible Speech (New York 1947).
8. *Steinberg, J. C.:* Application of sound measuring instruments to the study of phonetic problems. J. acoust. Soc. Amer. *6:* 16–24 (1934). The spectrogram referred to in the text is Fig. 9.
9. *Joos, M.:* Acoustic phonetics. Lang. Monogr. 23, Language *24:* 2. Suppl. (1948).
10. Kay Electric Company, Pine Brook, N. J., USA.
11. A spectrograph, based on a recent Bell Telephone Laboratories' design, has been announced for sale by Voiceprint Laboratories, Somerville, N. J.; for an account of the earlier spectrograph mentioned in the text, see *Prestigiacomo, A. J.:* Plastic tape sound spectrograph. J. Speech Dis. *22:* 321–327 (1957).
12. Personal communication; a description is to be published.
13. Quarterly Progress and Status Report, Speech Transmission Laboratory, R. I. T., Stockholm, STL-QPSR *4:* 1–6 (1963).
14. *Harris, C. M.* and *Waite, W. M.:* Gaussian-filter spectrum analyzer. J. acoust. Soc. Amer. *35:* 447–450 (1963). – Display of sound spectrographs in real time. J. acoust. Soc. Amer. *35:* 729 (L) (1963).
15. *Gill, J. S.:* A versatile method for short-time spectrum analysis in "real time". Nature *189:* 117–119 (1961).
16. *Weiss, M. R.* and *Harris, C. M.:* Computer technique for high-speed extraction of speech parameters. J. acoust. Soc. Amer. *35:* 207–214 (1963).
17. *Prestigiacomo, A. J.:* Amplitude contour display of sound spectrograms. J. acoust. Soc. Amer. *34:* 1684–1688 (1962).
18. *Kersta, L. G.:* Voiceprint identification. J. acoust. Soc. Amer. *34:* 725 (A) (1962). – Voiceprint-identification infallibility. J. acoust. Soc. Amer. *34:* 1978 (A) (1962).
19. *Wood, D. E.* and *Hewitt, T. H.:* New instrumentation for making spectrographic pictures of speech. J. acoust. Soc. Amer. *35:* 1274–1278 (1963).
   Wood, D. E.: New display format and a flexible-time integrator for spectral-analysis instrumentation. J. acoust. Soc. Amer. *36:* 639–643 (1964).
20. *Huggins, W. H.:* A phase principle for complex-frequency analysis and its implications in auditory theory. J. acoust. Soc. Amer. *24:* 582–589 (1952).

21. *Smith, C. P.:* Voice-communication method, using pattern matching for data compression. J. acoust. Soc. Amer. *35:* 805 (A) (1963). – Pitch intensity and spectrum stereotypes in vocoder-analysis data for readings of phonetically balanced word lists. J. acoust. Soc. Amer. *35:* 807 (A) (1963). – An approach to speech bandwidth compression. Proc. Seminar on Speech Compression and Processing. Electronics Research Directorate, Air Force Cambridge Research Center (1959). Avail. fr. U.S. Dept. of Commerce, Office of Technical Services, Washington, D.C.
   Kerby, A. E. and *Schoene, L. D.:* A digital voice data processer, ibid.
22. *Miller, R. L.:* Nature of the vocal cord wave. J. acoust. Soc. Amer. *31:* 667–677 (1959).
23. *Lawrence, W.:* Formant tracking by self-adjusting inverse filtering. J. acoust. Soc. Amer. *35:* 1113 (A) (1963).
24. *Holmes, J. N.:* An investigation of the volume velocity waveform at the larynx during speech by means of an inverse filter. Proc. Speech Communications Seminar, Speech Transmission Laboratory, RIT (Stockholm 1963).
25. *Noll, A. M.:* Short-time spectrum and "cepstrum" techniques for vocal-pitch detection. J. acoust. Soc. Amer. *36:* 296–302 (1964).
26. Proc. Speech Communications Seminar, Stockholm 1962, Speech Transmission Laboratory, RIT (3 vols.) (Stockholm 1963).
   Summaries of the papers presented at Stockholm are to be found in J. acoust. Soc. Amer. *35:* 1112–1117 (1963); a brief announcement about the meeting, ibid., *34:* 1793 (1962).
27. *Liberman, A. M.:* Some results of research on speech perception. J. acoust. Soc. Amer. *29:* 117–123 (1957).
   Delattre, P.: Les indices acoustiques de la parole: premier rapport. Phonetica *2:* 108–118; 226–251 (1958).
   Liberman, A. M.; Ingemann, F.; Lisker, L.; Delattre, P. and *Cooper, F. S.:* Minimal rules for synthesizing speech. J. acoust. Soc. Amer. *31:* 1490–1499 (1959).
28. Speech Analysis and Synthesis Final Report, Massachusetts Institute of Technology, Research Laboratory of Electronics, Cambridge, Mass. (1963), doc. no. AFCRL-64-300, avail. fr. U.S. Dept. of Commerce, Office of Technical Services, Washington, D.C.
   Various papers in Proc. Speech Communication Seminar, Speech Transmission Laboratory, RIT, Stockholm (1963).
   Dennis, J. B.; Whitman, E. C. and *Tomlinson, R. S.:* On the construction of a dynamic vocal-tract model. J. acoust. Soc. Amer. *36:* 1038 (A) (1964).
   Dennis, J. B.: Control language for experimental speech synthesis. J. acoust. Soc. Amer. *36:* 1039 (A) (1964).
29. *Abramson, A. S.* and *Cooper, F. S.:* Slow-motion X-ray pictures with stretched speech as a research tool. J. acoust. Soc. Amer. *35:* 1888 (A) (1963).
30. *Denes, P. B.:* On the statistics of spoken English. J. acoust. Soc. Amer. *35:* 892–904 (1963).
31. *Gold, B.:* Computer program for pitch extraction. J. acoust. Soc. Amer. *34:* 916–921 (1962).
32. *Mathews, M. V.; Miller, Joan E.* and *David, E. E., Jr.:* Pitch synchronous analysis of voiced sounds. J. acoust. Soc. Amer. *33:* 179–186 (1961).
33. *Pinson, E. N.:* Pitch-synchronous time-domain estimation of formant frequencies and bandwidths. J. acoust. Soc. Amer. *35:* 1264–1273 (1963); for another study employing comparable techniques, see:
   Manley, H. J.: Analysis-synthesis of connected speech in terms of orthogonalized exponentially damped sinusoids. J. acoust. Soc. Amer. *35:* 464–474 (1963).
34. *Harris, C. M.* and *Weiss, M. R.:* Pitch extraction by computer processing of high-resolution Fourier analysis data. J. acoust. Soc. Amer. *35:* 339–343 (1963).

35. *Forgie, J. W.* and *Forgie, Carma D.:* Scheme for utilizing formant-transition cues in the recognition of a class of prevocalic consonants; also Automatic method of plosive identification. J. acoust. Soc. Amer. *34:* 1979 (A) (1962).
36. *Pruzansky, Sandra:* Pattern-matching procedure for automatic talker recognition. J. acoust. Soc. Amer. *35:* 354–358 (1963).
37. *Denes, P. B.* and *Mathews, M. V.:* Spoken digit recognition using time-frequency pattern matching. J. acoust. Soc. Amer. *32:* 1450–1454 (1960).
38. *Golden, R. M.:* Digital computer simulation of a sampled-data voice-excited vocoder. J. acoust. Soc. Amer. *35:* 1358–1366 (1963).
39. Personal communication; see also
Miller, Joan E. and *Mathews, M. V.:* Investigation of the glottal waveshape by automatic inverse filtering. J. acoust. Soc. Amer. *35:* 1876 (A) (1963).
40. *Liberman, A. M.; Ingemann, F.; Lisker, L.; Delattre, P.* and *Cooper, F. S.:* Minimal rules for synthesizing speech. J. acoust. Soc. Amer. *31:* 1490–1499 (1959).
Peterson, G. E.; Wang, W. S-Y and *Sivertsen, Eva:* Segmentation techniques in speech synthesis. J. acoust. Soc. Amer. *30:* 739–742 (1958).
Estes, S. E.; Maxey, H. D. and *Walker, R. M.:* Speech synthesis from stored data. IBM Journal *8:* 2–12 (1964).
Cooper, F. S.: Speech from stored data. IEEE Convention Record *7:* 137–149 (1963).
41. *Gerstman, L. J.* and *Kelly, J. H.:* An artificial talker driven from a phoneme input. J. acoust. Soc. Amer. *33:* 835 (A) (1961); see also
Holmes, J. N. and *Shearme, J. N.:* Speech synthesis by rule controlled by a small, low-speed digital computer. J. acoust. Soc. Amer. *35:* 1911 (A) (1963).
42. *Denes, P.:* The use of computers for research in phonetics. Proc. 4th Int. Congr. Phonetic Sciences, Helsinki 1961, pp. 149–154 (The Hague 1962).
43. *Paul, A. P.; House, A. S.* and *Stevens, K. N.:* Automatic reduction of vowel spectra: an analysis-by-synthesis method and its evaluation. J. acoust. Soc. Amer. *36:* 303–308 (1964); also, the Final Report cited in Ref. 28, with its references.
44. *Stevens, K. N.* and *House, A. S.:* An acoustical theory of vowel production and some of its implications. J. Speech and Hear, Res. *4:* 303-320 (1961); for a complete treatment, see:
Fant, G.: Acoustic theory of speech production (Mouton, 's-Gravenhage 1960).
45. Speech Analysis and Synthesis: Final Report: 74–75. See Ref. 28.
46. Work in this field continues to be active, as evidenced by the research reported at this Congress by *Faure, Öhman* and *Lindblom, Ondráčková, Truby* and *Wierzchowska;* a further example is the work of one group concerned with quantitative measures of movement derived from x-ray movies:
Öhman, S. E. G. and *Stevens, K. N.:* Cineradiographic studies of speech: procedures and objectives. J. acoust. Soc. Amer. *35:* 1889 (A) (1963).
Heinz, J. M. and *Stevens, K. N.:* On the derivation of area functions and acoustic spectra from cineradiographic films of speech. J. acoust. Soc. Amer. *36:* 1037 (A) (1964).
Lindblom, B.E.F.: Articulatory activity in vowels. J. acoust. Soc. Amer. *36:* 1038 (A) (1964).
47. *Fujimura, O.:* Bilabial stop and nasal consonants: a motion picture study and its acoustical implications. J. Speech and Hear, Res. *4:* 233–247 (1961); see also Ref. 57.
48. *Subtelny, Joanne D.* and *Subtelny, J. D.:* Roentgenographic techniques and phonetic research. Proc. 4th Int. Congr. Phonetic Sciences, Helsinki 1961, pp. 129–146 (The Hague 1962).
Truby, H. M.: Synchronized cineradiography and visual-acoustic analysis, ibid. pp. 265–279; see also
Moll, K. L.: Cinefluorographic techniques in speech research. J. Speech and Hear, Res. *3:* 227–241 (1960).

49. *Liberman, A. M.:* Some results of research on speech perception. J. acoust. Soc. Amer. *29:* 117–123 (1957).
Liberman, A. M.; Cooper, F. S.; Harris, K. S. and *MacNeilage, P. F.:* A motor theory of speech perception. Proc. Speech Communication Seminar, Stockholm 1962, Speech Transmission Laboratory, RIT.
50. *Liberman, A. M.; Harris, Katherine S.; Hoffman, H. S.* and *Griffith, B. C.:* The discrimination of speech sounds within and across phoneme boundaries. J. exp. Psychol. *54:* 358–368 (1957).
Fry, D. B.; Abramson, A. S.; Eimas, P. D. and *Liberman, A. M.:* The identification and discrimination of synthetic vowels. Language and Speech *5:* 171–189 (1962).
Studdert-Kennedy, M.; Liberman, A. M. and *Stevens, K. N.:* Reaction time to synthetic stop consonants and vowels at phoneme centers and at phoneme boundaries. J. acoust. Soc. Amer. *35:* 1900 (A) (1963).
Stevens, K. N.; Öhman, S. E. G. and *Liberman, A. M.:* Identification and discrimination of rounded and unrounded vowels. J. acoust. Soc. Amer. *35:* 1900 (A) (1963).
51. *Lisker, L.; Cooper, F.S.* and *Liberman, A.M.:* The uses of experiment in language description. Word *18:* 82–106 (1962).
52. *Cooper, F.S.:* Research techniques: EMG. Proc. Conference on Communicative Problems in Cleft Palate 1963 (Amer. Speech and Hearing Ass.).
MacNeilage, P.F.: Electromyographic and acoustic study of the production of certain final clusters. J. acoust. Soc. Amer. *35:* 461–463 (1963).
53. *Ladefoged, P.:* Sub-glottal activity during speech. Proc. 4th Int. Congr. Phonetic Sciences, Helsinki 1961, pp. 73–91 (The Hague 1962). – Some physiological parameters in speech. Lang. and Sp. *6:* 109 (1963).
Ladefoged, P. and *McKinney, Norris P.:* Loudness, sound pressure, and subglottal pressure in speech. J. acoust. Soc. Amer. *35:* 454–460 (1963).
54. *Sonesson, B.:* On the anatomy and vibratory pattern of the human vocal folds (Lund 1960).
55. *Harris, Katherine S.; Rosov, R.; Cooper, F. S.* and *Lysaught, Gloria F.:* A multiple suction electrode system. J. Electroenceph. clin. Neurophysiol. (1964).
56. *Ladefoged, P.:* Physiological parameters of speech. J. acoust. Soc. Amer. *36:* 1037 (A) (1964).
Fromkin, Victoria A.: Parameters of lip position. J. acoust. Soc. Amer. *36:* 1037 (A) (1964).
Anthony, J.: True model of the vocal tract. J. acoust. Soc. Amer. *36:* 1037 (A) (1964).
57. *Lieberman, P.:* Some acoustic measures of the fundamental periodicity of normal and pathologic larynges. J. acoust. Soc. Amer. *35:* 344–353 (1963).
Soron, H.I. and *Lieberman, P.:* Some measurements of the glottal-area waveform. J. acoust. Soc. Amer. *35:* 1876–1877 (A) (1963).
58. Personal communication.
59. *Kozhevnikov, V.A.; Kuz'min, Yu.I.* and *Shapliakov, V. S.:* A system of transducers and electronic devices for analog recording and data processing in the study of speech articulation. Med. Electronics and Biol. Engineering *2:* 135–144 (1964).
60. Conference on Communicative Problems in Cleft Palate, Washington, D.C. (Amer. Speech and Hearing Ass. 1963).

Author's address: Dr. Franklin C. Cooper, Haskins Laboratories, 305 E. 43rd Street, *New York 17, N.Y.* (USA).

Proc. 5th int. Congr. phon. Sci., Münster 1964, pp. 172–200
(S. Karger, Basel/New York 1965).

From the Ohio State University, Columbus, Ohio

Juncture*

By Ilse Lehiste, Columbus

1. Introduction

The following report is concerned with juncture, which for the purposes of this paper is defined as a phonologically manifested boundary. A search for boundary signals implies the existence of certain linguistic units, whose boundaries may then be signalled by features associated with the term juncture. These linguistic units may be determined according to some non-phonological (i.e. grammatical) criteria, and the investigation may then be focussed on discovering the phonological manifestations that may accompany the grammatical boundaries. On the other hand, some recurrent phonological features may be observed without reference to the non-phonological structuring of the utterance; these features may suggest the presence of a boundary, and the units bounded by such junctures may be investigated in order to determine whether they constitute some linguistically significant building blocks of speech[1].

* This research has been supported by the National Science Foundation of the United States of America.

[1] I have considered various aspects of this problem in *An Acoustic-Phonetic Study of Internal Open Juncture* (s. Lit. 15). That publication also contains a bibliography of books and articles dealing with the problem of juncture. Lack of space makes it all but impossible to do full justice here to the work of all scholars who have concerned themselves with some aspect of this question. This is even more true with regard to the two language areas – Finno-Ugric and Slavic – from which the specific examples have been drawn that are discussed in the present paper. Since my primary aim was to present a research report, I have reluctantly decided to omit a critical discussion of previously expressed views concerning juncture, and not to attempt an exhaustive coverage of the relevant literature. The short bibliography presented at the end of this paper contains three types of materials. Certain of the references deal with the problem of boundary signals, especially at word level. A few references are included for each of the languages considered in

In two previous papers I have presented the view that there exists a relationship between units of the grammatical and phonological hierarchies, in which one may condition the other, but both have some degree of independence[2]. In this paper I shall discuss the phonological manifestation of boundaries between word-level units, and suprasegmental patterns characterizing such units. The specific examples are taken from Finnish, Czech, and Serbocroatian. The discussion is based on information obtained by acoustic-phonetic methods; distributional boundary signals are left out of consideration, although they may – and often do – constitute important criteria for the presence of a boundary. Spectrographic analysis of recorded test utterances constitutes the main research technique[3]. Most of the tapes have also been processed through a circuit designed to measure speech power[4]. The interpretation of intensity curves and oscillograms forms an important, although as yet incomplete, part of the investigation[5].

2. Units and their Boundaries in Finnish

During an acoustic-phonetic study of Finnish, a number of acoustic features was discovered to be associated with syllable

this paper; their selection was partly determined by the bibliographical information which they contain. Since the present paper constitutes a partial report of a series of studies in which I have been engaged for some time, occasional references are necessary to such earlier publications of which the current paper forms a continuation; these publications are likewise included in the bibliography.

[2] The relationship between syllable boundaries and morpheme boundaries was considered in "Acoustic studies of boundary signals" (Lit. 16). Some questions concerning word boundaries were brought up in "Compounding as a phonological process" (Lit. 17).

[3] Broad-band and narrow-band spectrograms were produced on the two Bell Telephone Laboratories' Model D spectrographs available at The Communication Sciences Laboratory of The University of Michigan, where most of the experimental work connected with this research was performed.

[4] The intensity circuit was designed and built in The Communication Sciences Laboratory by *G. E. Peterson* and *N. P. McKinney*. The signals were displayed on a multichannel Model 1108 Visicorder Oscillograph of the Minneapolis Honeywell Regulator Company. Measurement techniques employed in this and earlier studies are described in *Accent in Serbocroatian: An Experimental Study* (Lit. 18, pp. 2–12).

[5] The contribution of intensity toward establishing suprasegmental patterns characterizing phonological units is difficult to define. Intensity is often associated with stress, and differences in stress may indeed be expected to be accompanied by differences in intensity, all other factors being kept constant. The acoustic correlates of stress, however, are complex; the role of intensity as one of these correlates appears to be different in each of the three languages under consideration.

boundaries and word boundaries. Some of these will be discussed below[6].

There are numerous compound words in Finnish in which the first word ends in a vowel and the second begins with a vowel. In many instances the same V+V sequence may occur in otherwise similar noncompound words; in most cases the V+V sequence then contains a syllable boundary. (The word pair *lintuansa* – *lintu-ansa* may serve as an example.) In instances such as these, differences between the manifestations of the V+V sequence may be interpreted as signals of the presence of a word boundary. The test material analyzed in the course of the study contained 86 words of this type. The description of the boundary signals is based on productions of these pairs by all informants, although illustrations can be presented and actual measurements reported for only some of the speakers.

The phonetic quality of the second vowel in the V+V sequence was found to depend to a considerable degree on the presence or absence of a word boundary. Figure 1 presents an acoustical vowel diagram for certain of the vowels produced by speaker J.P. in words belonging to this type. On the diagram, the outer quadrangle (the solid line) connects points representing the average formant positions of the vowels /i ä a u/ occurring in the second position of a V+V sequence in instances in which the second vowel started the first syllable of the second element of a compound; the inner quadrangle (the dashed line) connects points representing the average formant positions of the same vowels in analogous position in noncompound words. As may be seen from the vowel diagram, the positions of the vowels occurring in noncompounds are considerably closer to the center of the acoustical vowel diagram than the positions of the same vowels occurring in noncompound words. The

[6] The data presented here are taken from an extensive study of boundary signals in Finnish, which will be reported in a forthcoming joint publication with *Kalevi K. Wiik*. In this study, the basic patterns were first established by studying the speech of several main informants; the generality of the patterns was determined by the study of smaller samples of utterances by a larger group of informants. Six main informants recorded from 567 to 1674 test utterances each, averaging 1000 utterances per informant. A restricted list of 118 test sentences was also recorded by seven informants each at the phonetics laboratories of Turku and Jyväskylä universities, and by four informants at the University of Helsinki. Several supplementary experiments were designed and carried out in order to follow up some hypotheses formulated on the basis of preliminary findings. Listening tests are being carried through to test the reaction of native listeners to the discovered boundary signals.

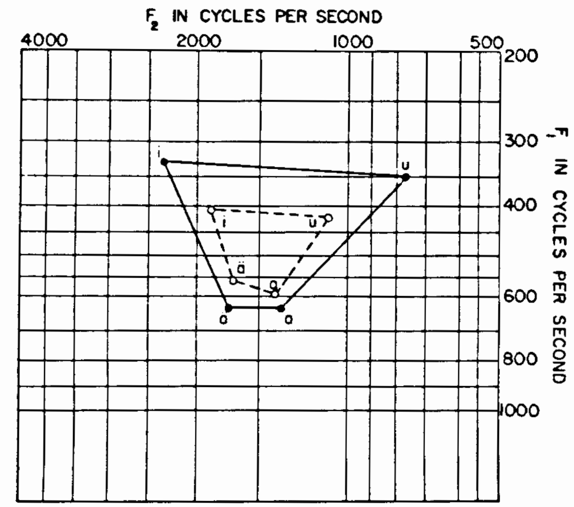


Fig. 1. Acoustical vowel diagram of four vowels occurring in Finnish test words produced by J.P. The solid line connects points representing vowels occurring in word-initial position in the second component of compound words; the dashed line connects points representing vowels occurring in analogous position in noncompound words.

occurrence of a reduced vowel indicates that no word boundary is present, while the occurrence of a non-reduced vowel serves as one of the phonetic features signalling the presence of the boundary.

A further cue to the presence of a word boundary in such sequences appears in the duration of the segments. The sequence V+V had a longer second component, if the second vowel started the next word. If the second word began with a closed syllable (as in the pair *lintuansa* – *lintu-ansa*), the additional duration was shared by the consonant closing the syllable. In the test materials recorded by speakers K-K.W. and J.P., the lengthening of the postjunctural vowel added an average of 4 centiseconds to the duration of the vowel, and 3 centiseconds to the consonant. The two speakers differed with respect to the effect of a following juncture on the vowel preceding the juncture. In utterances produced by J.P., the duration of the prejunctural vowel showed an average increase of 2.5 centiseconds, while K-K.W. had a negligible 0.5 centisecond addition.

A third clue to the presence of a word boundary consisted of the insertion of a brief period of laryngealization or a short glottal stop in cases where a word boundary occurred between two vowels. Such periods of laryngealization were also observed at word boundaries at other points in the utterances, not only setting off the test words in the frame. The glottal stop or laryngealization was likewise observed when the placement of a word boundary contradicted the expected syllabification pattern in sequences containing an intervocalic consonant. The period of laryngealization was particularly prominent in a subtype of test words involving a contrast between a long vowel and a sequence of two identical vowels containing a word boundary.

Figure 2 contains broad-band spectrograms of the pair *lintuansa – lintu-ansa*, spoken by informant S-L.K.; *pyryiltä – pyry-ilta*, spoken by P.K.; *asiasta – esi-aste*, spoken by J.P.; and *rantautua – ranta-utua*, produced by O.I. The feature of laryngealization is clearly observable in all pairs except the last; however, the difference between the

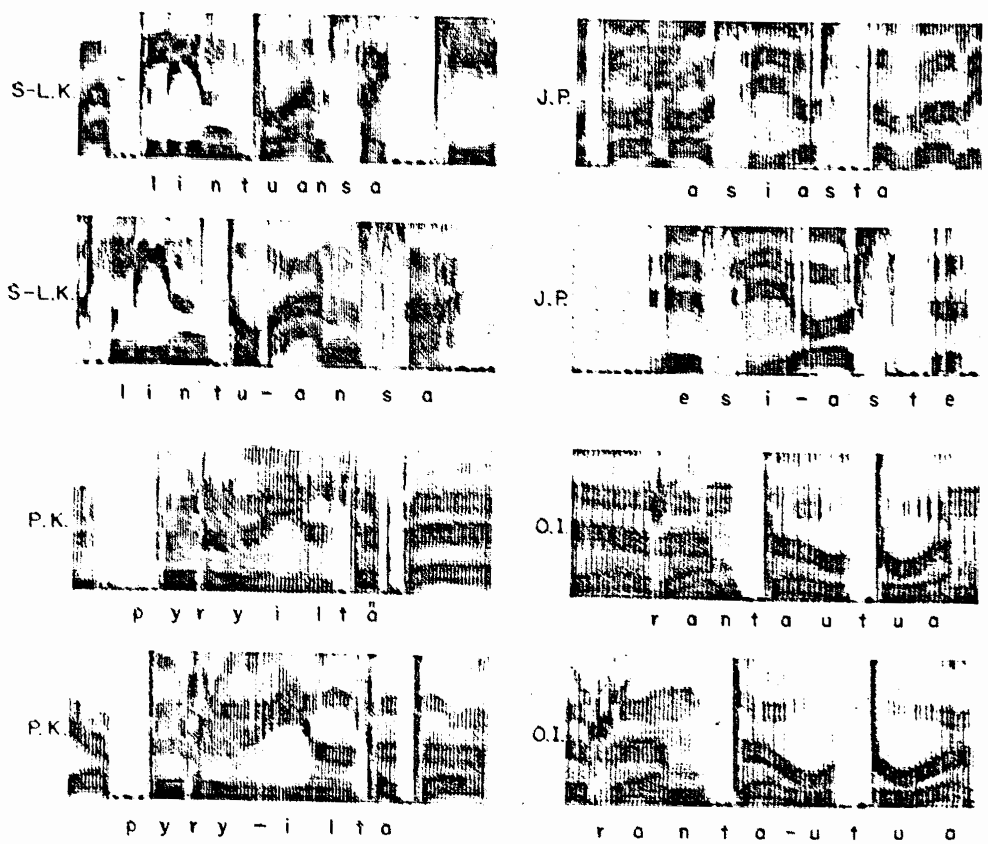


Fig. 2. Broad-band spectrograms of four pairs of Finnish test words, differing in the presence or absence of a word boundary. The utterances were produced by informants S-L.K., P.K., J.P., and O.I.

two /a/ + /u/ sequences as to length and phonetic quality is especially clear in this case. The increased duration of the vowels starting the second word of the compounds as well as of the consonants following these vowels is easily noticeable in every instance.

In this first set of words, the presence of a word boundary contrasted with the absence of a word boundary. The test materials were designed to contain twenty words representing a different type of contrast. In this case, both test words were compounds, i.e. both contained a word boundary, but the compound words differed in the placement of the boundary. In a V+C+V sequence of this kind, the consonant may either end the first word (if it is one of the dentals /s t n r l/) or begin the next word. In Finnish, a syllable boundary ordinarily occurs before an intervocalic consonant. When the word boundary placement contradicted the expected syllabification pattern, the boundary was usually manifested by a period of laryngealization or by a glottal stop. When the second word began with a consonant, no laryngealization was observed.

Since the consonant involved in such V+C+V sequences was frequently a nasal, a special experiment was conducted to study the effect of the placement of the word boundary upon the nasalization of the preceding and following vowel. Two informants (K-K.W. and K-A.W.) participated in the experiment, which involved the use of an oral and a nasal microphone, a two-channel tape recorder, and a multi-channel Visicorder oscillograph connected to a special speech power measuring circuit (cf. footnote 4). Each informant produced 95 frame utterances.

When the nasal consonant started the word, progressive nasalization of the vowel following the nasal consonant was always present. There was no appreciable anticipatory nasalization before a word-final nasal. When a word boundary occurred between a word ending in a nasal and one beginning with a vowel, progressive nasalization was not observed. Figure 3 shows oscillograms and intensity curves for the pairs *maa-nisäkäs – maan-isä* and *puu-neliö – puun-eliö*, produced by informant K-K.W. The oscillograms marked with O represent the signal from the oral microphone; the oscillograms marked N show the simultaneous output from the nasal microphone; the intensity graph corresponds to the nasal oscillogram. The test words were produced in the frame *Mitä. . . . tarkoittaa;* the first word of the frame thus offers additional evidence of the effect of a nasal on the following vowel. A comparison of the two

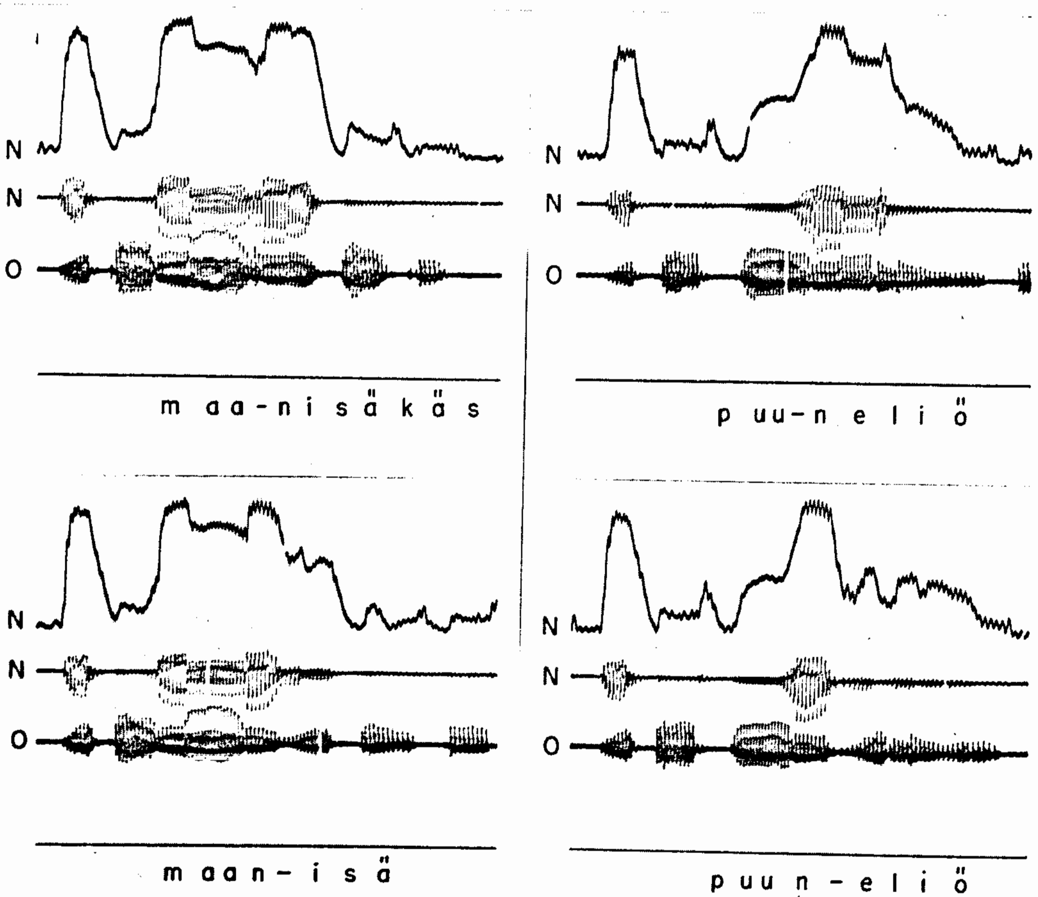


Fig. 3. Oscillograms and intensity curves of two pairs of Finnish test words uttered by K-K.W. Curves labeled N show the output from a nasal microphone, the oscillograms labeled O represent the output of an oral microphone.

oscillograms of each utterance shows the nasalization of the /i/ in *mitä* and the lack of nasalization of /ä/, regardless of whether the next word began with /m/ or with /p/. The /a/ in *maa* was nasalized in both cases, while /u/ in *puu* showed only a slight increase of the intensity of the nasal trace toward the end of its duration. The vowels /i/ in *nisäkäs* and /e/ in *neliö* were fully nasalized, whereas only a very slight degree of nasalization was present in *isä* and *eliö*. The oral oscillograms also suggest the presence of a period with reduced and irregular vocal fold activity before the onset of *isä* and *eliö*.

The boundary signals discussed thus far include differences in phonetic quality, segmental duration, vocal fold activity, and nasalization. All of these serve to establish a word boundary. The study of Finnish also revealed some patterns that serve to identify the units themselves between which those boundaries were observed. These patterns are predominantly of a suprasegmental nature.

Two contrastive degrees of quantity occur in Finnish. The duration of vowels is either phonemically short or phonemically long in every syllable of a word, while consonant quantity is contrastive in intervocalic position. In the speech of informants from Southwest Finland (a considerable number of the informants came from Turku), the realization of a phonemically short vowel in a non-first syllable depended on the nature of the preceding syllable. After a long initial syllable, the durations of short vowels fell within the duration range of stressed short vowels; after a short initial syllable, the duration of the phonemically short vowel was phonetically intermediate between the durations of stressed short and long vowels. The phonetic manifestation of the phonemic quantity of the vowel of the second syllable thus depends on the quantity of the first syllable, and the relationship between the durations of the successive syllables constitutes a unifying pattern that is part of the phonological structure of a word[7].

The duration of the vowels in a number of words of different types was investigated. Table I presents the average durations, in centiseconds, of the vowels occurring in words belonging to nine word types, contained in the test materials produced by K-K.W. The half-long vowel may be observed in the second syllable of every word whose first two syllables have the CV.CV pattern. The table reveals further that the same quantity ratio prevails, when the CV.CV sequence occurs as third and fourth syllable in a polysyllabic word. In essence, the Finnish words appear to be constructed of units larger than one syllable: all words consisting of more than three syllables seem to be built of dissyllabic or trisyllabic components, whose quantity patterns are similar to those of dissyllabic or trisyllabic words. A four-syllable word of the type ex-

[7] The occurrence of a half-long vowel in the syllable following an initial short syllable is a feature Southwestern Finnish shares with Estonian. In Estonian, however, the development of suprasegmental patterns characterizing words as phonological units appears to have proceeded farther than in Finnish: the quantity of the vowel of a non-first syllable is not independently variable, but depends entirely upon the quantity of the first syllable (the problem is discussed, and earlier literature cited, in *Segmental and syllabic quantity in Estonian*, Lit. 14, pp. 21–82). The relatively greater importance of the phonological manifestation of units rather than their boundaries is also reflected in the absence of laryngealization or a glottal stop as a boundary signal in Estonian.

Table I

Average durations, in centiseconds, of vowels in nine Finnish word types produced by K-K.W. In the formulae representing word types, C and V symbolize consonants and vowels; a period is used to indicate syllable boundaries. A representative word is given below each formula. The duration of the syllable-final consonant is included in the two syllables with a CVC structure. N = number of occurrences.

Word type	N	Duration of vowels in successive syllables 1	2	3	4	5
CV.CV *tuli*	44	8.2	12.4			
CVV.CV *tuuli*	64	19.7	7.4			
CVC.CV *vasta*	10	9.5+11.3	7.8			
CV.CVV *salaa*	25	8.9	27.8			
CVV.CVV *tienoo*	23	23.3	25.8			
CV.CV.CV *manala*	9	8.0	12.6	7.9		
CV.CV.CV.CV *manalana*	11	8.2	11.5	8.2	11.5	
CV.CV.CV.CV.CV *lakananani*	18	7.4	10.7	8.4	11.6	6.9
CV.CV.CV.CVC.CV *lakanana nsa*	10	7.8	10.6	8.4	9.9+11.8	7.0

emplified by *manalana* consists of two CV.CV units; words of the type *lakananani* consist of a CV.CV unit and a CV.CV.CV unit; and words of the type *lakanana nsa* consist of a CV.CV.CV unit and a CVC.CV unit (type *vasta*). In the last two word types, the duration of the consonant in the CVC syllable is included in the table. The quantity relationships between successive syllables serve not only to establish word patterns, but also to subdivide words into phonological components that are intermediate between a syllable and a word [8]. No immediate connection between these intermediate units and the morphological structure of the words could be established.

[8] Finnish words are traditionally stated to be stressed on the first syllable; in words of the CV.CV.CV.CV type, secondary stress is assumed to fall on the third syllable. Intensity curves of such Finnish words showed that the intensity of the second syllable of the first CV.CV unit was higher than that of the first syllable of the following CV.CV unit. All listeners, however, agreed that the second syllable was unstressed, while the third syllable (the first syllable of the second CV.CV unit) carried secondary stress. The fundamental frequency of nonfirst syllables in words of this type was usually more or less level. The unstressed second syllable was thus longer and had greater intensity than the stressed third syllable, while fundamental frequency failed to differentiate between the two syllables. The suprasegmental features associated with a single syllable evidently lose their individual significance when they become part of a higher-level pattern.

3. Some Boundary Signals in Czech

It was found in Finnish that the suprasegmental feature of quantity contributed both to identifying the segments immediately adjacent to a word boundary, and to the establishment of the phonological structure of the units themselves which were set off by the boundaries. The use of a modification in the phonatory pattern (laryngealization or a glottal stop) as a manifestation of a word boundary was quite general when the boundary occurred in a V+V sequence. In Finnish, vowel quantity is significant in any syllable of the word; the use of a glottal stop as a boundary marker may therefore be conditioned by the need to provide a point of reference for establishing the phonemic durations of the vowels in prejunctural and postjunctural position. As the Finnish example shows, the use of laryngealization (or other phonatory modification) need not imply the lack of a unifying pattern that could serve to identify phonological units. It appeared interesting to investigate a language with a somewhat similar segmental quantity structure to discover whether this type of boundary signal was indeed present with a comparable relative prominence as in Finnish, and to establish whether quantity contributes in a similar manner to the establishment of word patterns.

The phonological structure of Czech is similar to that of Finnish with respect to stress and quantity: words are stressed on the first syllable, and contrastively short and long vowels may occur in every syllable. Czech has two syllabic consonants, [l] and [r], that do not share in the long-short opposition. A brief study was designed to determine whether the boundary in a V+V sequence would be signalled with equal consistency in cases where the first element of the sequence (i.e. the last syllabic sound of the word preceding the juncture) is a syllabic consonant, whose duration is clearly noncontrastive and need not be signalled. A comparison of syllabic and nonsyllabic /l/ and /r/ in the same position might also contribute some information about syllabicity [9].

The Czech materials analyzed in this connection consist of a set of 32 phrases, in which syllabic and nonsyllabic /l/ and /r/ are followed, in identical consonantal environment, by the four vowels /a o u ū/ that may occur in word-initial position. The sequences also

[9] The Czech materials have been produced and analyzed in collaboration with *Ladislav Matejka*. More details will be presented in a separate publication.

contain a word boundary, whose position relative to /r/ and /l/ determines their syllabic or nonsyllabic nature. Four phrases may serve as examples. The phrase *Petr apoštol* contains the sequence symbolized as *tr + Vp; Petra poštvali* represents the sequence *trV + p*, *pět raportů* the sequence *t + rVp*, and *Petr raportuje* the sequence *tr + rVp*. Similar phrases were constructed for the other vowels with /r/, and for all four vowels with /l/. The 32 phrases were recorded twice by each of two informants. The tapes were processed acoustically using techniques described in footnotes 3 and 4. Figure 4 presents continuous intensity curves and oscillograms for the four quoted utterances produced by speaker L.M.

Table II shows the results of measurements for informant L.M. As a rule, each value given in the table represents the average of

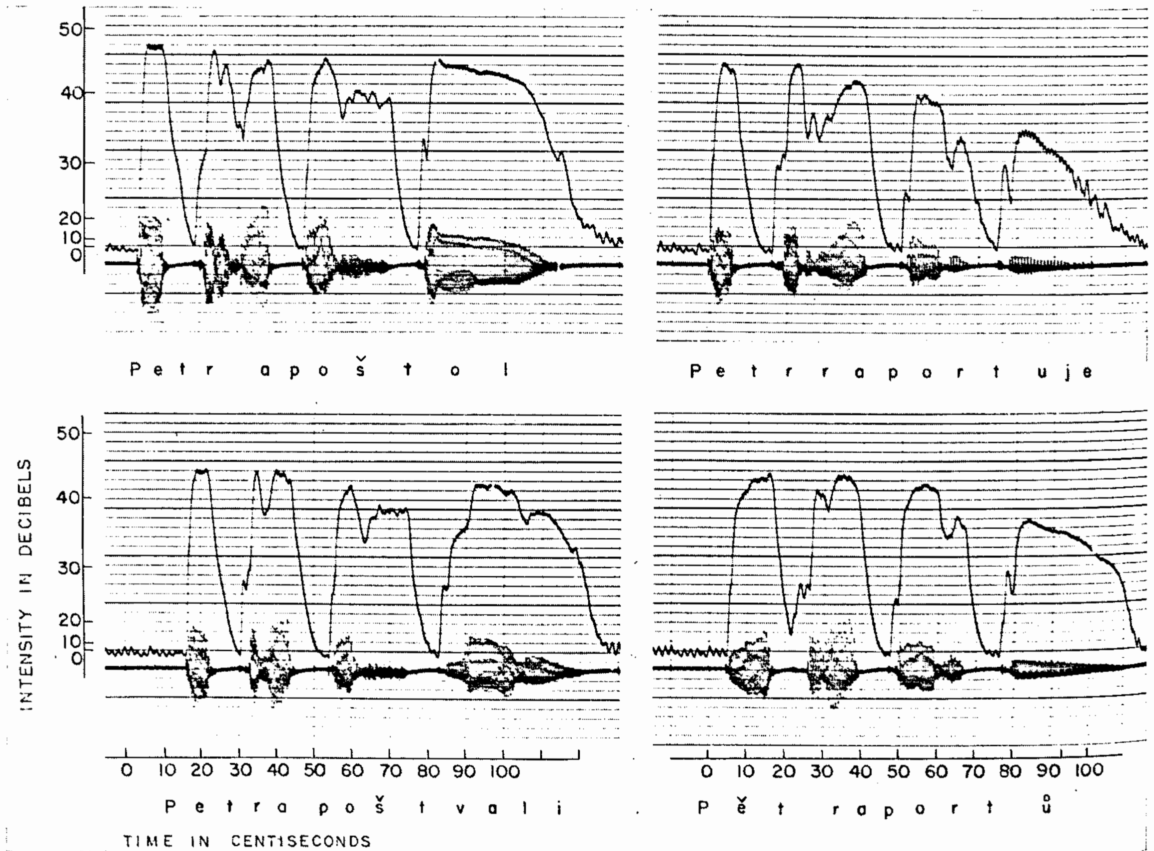


Fig. 4. Continuous intensity curves and oscillograms of four Czech utterances produced by speaker L.M.

Table II

Average formant frequencies, durations, and intensities of prejunctural and postjunctural sounds in a set of Czech utterances produced by speaker L.M. Formant frequencies are given in cycles per second, durations in centiseconds, and intensities in decibels relative to a constant reference level.

Feature under study	1 tl+Vp	2 tlV+p	3 t+lVp	4 tl+lVp	5 tr+Vp	6 trV+p	7 t+rVp	8 tr+rVp
1. Duration of /t/	5.1	9.1	8.1	8.1	7.5	8.4	7.4	8.8
2. Duration of release	4.3	3.1	3.3	3.3	3.1	3.1	4.3	3.6
3. Duration of syllabic /l/ and /r/	4.3+6.1			3.3+6.6	3.1+6.9			3.6+8.7
4. Intensity of syllabic /l/ and /r/	45.4			48.2	52.3			53.8
5. Frequency of F_1	415			375	495			490
6. Frequency of F_2	1265			1395	1370			1405
7. Frequency of F_3	2395			2350	2415			2365
8. Duration of boundary segment	4.9				4.6			
9. Intensity of boundary segment	35.8				34.3			
10. Duration of consonantal /l/ and /r/		3.1+3.7	3.3+5.1	4.8		3.1+4.8	4.3+5.7	6.9
11. Intensity of consonantal /l/ and /r/		42.3	44.8	46.8		49.7	48.2	41.5
12. Frequency of F_1		415	360	360		450	445	465
13. Frequency of F_2		1150	1200	1225		1310	1245	1215
14. Frequency of F_3		2570	2570	2375		2295	2195	2250
15. Duration of V	6.5	6.0	6.3	6.8	6.6	6.8	7.0	7.2
	15.0	10.0	7.5	10.0	14.5	9.0	20.0	15.0
16. Intensity of V	45.5	45.9	51.0	51.0	43.5	48.4	48.3	49.3
17. Duration of /p/	8.9	11.0	10.5	8.8	8.1	10.9	10.7	10.2

eight different occurrences. In the table, durations are given in centiseconds, intensities in decibels relative to an arbitrary, but constant reference level, and formant frequencies in cycles per second. In the case of /r/, the intensity measurements were made from the peak intensities of the vocalic portions between the individual tongue-tip closures; the formant frequencies were measured for the same segments from broad-band spectrograms. The symbol V stands for the vowels /a o u ū/; in row 15 the values for occurrences of /ū/ are given separately. In rows 3 and 10, the duration of the release of /t/ is repeated from row 2.

Although the duration of syllabic /l/ and /r/ in the sequences *tl + Vp* and *tr + Vp* is not contrastive, each occurrence of these sequences was found to contain a separate boundary segment (cf. row 8 of table II and figure 4). This boundary segment, like the segment found in Finnish utterances in analogous environments, evidently represents the acoustic correlate of a modification of the activity of the vocal folds. In traditional descriptions, the boundary segment is often referred to as a glottal stop. Since a glottal stop by definition implies absence of any airflow through the glottis, the intensity may be expected to drop to zero during the segment identified as a glottal stop. Only four such segments were observed in the 16 instances described here. In the other twelve cases, the boundary segment was realized as a period of irregular or breathy phonation or as a voiceless vocoid, acoustically manifested as a period of noise with energy concentrations at approximately the formant positions of the following vowel. The feature common to all these manifestations was a decrease in intensity of approximately 10 db from the level of a preceding syllabic /l/ and approximately 18 db in the case of /r/ (cf. rows 4 and 9 of the table).

The sequences *tl + Vp* and *tr + Vp* in which the boundary segment occurred contained a syllabic consonant followed by a vowel A syllabic consonant may occur either between two nonsyllabic consonants or before a word boundary. In the sequences referred to here, either the syllabicity of the prejunctural /r/ and /l/ or the presence of the word boundary must be indicated. If the syllabicity is manifested within the segments themselves, the presence of a word boundary may be deduced from the syllabicity of the prejunctural consonants; if the word boundary is phonetically manifested, the syllabicity of word-final /l/ and /r/ may be deduced from the presence of the juncture. However, in these sequences the boundary segment may also serve to provide a point of reference for the phonemically significant duration of the postjunctural vowel. This seems indeed to be the case, since a separate boundary segment was never found in the sequences *tl + lVp* and *tr + rVp*. The lack of a separate boundary segment in sequences of this type implies also that either the syllabicity of the word-final manifestations of /l/ and /r/ or the lack of syllabicity of word-initial /l/ and /r/ must be phonetically signalled by some other means. The presence or absence of syllabicity may be manifested during the segments themselves, or boundary signals of a different type may be present.

In actual manifestations of these sequences, it was in fact possible to establish the approximate point in time at which the first member of the sequence (i.e. the syllabic consonant) was followed by the second member of the sequence (the nonsyllabic consonant). In almost every instance, a change in the resonance patterns could be observed on broad-band spectrograms. The intensity curves showed a slight drop in intensity in the transition from syllabic /l/ to nonsyllabic /l/, and a rather considerable decrease in intensity in the case of /r/. Some characteristics of syllabicity were also associated with the segments themselves. In general, syllabic /l/ and /r/ were found to be longer than their nonsyllabic counterparts: the average duration of syllabic /l/ and /r/ was approximately 10.7 csec (including the release of the preceding consonant), that of nonsyllabic /l/ and /r/ – 7.5 csec. The intensity of syllabic /l/ and /r/ was also somewhat higher than the intensity of nonsyllabic /l/ and /r/ (cf. rows 4 and 11 in the table). The average F_2 positions for both syllabic /l/ and syllabic /r/ were approximately 150 cps higher than for nonsyllabic /l/ and /r/ (cf. rows 6 and 13 in the table).

The manifestation of the boundary was less obvious in those cases in which the word boundary was either preceded or followed by a single consonant (cf. columns 2, 3, 6, and 7 in the table). In the sequences *t + lVp* and *t + rVp* a longer release of the word-final /t/ might be expected than in those cases where /t/ is not followed by a word boundary. The average durations of the /t/ releases, given in row 2 of the table, show no significant differences between the different positions. Neither the duration of the release nor the presence or absence of a voiceless aspiration seemed to function as a boundary signal. A small, but rather regular difference was observed between the durations of initial consonants as compared to medial or final occurrences of the same consonants: initial consonants tended to be somewhat longer. Thus /l/ and /r/ were longer in initial position (cf. row 10, columns 3, 4, 7, and 8) than when they occurred medially (cf. row 10, columns 2 and 6), although they did not quite reach the duration of syllabic /l/ and /r/ (cf. row 3).

The intensities of the vowels preceding and following the boundary appeared to contribute but little to the establishment of the boundaries. The most remarkable feature was the relatively low intensity observed in word-initial vowels following the boundary segment (cf. row 16, columns 1 and 5 and the other occurrences). A possible reason for this low intensity is the modification of the phonatory activity associated with the boundary segment immediately preceding these vowels. The contribution of intensity toward identifying the stressed vowel (i.e. the vowel of the first syllable of a word) is not obvious from these data.

A comparison of V+V sequences for the contrastive presence of syllable boundaries and word boundaries is less fruitful in Czech than in Finnish, since compounds of words ending and beginning in a vowel are rare. Some comparisons were nevertheless carried through between words containing the diphthong /ou/ and sequences of /o/ + /u/ containing a word boundary. Fourteen such utterances were recorded twice by each informant, as well as two compound words containing the same /o/ + /u/ sequence. Figure 5 illustrates the materials and the manifestations of these sequences. The figure contains continuous intensity curves and oscillograms for productions of *mouka*, *kradmo ukazuje*, and *prvoúcta* by speaker L.M. The consonant environment was kept constant for pairs of the kind represented by the first two items.

The boundary segment that appears in the production of *kradmo ukazuje* was present in every such sequence. The average duration of this segment was 5.3 csec. In one out of 14 cases, the segment was manifested as a glottal stop. In the remaining instances the segment was characterized by breathy and irregular phonation, accompanied by a drop in intensity to an average level of 29.2 db from an average of 46.3 db for /o/, 41.9 for /u/, or approximately 15 db. The phonetic nature of the boundary segment observed in these sequences was in every respect similar to that observed in sequences involving syllabic consonants followed by words beginning with a vowel.

The peak intensities of the vowels in the /o/ + /u/ sequence could be established with less difficulty than the intensities of the two components of the diphthong /ou/, where the changes in in-

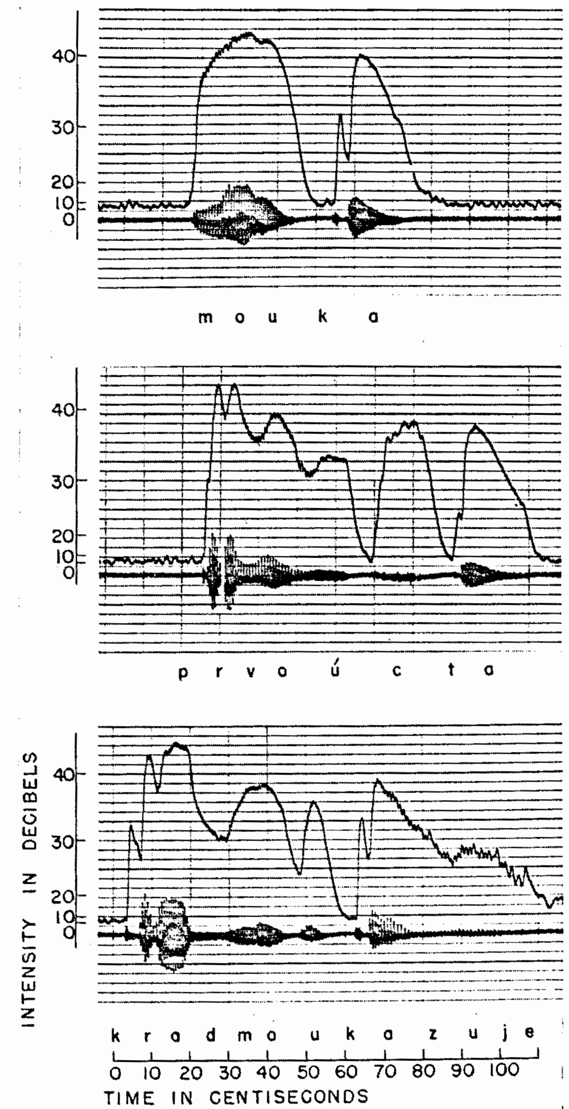


Fig. 5. Continuous intensity curves and oscillograms of three Czech utterances produced by speaker L.M.

tensity were more gradual. However, as may be seen from the reproduction of the intensity curve and the oscillogram of *mouka*, the boundary between the two components can be observed and their durations established with fair accuracy. The average duration of 14 productions of the diphthong /ou/ in such words as *mouka* was

19.5 csec, of which the duration of the first component constituted approximately 9 csec, that of the second 10.5 csec. The average duration of the total /o/ + /u/ sequence, including that of the boundary segment, was 18.2 csec; the duration of /o/ was 5.8 csec, that of the boundary segment 5.3 csec, and that of /u/ 7.1 csec. Since the second component of /ou/ was longer than the first by an approximately equal amount, the greater length of /u/ here does not contribute to the boundary signals.

The two compounds presented an intermediate pattern. The number of examples is clearly too small for any valid conclusions; it is nevertheless interesting to note that a breathy period, with a corresponding drop in intensity, was present in three out of four instances. However, the average decrease in intensity in compounds such as *prvoúcta* was only 7.3 db, or about half of that of sequences like *kradmo ukazuje*. The smooth intensity curve characteristic of the diphthong /ou/ was never observed in manifestations of compound words of this type.

The Czech materials thus have yielded evidence for the systematic use of a special segment as boundary marker. A word boundary between two syllabic sounds was manifested by a modification of the phonatory process even in cases where there was no need to indicate the phonemic length of the prejunctural sound. In Finnish, the presence of this type of boundary signal was accompanied by various qualitative and quantitative differences in the segmental sounds adjacent to the word boundary; quantity was also involved in establishing a unifying suprasegmental word pattern. The situation in Czech appears to be quite different.

In an attempt to determine whether vowel quantity and quality play any part in combining the syllables of a polysyllabic word into a higher-level phonological unit, 642 dissyllabic test words, embedded in frame utterances, were recorded by two informants and analyzed by techniques described in footnotes 3 and 4.

Table III presents the average formant positions and durations of the syllabic sounds in the first and second syllable of this set of words produced by one of the informants. As becomes apparent in studying the table, the short and long vowels differed among themselves in both positions; this is particularly evident in the case of /i/ – /ī/ and /u/ – /ū/. There was, however, no appreciable difference in the phonetic value of the vowels that could be caused by position in either the first or the second syllable.

Table III

Average formant positions and durations of syllabic sounds in the first and second syllable of 642 dissyllabic Czech words produced by L.M. Formant frequencies are given in cycles per second, durations in centiseconds. N = number of occurrences.

Syllable nucleus	First syllable N	F_1	F_2	F_3	D	Second syllable N	F_1	F_2	F_3	D
i	69	390	1835	2465	7.6	67	380	1800	2510	9.0
ī	72	270	2030	3085	15.3	71	270	1990	2970	15.0
e	59	490	1695	2425	7.6	55	475	1665	2360	9.4
ē	50	440	1895	2525	17.2	54	505	1800	2420	15.5
a	51	640	1235	2350	8.7	58	640	1290	2365	9.4
ā	54	725	1325	2405	19.6	58	700	1300	2425	19.4
o	56	465	895	2265	8.9	51	480	985	2235	8.7
ō	42	410	725	2335	19.4	34	440	810	2325	20.7
u	46	390	940	2160	7.1	51	385	900	2045	8.3
ū	42	300	710	2195	13.2	41	340	710	2125	15.4
ou	51	420–355	905–670	2095–2200	6.5–8.2	58	465–355	955–680	2150–2110	9.1–9.6
					14.7					18.7
l	20	485	1260	2300	9.9	24	440	1180	2200	10.1
r	30	540	1260	2265	8.6	20	485	1350	2285	10.2

Vowel quality thus seems to play no part in establishing the first syllable of a word. The same may be said about quantity: no systematic differences could be observed in the durations of short and long vowels occurring in the two syllables of a dissyllabic word[10].

Czech thus appears to be a language in which the boundaries are indicated primarily by modifications of the phonatory pattern. Neither segmental nor suprasegmental features emerged which could be identified with certainty as contributing toward the establishment of phonological units, and the boundary signals were primarily of a segmental nature.

[10] Czech words are traditionally assumed to be stressed on the first syllable. The data presented here show that neither vowel quantity nor vowel quality contribute appreciably toward the identification of a stressed syllable. It may therefore be assumed that intensity and/or fundamental frequency carry greater significance as acoustic correlates of stress in Czech than in some other language. The limited intensity data included in table II are inconclusive. The analysis of the intensity and fundamental frequency patterns of a larger set of test items is in progress; it remains to be seen whether these features serve only to characterize the stressed syllable, or play some part in a larger pattern.

4. *Word Patterns in Serbocroatian*

The situation appears considerably different in Serbocroatian. Here the suprasegmental features of intensity, duration, and fundamental frequency combine with a qualitative difference between accented and unaccented vowels to produce phonological patterns which characterize the different word types[11]. The units themselves being clearly phonologically determined, segmental boundary signals have less significance. The possibility exists, however, that modifications of these suprasegmental word patterns may be used by the speakers to indicate the presence of certain morphological and lexical boundaries. A brief consideration of the problem of proclitics may yield some relevant information.

The domain of an accentual pattern in Serbocroatian is, as a rule, a word; however, certain proclitics may form an accentual unit with a following word, which in turn loses its separate accent and becomes part of the larger accentual unit. There are other sequences in which a proclitic, followed by a word with a falling accent on the first syllable, may lack any accent of its own. The phrases *u rȁt* and *u grâd* differ from *ù rat* and *ȕ grād* with respect to the domain of the accentual patterns. In the latter sequences, the accentual pattern embraces the preposition as well as the noun; the sequences *u rȁt* and *u grâd*, consisting of the same segmental phonemes, contain an accentually indeterminate proclitic and a word with a monosyllabic accentual pattern.

The same applies, mutatis mutandis, to words and phrases such as *da bȉje*[11a] – *nȁbije* and *da mȍli* – *zàmoli*. The accentual patterns of monosyllabic, dissyllabic and trisyllabic words having been established, productions of phrases containing unaccented proclitics may be compared with otherwise similar sequences serving as the domain of an accentual word pattern. Any differences between the suprasegmental patterns of the respective pairs may be assumed to signal the presence of a word boundary within the sequences containing an unaccented proclitic.

[11] The Serbocroatian data presented here are drawn from a forthcoming joint publication with *Pavle Ivić*. The patterns referred to in the introductory remarks are described in detail in *Accent in Serbocroatian* (Lit. 18), which also contains a selected bibliography on pp. 136–142, and in *Pavle Ivić* and *Ilse Lehiste* (Lit. 9).

[11a] The symbol [ȉ] is used to indicate the vowel [i] pronounced with a short falling accent.

The eight dissyllabic and trisyllabic utterances quoted above were produced by twelve speakers in the course of a more extensive recording session, embedded in a frame utterance and randomly inserted in a larger set of test sentences[12]. Broad-band and narrow-band spectrograms, intensity curves, and oscillograms were produced from the recorded tapes, and the formant positions of the vowels, the duration of all vocalic segments, the fundamental frequencies at the onset, peak, and termination of each vowel, and the peak intensities of all syllabic sounds were measured. While space does not permit the presentation of the results in detail, certain relevant observations will be pointed out.

The difference in the phonetic quality of accented and unaccented /a/ furnished an important clue for the identification of the stressed syllable in such contrastive pairs as *u rȁt* vs. *ù rat* and *da bȉje* vs. *nȁbije*. For example, the average positions of the first three formants of /ȁ/ in *u rȁt*, produced by seven female speakers, were 910 – 1715 – 2800 cps; in *ù rat*, the formant positions of posttonic /a/ were 730 – 1780 – 2625 cps. In *nȁbije*, the average formant positions of /ȁ/ were 915 – 1720 – 2825 cps, whereas the values for pretonic /a/ in *da bȉje* were 665 – 1825 – 2650 cps. The comparable averages for five male speakers were as follows: /ȁ/ in *u rȁt* 660 – 1420 – 2400 cps, /a/ in *ù rat* 590 – 1420 – 2515 cps; /ȁ/ in *nȁbije* 720 – 1415 – 2565 cps, /a/ in *da bȉje* 595 – 1460 – 2570 cps. Accented /a/ thus always had a higher first formant value than unaccented /a/, whereas unaccented /a/ showed a certain amount of centralization. In this limited set of data, very little difference could be observed between pretonic and posttonic /a/, although both were clearly different from an accented /a/[13].

Table IV presents the fundamental frequency, intensity, and duration data, arranged according to the average fundamental frequency ranges of the informants into low, medium, and high-pitched

[12] The informants and their dialectal background are described in detail in *Accent in Serbocroatian* (Lit. 18), pp. 31–38.

[13] The average positions of the first three formants of short /a/ in pretonic position (in *da mȍli* and *da bȉje*) were 625 – 1430 – 2550 cps for the men and 690 – 1760 – 2800 cps for the women. In posttonic position (in *ù rat* and in three productions of *ȕ grăd* with short /a/) the averages were 630 – 1450 – 2445 cps for the men and 765 – 1765 – 2625 cps for the women. In accented position (in *u rȁt*, *nȁbije*, and *zàmoli*) the positions were, respectively, 710 – 1400 – 2490 cps for the men and 915 – 1710 – 2765 cps for the women. These values fall within the allophonic ranges for stressed and posttonic /a/, established for these speakers during a previous stage of the study and reported in *Accent in Serbocroatian* (Lit. 18), pp. 95–127.

Table IV

Fundamental frequency, intensity, and duration of syllable nuclei occurring in eight Serbocroatian test items uttered by twelve informants, averaged separately for speakers with low, medium, and high pitch. Fundamental frequencies are given in cycles per second, durations in centiseconds, and intensity in decibels relative to a constant reference level.

Test item and group of speakers	First vowel Fund. freq. Beg.	Peak	End	Dur.	Int.	Second vowel Fund. freq. Beg.	Peak	End	Dur.	Int.	Third vowel Fund. freq. Beg.	Peak	End	Dur.	Int.
Low															
u rȁt	106	107	100	11.8	39.5	108	122	113	12.6	42.7					
ù rat	119	128	125	9.8	42.5	131	132	111	9.4	42.5					
u grâd	104	107	96	10.2	41.3	113	125	92	20.0	40.3					
ȕ grād	126	132	117	9.3	43.7	114	114	97	12.3	37.8					
ȕ grăd						80	80	70	9.5	29.5					
Medium															
u rȁt	211	221	214	10.7	40.7	181	217	213	16.3	42.0					
ù rat	202	215	215	11.8	42.7	218	225	225	11.0	43.7					
u grâd	197	207	205	10.8	42.5	225	225	154	24.5	33.0					
ȕ grād	205	234	210	11.8	44.7	171	171	158	13.3	36.0					
High															
u rȁt	260	260	254	8.5	36.0	247	321	319	18.8	46.0					
ù rat	282	318	313	11.7	38.7	308	342	320	11.2	43.0					
u grâd	260	263	246	8.5	38.3	257	292	197	22.7	42.3					
ȕ grād	302	354	305	9.7	41.3	204	204	193	13.3	37.5					
ȕ grăd						200	200	178	10.0	33.0					
Low															
da mȍli	100	102	99	8.0	40.7	109	119	105	11.1	42.3	89	90	83	7.3	36.2
											(1 laryng.)				
zàmoli	110	111	108	10.1	41.0	115	119	103	7.6	41.7	102	102	91	8.1	35.8
da bȉje	104	106	102	7.5	41.5	119	128	115	10.0	42.8	100	100	84	9.2	39.0
											(1 laryng.)				
nȁbije	108	120	114	9.8	41.5	102	102	94	6.8	37.0	93	94	89	6.8	36.7
Medium															
da mȍli	199	199	195	8.2	43.7	218	232	205	15.3	42.0	159	165	165	6.8	36.7
zàmoli	179	181	177	12.5	43.7	194	218	213	7.3	43.7	211	220	211	8.8	38.0
da bȉje	154	154	144	8.5	43.5	156	165	160	9.8	47.0	151	151	138	6.8	41.0
nȁbije	205	210	200	11.0	43.0	161	161	156	6.3	36.0	152	157	155	6.3	35.5
High															
da mȍli	265	267	262	7.8	42.7	293	327	274	14.2	43.0	207	207	201	6.8	32.7
zàmoli	258	269	267	13.5	40.7	294	321	276	8.8	37.3	218	218	187	7.0	34.0
da bȉje	243	254	238	8.5	42.3	297	315	293	9.7	38.3	202	202	190	6.8	37.0
nȁbije	275	309	291	12.7	42.0	209	209	192	6.7	30.0	191	192	190	5.0	35.3

groups. The contribution of intensity toward the identification of the stressed syllable will be considered first.

For informants in whose speech features of intensity are patterned with the same regularity as features of fundamental frequency[14], the intensity relationships between the two syllables of a dissyllabic word with a rising accent on the first syllable differ in a predictable manner from those of comparable words with a falling accent: all other factors being kept constant, the two syllables of a word with rising accent have approximately equal intensity, whereas in a word with a falling accent the second syllable is considerably lower in intensity. Such a step-down decrease in intensity constitutes a characteristic of a word with a falling accent, and may serve as a signal for the presence of such a word. This step-down pattern was indeed present in the phrases *da mȍli* and *da bȉje*, where the stressed syllable, with an average intensity of 42.6 db, was approximately 5.5 db higher in intensity than the posttonic syllable, whose average intensity was 37.1 db. In these two phrases, the intensity of the proclitic was comparable to that of the stressed syllable. The intensity pattern of *nȁbije* showed a similar decrease from the first to the second syllable, while the third syllable remained at approximately the same intensity level as the second. In *zàmoli*, however, the first two syllables had approximately equal intensity, while a drop of 5 db (from an average of 40.9 db to 35.9 db) took place between the second and third syllable. Intensity patterns alone are thus not sufficient for distinguishing between sequences with proclitics and utterances with a rising accent on the first syllable.

In the pairs involving accented and unaccented /u/, an increase in intensity was found to be associated with the placement of accent, amounting to an average of 2.6 db. The change in the intensity of the syllable from which the accent was shifted to the proclitic depended on the nature of the accent. In cases of rising accent, the intensity of the originally stressed vowel decreased by an average of 0.5 db (remaining, for all practical purposes, unchanged); in cases of falling accent the decrease was considerably greater. Details are presented in table IV.

[14] It was found during the earlier study (Lit. 18 and 9) that intensity features were not unambiguously present in the speech of all informants. For the sake of comparability with the earlier materials, the data presented in table IV are organized according to the same pattern that was used in the earlier study. The grouping of speakers according to their average fundamental frequency ranges rather than according to the relative significance of the intensity features makes the fundamental frequency patterns clearer, but obscures partially the intensity patterns. In the speech of those informants who are consistent in their use of intensity features as regular concomitants of the fundamental frequency patterns, the intensity relationships are much more clearly defined.

Figure 6 shows intensity curves and oscillograms of the four utterances *u ràt*, *ù rat*, *u grâd*, and *ȕ grād*, produced by speaker D 1. The words are preceded by the first word of the frame in which the test items were commuted (*Forma . . . data je kao primer*). The different effect of the accent shift to the preposition, depending on whether the shifted accent is falling or rising, is clearly evident from the intensity curves. The figure also illustrates the occasional use of a period of laryngealization between the first word of the frame and the test word. However, the use of this laryngealized period did not serve to differentiate between the two types of sequences: the period

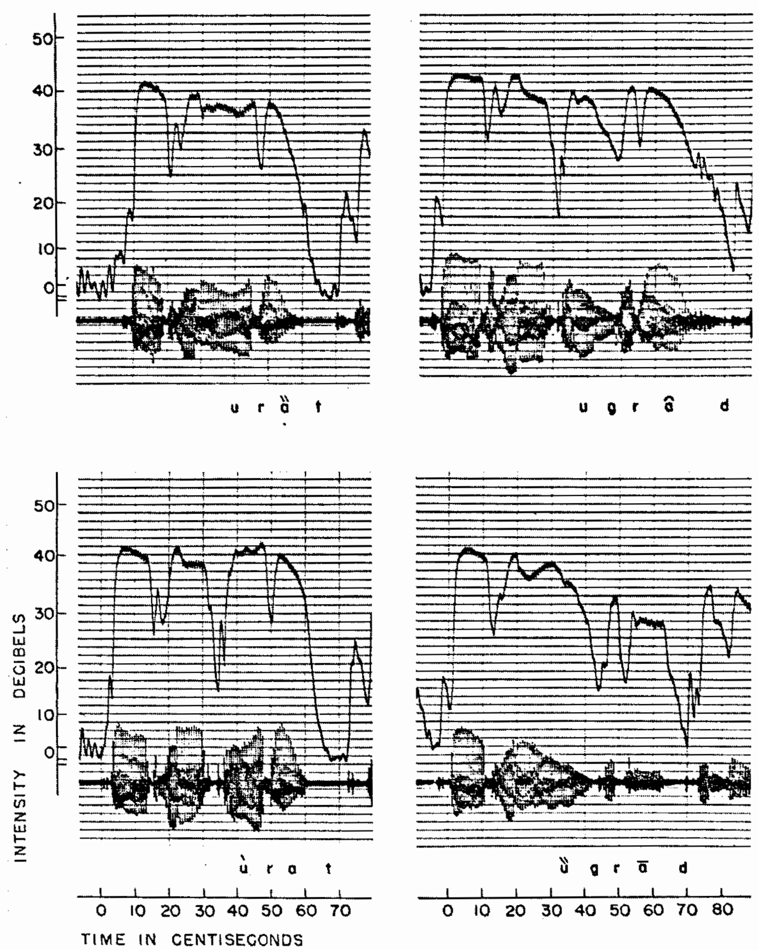


Fig. 6. Intensity curves and oscillograms of four Serbocroatian test items produced by speaker D 1.

of laryngealization was used by all speakers with equal frequency before the unstressed proclitic in *u ràt* and *u grâd* as before the stressed preposition in the sequences *ù rat* and *ȕ grād*, and there were instances where laryngealization was absent before either type of sequence. In the utterances shown on figure 6, laryngealization occurred before *u grâd* and *ù rat*, and was absent in *ȕ grād* and *u ràt*. The presence of the period of laryngealization also precludes the possibility of assuming that the unstressed proclitic forms an accentual unit with the preceding word.

The contribution of duration toward the identification of the stressed syllable appears relatively greater than that of intensity. During the previous investigation, the ratio between stressed and posttonic short vowels was found to be approximately 3:2, regardless of accent type. This result was confirmed by the present materials, in which the average duration of stressed short vowels was 12.3 csec, that of posttonic short vowels 8.2 csec. As was noted above, the intensity drop between the second and third syllable of *zàmoli* was comparable to that in *da mȍli*, and intensity failed to distinguish between the two sequences. The relatively greater duration of the first syllable in *zàmoli*, as compared to that of the proclitic in *da mȍli*, serves as an unambiguous cue for identifying this syllable as the bearer of accent.

The fundamental frequency patterns of polysyllabic words with falling accents involve a high frequency on the stressed syllable, followed by a posttonic syllable with relatively low frequency. In words with rising accents, the syllable following the stressed syllable has either the same or even a slightly higher fundamental frequency. As may be seen from the data summarized in table IV, the fundamental frequency pattern characterizing the short falling accent in *da bȉje* and *da mȍli* is comparable to the pattern occurring on the first two syllables of *nȁbije*. A comparison of the fundamental frequency values for *zàmoli* with those of *da mȍli* and *nȁbije* is instructive. The fundamental frequency movement of *zàmoli* resembles that of *da mȍli* much more than that of *nȁbije*. Nevertheless, there are some differences that make it possible to distinguish between all three patterns. *Zàmoli* differs from *nȁbije* mainly with regard to the fundamental frequency pattern, and from *da mȍli* mostly with respect to duration. In *da mȍli* the stressed syllable is longer than the pretonic syllable; in *zàmoli* the first syllable is longer than the second. *Nȁbije* shares this feature with *zàmoli;* the duration pattern of *da bȉje* resembles that of *da mȍli*. In addition, the fundamental frequency of the first syllable of *zàmoli* appears somewhat higher than that of the pretonic syllable of *da mȍli*.

The differences between the phrases containing unstressed proclitics and similar sequences serving as the domain of accentual word patterns thus can be described in terms of modifications of the suprasegmental features of fundamental frequency, duration, and intensity. The qualitative difference between accented and unaccented vowels plays a part in establishing the word patterns. No unambiguous boundary signals were provided by modifications in the phonatory pattern.

5. *Summary and Outlook*

Certain types of boundary signals have been identified in the course of this investigation. These include modifications of the phonatory pattern (laryngealization, breathy phonation, insertion of a glottal stop); modifications of nasalization; articulatory modifications; and modifications of suprasegmental patterns of fundamental frequency, duration, and intensity. Considerable differences in the use of these boundary signals exist between languages; no one single feature could be found which would be common to all manifestations of a word boundary. The manner in which boundaries are realized in a language constitutes an integral part of its structure, and has to be included in its phonological description.

A first approximation may nevertheless be attempted in the classification of languages according to their use of boundary signals. There appear to exist two general types: languages in which boundary signals are primarily of a segmental nature, and languages with well-developed suprasegmental patterns characterizing units of the phonological hierarchy. In languages of the latter type, the presence of these phonologically definable units implies the presence of junctures in sequences of segmental phonemes, which need not be signalled by separate boundary segments. Elements of both types may be present in a language. In the case of Czech, the contribution of suprasegmental features toward establishing word patterns was small[15]; the boundary signals were predominantly segmental in

[15] The acoustic correlates of stress in Czech have not yet been exhaustively investigated. The data presented in the course of this paper make it clear that with respect to the effect of stress on vowel quality and the manifestation of phonemic quantity, Czech differs basically from both Finnish and Serbocroatian.

nature. In the case of Serbocroatian, word patterns were established mainly on the basis of suprasegmental features; the presence of words could be deduced from the presence of these patterns, which in turn served to imply the presence of boundaries. In Finnish, segmental and suprasegmental features were combined in a system in which boundaries were indicated by predominantly segmental features, but word patterns were established by suprasegmental features. The Finnish materials also contributed some evidence for the existence of phonologically definable building blocks of speech occupying an intermediate level between syllables and words in a hierarchy of phonological structures.

The present study has been devoted to boundaries between units not larger than a phonological word. The investigated materials have also brought forth evidence (not reported here) that certain of the boundary signals may be further modified or superseded, when the phonological words themselves become part of larger phonological units. In those instances, the absence of boundary signals may become a higher-level signal. Assimilations were found to take place in Czech and Finnish across word boundaries; neutralizations of contrasts between tonal movements were observed in Serbocroatian utterances in positions removed from primary sentence stress. It is hoped that this study of the phonological structure of word-level units and their boundaries may serve as a basis for future studies of units and boundaries at higher levels in the phonological and grammatical hierarchies.

References

1. *Bolinger, D. L.* and *Gerstman, L. J.:* Disjuncture as a cue to constructs. Word *13:* 246–255 (1957).
2. *Chomsky, N.; Halle, M.* and *Lukoff, F.:* On accent and juncture in English. For Roman Jakobson, pp. 65–80 (The Hague 1956).
3. *Dietrich, G.:* Das Wort als phonetische Einheit. Le Maître phonét. *38:* 31–33 (1932).
4. *Feltes, J.:* The word as a phonetic entity in Luxemburg dialect. Le Maître phonét. *38:* 28–31 (1932).
5. *Hakulinen, L.:* The Structure and Development of the Finnish Language (translated by *J. Atkinson*), (Bloomington, Ind., 1961).
6. *Hála, B.:* The Slavonic languages, in: *L. Kaiser's* Manual of Phonetics, pp. 302–311 (Amsterdam 1957).
7. *Halle, M.:* The Sound Pattern of Russian ('s-Gravenhage 1959).
8. *Hill, A. A.:* Juncture and syllable division in Latin. Language *30:* 439–447 (1954).
9. *Ivić, P.* and *Lehiste, I.:* Prilozi ispitivanju fonetske i fonološke prirode akcenata u savremenom srpskohrvatskom književnom jeziku. Zbornik za filologiju i lingvistiku *6:* 33–73 (1963).

10. *Joki, A. J.:* Die phonetische und lautgeschichtliche Erforschung der finnisch-ugrischen Sprachen in den Jahren 1941–1955 (I. Teil). Phonetica *4:* 151–160 (1959).
11. *Jones, D.:* The 'word' as a phonetic entity. Le Maître phonét. *36:* 60–65 (1931).
12. *Jones, D.:* The hyphen as a phonetic sign. Z. f. Phonetik u. allg. Sprachw. *9:* 99–107 (1956).
13. *Kučera, H.:* The Phonology of Czech (Mouton, 's-Gravenhage 1961)
14. *Lehiste, I.:* Segmental and syllabic quantity in Estonian. Amer. Stud. Uralic Linguist., pp. 21–82 (Bloomington 1960).
15. *Lehiste, I.:* An acoustic-phonetic study of internal open juncture. Suppl. to Phonetica *5* (1960).
16. *Lehiste, I.:* Acoustic studies of boundary signals. Proc. IVth int. Congr. phonet. Sci. Helsinki, 1961, pp. 178–187 (The Hague 1962).
17. *Lehiste, I.:* Compounding as a phonological process. Proc. IXth Congr. Linguists, Cambridge 1962, pp. 331–336 (The Hague 1964).
18. *Lehiste, I.* and *Ivić, P.:* Accent in Serbocroatian: An experimental study (Michigan Slavic Materials *4*, Ann Arbor 1963).
19. *Moulton, W. G.:* Juncture in modern standard German. Language *23:* 212–226 (1947).
20. *Panov, M. V.:* O razgraničitel'nyx signalax v jazyke. Voprosy jazykoznanija *1:* 3–19 (1961).
21. *Pike, K. L.:* Grammatical prerequisites to phonemic analysis. Word *3:* 155–172 (1947).
22. *Pike, K. L.:* Language in relation to a unified theory of the structure of human behavior. Parts I, II, and III (Glendale, Calif., 1954, 1955, and 1960).
23. *Sadeniemi, M.:* Metriikkamme perusteet (Helsinki 1949).
24. *Sovijärvi, A.:* The Finno-Ugrian languages, in *L. Kaiser's* Manual of Phonetics, pp. 312–324 (Amsterdam 1957).
25. *Stockwell, R. P.: Bowen, J. D.* and *Silva-Fuenzalida, I.:* Spanish juncture and intonation. Language *32:* 641–665 (1956).
26. *Trager, G. L.:* Some thoughts on 'juncture'. Stud. Linguistics *16:* 11–22 (1962).
27. *Trim, J. L. M.:* The identification of phonological units. Proc. IVth int. Congr. phonet. Sci., Helsinki, 1961, pp. 773–778 (The Hague 1962).
28. *Trubetzkoy, N. S.:* Die phonologischen Grenzsignale. Proc. IInd int. Congr. phonet. Sci., pp. 45–49 (Cambridge 1936).
29. *Trubetzkoy, N. S.:* Grundzüge der Phonologie (Travaux du cercle linguistique de Prague *7*, Prague 1939).

Author's address: Professor Dr. Ilse Lehiste, The Ohio State University, Division of Linguistics, 216 North Oval Drive, *Columbus 10/Ohio* (USA).

Discussion

Sovijärvi (Helsinki): Fräulein *Lehiste* hat u. a. den verschiedenen Grad der Nasalisation bei ihren Untersuchungen über die finnischen Juncture-Erscheinungen als einen distinktiven Faktor betrachtet. Ich möchte sie fragen, wieviel Vpn. sie in ihren Nasenkurvenversuchen gehabt hat; weil es meines Erachtens nicht genügen würde, nur 1–2 Vpn. für die Nasalisationsuntersuchungen zu verwenden. Es ist ja bekannt, daß die individuelle und regionale Nasalisation der Vokalartikulationen sehr verschieden sein kann. Meines Wissens hat die Vortragende nur einen Mann und seine Frau aus Turku bei ihren Nasenkurvenversuchen zur Verfügung gehabt. Ich habe nur bei den gewöhnlichen Tonbandaufnahmen als Vp. fungiert, aber ich habe nicht an den Nasenkurvenversuchen teilgenommen.

Gårding (Lund): I have studied internal open juncture in Swedish. My material consists of minimal pairs uttered at various rates of speech. So far my observations are obtained from one informant only.

The most consistent differentiating factor between the pairs seems to be intensity.

Duration does not behave in the same way as it does in Miss *Lehiste's* material. The final prejunctural allophone is most often considerably longer than the postjunctural initial allophone.

Romportl (Praha): Ich hätte nur zwei kleine Bemerkungen:

1. Ich bedaure, daß von den im gedruckten Referat von Fräulein *Lehiste* behandelten Problembereichen aus dem Tschechischen in der vorgetragenen gekürzten Fassung nicht das Beispiel der «Juncture» in der Nachbarschaft von Liquiden gewählt worden ist. Es könnte interessantere Tatsachen zeigen als die gewählte Frage der Gruppen von zwei Vokalen.

2. Als Grenzsignal ist «coup de glotte» im Tschechischen potenziell. Wie häufig er vorkommt, hängt – auch in schriftsprachlichen Äußerungen – von der lokalen Herkunft des Sprechers, von dem Stil der Äußerung usw. ab. Es wäre nützlich, diese Tatsache bei den weiteren Untersuchungen unter Anwendung eines reicheren Materials zu berücksichtigen.

Pulgram (Ann Arbor): Whether internal boundaries are marked or not depends on the phonological condition of the language. Miss *Lehiste* chose for her argument, as indeed she noted, languages which by some means (especially stress) mark word boundaries (whether stress is distinctive or, as in Czech, merely a boundary marker, is irrelevant). In amplification, though not in correction, of Miss *Lehiste's* remarks it should be said that some languages, notably French, while giving boundary markers for words in isolation (French stresses the last syllable), eliminate all such markers of the lexical unit in utterances larger than a single lexical unit, with the result that word boundaries are eliminated since the longer utterance is now the phonological word *aux Etats-Unis* is /ozetazyni/, and the further result that all boundary markers ("junctures") within such a phrase become impossible. It would be useful to see in this distinction between languages which do and languages which do not allow of boundary markers in an utterance longer than the lexical word, a typological distinction of great importance. This peculiar feature has also its ramification in the syllabation, for in languages that obliterate word boundaries syllabation takes places in complete disregard for word boundaries, for example /o-ze-ta-zy-ni/. In this context it should also be noted that the division of Russian *eta kniga* by English students who do not know the morphological seams in the phrase, into *etak niga* (an example cited by Miss *Lehiste*) is exactly what is to be expected from speakers if a language which marks word boundaries, but which does not allow of a post-pausal, hence also not of a syllable-initial /kn-/.

Jassem (Poznań): 1. It is common knowledge, and Miss *Lehiste* has more knowledge of this than anyone else, after she has studied the phenomenon for several years, that the occurrence of juncture phenomena, just like the occurrence of other phonetic signals, in actual speech depends heavily on such factors as style, tempo, length of the utterance under investigation, etc. Have you, Miss *Lehiste*, a programme directed towards finding differences in the occurrence of junctural signals in various types of speech?

2. An informal experiment has shown to me that a large pannel of naive listeners may almost unanimously mark juncture on their answer sheets, whilst an analysis of the spoken text which they have been asked to mark appropriably with juncture signs, has not revealed any acoustic juncture signals. The junctures are often just "in the listeners' heads".

Kiparsky (Helsinki): Is there a "phonological boundary" in Czech *má úcta* ('my regards') or *je den* 'there is a day' or in similar words? The "phonological boundary" in Russian should be investigated.

Vachek (Praha): The important and convincing arguments should be complemented in two small points. First; the title of the paper should rather have been "Boundary signals" than "Juncture": the idea of juncture, involved by the American descriptivists, arose from the obstinate intention to disregard meaning in language. Second, the instances quoted from Czech look somewhat artificial or bookish; there is no doubt that more suitable specimens of Czech would have demonstrated the speaker's idea just as (or rather more) convincingly.

Rudnyckyj (Winnipeg): Belorussian and Ukrainian with their sandhi-phenomena can contribute much to the problem; especially in diphthongisation of boundaries of lexical units.

Daneš (Praha): I think that it is not very important whether some acoustic signal (clue) is present in each case, is implemented in every particular speech-act. But what is relevant is the fact that in one class of instances (e.g. in Czech compounds *nedouk, poukaz*) the glottal stop m a y be implemented, while in an other class (e.g. *mouka, louka*) it may n o t. It is just this possibility versus the impossibility of implementation of such a signal that constitutes the phonological opposition. It may be said that such a p o t e n t i o n a l i t y of language phenomena (to use *Mathesius*' expression) belongs to the set of general characteristic features of human language.

Answer *Lehiste:* I am grateful to the commentators for pointing out various interesting problems connected with juncture that should be studied in more detail in languages treated in the paper as well as in other languages (comments by *Kiparsky, Rudnyckyj, Sovijärvi,* and others). I agree that the various phonetic factors contributing to the identification of the presence of boundaries may be realized differently in languages other than those described in my report. The manifestation of boundaries constitutes a part of the phonological structure of every language, and languages may differ in this respect as in other aspects of their structure. For example, quantity may be expected to function differently as a boundary signal in languages without phonemic quantity on a segmental level on the one hand and in languages with varying types of phonemically significant quantity on the other (comment by *Gårding*).

It is one of the points of the paper that there exists a hierarchy of phonological units, whose boundaries may be manifested in various ways, or which may merge with other units of the same level to form higher-level units. The lower-level units retain the property that their boundaries may, under certain conditions, be signalled by phonological boundary markers (comments by *Romportl* and *Daneš*). The phonological units, whose boundaries are in fact manifested, may be coterminous with morphological, lexical, or syntactic units, but need not be so at every level, in every case, or in every language (cf. comment by *Pulgram*).

The term *juncture* was redefined in the paper to apply to phonologically manifested boundaries (cf. comment by *Vachek*). There are several theoretical consequences of this redefinition which could not be treated in detail either in the paper or in this brief reply. For example, the application of this definition precludes the possibility of re-labeling morphological boundaries as zero allophones of juncture phonemes.

Replying to comments by Professors *Sovijärvi* and *Jassem*, I would like to add that we are currently engaged in a considerably more detailed study of various sentence types in Serbocroatian as well as Finnish.

Papers of Section Sessions

Proc. 5th int. Congr. phon. Sci., Münster 1964, pp. 201–206
(S. Karger, Basel/New York 1965).

The Relation of the Phoneme to Other Phonological Elements

By D. J. ALLERTON, Manchester

It is generally felt that the phonology[1] of a language may be analyzed in terms of phonemes, i.e. distinctive sound segments, on the one hand, and non-segmental or long sound elements, or prosodies, on the other. There is a tendency to regard the segmental phonemes, characterized by features of what Professor *Daniel Jones* would call *tamber*[2], as having a mainly distinctive or lexical function: they serve to distinguish words or morphemes. The non-segmental prosodies (normally length, stress, and pitch features) for their part have the principal function of marking off morphemic boundaries or types, or characterizing certain utterance types, such as Yes-no questions. This view seems in some ways to represent an oversimplification.

One pointer to the inadequacy of this simple dichotomy is the fact that some segmental elements seem to have no distinctive function but only demarcative value[3]; an example is the glottal stop in German, which, although a segment, is not distinctive, but does act as a marker of morpheme-initial position[4]. On the other hand some non-segmental elements do have distinctive value, e.g. Swedish tone patterns, Russian stress patterns; in such cases lexical meanings may be distinguished through prosodic features. Here we may justifiably speak of prosodemes.

A very important question is the length or scope of a prosodeme of prosodic feature. Some linguists invariably cut down their pro-

[1] In the narrower sense; also called by *Trubetzkoy* Darstellungsphonologie, i.e. excluding Lautstilistik, cf. Lit. 5, p. 17–29.

[2] Cf. Lit. 2, p. 108f.

[3] Cf. Lit. 5, p. 29–30; *demarcative*, as used here, includes *Trubetzkoy's* delimitative and culminative.

[4] Cf. *J. R. Firth*, Lit. 1, p. 134.

sodic elements as far as possible, i.e. to cover the shortest possible segments. What criteria should influence our decision to select a certain length of component? Our treatment of segmental phonemes may give us an insight. We class Spanish [tʃ] as a single phoneme, for example, because [ʃ] does not occur in Castilian Spanish anywhere except after [t], so that, as *Martinet* says[5], [tʃ] represents one choice and not two successive choices. The case of English [tʃ] is not so clear-cut (although the analogy of [dʒ] is vital), but the criterion remains the same, the extent of the limitation in our choice. The most severe limitation in the paradigm at a given point in the syntagma occurs when one element implies another. In the case of segmental phonemes, we regard two such segments as one phoneme, normally with the proviso that they are phonetically similar or at least comparable.

We can follow a similar procedure with prosodic elements. If we have great freedom of operation, and all or most of the oppositions are valid for each position, then we make our cut, i.e. select elements, at that level. If, for instance, in a language most (root) morphemes are disyllabic, and the tone patterns [¯ _], [_ ¯], [_ _] and [¯ ¯] all occur, then we must select elements at the syllable level, i.e. [¯] and [_]. If, however, only [_ ¯] and [¯ _] are possible, then we should take these as compound elements, since the initial low tone implies a following high tone, and the initial high tone a following low tone.

Quite another question (although admittedly a related one) is the role or function of the phonological unit in question. It may have a mainly distinctive role, in which case we should want to class it as an oppositional phonological element, or *phoneme:* in this case it must be capable of distinguishing lexical or intellectual meaning. On the other hand, its role may be principally demarcative, and we should then simply class it as a contrastive[6] unit, for which there is no generally agreed term. It is important to note that contrastive phonological elements do not *a priori* have to be any longer than a phoneme (witness the German glottal stop, or [ə] in some varieties of French); thus the term "long components" used by some linguists is not in many cases equivalent to our contrastive unit.

Yet a third dimension in the classification of phonologica- elements is introduced by the division between segments and feal

[5] Cf. Lit. 3, sections 2.6 and 3.8.

[6] In *Martinet's* sense, i.e. with syntagmatic function – as opposed to oppositional.

tures, e.g. between, for instance, [p] or [a] or [ʔ] on the one hand, and, say, a rising tone, on the other. It might be objected that a segmental phoneme is no more a segment than a feature such as high pitch, especially if a suprasegmental contour has already been "extracted" which cuts right through the "segment" in question. Fòr example are [a] and [ə] segments in Swedish [ˇandən] *anden*, 'the spirit', when the [ˇ] has been cut out? We can perhaps best overcome this difficulty by reinterpreting the distinction as one between a complex of distinctive features (applying to, or characterizing, the same segment) and a single distinctive feature.

What criteria do we consider in coming to a decision between a segment (or feature complex) and an independent single feature, i.e. non-segmental unit? Let us consider some concrete examples. No one apart from the thorough-going prosodist seriously considers extracting voice as a phoneme: we prefer to keep it as a distinctive feature applying to phonemes. If, in a language, pitch or length operates at the phoneme level [i.e. its recurring distinctive patterns extend over one (vocalic) phoneme], can we ever be justified in extracting tonemes or chronemes? We must not demand that the prosody in question should apply to all phonemes (i.e. to the consonants as well), since voice is only normally distinctive with stops and fricatives. So why should it not be possible in the assumed conditions to regard tone (high, low, falling, etc.) simply as a distinctive feature of a whole series of vowel phonemes? We are not even as justified in extracting it as a separate phoneme (or toneme) as we are when we isolate /h/ in a language which possesses a complete series of aspirated and unaspirated stops, because (i) /h/ would also have to occur independently[7] for us to accept this interpretation (this is hardly possible for a tone), (ii) /h/ can be cut off in time as a segment (it might be considered as a voiceless vowel), while tone is a simultaneous feature. (There is more of a case here for units of length, i.e. chronemes[8].) We may generalize and say that a distinctive feature is only to be considered as an independent element when it applies to a segment to which no other distinctive feature applies, i.e. to a supraphonemic segment. A complex of distinctive features which characterize the same segment form a segmental phoneme.

Thus although there are apparently three intersecting criteria for classifying phonological units, viz. distinctive v. demarcative,

[7] I.e. not automatically attached to one particular type of phoneme, e.g. vowel, stop.

[8] Cf. *Moulton*, Lit. 4, p. 379.

minimal v. long, and "segment" v. suprasegmental feature, there are cases where one criterion does not apply. For example, prosodic features should only be considered as prosodemes, when they are long; otherwise they should be classed simply as distinctive features of segmental phonemes. Demarcative units, when long, are never segments but always features, e.g. stress patterns in the word (in languages with fixed stress); but when short, i.e. subphonemic, they may be either single features, e.g. the frontness of English /l/ initially, or segments, e.g. the aspiration of English /p, t, k/ initially, although both types come under the head of allophonic variation besides their demarcative role.

An overall classification might be presented diagrammatically as follows:

Phonological units with distinctive function

suprasegmental features (long) (*prosodemes*)

segments (= complexes of distinctive features) (*phonemes*)

minimal (*simple phonemes*)

long (*compound phonemes*) e.g. diphthongs, affricates

Phonological units with demarcative function

short (allophonic variation)

long (prosodic patterns)

It is important to observe that one and the same unit may have both distinctive and demarcative function, e.g. the English /h/ phoneme, which beside its purely oppositional role, is important as a marker of morpheme-initial position in all native English words.

References

1. *Firth, J. R.:* Papers in linguistics 1934–1951 (London 1957).
2. *Jones, D.:* The phoneme, its nature and use (Cambridge 1950).
3. *Martinet, A.:* Eléments de linguistique générale (Colin, Paris 1960).
4. *Moulton, W. G.:* Syllable nuclei and final consonant clusters in German, in: For *Roman Jakobson* (Mouton & Co., s'Gravenhage 1956), p. 372–381.

5. *Trubetzkoy, N. S.:* Grundzüge der Phonologie. 3. Aufl. (Vandenhoeck & Ruprecht, Göttingen 1962).

Author's address: David J. Allerton, Department of General Linguistics, The University, *Manchester 13* (England).

Discussion

Valdman (Bloomington): It is often difficult in linguistic description to distinguish amongst the various functions of phonic material. In French, for instance, vowel length seems to operate as a segmental phoneme since in careful style it distinguishes *maître* from *mettre*, but for many vowels it is allophonic. Thus, the vowels /a o œ/ are long in checked final syllables as are all nasal vowels. Furthermore since in French [:] occurs only in the final syllable of a phonemic phrase it also has demarcative function.

The same distinctive feature can also serve as a short or a long component. In Haitian Creole nasalization differentiates segments, e.g. [še] *chien* us. [šè] *cher*, but it also appears on segments adjoining a nasalized segment. For instance *la chaîne* is [šen la] in citation style but [šennə] in normal style.

Answer *Allerton:* Perhaps I did not stress enough that distinctive and demarcative functions often apply to the same phonological element. I concede that this may even be so in the majority of cases.

With regard to Mr. *Valdman's* claim that nasalization is both a long component and a distinctive feature of phonemes in Haitian Creole, I feel that this is in no way in conflict with the proposed classification.

Krámský (Praha): I should like to express my strong approval with the speaker's opinion on the so-called tonemes and chronemes. Why should we complicate the phoneme inventory by separating the distinctive features of tone or length? There is no reason for it and there were some good arguments against it in the lecture. In American structural descriptions of languages it is quite current to speak about a "length phoneme". It can be argued that if length is a phoneme then the other distinctive features are phonemes too, because nobody can deny that length is a distinctive feature. Further, phonemes must be separable: how can we separate length from a sound? Similar is the problem of the so-called prosodic phonemes, i.e. pitch phonemes, stress phonemes, and juncture phonemes. This problem, however, differs from the problem of length by the fact that length pertains to one phoneme only, whereas prosodic phonemes can extend over more than one phoneme. Of course, this is not characteristic of either phonemes or distinctive features. Formerly we spoke simply about "prosodic qualities" and did not try to make phonemes of them. At the end the lecturer has made a compromise in the case of prosodemes: as he writes, "prosodic features should only be considered as prosodemes, when they are long, otherwise they should be classed simply as distinctive features of segmental phonemes". I do not regard this compromise as necessary. The problem evidently needs a more thorough examination.

My second remark concerns the classification of phonemes into simple and compound phonemes (e.g. diphthongs and affricates). According to my opinion, the question of complexity or compoundness cannot be the criterion of classification. It does not play any distinctive role in the classification, apart from the fact that it is not adequate to the classification into suprasegmental features and segments. It is not on the same level.

Answer *Allerton:* Mr. *Krámský* has raised the question of juncture, and I ought to explain how it would fit into my scheme. Since juncture phenomena are demarcative in their function, they should be considered under this head, whether they are examples of conditioned allophonic variation or of long prosodic features such as accentual patterns.

I do not feel my treatment of prosodic features to be a compromise. When their patterns extend over segments longer than a syllable they must be considered as occurring independently of phonemes.

The term "compound" phonemes is in some ways unsatisfactory. All that is meant is a segment which is considered as one phonemic segment, even though it might be analyzed into two elements either in the language in question or in some other language. It is not my term but a conventional one which I have explained in terms of my own classification of phonological elements.

Buyssens (Bruxelles): In tone languages distinguishing between 4 tones in disyllabic words you consider the pitch as a feature of the vowel; but I do not see how that is possible, for the pitch can only be determined if another syllable precedes or follows; provisionally I can only consider tone as a prosodic feature.

Answer *Allerton:* I do not consider the pitch of a syllable to be determined in such a language, since the selection of high or low onet for one syllable or vowel still leaves the possibility of high or low in the other.

Verh. 5. int. Kongr. Phon. Wiss., Münster 1964, pp. 207–210
(S. Karger, Basel/New York 1965).

Zur Biophonetik des rhetorischen Ausdrucks

Von HERBERT ARNDORFER, Wien

Wegen der seit der Aufklärung in Deutschland geltenden einseitig logisch-grammatischen Sprachbetrachtung wurde die Rede als reine Sachdarstellung aufgefaßt und ihr bisher noch kein angemessener Platz im Kreise der phonetischen Wissenschaften eingeräumt. Wenn man von der Phänomenologie ihrer lautlichen Gestalt ausgeht, wird es jedoch klar, daß nur die Kriterien und Aspekte der Phonetik die relationstreue Interpretation der Sprechweise und darüber hinaus – auf dem Wege über die physiologische Fundierung des sprecherischen Ausdrucks im Sinne von *F. Trojan* – auch zufolge ihrer Rolle als Appell und Ausdruck die Erfassung der Sprecherpersönlichkeit ermöglichen. Biophonetisch gesehen hat also eine politische Rede den gleichen Ausdrucksgehalt wie ein lyrisches Gedicht.

Dies sei an Zitaten der drei erfolgreichsten Redner der ersten Hälfte unseres Jahrhunderts bewiesen:

(Drei Sprechaufnahmen.) Phänomenologisch treten uns hier die Repräsentanten von drei verschiedenen phonetischen Ausdrucksrichtungen entgegen. Im ersten und zweiten Beispiel (A und B) fallen sofort die Dominanz des Konsonantismus, die Verhärtung der Einsätze und das Fehlen des Näselklanges auf. Die Längung der Tonworte läßt sich spektrographisch gut nachweisen.

Beide Redner bedienen sich in reichem Maße des emphatischen Akzents, dem im syntaktischen Bereich die Umstellung der Wortfolge entspricht ("...unless we conquer, as conquer we must, as conquer we shall...").

Der Redner C dagegen scheint trotz volltönender, von Dominanz des Vokalismus geprägter Stimme innerlich unbeteiligt, was durch Unveränderlichkeit von Melos, Atemführung und Stimmlage, vor allem aber durch den streng sprachlogischen Akzent unterstrichen wird.

Äußerst bemerkenswert ist jedoch die Differenzierung und Modifizierung von A gegenüber B in Kriterien, in denen die oben zitierte sprecherische Grundtendenz durch Bildung und Erziehung beeinflußt wurde. Es sind dies vor allem:

	A	B
Atem	normal fließend	stoßweise pulsierend
Kompression	optimal	übermäßig
Artikulationsspannung	gering	auffallend verstärkt
Silbenschnitt	schwächer	scharf
Melos	ruhig	Dynamik
Faukale Distanz	relative Weite	Enge
Stimmführung	Legato	Staccato

Es differenzieren sich somit auf Grund distinktiver Merkmale drei phänomenologisch unterschiedliche sprecherische Ausdrucksgruppen, denen – unter allen Vorbehalten des Typenbegriffs – durch entsprechende biophonetische Interpretation auch alle nur mehr schriftlich erhaltenen rhetorischen Zeugnisse zugeordnet werden können.

Die Validität dieser Gruppenteilung wird ferner durch die inhaltliche Deutung wie auch durch die exakte psychologische Analyse erhärtet, die uns über die konstitutionsbiologischen Typen von *Kretschmer* und *Sheldon* zu den drei Grundformen des stimmlichen Ausdrucks im Sinne *F. Trojans* führen.

Es ist bemerkenswert und steht mit dem modernen Geschichtsbild vollkommen in Einklang, daß der Trophotrope (der Zyklothyme *Kretschmers*) in der durch Kampf und List charakterisierten Politik der Neuzeit nicht in Erscheinung tritt. Nur einzelne Gestaltzüge zeigen sich in der englischen Rhetorik, so wenn etwa Sir Winston Churchill (Beispiel A) durch besonnene Ruhe auch im sprecherischen Ausdruck das Vertrauen seiner Zuhörer ebenso gewinnen will, wie das einst William Pitt d.J. getan hatte. Churchills Grundtendenz ist aber die kämpferische Ergotropie, die nur durch Bildung und Erziehung trophotrop modifiziert wird.

Die reine Ausprägung des von aggressivem, ichbetontem Machtstreben bestimmten Ergotropen erkennen wir dann im phonetischen Ausdruck des zweiten Sprechers – in Adolf Hitler. Der inhaltlichen Wiederholung von Schlagworten und der einseitigen Alternativhaltung Sein–Nichtsein entsprechen im lautlichen Bereich die ungezügelte Verwendung der Kraftstimme und das Vorherrschen des Rhythmus gegenüber der Melodik. Die Stimme steigert sich bis zum Staccato und betont in der Steigerung der animalischen Leistungen durch die Sympathikuserregung das imponierende konsonantische Element; konform damit geht die Verhärtung der Einsätze, die Zunahme des Muskeltonus und die stoßweise Atmung. Cromwell und Bismarck lassen sich auf Grund phonetischer Analyse ebenfalls dieser Ausdrucksgruppe zuordnen, wobei sich die Bestätigung aus zeitgenössischen Kritiken ihrer Sprechweise erbringen läßt.

Im sprecherischen Ausdruck des durch Beispiel C charakterisierten Ideotropen dagegen dominiert der von gefühlsmäßigen Schwankungen unbeeinflußte verstandesmäßige Ausdruck. Seine vitale Schwäche sucht ihre Kompensation im volltönenden Klang und in dem streng logischen Akzent. Politisch ist er durch seinen Idealismus der Gefährlichste – der «Demagog». Kalt und grausam wird er die ihm gleichgültigen Mitmenschen der kompromißlosen Durchsetzung seiner Ideen opfern. Robespierre hatte versucht, durch pausenlosen Einsatz der Guillotine den tugendhaften Idealstaat zu schaffen, 150 Jahre später verspricht Goebbels Deutschland die Weltherrschaft.

Es sei hier jedoch nachdrücklichst darauf hingewiesen, daß eine solche biophonetische Analyse nur objektiv registriert und sich keinerlei Werturteil über Intelligenz oder moralische Haltung anmaßt. Es besteht daher solcherart die Möglichkeit, zur Ermittlung der historischen Wahrheit auch politische Reden leidenschaftslos nur von ihrem phonetischen Ausdrucksgehalt her zu interpretieren und so ein objektives Bild der Rednerpersönlichkeit zu gewinnen. Durch eine solche biophonetische Sicht wird es möglich sein, die Rhetorik wieder in den Kreis der phonetischen Wissenschaften zurückzuführen.

Literatur

Abbot, W. C. (ed.): The writings and speeches of Oliver Cromwell (Harvard University Press, Cambridge 1937).

Arndorfer, H.: Typen rhetorischer Gestaltung (unbedruckte, maschinenschriftliche Dissertation, Wien 1960).

Bismarck, O. von: Die politischen Reden; hg. von *H. Kohl* (Cotta, Stuttgart/Berlin 1892–1905).

Gauger, H.: Die Kunst der politischen Rede in England (Niemeyer, Tübingen 1952).

Kretschmer, E.: Körperbau und Charakter; 22. Aufl. (Springer, Berlin 1955).

Pitt, W.: Speeches in the House of Commons; 2nd ed. (Longman, London 1808).

Robespierre, M.: Discours et rapports (Charpentier et Fasquelle, Paris 1908).

Trojan, F.: Der Ausdruck der Sprechstimme; 2. Aufl. (Maudrich, Wien/Düsseldorf 1952).

Ein Teil der Sprechaufnahmen stammt aus dem Archiv des British Council Wien.

Adresse des Autors: Dr. Herbert Arndorfer, Siebenbrunnengasse 5a, *Wien 5* (Österreich).

Discussion

Höffe (Dortmund): Wir haben dem Referenten herzlich dafür zu danken, daß er überzeugend dargelegt hat, in welcher Weise sich Inhalt, stilistische Form, stimmlich-lautlicher Ausdruck und Typus bedingen. Versuche dieser Art sollten fortgesetzt und so die ganzheitliche Schau sprachlicher Äußerungen gefestigt werden.

Proc. 5th int. Congr. phon. Sci., Münster 1964, pp. 211-214
(S. Karger, Basel/New York 1965).

On the Linguistical Characteristics of the American Hungarian. (An Outline)

By E. Bako, Silver Spring

The establishment of the first sizable Hungarian settlements in the United States of America, beginning with the 1870's, has coincided with the issuance of the first set of standard handbooks on the history, grammar, lexical elements, dialects, etc. of the Hungarian language. In this period, attention of the scientific community was focused upon linguistic problems as they appeared in a historical setting. Being of short duration, Hungarian immigration to the United States has been regarded for decades as temporary and beyond any sort of research value. Accordingly, not much significance, if any, was attributed to the appearance of the first symptoms of an English (American) – Hungarian bi-lingualism. While during the life span of three generations the speech habits of American Hungarians have gained a fairly constant and firm character, and the number of Hungarian speaking persons on the North-American continent was climbing close to 800,000 (totaling more than 700,000 in the United States and about 80,000 in Canada at the time of this writing), linguistic research in Hungary has completely overlooked this new area of research objectives and all attempts outside Hungary trying to initiate a positive approach to them[1].

Mostly in the course of the past seven years, this writer was

[1] The Eastern American Dialect of Hungarian; An Analytical Study, by *Pierre Ervin Szamek*, a 145-page (typed) dissertation presented to and accepted by the Faculty of Princeton University in candidacy for the degree of Doctor of Philosophy in 1947, went annoticed by Hungarian linguists. In this connection, it has to be stated that the method applied by Mr. *Szamek* does not appear to this writer as correct or purposeful. Since his findings were based upon materials derived from a very small group of speakers of a couple of communities in New Jersey, such a material does not permit the drawing of conclusions regarding analysis and definition of American Hungarian dialect speech phenomena.

conducting tape-recordings of Hungarian dialect materials in the Northeastern United States. His trips covered the area between Milwaukee, Wisconsin, and Pittsburgh, Pennsylvania, in the West, and Boston, Massachusetts, and the City of Washington in the East. A total of about two hundred selected speakers were interviewed, the 150 hours of their recorded dialogs and narratives representing about 3,000 pages of dialect texts when fully transcribed[2].

The present occasion does not permit the inclusion of detailed sample material. However, based upon the above mentioned collection of American Hungarian dialect texts (a considerable part of which is already transcribed and indexed), and of other observations and studies closely connected with it, the following characteristics became apparent to this writer:

1. As a general rule, the overwhelming majority of the oldest generation of American Hungarians speak one of the dialects instead of the literary (or educated middle-class) form of the Hungarian language. Consequently, any type of phonetical research directed at American Hungarians has to set out under the guidance of methods and principles genuine to dialectology.

2. Hungarian immigrants to the United States came from all regions of Hungary's pre-World War I territory, representing, practically speaking, all main dialects and many of the sub-dialects of the Hungarian language.

3. Accordingly, all Hungarian dialect units were and are exposed to the North-American variants of the English language, and to various American versions of other, non-English languages. Such a multi-lingual environment creates an abundance of interlingual phonetical, lexical, structural, etc. interferences.

4. Consequently, studies of the American Hungarian dialects offer a full set of subjects and objectives for the researcher, comparable only to those known to and pursued by phoneticians and dialectologists active in Hungary itself. More than that, interlingual borrowing of phonetical, lexical and structural elements between Finnish and Hungarian as it is practiced in communities of mixed

[2] See: *Bako, Elemer:* Goals and methods of Hungarian dialectology in America. *Az amerikai magyar népnyelvkutatás célja és módszere*. Published for the American Hungarian Institute by the American Hungarian Studies Foundation at Rutgers, The State University, New Brunswick, N. J. (USA), 1962, 24 p. (American Hungarian Dialect Notes, 1), and, by the same writer, a conclusive report on tape-recording trips made in 1962 and 1963 under a grant received from the American Philosophical Society which was published in the Society's *Yearbook 1963*, Philadelphia, Penn., 1964, pp. 494–497.

Finnish-Hungarian majority in Ashtabula County, in the State of Ohio, or similar contacts of Hungarian dialect speakers with persons speaking Italian, Polish, Spanish, Swedish, Flemish, Greek, Welsh, and other languages in the giant cities or areas of varied ethnic character all over the United States and Canada, resulted in a wealth of unique cases for phonetical analysis and comparative phonology.

5. Since the presently living oldest group among Hungarian dialect speakers in America has left the European homeland some 50–70 years ago, their dialect forms may be more archaic than those of the people residing in their land of origin. Consequently, the possibility that American Hungarian dialect forms may be regarded as possible links in a chain of transition of Hungarian dialect phenomena, cannot be excluded.

6. It is recognized that the majority of Hungarian dialect speakers was never exposed to the influence of recently developed vocabulary and phraseology in Hungary. Thus, the American Hungarian dialects have retained numerous elements of their original vocabularies (with phonetical characteristics attached to them) which may not be anymore among the actively functioning speech elements of their mother dialects.

7. Most dialect speakers interviewed by this writer have exhibited a sort of "dual personality": in their Hungarian conversation, they usually appear as bearers of the traditional speech habits of the rural Hungarian types, whereas, in forms of conversational self-expression as well as in the light of subjects selected for conversation in English, the same persons often appear as urban citizens of a highly industrialized society. This shift is accompanied by changes reflecting a motivation of social attitudes, characteristic features which distinguish the peasant from other socio-economic types slightly below the middle-class level. Such shifts usually affect speech production, both when speaking in Hungarian or in English.

8. A number of the persons interviewed use English as a second home language, especially in the presence of their children, grandchildren or their American friends or spouses. Most members of the second generation speak Hungarian in a form which unveils many compromise solutions or borrowings between Hungarian and English. Third generation speakers if they had enough opportunity to live with their grandparents during their early childhood and their teens, have often developed amazing consistency in their

Hungarian speaking abilities and habits, often surpassing those of their own parents.

9. As a general rule, there is a numerical and functional predominance of the so-called Northeastern Hungarian dialect among dialect speakers in the Northeastern part of the United States; this fact being attributable to the circumstance that persons from the Northeastern regions of pre-World War I Hungary were in majority among the immigrants in those times.

10. Another contributing factor to the emergence of the Northeastern dialect as a dominant idiom and formative element in the development processes of the American versions of Hungarian dialects was the presence and activity of a considerable number of Hungarian Reformed (Calvinist) ministers who came mostly from the same Northeastern parts of Hungary.

Since the Hungarian literary style and orthography were developed largely on the same historical-geographical basis (beginning with the first complete Bible translation by *Gáspár Károli* in 1590 and including most of the influential poets and writers of the 19th and 20th centuries such as *Ferenc Kazinczy*, *Ferenc Kölcsey*, *János Arany*, *Endre Ady*, *Zsigmond Móricz*, and others), the weight of the printed material used by American Hungarian readers was constantly reaffirming the trend toward the acceptance of the Northeastern dialect as the structural pattern for generally recognizable new forms of the American Hungarian speech.

Author's address: Elemer Bako, 810 Loxford Terrace, *Silver Spring*, Md. (USA).

Proc. 5th int. Congr. phon. Sci., Münster 1964, pp. 215–217
(S. Karger, Basel/New York 1965).

Predicting the Intelligibility of Words III

John W. Black, Columbus, Ohio

The degree and manner in which the intelligibility of a word relates to the intelligibility of the phonemes that comprise the word has been the topic of two investigations*. The first was of two parts. In each the written responses of 200 listeners were analyzed. One group had attempted to identify a set of approximately 600 words, about 1940; the other, a similar set, about 1950[1]. Both sets of words were heard through electronic communication systems by military personnel in the presence of high-level noise (108–110 db, resetting *C* on a General Radio sound level meter). In the earlier instance the listeners heard direct transmissions by the full complement of their 200 peers; in the later case, the listeners heard material that had been recorded by persons who were experienced in conducting researches in voice communication and were aware of the difficulties of communicating in noise. There was no overlap in speakers or listeners and minimal overlap in the words that were spoken.

A second study was conducted (1960) as an economical replication and refinement of either part of the first experiment and analysis[2]. Each of five workers in voice communication selected randomly twenty lists of twenty-five words from a 300-word master list. He recorded one list daily and then in the same laboratory period listened to the recordings made by his four colleagues. In this instance the listener, working from a voluntarily screened script of the speaker's words, exposed a written word subsequent to hearing it and wrote only his responses (receptions) that were in error. The economy in time is apparent. The refinement included the fact that all responses were made and, in turn, utilized. In the earlier instance there were many omissions among the responses; also single instances

* These studies and the present one were conducted at The Ohio State University, Columbus, Ohio, under a contract between the Office of Naval Research and The Ohio State University Research Foundation [Contract No. Nonr–495 (18) NR 145–993].

of particular responses – those made by one listener only. These were treated as bizarre and discarded.

The present experiment (1963) was in the main a replication of that portion of the foregoing one that fell subsequent to the recording stage. The same five listeners who heard the recordings of the earlier experiment listened a second time after a lapse of three years. In the present instance each listener could also hear the material that he had recorded earlier. This was not permitted in the original instance lest the effects of short-term memory contaminate the values of intelligibility. Additionally in the present study five new listeners provided a set of comparison values.

The rationale for the study lay particularly in reliability:

a) a comparison of relative intelligibility values of phonemes and words yielded by five listeners hearing the same verbal material a second time with three years intervening between the listening sessions, and

b) a comparison of the results yielded by a second group of listeners of the same limited size as the first.

Procedure

The recorded lists of one- and two-syllable words described above were heard by ten listeners at a O–db signal-to-noise ratio (white noise). Five of the listeners had recorded the words three years previously and each had listened to the recordings made by his colleagues; five were hearing them for the first time. The speakers had contributed phonetic transcriptions of their word lists at the time of recording. The listeners' phonetic transcriptions of the heard material were compared to these. This led to measures (a) of correct identifications of words (and concomitantly of the phonemes comprizing the words), (b) of correct identifications of phonemes in words that were misidentifications of a stimulus – called preservation-in-error (P–I–E) values, and (c) of the relative concurrence among substitutions of one phoneme for another.

The procedures for the former study utilized five methods of predicting the intelligibility of words from empirically derived values: 1. initial sound squared; 2. joint probability of the first two sounds, retaining the values peculiar to each position; 3. joint probability of the first two sounds, using the value for the initial position only; 4.–5. joint probability of the two and three most intelligible sounds in the word. None of these appeared to hold an advantage over the recommendation of the earlier study, that the intelligibility of a word is related beyond chance to the joint intelligibility of the first two phonemes. Accordingly in the present study this method was employed.

Results

The three sets of data showed a significant correspondence among the relative intelligibility values of the various phonemes. The consonants varied in intelligibility from 60 to 90%; the vowels, 65 to 90%. The comparison group of listeners yielded the highest of the three sets of scores for each of one half of the consonants and for 10 of the 16 vowels and diphthongs.

On the basis of scores predicted from the joint intelligibility of the first two phonemes of a word, the predicted values would be the same for one- and two-syllable words. The obtained values of two-syllable words exceeded the values of one-syllable words. Thus a constant error is built into this method of predicting the intelligibility of words. The method accurately predicts the mean value of two-syllable words and over estimates the mean intelligibility value of one-syllable words.

Not only does the joint-intelligibility of the first two phonemes yield an acceptable mean value, it provides a distribution of scores that is comparable to the distribution of obtained scores. The standard deviations of the obtained scores in categories of at least 35 words ranged from 7.2 to 18.9; for the predicted values, 6.7 to 12.9.

Conclusions

The present data suggest that the relative intelligibility of a phoneme is a stable phenomenon; further that this characteristic can be demonstrated with few listeners, few speakers, and a limited number of words. Currently, this outcome is based on five speakers, five listeners, and 300 words, a combination that yielded 10,000 responses to simple words. Possibly more results are to be expected.

References

1. *Black, J. W.:* Predicting the intelligibility of words. Folia phoniat. *12:* 260–272 (1960).
2. *Black, J. W.:* Predicting the intelligibility of words II. Proc. Fourth Int. Congr. Phonetic Sciences; pp. 347–354 (Mouton, s'Gravenhage 1962).

Author's address: Prof. Dr. John W. Black, The Ohio State University, Department of Speech, 154 North Oval Drive, *Columbus 10*, Ohio (USA).

Verh. 5. int. Kongr. Phon. Wiss., Münster 1964, pp. 218–221
(S. Karger, Basel/New York 1965).

Zur phonologischen Behandlung von Fremdwörtern

Von H. Bluhme, Amsterdam

Bei der Aufstellung von Phoneminventaren und den dazugehörigen Verwendungsregeln, der sogenannten Distribution, beziehen manche Vertreter der Phonologie Fremdwörter mit ein, während andere sich auf den gemäß den Gesetzen der historischen Sprachwissenschaft ererbten, einheimischen Wortschatz der Sprache oder Mundart beschränken. Zu letzteren kann mit Fug *N. S. Trubetzkoy* gerechnet werden, auch wenn er sich in den «Grundzügen der Phonologie» nicht explicite zu diesem Problem geäußert hat; dies geht aber aus der Aufstellung seiner Phoneminventare hervor.

Für die definitive phonologische Analyse einer Sprache kann man den Standpunkt beziehen, daß auch die Fremdwörter, die in der täglichen Umgangssprache vorkommen, einbezogen werden sollten, da sie Oppositionen zu anderen Wörtern bilden können. Diesen Weg beschreitet die Informationstheorie: Sie interessiert sich unterschiedslos für alle Phoneme und Phonemverbindungen einer Sprache und versucht durch Feststellung ihrer Häufigkeiten den Informationswert eines Textes zu bestimmen. Auf die Phonemverbindungen wird dabei größerer Wert gelegt als auf die einzelnen Phoneme selbst, da sich zeigen läßt, daß Spracherkennung weniger von den Phonemen als von den Phonemfolgen, den clusters, abhängig ist*. Spracherkennung Phonem um Phonem ist überhaupt unmöglich. Wenn wir jedoch die Phonemverbindungen, die in ihrer Art und Zahl sehr beschränkt sind, zugrunde legen, dann reicht die menschliche Kanalkapazität sehr wohl aus, um das Gesprochene zu erkennen. Daraus läßt sich folgern, daß die Phonemclusters noch mehr Aufmerksamkeit verdienen als die Phoneme selbst. Mit anderen

* Die Kanalkapazität des menschlichen Nervensystems beträgt ca. 40 bits/sec. Um einen mit einer Sprechgeschwindigkeit von 15 Phonemen pro Sekunde gesprochenen Text aufzunehmen, ist jedoch, wenn wir im Deutschen ein Inventar von 35 Phonemen zugrunde legen, eine Kanalkapazität von 15 ld 35 = 54 bits/sec erforderlich.

Worten, es geht nicht mehr allein darum, ob zum Beispiel im Deutschen /ã/ von *Restaurant* oder im Polnischen /f/ von *farmaceuta* Fremdphoneme sind, sondern auch darum, welche Phonemverbindungen und prosodische Eigenschaften allein auf Fremdwörter beschränkt und vom übrigen System getrennt zu behandeln sind.

Wir sind zwar imstande, die intellektuelle Bedeutung zweier Wörter durch distinktive Schallgegensätze zu unterscheiden, zum Beispiel im Deutschen *Rose* und *Riese*, *Lamm* und *lahm*, doch lehrt uns ein Blick in das Homonymenlexikon, daß die intellektuelle Bedeutungsdifferenzierung nicht der einzige Grund sein kann, diese und andere Wörter auseinanderzuhalten. So verzichten wir zum Beispiel bei dem Wort [re:də] auf jede phonologische Differenzierung um die beiden Bedeutungen «sprecherische Darbietung» und «Ankerplatz vor dem Hafen» zu unterscheiden, und doch ergeben sich daraus kaum je irgendwelche Schwierigkeiten. Es ist nämlich so, daß wir die oben erwähnten Vokalqualitäten und -quantitäten unterscheiden, nicht weil dies für die Verständigung unumgänglich wäre, sondern weil die sprachliche Überlieferung uns dazu zwingt. Auf einer solchen Tradition beruht das Verharren von Fremdwörtern in der Form, die sie in der Fremdsprache hatten, aus der sie übernommen worden sind. Zur Zeit der Übernahme werden sie noch im Rahmen ihrer Ursprungssprache gesehen. Je nach den sozialen und politischen Umständen, den sprachlichen Beziehungen und der Zahl der Wörter, die ungefähr gleichzeitig übernommen werden, bleiben diese Fremdwörter in ihrer ursprünglichen Form erhalten oder gleichen sich im Laufe der Zeit der aufnehmenden Sprache an, werden also zu Lehnwörtern. Die Angleichung selbst ist in erster Linie durch strukturelle Prinzipien bestimmt.

Nach Möglichkeit haben wir also Fremdwörter zuerst mit anderen Fremdwörtern aus derselben Sprache und derselben Schicht zu vergleichen, wobei logisch privative oder graduelle Oppositionen besonders wichtig sind. Erst anschließend ist ein Vergleich mit dem einheimischen Phonemsystem am Platze. Die dabei neu auftretenden Phoneme und ihre Distributionen können dann dem einheimischen System als Fremdphoneme und Fremddistributionen hinzugefügt werden. Auf diese Weise können auch die prosodischen Eigenschaften besser systematisiert werden. Fremdwörter wie *Radieschen*, *Kaninchen* oder *Rakete* unterscheiden sich nämlich nicht nur in ihrem Vokalismus von ursprünglich deutschen Wörtern, sondern auch im dynamischen Akzent: Sie bewahren den Akzent auf derselben Stelle wie in ihrer Ursprungssprache und weichen dadurch prosodisch vom Deutschen ab.

Besonders wichtig scheint mir, daß die Distributionen zuerst ohne alle Fremdwörter untersucht werden, da die Phoneme der Fremdwörter nach den Gesetzen ihrer Herkunftssprache distribuieren, also Phonemverbindungen aufweisen können, die sonst in der Sprache nicht vorkommen. Durch die Einbeziehung von Fremdwörtern werden die Distributionsregeln einer Sprache nicht konsolidiert, sondern zur Distribution der einheimischen Wörter kommen noch hinzu die Distributionen der Sprachen, aus denen die Fremdwörter entlehnt sind. Mit ihrem Zusammenwerfen ist sprachwissenschaftlich nichts gewonnen, so interessant dieses Verfahren auch informationstheoretisch sein mag. Von den strukturellen Besonderheiten auszugehen, wie *Pilch* es vorschlägt, ist keine endgültige Lösung, da sich die Frage, was eine strukturelle Besonderheit ist, erst lösen läßt, wenn bekannt ist, daß diese auf eine bestimmte Sprachschicht beschränkt bleibt.

Bei der sprachwissenschaftlichen Betrachtung sind wir darauf angewiesen, Fremdwörter zu eliminieren, da wir anders den Wortschatz der zu untersuchenden Sprache nicht vom Wortschatz fremder Sprachen zu scheiden vermögen. Die Gebräuchlichkeit von Fremdwörtern in der Umgangssprache als Kriterium zu verwenden, ist unannehmbar, da dies auf einer psychologischen Wertung beruht und den Begriff 'Umgangssprache' völlig unbestimmt läßt. Nach unserer Auffassung sollten wir das Material nicht nach dem Prinzip 'Umgangssprache gegenüber Nichtumgangssprache' sichten, sondern vielmehr beide gleichermaßen nach den obengenannten Gesichtspunkten untersuchen. Wir ziehen lieber die bewährten Mittel der historischen Sprachforschung heran und halten zunächst alle Fremdwörter vom einheimischen Vokabular getrennt, um deren Phoneme erst dann als Fremdphoneme und Fremddistributionen wieder ins System einzubeziehen, wenn die Analyse des einheimischen Wortschatzes vollendet ist. Ebenso werden wir in den meisten Fällen mit den in der Sprache vorkommenden Namen, d.h. den Orts-, Flur- und Flußnamen sowie den Vor- und Familiennamen, verfahren müssen, da die Namenforschung nur zu oft zeigt, daß diese Wortgruppen anderen Gesetzen unterworfen sind als der allgemeine Wortschatz der Sprache; außerdem läßt sich bei Namen noch viel schwerer definieren, was zur Sprache gehört als bei den Fremdwörtern. Daß auch Interjektionen, Appellativa und schall-

nachahmende Wörter in diesem Zusammenhang ausgegliedert werden müssen, versteht sich seit *Trubetzkoys* «Grundzügen» wohl von selbst.

Schwieriger ist das Problem der Lehnwörter, wie z.B. nhd. *Keller* aus lat. *cellārium:* Sie dürfen erst herangezogen werden, wenn durch Vergleich mit einheimischen Wörtern gesichert ist, daß sie in jeder Hinsicht assimiliert sind, z.B. wäre für *Keller* der Beweis erbracht durch den Vergleich mit *ke*nnen, *El*le und M*áler*.

Adresse des Autors: Dr. Hermann Bluhme, Rooswijck 82, *Amsterdam 11* (Niederlande).

Discussion

Hoffmann (Hamburg): Herr *Bluhme* hat, wenn ich ihn richtig verstanden habe, gesagt, daß Fremdwörter mit der ihnen eigenen Betonung übernommen werden. Das stimmt nicht. In den Nachkriegsjahren konnte man beobachten, daß z. B. lettische Namen trotz einer dem Deutschen kongruenten Betonung mit einer Verschiebung des Akzents übernommen wurden, z. B.

lett. Fréimânis (cf. dt. Fréimann),	von Deutschen gesprochen als Freimánis
lett. Mángulis	gesprochen als Mangúlis.

C.R. 5e Congr. int. Sci. phon., Münster 1964, pp. 222–226
(S. Karger, Basel/New York 1965).

Le phonème et sa réalisation

Par Eric Buyssens, Bruxelles

Le mot *réalisation* dont on s'est servi pour annoncer le thème de ce Congrès, rappelle la conception du phonème qui avait cours il y a un quart de siècle: à cette époque, certains phonologues déclaraient que le phonème était ce que l'on voulait prononcer, et le son ce qu'on prononçait réellement. Cette conception ne tenait compte que du locuteur. L'auditeur, lui, perçoit le flot sonore et doit y retrouver les phonèmes: pour lui, le phonème est une abstraction. On se trouve donc devant un cercle qui rappelle un peu la célèbre dispute médiévale au sujet de l'œuf et de la poule.

Le problème se présente autrement pour le linguiste qui se trouve devant des locuteurs dont il ne connaît pas la langue, et qui cherche à découvrir le sytème phonologique de cette langue: le linguiste doit-il connaître le son pour identifier le phonème, ou – au contraire – connaître le phonème pour identifier le son? S'il doit connaître le phonème pour trouver le son, il a le droit de considérer le son comme la réalisation du phonème; sinon, il considèrera le phonème comme ce qui est commun à un certain groupe de sons ou bien comme ce groupe même. On peut formuler le problème autrement: le son est-il une donnée acoustique, c'est-à-dire une unité acoustique que l'acousticien identifie sans recourir à des critères linguistiques, ou – au contraire – le son est-il une donnée linguistique, c'est-à-dire une unité qui n'apparaît que lorsqu'on connaît la langue considérée?

L'Ecole phonologique de Prague a dès le début séparé radicalement le phonème du son; elle définissait le phonème comme un fragment du signifiant, c'est-à-dire sans faire allusion au son; cela impliquait que la connaissance du phonème précédait celle du son. Mais aujourd'hui certains linguistes américains prétendent qu'il est possible d'analyser la parole sans la comprendre, ce qui implique

que l'on peut partir du son pour aboutir au phonème. Il faut donc examiner qui a raison.

Le flot sonore qui constitue la parole est continu; celui qui cherche à le segmenter sans tenir compte de sa fonction communicative aboutit à des résultats déroutants. *Z.S. Harris* en donne un exemple dans son excellent livre «Structural Linguistics»; il montre que le son *s* comprend trois parties: 1. celle qui correspond à la mise en place des organes phonatoires, 2. celle qui correspond ou maintien des organes en place, 3. celle qui correspond à l'abandon de cette place. Cela veut dire que du point de vue acoustique, le son *s* ne présente pas d'unité.

Comme le dit *Harris*, le linguiste ne pousse pas la segmentation aussi loin:
"The point at which segmentation stops may be stated as follows: We associate elements with parts or features of an utterance only to the extent that these parts or features occur independently (i.e. not always in the same combination) somewhere else. It is assumed that if we set up new elements for successive portions of what we had represented by *s*, and then used them in representing various utterances, these new elements would not occur except together. We therefore do not subdivide *s* into three parts" (p. 21).

Comme on le voit *Harris* n'admet que la méthode suivante: comparer des paroles pour voir quels sont les éléments qui se présentent toujours ensemble. Mais cette comparaison n'est pas praticable d'un point de vue purement acoustique; et *Harris* l'a dit:

"The question of setting up elements may be approached with little initial sophistication. It is empirically discoverable that in all languages which have been described we can find some part of one utterance which will be similar to a part of some other utterance. Similar here means not physically identical but substitutable without obtaining a change in response from the native speakers who hear the utterance before and after the substitution: e.g. the last part of *He's in* is substitutable for the last part of *That's my pin.* In accepting this criterion of hearer's response we approach the reliance on meaning usually required by linguists. Something of this order seems inescapable, at least in the present state of linguistics: in addition to the data concerning sounds we require data about the hearer's response" (p. 20).

Harris a cherché analyser le flot sonore sans le comprendre, mais il reconnaît explicitement que cette analyse du flot sonore en éléments fonctionnels ne peut se faire sans comprendre ce flot, c'est-à-dire sans connaître d'avance l'articulation de la langue en question. Le son n'est donc pas une donnée objective de l'acoustique.

On est tenté de définir le son comme le plus petit segment sonore permettant de distinguer formellement deux signifiants ou deux phrases ayant des significations différentes; mais dès que l'on fait allusion à la signification, on se place sur le plan social de la communication; et dès lors on doit tenir compte des variantes individuelles: il n'y a pas deux individus produisant exactement les mêmes sons; le linguiste se trouve devant un groupe de sons différents.

Ce qui, aux yeux du linguiste, distingue deux signifiants donnés, ce n'est pas un son, ce n'est pas une réalité acoustique concrète; c'est un groupe de caractéristiques acoustiques présent à un certain endroit du flot sonore: c'est le faisceau des traits pertinents. Par exemple, si j'entends un Français prononcer la phrase *Il est rentré* en roulant les *r*, je comprends la même chose que lorsqu'un autre Français prononce cette même phrase en grasseyant: l'*r* uvulaire remplit la même fonction que l'*r* apical; ou plutôt, c'est un ensemble de traits phonétiques communs aux deux sons qui confère la même fonction aux deux sons. Cet ensemble de traits est l'unité fonctionnelle qui distingue les signifiants.

Quant au son, c'est ce qu'on découvre dans le flot sonore individuel à l'endroit où fonctionne l'unité distinctive. Autrement dit, le son concret est identifié à partir d'une unité fonctionnelle abstraite. Cette unité n'est pas le phonème; on entend par phonème un groupe de telles unités fonctionnelles ou ce que ces unités ont en commun; ce groupement se fait au nom de critères distributionnels bien connus. L'unité à laquelle correspond le son concret est l'allophone; autrement dit, la réalisation de l'allophone s'effectue dans le son. Pour réaliser un allophone, le locuteur ajoute aux traits de l'allophone les traits dus à l'intonation et enfin ses caractéristiques personnelles.

Le son n'est donc pas simplement la réalisation de l'allophone; c'est plus que cela: le son est un segment du flot sonore dans lequel on trouve la réalisation de l'allophone.

Cela ne signifie pas qu'il faille renoncer à affirmer qu'un allophone est constitué par ce qui est commun à un groupe de sons; mais il faut renoncer à voir dans cette affirmation la définition de l'allophone. Une définition est l'énumération des faits nécessaires et

suffisants pour reconnaître une chose; l'allophone se reconnaît essentiellement à sa fonction.

Quant au phonème il se définit comme ce qui est commun à un groupe d'allophones complémentaires. Certains auteurs conçoivent le phonème comme le groupe même des allophones complémentaires; mais cela revient au même, car pour établir quels allophones entrent dans ce groupe il faut établir ce qu'ils ont de commun.

Si l'on opte pour la conception du phonème comme un groupe, on ne peut pas dire que le phonème se réalise. Si au contraire on voit dans le phonème ce qu'il y a de commun à certains allophones, on peut dire que le passage du phonème à l'allophone est une première étape vers la réalisation; mais la réalisation véritable ne se produit que dans le son, qui complète l'allophone par l'intonation et par les caractéristiques individuelles.

Adresse de l'auteur: Prof. Eric Buyssens, Université de Bruxelles, 99, rue de l'Abbaye, *Bruxelles 5* (Belgique).

Discussion

Di Pietro (Georgetown): In English, we may make a three-way distinction between "phone", "phoneme" and "allophone". It seems to me that "allophone" as well as "phoneme" refers to abstractions. That is to say, the decision that a speech sound is an allophone indicates that one already knows what phoneme class it belongs to. Without such a decision, the speech sound remains a "phone" – which appears to be the only reality at the first stage of analysis.

Fourquet (Paris): Il peut y avoir des signifiants dépourvus de signifié. Un exercise de contrôle de la possession d'une langue consiste à dicter des mots sans signification, mais conformes au système phonologique de cette langue, et à en demander la transcription dans l'alphabet phonétique international. Les élèves ne peuvent plus identifier le mot *globalement*, puis transcrire d'après leur propre prononciation ou d'après la graphie ordinaire. Ils doivent identifier correctement les *phonèmes* dont le mot est composé, de façon entièrement indépendante de la fonction de communication, de l'identification du mot comme unité signifiant/signifié. L'opération consiste à mettre en correspondance les sons entendus avec les phonèmes du système allemand, considéré comme réseau de référence.

Mikuš (Zadar): Les trois phases phonatoires – implosion, tenue, explosion – sont présentés à la réalisation de tout phonème, mais elles ne sont pas également relevantes. Dans un cas, c'est l'implosion ou l'explosion, dans d'autres cas c'est la tenue qui est relevante comme fait acoustique.

Martinet (Paris): L'implosion, la tenue et l'explosion sont presque nécessairement conçues comme pouvant chacune représenter le même phonème, car l'auditeur est également un locuteur qui sait que l'un de ces traits n'existe pas sans les autres, même si ceux-ci ne sont pas perceptibles.

L'idée d'identifier un phonème dans différentes positions en coupant une bande magnétique est classique. Mais l'opération peut fort bien ne pas donner les résultats escomptés: Si l'on remplace le *e* de danois *net* par celui de danois *ret* on obtient un autre mot *nat*, bien que le *e* de *net* et celui de *ret* soient le même phonème (3e degré d'ouverture à l'avant).

Il faut bien se rendre compte que les linguistes tombent aisément d'accord pour déterminer le nombre d'unités distinctives qui existent dans un contexte déterminé, mais qu'ils sont moins unanimes lorsqu'il s'agit de savoir si telle unité dans telle position est le même phonème que telle unité dans telle autre position: tout le monde sait que les voyelles d'anglais *bed* et *bad* sont distinctes, mais il est impossible de rien dire de précis relativement à l'identité ou le non-identité des deux voyelles de *kitchen*.

Il faut établir les systèmes de phonèmes pour chaque position et cela est relativement facile; la difficulté et les divergences commencent lorsqu'on cherche à rapprocher ces systèmes et à identifier une à une les unités des différents systèmes. Certains linguistes opèrent, par example, dans ce cas avec le compte de neutralisation, d'autre l'évitent. Il n'est pas nécessaire que les systèmes aient réellement les mêmes unités: dans le français tel que je le pratique, il y a, en syllabe finale couverte, les phonèmes vocaliques antérieurs suivant: /i ɛ ɛ· a/ en syllabe finale non couverte /i e ɛ a/ et il serait ridicule d'essayer de réduire ces deux systèmes à un seul: /ɛ ∾ ɛ·/ est un type d'opposition; /e ∾ ɛ/ en est un autre.

On peut fort bien arriver au phonème en partant des «allophones» sans pour cela se refuser de parler de «réalisations» du phonème une fois le phonème bien dégagé.

Dans l'emploi qu'on fait souvent du terme d'«allophone» est inhérent le danger de parler d'un nombre déterminé d'«allophones» pour un phonème alors que chaque réalisation d'un phonème a des chances d'être, d'une façon ou d'autre, différente de toutes les autres. On est, dans la répartition des diverses réalisations entre les «allophones», amené à se laisser guider par des considérations qui n'ont pas de rapport avec la structure de la langue à l'étude («le phonème français /ɔ/ a deux allophones» parce que ses réalisations ressemblent tantôt à celle de /ʌ/ ou celle de /ɔ/ de l'anglais d'Amérique).

Answers *Buyssens:* 1. *B.* agrees that allophones are abstractions.

2. *B.* se déclare d'accord avec M. *Mikuš.*

3. *B.* ne voit aucune objection dans ce qu'a dit M. *Martinet.*

Proc. 5th int. Congr. phon. Sci., Münster 1964, pp. 227–230
(S. Karger, Basel/New York 1965).

The Perceptual Value of Sibilant Transitions

By EDWARD CARNEY, Manchester

Spectrograms of clusters made up of voiceless sibilant + voiceless stop show clearly marked shifts in the latter part of the noise spectrum which seem analogous to the formant inflections of vowels in such a position; they would seem therefore to have potential perceptual value. Sporadic tests in the classroom showed that the perception of an unreleased post-sibilant voiceless stop fluctuated considerably. It seemed therefore worthwhile to test this perceptibility on a good statistical basis. A series of perception tests was devised for this purpose, using tape-cutting, to show the extent to which the following stop might be predicted from the allophonic variation in the sibilant, and to show the contextual variations in this probability.

Material

A series of nonsense words (English non-words) was constructed for the purpose of the test in which the six consonant groups [sp, st, sk, ʃp, ʃt, ʃk] were placed after and varied with the five short vowels [ɪ, ɛ, a, ɒ, ʊ], thus providing 30 different basic combinations. An initial [n] and a final [ə] were added to form the "words" (e.g. naspə, nɔʃkə, nʊstə, etc.). The 30 words were recorded by the author 6 times each in normal, fairly quick, unexaggerated speech, with pauses between words and no catenation features. The pronunciation was standard English except for [a], where a fully open vowel was used. The vowels were chosen to provide a range of allophonic variation in the sibilants. The [n] and the [ə] were added to provide phonetic bulk. Nonsense words were used so as to avoid semantic loading and to ensure statistical symmetry.

Cuts were made during the hold phase of the plosive without including any part of the release and without encroaching on the sibilant. Jockeying the tape past the playback head was too unreliable a method of locating the cut. Instead, preliminary cuts, located by jockeying, were made to fall within the [ə] just after the release, and leader tape inserted. Oscillograms, intensity displays and spectrograms were then made and the place of the final cut located by linear measurement. The resultant monosyllables consisted of [n] + V + [s/ʃ], with the sibilant terminated by the closure phase of the "unreleased" stop.

Testing

Each of the 180 resultant segments was then inserted on a loop and re-recorded on testing tapes four times in succession at 4 second intervals and with 10 seconds between items. The items were thoroughly randomised in the process. There were four separate test tapes, each with 45 items, since this number could be heard at a sitting without fatigue. These four tests were presented to two groups of Swedish listeners: a) a group of 45 language students at the University of Lund (= "Gen. Gp.") and b) a group of 7 qualified and experienced teachers of the deaf from Dövskolan, Lund, plus one postgraduate student of phonetics (= "T.D. Gp."). The phonological restrictions on cluster types in Swedish are not such as to affect the test by linguistic prejudice or interference. The Gen. Gp. heard each test tape twice (in several small sub-groups) – total 16,200 responses; the T.D. Gp. heard each test list three times – total 4,320 responses. Tests were given at weekly intervals. Playback was via a single good loudspeaker in ordinary classrooms.

Since the material was strictly symmetrical and non-meaningful, there was no risk of influencing judgements by explaining the purpose of the tests, and so this was done. Listeners were asked to "predict" the following "unreleased" plosive. They had in effect to identify one of three allophonic types and a choice was compulsory. Absolutely no information about performance results was given until the series was complete.

Results

With three variables one would expect a chance distribution of 33 %. The overall percentage of correct responses was 59.99 % (Gen. Gp. 58.63 %; T.D. Gp. 65.09 %) and is clearly significant. The distribution of both correct and incorrect responses in table I is remarkably even, but table II shows that there is an underlying skewed distribution, with a tendency to favour [t] in –s– contexts, largely at

Table I

Confusion Matrices of Overall Distribution

Gen. Gp.				*T.D. Gp.*			
%	p	t	k	%	p	t	k
p	*58*	23	19	p	*59*	25	16
t	22	*58*	20	t	20	*70*	10
k	24	16	*60*	k	19	15	*66*
	35	32	33		33	37	30

Table II

Confusion Matrices Showing Distribution in –s– and –ʃ– Contexts

Gen. Gp.	*–s– contexts*			*T.D. Gp.*	*–s– contexts*		
%	p	t	k	%	p	t	k
p	*55*	32	13	p	*61*	34	5
t	17	*71*	12	t	19	*77*	4
k	24	18	*58*	k	20	17	*63*
	32	40	28		34	42	24
Gen. Gp.	*–ʃ– contexts*			*T.D. Gp.*	*–ʃ– contexts*		
%	p	t	k	%	p	t	k
p	*61*	14	25	p	*57*	17	26
t	27	*46*	27	t	21	*64*	15
k	25	14	*61*	k	18	13	*69*
	38	24	38		32	31	37

the expense of [k]. The Gen. Gp. shows a corresponding disclination to hear [t] in –ʃ– contexts, though this is not so in the T.D. Gp. This can best be seen in the bottom line of the matrices, which shows the extent to which all responses, correct and incorrect, vary from a 33 % norm. This unevenness may have some connection, though not an immediate one, with the known fact that an abrupt tape-cut made during the course of a fricative gives the auditory impression of the homorganic stop.

The contextual variations in correct prediction do not always form an explicable pattern, but one pattern to emerge is that in –s– contexts the closer the front vowel, the more efficient the transitions.

Table III

Distribution (in %) of Correct Responses in Front Vowel + –s– Contexts

	Gen.	(T.D.)		Gen.	(T.D.)		Gen.	(T.D.)
nɪsp	75	(84)	nɛsp	57	(67)	nasp	48	(56)
nɪst	77	(80)	nɛst	68	(67)	nast	57	(66)
nɪsk	72	(76)	nɛsk	69	(81)	nask	59	(60)

Perhaps the chief interest of the results of the tests is to show that in calculating the perceptual efficiency of a given factor, one should give full and symmetrical allophonic coverage in the material.

The hiss transitions probably play a minor part in the perception of sibilant + stop clusters and may well prove necessary in an adequate synthesis.

Author's address: Mr. Edward Carney, Department of General Linguistics, University of Manchester, *Manchester 13* (England).

Discussion

Penchoen (Paris): Question: Did you try to put in control items without "unexploded stops" at the end? It would be interesting to see if the percentages corresponded with the overall distribution responses of [p, t, k].

Answer *Carney:* The use of a control group in this particular experiment would be difficult. What would the control material consist of? Presumably of sibilants without p–t–k transitions; but the "plain" hiss must terminate somehow. A vertical cut without attenuation would be a quasi-transition to [t]. On the other hand, the final sibilant of an independent syllable would differ markedly in length and intensity from the final sibilants in the segments. A vertical cut could be made with imposed attenuation so as to produce a syllable equivalent in length and other respects to the segments, but its validity as control material would be questionable. Moreover, I doubt the basic validity of giving listeners an unreal choice of this kind. Presumably the result would be a chance distribution of 33%, or some degree of divergence from this. Given either result, I would not know how to interpret it and I do not see how it would constitute a control basis for this experiment.

Proc. 5th int. Congr. phon. Sci., Münster 1964, pp. 231–234
(S. Karger, Basel/New York 1965).

Instituut voor Perceptie Onderzoek, Eindhoven

The Perception of Phonemes as a Function of Acoustic and Distributional Cues

By A. COHEN and A. F. V. VAN KATWIJK, Eindhoven

Do phonemes have any kind of existence to the extent that they can be perceived? One may well argue that phonemes are the outcome of a linguistic operation on language material and may be described in terms of their distinctive character. It is by no means certain that this distinctive character has a perceptual correlate as such in the listener. On the contrary, there is substantial evidence that in speech perception a number of perceptual cues are operative that need not coincide with the distinctive features as postulated by *Jakobson* and *Halle*[1]. In fact, some phonemes can be recognized in isolation on the strength of inherent perceptual cues, such as colour and duration in the case of vowels.

That linguistic elements belonging to the same class of phenomena, in this case phonemes, should show a certain differentiation in the degree of autonomy need cause no surprise. On the morphemic level one generally distinguishes between free and bound forms. A similar observation may be made regarding the meaning of words. Some linguists hold that word meanings can be established only by studying the contexts in which these words are used. But if we ask: what is the meaning of the word "transubstantiation", do we necessarily force our informant to elicit a number of sentences in which the word may be used? On the other hand, there are words, e.g. functional words, which will put an informant in a quandary if he is not allowed to generate examples. We believe that much the same applies to the phonemes of a language and that they may be "known" by the language users either in isolation or in specific contexts.

To substantiate this belief one should try to create conditions in which language users are forced to make use of a type of knowledge that they are not explicitly aware of in normal speech situations.

Just as *Sapir*[2] was able to elicit from his analphabetic native informants the segmentation of whole utterances into single words, so all phoneticians work on the assumption that it is quite feasible to ask subjects, who may be phonetically naive, to respond to questions hinging on the perception of individual phonemes either in isolation or in context. We believe that this perception of individual phonemes is operative whenever a situation is created in which acoustic speech forms are presented divested of meaning. This is the case with PB lists, as used by communications engineers (who may be unaware of the linguistic implications of their work) in order to test the efficiency of a speech transmission channel. These involve e.g. CVC forms that contain segments, representing all the phonemes of a particular language in a way that the actual frequency of occurrence of separate phonemes is faithfully reflected in the total material. The whole burden of identification is therefore carried by the acoustic information contained in the nonsense forms.

So far these examples are largely concerned with the knowledge of phonemes by virtue of acoustic cues, but there are quite a number of data testifying to the existence of some kind of knowledge of the distributional rules of the occurrence of phonemes on the part of naive users of the language. Following *Shannon*[3] one may engage listeners in a guessing game by asking them to establish an utterance as a series of consecutive speech sounds while they are only told whenever they have made the right guess. It turns out that they show themselves well aware of the contextual constraints of their language by the decreasing number of guesses as more context becomes available. Evidence from a few more sources can be adduced to indicate knowledge in native speakers of word constituents devoid of meaning as such, as well as of the distributional rules prevailing among them: the acquisition of the mother tongue by children, where often words that are not acceptable have to be corrected mainly on an acoustic basis; a person solving a crossword puzzle or the occasional versifier on the look-out for rhyme words; evidence from diachronic linguistics and lapses in adult speech. Instances of sound shifts may be observed in everyday usage whenever native speakers are confronted with dialectal speech.

As for the lapses in speaking it may be observed that these can

be perceived either by the speaker, in which case he generally hastens to correct them, or by the listener acting as an observer on the look-out for cues about the general programming strategy applied in normal speech (see e.g. *E. H. Sturtevant*[4]). Such errors may lay themselves open to some kind of systematisation.

In short, the evidence of a language user's knowledge of linguistic elements, such as phonemes, is there for the asking. One may even go further and state that in some respects this knowledge obtrudes as is shown in the case of the obvious difficulties encountered by adult speakers when first confronted with a foreign language. Even though the individual phonemes are known, such as /s/ and /x/ in German, their unusual combination into a sequence such as in Dutch *Scheveningen* has traditionally been used as a highly efficient shibboleth to trap the unwary foreigner.

Now a rough classification can be made of the phonemes of a language as to their perceptibility on an acoustic or combined acoustic and distributional basis. Vowels stand out as basically identifiable on their own strength, with the exception of /ə/, as shown in experiments with synthetic speech. These experiments involved a firm knowledge on the side of the listeners of both spectral and temporal characteristics.

Consonants like /l, m, n, j, w/ which may occur in some cases only in conjunction with vowels are a lot harder to identify by themselves. It takes a tiny fraction of an adjoining vowel to bring about correct identification.

As for such consonants as plosives and fricatives which do not necessarily occur in languages of this type in direct contact with vowels, it seems doubtful if experiments limited to studying these consonants in this context can be considered conclusive.

In conclusion, phonemes may be seen on the one hand as the outcome of a purely linguistic analysis, on the other hand there seems to be ample reason for accepting them as entities, perceptually marked by a number of cues. The latter view implies that speech perception is hard to envisage as a unidimensional process in terms of distinctive features. These features are themselves an outcome of a phoneme analysis and have to be tested empirically before they can be accepted as perceptual cues.

References

1. *Jakobson, R.* and *Halle, M.:* Fundamentals of language; ch. I (Mouton, The Hague 1956).
2. *Sapir, E.:* Language; pp. 33–34 (Harcourt, Brace & Co., New York 1959).
3. *Shannon, C. E.:* Prediction and entropy of printed English. Bell Syst. Techn. J. *30:* 50–64 (1951).
4. *Sturtevant, E. H.:* An introduction to linguistic science; ch. IV, IX (Yale Un. Press, New Haven 1963).

Authors' address: A. Cohen, Technische Hogeschool, *Eindhoven* (The Netherlands).

Proc. 5th int. Congr. phon. Sci., Münster 1964, pp. 235–238
(S. Karger, Basel/New York 1965).

Phonetic Information and Misinformation in 'Dead' Languages

By N. E. Collinge, Durham

It is well known that 'dead' languages can be made to yield indications of their pronunciation. Among the pieces of evidence are foreign transcriptions, orthographic variants, assimilatory and dissimilatory processes, expressive effects (animal cries spelt out etc.) and play on words. There is also commonly extant the direct testimony of contemporary native speakers (writers), of widely varying value. Their prima facie statement can be reliable; their obvious folly can be instructive. The description offered may be cryptic, inadequate, inconsistent, or plausible in itself and yet at variance with inscriptional evidence, with deductions from later developments or with the total phonological pattern of the language. Examples of all these possibilities are easily found in the standard handbooks. The purpose of this paper is to add some detail to the reasons which have been put forward for vagaries of contemporary phonetic analyses and to the signs thought to indicate trustworthiness.

Sturtevant spotlighted three main types of, and reasons for, ancient inaccuracy: a) lack of phonetic training causes inexact apprehension of sounds, b) theoretical considerations cause blindness to visible or audible data, c) adherence to traditional terms causes loss of contact with actual, developed, phenomena. He saw as the most reliable ancient witness an educated and interested speaker with no systemic motive; and as the most credible description one which runs counter to tradition. To this latter recommendation one may add this: it is advisable to prize, as symptoms of clarity and honesty, abrupt changes of mind, homeliness of expression and ingenuousness in description. Cicero rejects the academic question of hiatus and its aesthetic effect by a sudden realization that it applies only to Greek after all, and opines that 'nobody is such a boor' as not to run contiguous Latin vowels together; Terentius Scaurus makes it clear that the middle vowel of *artubus* was not at any rate pronounced [u] by saying 'nobody would be such a fool' as to try it. Equally credibly, Sextus Empiricus underlines his assertion of the monophthongal nature of Greek *ai* in his day not only by his break with tradition but also by contrasting sequences like [ra], with a beginning and an end which do not sound the same, and by using terms like 'homogeneous' – so like the denial of 'sequence' and assertion of 'coalescence' for Skt. *e/o* (< *a–i/a–u*) in *Ṛk-Prātiśākhya*, and the commentator's striking comparison of the diphthong's 'solution' to milk and water (*kṣīrodakavat*).

On the other side, *Sturtevant's* first cause of ancient inaccuracy (failure to hear correctly) may be supplemented by what seem to be cases of simple failure to listen to the relevant phone. Notoriously, Terentianus (and his paraphraser, Victorinus) establishes for Latin a dental [d] and an alveolar [t] – unbelievably and inconsistently with the report of Martianus Capella. Now it is probable that it was the choice of environmental words for Terentianus' personal testing of the sounds which led him to compare and state contrastively what were in effect relatively different allophones from each of these phonemes. The same may apply to the statement that Latin [g] is more 'back' than [c]; Victorinus' use of 'lenius' scarcely supports a fortis/lenis distinction here, although it is not impossible (cf. 'oris molimine nisuque'). There is quite certainly no such indication in the equivalent discussion of [b] and [p], pace *Sturtevant* and those he has convinced: [b] is actually said to have complete lip-closure, [p] to have lip-central release. This analysis may derive its curious differentiation from the inequality of expelled breath as between exponents of these phonemes in comparable environments; but the odd description of [b] is almost certainly the result of misleading personal testing of the sounds in whispered soliloquy, where the voiceless sound remains intact but the 'voiced' to maintain differentiation replaces vocal cord movement by lip-tension. *Sturtevant's* second diagnosis (that theory precludes observational precision) applies to Priscian's treatment of Latin *f* as an aspirated stop or to Varro's denial of the graphic status of Latin *h*. But theory may also occasion wilful distortion of phonetic fact, as when Pāṇini equates Skt. [a] and [a:] in aperture quality, implicitly admitting the fiction in his final aphorism. The third point (that tradition blinds the analyst to development in the phenomena, as in statements on the rough breathing in late Greek) may be enlarged by recognition of the effect of archaisms of orthography, which is often inaccurate even when first applied. Aside from the words of ancient scholars who are thinking about music and of modern scholars who are theorizing, there is no compelling evidence that the Greek acute accent involved a necessitous *rise* (except in the loose sense that a sequence of lower point–higher point is so called, like a 'rise' in price). Restriction on accent-recession suggests that there was a necessitated true *fall* from high pitch in post-acute morae, lasting beyond the immediate syllable-boundary; but a true preceding rising movement on pre-tonic syllables is ruled out by e.g. the succession of grave (i.e. low) markers on all pre-acute syllables in one Greek system (cf. anudāttas preceding the first udātta in Vedic verse notation), and rising motion at least on the commonly short acute syllable itself is far less likely than that the voice hits a simple high point. Fall within a syllable needs 'double length' (two morae) if the syllable contains effective accent (circumflex); rise may be credited to long syllables likewise (long acute), but therefore not to short. 'Fall' as the meaning of final grave sign is accepted by few nowadays, which means that scholars are not swayed by the diacritic ` despite the value of the second half of ^; but still the shape of ´ is over-persuasive. Probably ´ has misled analysts from ancient times, although the sign is indistinguishable from a common sign for place of stress, and the latter implies no 'contour' at all. And how many of us still write the stress sign from the top down and the acute from the bottom up?

Sporadic inscriptional testimony of lengthened [i] in Latin before the cluster *–gn–* seems to support Priscian's general statement about vowels in that environment. Now spelling (*ignosco*, not **ingnosco; sinnu* etc.), sound-shift conditions (*e* > *i* in *dignus* as in *lingua*), and perhaps grammarians' readiness to hear agma elsewhere, all show a dorsal nasal as the first consonant in this sequence; *ign–* represents [iŋn–]. Introspective ancients may have misheard this as [i:n–] – if ŋ in fact gave way to a nasalization, [ĩn] or [ĩ:n], one prosody may have ousted the other – and this, the first of *Sturtevant's* 'faults', was aided by the third, for orthography held them fast, and *ign–* is spoken of as if it were [i:gn–] and even written *Ign–*. Romance reflexes show a return to a spelling pronunciation [ign–]. They also show no loss of the /n/ phoneme as such (any more than does Latin orthography itself) in the sequence [–Vnf–], whereas earlier [–Vns–]

has passed to [–Vs–]. Already in Latin spellings like *cosol*, *cesor* are frequent. Then why does Cicero say that in *both* environments the phonetic value is [V:ns/f–]? Epigraphic testimony often supports him, and other analysts generalize his ruling beyond the mere *ins–*, *inf–*, *cons–*, *conf–* of which he speaks. It is hard, too, to prove or disprove vowel lengthening before *–nf–* from the Romance results. But co-existence of secondary vowel length and full nasal phoneme is unparalleled in Latin, for barring analogy such length compensates for loss of weight in syllables affected by reduction of consonant clusters. Further, the assignment in quantitative verse of syllabic length to the sequence *–Vm C–* may reflect an addition of duration-prosody to nasalization-prosody to maintain weight in the absence of oral release of *m* ([–V̰:(#) C–]); but one is not justified in combining traditional spelling and (partial) phonetic analysis, and describing the result as [–V:m(#) C–]. In fact, there is no convincing proof of [–V:ns/f–] in Latin. It looks as if in *inf–*, *cons–*, *iunx–*, *iunct–* mishearing and traditional spelling (the *cesor/censor* variation really shows only the instability of /n/ before [s]) may have induced the stating of false co-existences and the equating of environments whose effects were diverse.

Postscript

A relevant point made in subsequent discussion was the tendency of nasalized vowels to display measurably longer duration than non-nasalized equivalents. This affects assessment of Cicero's report (apart from his equation of *–ns–* and *–nf–*, not justified by the balance of orthographic evidence). He may be credited with some phonetic perspicacity; but with accuracy only if the following grouping of phenomena can be paralleled: (1) vowel length associated with an apparent feature of prenasalization before a nasal phoneme, plus (2) limitation of this feature to strings defined not phonologically but morphologically (some lexical derivatives), plus (3) subsequent loss of (vowel length and) nasality, but not of the nasalphoneme, in one sub-group alongside loss of the nasal phoneme and prenasalization, but not of vowel length, in another.

Author's address: Dr. N. E. Collinge, University of Durham, 33 Western Hill, *Durham* (England).

Proc. 5th int. Congr. phon. Sci., Münster 1964, pp. 239–243
(S. Karger, Basel/New York 1965).

The Lisp of Children, its Cause, its Reason, its Meaning

By P. H. Damsté, Utrecht

When a child has his stock of speech sounds nearly completed, there often remain – for quite a long time – defects in the production of the s-sound. Mindful of the auditory control on the articulation, we may think a missing link in the afferent branch of the feed-back circuit to be the cause. A high tone-deafness would be a sufficient explanation of sigmatism, but it is very seldom found. In Dutch there is no phonemic opposition between one s and another; but in other languages the confusion belonging to an early phase of language development may well be at the root of sigmatism[7]. A further theoretical possibility is abnormality of the perceptive area: a weak representation of the s-norm or insufficiency of discriminative power; they are difficult to prove. One thing is certain, that a constant difference between the s perceived from others and from oneself tends to get so familiar, that the difference no longer has any corrective effect, even tends to remain unnoticed by the speaker. This accounts for the fact that a peculiar way to produce an s may persist long after the factor which started this articulation habit has ceased to operate*.

We shall have to explain how the abnormal s-sound came into existence before it has become a persistent habit. We will confine ourselves to the interdental lisp that is by far the most frequent. We do not find hearing losses in these infants, but we do find that many of them have an anterior open bite (a gap between the front incisors).

* I remember a 12 year-old boy with a most remarkable velar s-sound. He had acquired this during a temporary velar insufficiency after adenoidectomy, and nobody (including the patient) had made an attempt to correct it after that. It was cured in three weeks exercising.

To take this for the cause of interdental articulation habits would be a too hasty conclusion. Granted that gross irregularities of the dental arches may predispose for speech defects[8], it is surprising how important abnormalities in the mouth after a while can be met with a little change in the action of the tongue[4]. The adaptation to a dental plate may serve as an example. If this holds true for an aged subject, how much easier is it for a child.

In many cases of open bite we observe a reversed chain of cause and effect. The dental irregularity is not the cause for the tongue to leave its normal point of articulation. On the contrary, it is the forward thrust of the tongue between the teeth during swallowing that checks the vertical outgrowth of the alveolar processes, resulting in an open bite. I do not pretend to bring something new: in recent years there have appeared a great many articles about this subject in the orthodontic and speech pathology literature[1,2,3,5]. The reliability of this view has been proved by the experience that many cases of open bite have been cured with only tongue- and swallowing exercices, without the use of orthodontic appliances (see Fig. 1).

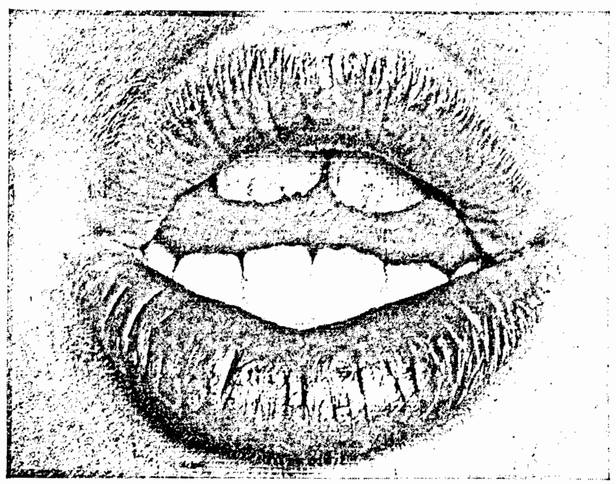


Fig. 1. Open bite associated with tongue thrust in swallowing and interdental sigmatism.

Changing the tongue position in swallowing sufficed to restore the normal outgrowth of the alveolar processes, and close the bite; moreover the lisping disappeared. It should be noted however, that also in the presence of a vertical open bite the possibility to produce a normal s-sound is already there, but is not used. Inter-

dentalism and open bite are both caused by an abnormal motor behaviour of the tongue. So there is a brother and sister relation between the two:

COMMON CAUSE
(infantile motor
pattern of the tongue)

TONGUE THRUST IN SWALLOWING (anterior open bite)	LOW TONGUE POSITION IN ARTICULATION (multiple interdentalism)

Implications for Phonetics

The structures for articulation do only part-time work for speech. The rest of the time they have to eat, chew, drink and swallow.

In an investigation of speech-characteristics it is sensible to take into consideration these elementary ("animal") functions of the organs for speech. They are phylogenetically much older, and of vital importance for the organism. A person with a paralysis of the palate is clearly not able to realize his speech-intentions. With our lisping children, however, matters are different. They have a primarily functional defect. The normal auditory controlling mechanism for articulation is being overruled by a deviated "animal" function of the organs for articulation. From a close analysis of the "deviation" we learn the following.

The movements of the tongue in these children, when they speak, drink or swallow, bear a reminiscence of sucking. They are a remains of the motor behaviour of the suckling, persisting long after this phase is over. The motor pattern of the tongue is still controlled by the tactile sensations of the tongue and the mucosal lining of the mouth. The desired release of pleasure cannot disengage itself from the infantile pattern. (If nursing habits are to be blamed for this does not matter here.)

For the phonetician the question is relevant if such a tongue habit (like thumb sucking it is an atavism from the nursery) has no meaning at all in connection with speech or if it has significance as subconscious but intentional expressive behaviour[9]. Lisping is well known as something belonging to the small child. It is conceivable that a child when growing up would take advantage of this and stick to its old pattern to deceive his surroundings about his age, dodge the responsibilities of growing up and invite pampering. The psychology of animals and children shows examples of a similar conduct.

However we are not inclined to accept this as the main aspect of lisping for the simple reason that the infantile swallow (with tongue-thrust) does not fit into the picture of expressive behaviour. We feel that the syndrome of infantile swallowing and interdental sigmatism should be labeled a pleasure-determined physiologic act with some sign-value for other people. It may be compared to our gait or handwriting that are equally difficult to change voluntarily.

If we may trust a popular saying: "girl that lisps, good to kiss" we find there to our surprise a confirmation of the reversed hypothesis that oral-tactile pleasure sensation goes together with the childish s-sound.

In the end when all these different aspects of an apparently simple phenomenon fall into their place, they lose their complexity. We will see them once more apart, and then realise how they fit together.

1. Sigmatism cannot be explained from its acoustical or audiological aspect only.

2. The phonetical explanation of this phenomenon has to reckon with the functioning of the tongue in swallowing; the habit of suck-swallowing may prevent a normal dental development.

3. It appears that interdental sigmatism is a mode of expression suiting an organism whose tongue-motoricity functions on an infantile level.

4. The failure to conform to the s-norm that has been discussed, is the result of contamination of the articulation pattern with an archaïc tongue-function. The latter has no communicative purpose, but by influencing the articulation it has an unintended expressive value.

It is not possible to draw an exact line between the two: any motor body function may be seen from a communicative aspect.

References

1. *Cauhépé, J.; Fieux, J.* et *Bouvet, J. M.:* Déglutition et troubles d'occlusion. Rev. Stomat. *54:* 905–912 (1953).
2. *Elffers, P. G.:* Infantiele slikgewoonte. Logoped. Foniat. *33*/9 (1961); *34*/3 (1962).
3. *Francis, F. R.:* The articulation of English speech sounds in arterior open bite accompanied by tongue thrusting. Speech Path. Therap. *1:* 70–72 (1958); *3:* 8–26 (1960).

4. *Heese, G.:* Zur Problematik der mechanisch bedingten Dyslalien. Z. Laryng. Rhinol. Otol. *37:* 622–625 (1958).
5. *Jann, H. W.:* Tongue thrusting as a frequent unrecognized cause of malocclusion and speech defects. N. Y. State dent. J. *26*/2 (1960).
6. *Kainz, F.:* Die «Sprache» der Tiere (F. Enke, Stuttgart 1961).
7. *Krech, H.:* Die Behandlung gestörter s-Laute (VEB M. Niemeyer, Halle 1955).
8. *Reichenbach, E.* und *Meinhold, G.:* Neuere Beobachtungen und Untersuchungen über orale organische Sigmatismen in Zusammenhang mit Zahnstellungs- und Kiefer-anomalien. Fortschr. Kieferorthopäd. *24*/1 (1963).
9. *Trojan, F.:* Die Ausdruckstheorie der Sprechstimme. Phonetica *4:* 121–150 (1959).

Author's address: Dr. P. H. Damsté, E.N.T. Clinic, University of Utrecht, Catharijnesingel 101, *Utrecht* (Holland).

C.R. 5e Congr. int. Sci. phon., Münster 1964, pp. 244–251
(S. Karger, Basel/New York 1965).

Université de Californie et Laboratoires Haskins

De la hiérarchie des indices acoustiques pour la perception de la parole

Par Pierre Delattre, Santa Barbara, Calif.

La connaissance des indices acoustiques dont se composent les éléments distinctifs d'une langue est maintenant assez avancée pour permettre de faire le point sur le problème général du pouvoir distinctif de chacun de ces indices. La relativité de ce pouvoir a été observée indirectement ou partiellement dans nombre de travaux, faits par des techniques variées. Il reste ici à voir comment l'examiner systématiquement.

C'est par le procédé de coupe et recollage de bandes magnétiques sur lesquelles était enregistrée de la parole humaine qu'*André Malécot* a étudié les résonances des consonnes nasales pendant l'occlusion buccale (ce que nous appèlerons ici les *résonances fermées*) pour déterminer leur rôle dans la distinction du lieu d'articulation [1]. Par ce même procédé, *Katherine Harris* a comparé entre eux les bruits des consonnes fricatives pour leur contribution à l'identification du lieu d'articulation [2]. *Lee Lisker* a fait varier la durée de l'occlusion buccale pour découvrir son rôle dans le voisement des consonnes occlusives [3]; *Frederick Householder* [4], *Morris Halle* [5], et d'autres ont comparé la perception des occlusives finales avec et sans leur détente explosive. Mais le procédé par coupe de bandes magnétiques est limité de deux manières. D'abord il ne permet que l'omission et l'addition, il ne permet pas la modification d'un indice acoustique. Ensuite il ne permet la segmentation que dans le temps et non parallèlement au temps. Pour omettre, modifier, manipuler, isoler même, les indices acoustiques de la parole parallèlement au temps, il faut opérer par synthèse de parole artificielle. C'est donc surtout au moyen de la parole synthétique que la contribution de chaque in-

dice à la perception des lieux et modes d'articulation peut être révélée. Ici la technique de recherche la plus élaborée consiste à faire varier un seul indice dans le temps, la fréquence et l'intensité pendant que les autres indices sont «neutralisés». Ainsi les chercheurs des Laboratoires Haskins ont pu étudier pour l'anglais les explosions des occlusives orales [6], les transitions des occlusives orales et nasales [7], les résonances fermées et les transitions des sonantes [8], les transitions et les bruits des fricatives [9], les formants des voyelles [10], les effets de tempo [11], etc.

Par suite de ces imposants travaux, de nombreux faits de hiérarchie nous sont connus. On sait par exemple que, pour la perception du lieu d'articulation anglais, le bruit contribue plus dans les consonnes sourdes que dans les sonores, et que par contre les transitions jouent un rôle plus important pour les sonores que pour les sourdes. Mais ces bruits ont des valeurs distinctives très différentes: les bruits de friction sont plus efficaces que les bruits d'explosion; parmi les frictions, les plus utiles sont les plus intenses, celles de /ʃ/, /s/; quand les frictions sont faibles, comme pour /θ/, /f/, le lieu d'articulation est surtout connu par les transitions. Parmi les explosions, les plus utiles à la perception sont celles des vélaires devant voyelles arrondies – ainsi dans /go/, c'est principalement le bruit qui identifie le lieu d'articulation vélaire, dans /bo/ c'est principalement le deuxième formant. On sait que la troisième transition joue un plus grand rôle dans les *r* que dans les autres sonantes, dans les dentales que dans les labiales ou les vélaires. On sait que la deuxième transition identifie mieux le lieu d'articulation vélaire devant voyelle écartée que devant voyelle arrondie. On sait que les consonnes nasales font connaître leur lieu d'articulation par les transitions beaucoup plus que par les résonances fermées. On sait que le tempo des transitions aide à distinguer /j/ de /g/, /ð/ de /v/, mais pas /z/ de /ʒ/. Et ainsi de suite. Les indications isolées du pouvoir relatif des indices acoustiques abondent; il faut maintenant y mettre de l'ordre.

Pour essayer d'atteindre une vue d'ensemble de cette hiérarchie, nous dressons ici des tableaux où tous les indices connus entrent en jeu pour distinguer l'un de l'autre les deux termes de toutes les oppositions de phonèmes, de lieux d'articulation, de mode d'articulation et de voisement d'une langue donnée. Ces tableaux (voir figures) se présentent sur différents plans. La place nous manque pour offrir plus d'un ou deux échantillons de chaque plan. Le pouvoir distinctif d'un indice donné pour une opposition donnée est noté par des x dans la case appropriée. La notation est arbitrairement en cinq degrés, de zéro x à quatre x. Les jugements de pouvoir distinctif sont en grande partie basés sur des travaux publiés, mais aussi sur les recherches exploratoires de l'auteur. Les exemples de consonnes sont pris à l'anglais, sauf que les affriquées sont omises à cause de l'ambiguité d'unité qu'elles présentent. Il reste donc quatre modes: les plosives, les fricatives, les nasales et les sonantes, ce dernier mode comprenant les liquides et les semi-voyelles. Les exemples de voyelles sont pris au français de façon à inclure les indices de nasalité.

Sur le premier plan nous opposons successivement chaque mode d'articulation à tous les autres modes (figure 1). Pour quatre modes

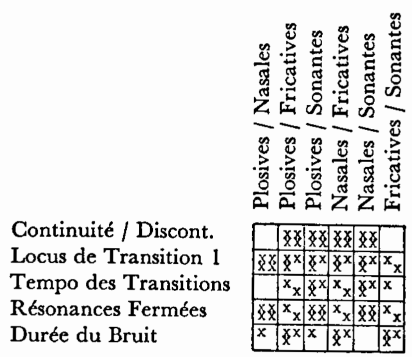


Fig. 1. Pouvoir distinctif des indices acoustiques de *mode d'articulation* en anglais. Comparaison de quatre modes entre eux, formant six paires d'opposition.

cela fait six oppositions. On peut lire, par exemple, sur la ligne horizontale du milieu, que le tempo des transitions sert plus à la distinction occlusives/sonantes qu'aux autres distinctions (occlusives = plosives + nasales); et sur la première colonne verticale que les plosives se distinguent surtout des nasales par la résonance fermée et la première transition.

Sur le deuxième plan, nous opposons les lieux d'articulation entre eux, à l'intérieur de chaque mode, séparément (figure 2). On est obligé de séparer les vélaires devant voyelles écartées (ge, ŋe) des vélaires devant voyelles arrondies (ga, ŋa), ces dernières n'utilisant pas de locus vélaire de deuxième formant pour la perception du lieu d'articulation vélaire. Plusieurs expériences démontrent cela. La plus simple consiste à retrancher l'explosion à un enregistrement de /go/. On entend alors /bo/, parce que la deuxième transition de /go/

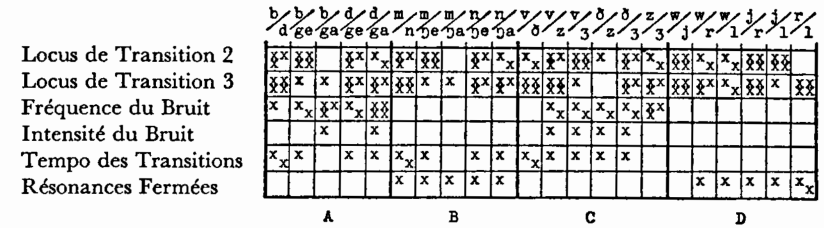


Fig. 2. Pouvoir distinctif des indices acoustiques de *lieu d'articulation* en anglais. A. Opposition des consonnes occlusives sonores entre elles. B. Opposition des consonnes nasales entre elles. C. Opposition des consonnes fricatives entre elles. D. Opposition des consonnes sonantes entre elles. ge = g devant voyelles écartées. ga = g devant voyelles arrondies.

se dirige vers le locus labial. Cela démontre aussi que l'explosion de /go/ joue un rôle prépondérant dans la perception du lieu d'articulation vélaire, ce que nous trouvons noté sur la figure 3, ligne 3, sons /ka/ et /ga/.

Sur le troisième plan, nous opposons une seule consonne, /m/, à toutes les autres (figure 3). Cela met en lumière les nombreux indices qui entrent en jeu en anglais: 6 pour le lieu d'articulation (en français, allemand, espagnol, il faudrait en ajouter un 7^e^, le locus de transition 1, ces langues ayant des consonnes pharyngales à locus 1 très élevé), 5 pour le mode, et 7 pour le voisement (d'après une investigation en cours).

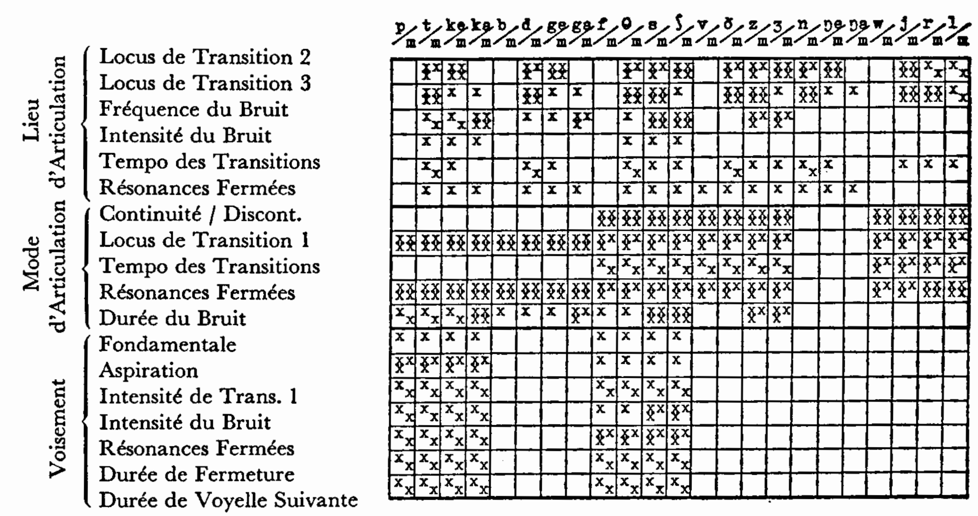


Fig. 3. Pouvoir distinctif des indices acoustiques de *lieu d'articulation*, de *mode d'articulation*, et de *voisement* en anglais. Opposition d'une seule consonne à toutes les autres.

Sur le troisième plan, nous opposons également une voyelle, la nasale française /ɛ̃/, à toutes les autres voyelles du français (figure 4).

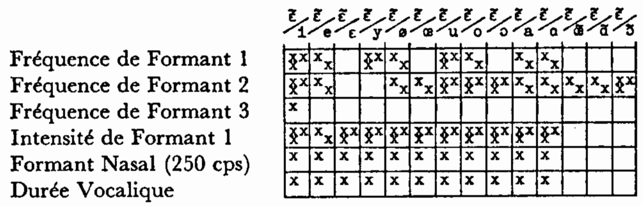


Fig. 4. Pouvoir distinctif de six indices acoustiques des *voyelles* françaises. Opposition d'une seule voyelle à toutes les autres.

Le premier formant ressort de là comme très important, non qu'il soit plus fort que le deuxième mais parce qu'il joue un double rôle, tant par ses variations de fréquence dans la détermination de la couleur que par ses variations d'intensité dans la perception de la nasalité.

Sur le quatrième plan, nous opposons toutes les consonnes entre elles pour un seul indice acoustique à la fois. Ici nous avons choisi l'indice du *bruit* (figure 5) et l'indice de la *deuxième transition* (figure 6).

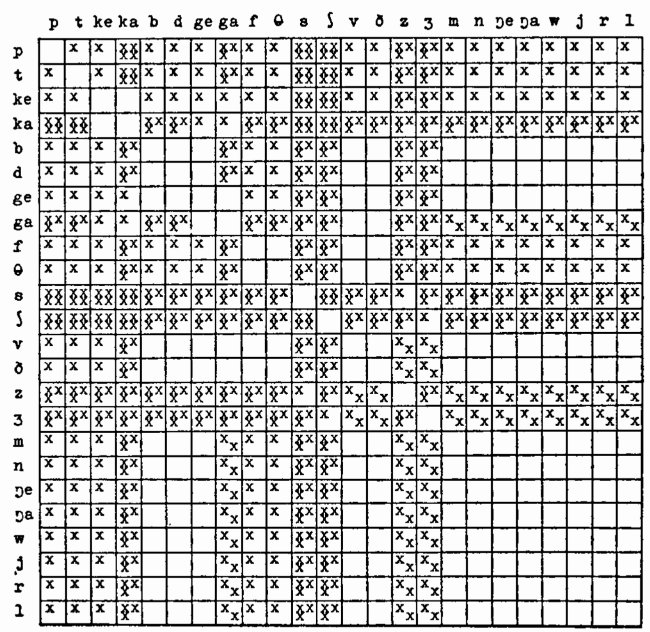


Fig. 5. Pouvoir distinctif d'un seul indice acoustique, celui du *bruit* (explosion ou friction) – indice de mode et de lieu – pour toutes les oppositions de deux consonnes en anglais.

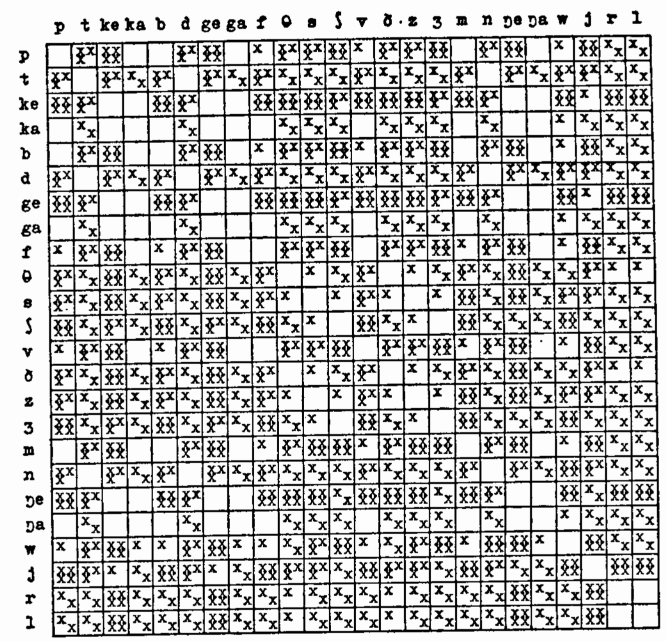


Fig. 6. Pouvoir distinctif d'un seul indice acoustique, celui de la *transition du deuxième formant* – indice de lieu et de mode – pour toutes les oppositions de deux consonnes en anglais.

La comparaison de ces deux tableaux ne laisse pas de doute sur le pouvoir distinctif plus vaste de la deuxième transition. Le pouvoir distinctif des bruits n'est grand que pour quatre fricatives: /ʃ/, /s/, /ʒ/, /z/, et deux explosives: /g/ et /k/ devant voyelles arrondies.

Sur le quatrième plan, nous opposons encore toutes les *voyelles* françaises entre elles pour un seul indice acoustique à la fois. Ici nous avons choisi la fréquence du premier formant (figure 7) et du deuxième formant (figure 8). De la comparaison de ces deux tableaux, il ressort que le pouvoir du deuxième formant (fréquence) est légèrement plus étendu que celui du premier formant (fréquence), à cause des voyelles nasales qui ne se distinguent entre elles que par la fréquence du deuxième formant, et de la coïncidence moins fréquente des deuxièmes formants que des premiers.

Concluons. Le problème de la hiérarchie des indices acoustiques est vaste. Nous n'avons pu que le parcourir à vol d'oiseau en montrant quels genres de tableaux il faut dresser pour s'y reconnaître. Ces tableaux sont sur quatre plans. Pour compléter la présente ébauche sur le troisième plan, il faudrait autant de tableaux qu'il y a de consonnes et de voyelles dans la langue analysée. Sur le qua-

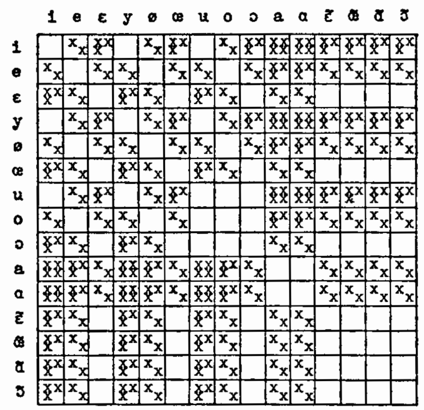


Fig. 7. Pouvoir distinctif d'un seul indice acoustique, la *fréquence du premier formant*, pour toutes les oppositions de voyelles en français.

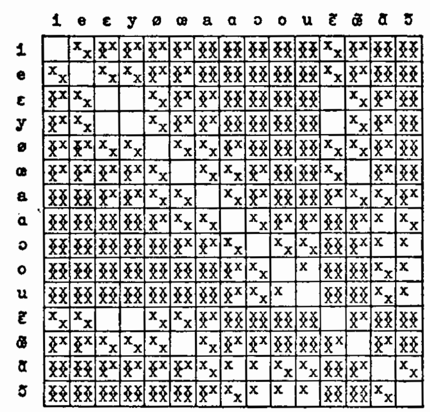


Fig. 8. Pouvoir distinctif d'un seul indice acoustique, la *fréquence du deuxième formant*, pour toutes les oppositions de voyelles du français.

trième plan, il en faudrait autant qu'il existe d'indices acoustiques séparables. Nous n'avons d'ailleurs examiné ici que les oppositions de phonèmes segmentaux. Le travail complet comprendrait également les indices acoustiques des phonèmes prosodiques d'accent, d'intonation et de jonction, ainsi que ceux de la dichotomie consonne/voyelle qu'elle soit phonémique ou non.

Enfin il sera intéressant, face aux tableaux complets, de grouper les indices en familles – de comparer, par exemple, les indices de fréquence avec ceux d'intensité, de durée et de tempo; ou encore de les comparer d'après leurs corrélatifs articulatoires.

Bibliographie

1. *Malécot, A.:* Acoustic cues for nasal consonants; an experimental study involving a tape-splicing technique. Language *32:* 274–284 (1956).
2. *Harris, Katherine:* Cues for the discrimination of American English fricatives in spoken syllables. Language and Speech *1:* 1–7 (1958).
3. *Lisker, L.:* Closure duration and the intervocalic voiced-voiceless distinction in English. Language *33:* 42–49 (1957).
4. *Householder, F.:* Unreleased ptk in American English. For *R. Jakobson*, pp. 235–244 (1956).
5. *Halle, M.; Hughes, G.* and *Radley, J.-P.:* Acoustic properties of stop consonants. J. Acoust. Soc. America *29:* 107–116 (1957).
   Delattre, P.: Unreleased velar plosives after back-rounded vowels. JAS 30, *6:* 581–582 (1958).
6. *Liberman, A.; Delattre, P.* and *Cooper, F. S.:* The role of selected stimulus-variables in the perception of the unvoiced stop consonants. Amer. J. Psychol. *65:* 497–517 (1952).
7. *Delattre, P.; Liberman, A. M.* and *Cooper, F. S.:* Acoustic loci and transitional cues for consonants. JAS *27:* 769–774 (1955).
   Harris, K.; Hoffman, H.; Liberman, A.; Delattre, P. and *Cooper, F.:* Effect of third-formant transitions on the perception of the voiced stop consonants. JAS 30, *2:* 122–126 (1958).
   Hoffman, H. S.: Study of some cues in the perception of voiced stop consonants. JAS 30, *11:* 1035–1041 (1958).
   Liberman, A.; Delattre, P.; Cooper, F. and *Gerstman, L.:* The role of consonant-vowel transitions in the perception of the stop and nasal consonants. Psychol. Monogr. *379:* 1–14 (1954).
8. *O'Connor, J.; Gerstman, L.; Liberman, A.; Delattre, P.* and *Cooper, F.:* Acoustic cues for the perception of initial /wjrl/ in English. Word *13:* 24–44 (1957).
9. *Delattre, P.; Liberman, A.* and *Cooper, F.:* Formant transitions and loci as acoustic correlates of place of articulation in American fricatives. Studia Linguistica (1963).
10. *Delattre, P.; Liberman, A.; Cooper, F.* and *Gerstman, L.:* An experimental study of the acoustic determinants of vowel color; observations on one- and two-formant vowels synthesized from spectrographic patterns. Word *8:* 195–211 (1952).
11. *Liberman, A.; Delattre, P.; Gerstman, L.* and *Cooper, F.:* Tempo of frequency change as a cue for distinguishing classes of speech sounds. J. Exp. Psychol. *52:* 127–138 (1956).

Adresse de l'auteur: Prof. Dr Pierre Delattre. University of California, Department of French, *Santa Barbara, California* (USA).

Proc. 5th int. Congr. phon. Sci., Münster 1964, pp. 252–258
(S. Karger, Basel/New York 1965).

Bell Telephone Laboratories

On the Motor Theory of Speech Perception

By Peter B. Denes, Murray Hill, N.J.

Our understanding of how we perceive speech has, in many ways, decreased rather than increased over the years. The better understanding of vocal tract acoustics and advances in instrumentation in recent times, seem to have uncovered as many new problems about the process of speech recognition as they have solved. Years ago, speech recognition was thought to be a simple process in which distinct and unique acoustic-auditory features are interpreted as specific phonemes. Modern technology enabled us to check these theories by incorporating them into models of the speech recognition process. The first models failed to recognize speech successfully, showing that our theories were inadequate. As the years passed our ideas about speech recognition – and the models built to implement them – became more and more sophisticated. Yet, the results are still unsatisfactory. As matters stand today, we have, on the one hand, the human being, who can recognize speech with ease even under conditions of severe noise and distortion, and on the other hand, models of the speech recognition process which take into account all we know about human speech recognition and yet are able to deal only with fewer than a dozen or so words and only when spoken in isolation rather than connected text. It is only natural therefore that new and promising theories should be continuously proposed about various aspects of speech perception. One of these is the *motor* theory of speech perception. The motor theory proposes that, during speech recognition, we do not directly associate the sound qualities we perceive with linguistic units, the phonemes, words, etc., but that, instead, we first interpret our auditory percepts in terms of the articulatory movements needed to produce these sounds and, in a second stage, we recognize the language units by

association with these articulatory movements. A corollary of this theory is that an essential part of the process of learning to *recognize* speech is training in *producing* speech ourselves. The purpose of the experiment to be described in this paper was to observe how far being able to listen to our own voice and thereby getting a chance of associating the articulatory movements we make with the sounds produced by these movements makes learning to recognize speech easier. In this way it was hoped to learn more about the motor theory of speech perception.

In order to carry out the experiment, a method was needed for making naturally produced speech sounds sufficiently unlike normal speech that it could not be recognized without learning. At the same time, the sounds had to retain enough information about articulatory movement to make them learnable. This was achieved by modifying an eleven channel vocoder to compress the spectrum of speech in the ratio of roughly 3 to 1. The system is explained in Fig. 1. The 180 cps to 4500 cps wide speech spectrum is split into eleven

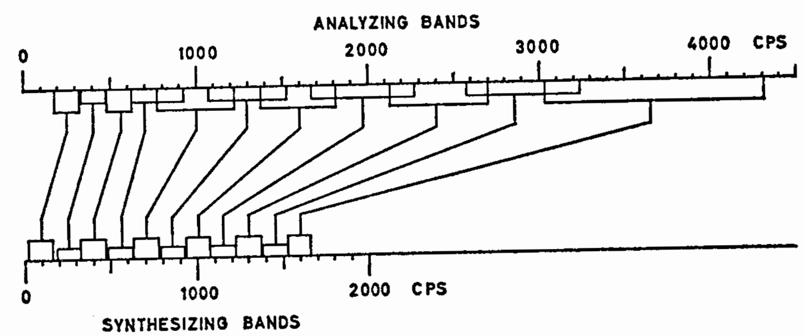


Fig. 1. Transposition of speech spectrum, as used in this experiment.

bands; the energy from each band controls the output of eleven synthesizing channels which, between them cover the frequency band of 50 cps to 1600 cps. The pitch of the speech input is also monitored and is used to control the frequency of the buzz source applied to the synthesizing filters. Again the pitch of the output, compared with that of the original speech input, is reduced in the ratio of about 3 to 1. The pitch of the output was reduced in order to have sufficient density of harmonics to excite the narrower band synthesizing filters. It will be noted that there are four analyzing filters to cover the first formant range and four to cover the second formant range. The first formant filters are about 150 cps wide and the second formant filters about 450 cps wide, whilst the three highest filters are even wider. The speech input could be obtained either from a prerecorded magnetic tape or from a microphone. The output could be heard over earphones.

Ten senior high school students, five men and five women, were used in the experiment. Each session consisted of a twenty minute training period followed by a five minute test. The speech material used was taken from a library of 150 words. The words were pronounced in isolation, by one speaker, prerecorded on magnetic tape and rerandomized every time they were used. During the training period, the subjects listened to the processed words over earphones and, at the same time, they had a printed list of the words in front of them. In this way they could associate the sounds they heard with the words they represented. During the test period, a different list of words was used, the subjects still heard the processed speech, but they did not have the printed list and instead had to write down the words they could recognize.

The subjects were divided into two groups: the so-called "Listeners" and the "Speakers". During the training period the "listeners" could learn solely by associating the processed sounds of the one, prerecorded speaker with the words on their printed list. The "speakers", on the other hand, in addition, had a microphone. They first listened to the prerecorded voice and then had to repeat the word into their microphone. Their voice was processed by the identical device as the voice of the prerecorded speaker and they heard the processed version of their own voice over their earphones. The sounds from the earphones had a sufficiently high level to ensure that they masked any unprocessed version of their own voice which may have reached their ears directly, by either air or bone conduction. In this way they could learn to relate 1. the words shown on the printed list, 2. the associated articulatory movements they themselves made when they pronounced these words, and, 3. the processed sounds which they heard as a result of their articulatory movements. The five minute test which followed each training period was of course the same for "speakers" and for "listeners". The test lists consisted of 50 randomized words and the proportion of words recognized correctly was taken as a measure of the learning achieved. A comparison of the learning progress of the two groups was considered as an indication of the value, in recognizing speech,

of the articulatory associations available to the "speakers" only and thereby furnish an indication of the validity of the motor theory. The learning progress observed in 15 consecutive training periods by "speakers" and by "listeners" is shown in Fig. 2. No clear-cut

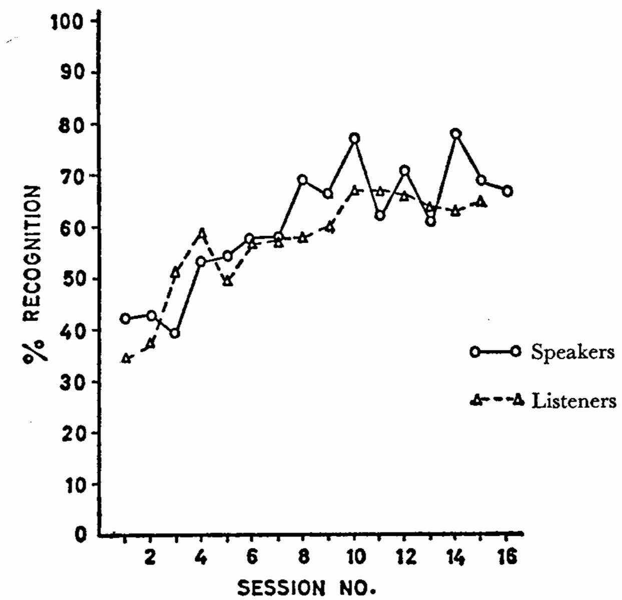


Fig. 2. Learning progress.

difference between the two groups is evident: both curves rise from roughly a 40% word recognition score to about 70%. The results therefore *do not* support the motor theory of speech perception. Just the same, a comparison of the speakers' and of the listeners' learning curves show certain dissimilarities that *may* indicate that the speakers' learning progress *was* influenced by the sound of their own articulations. There are, for example, the marked oscillations in the speakers' curve as compared with the relative smoothness of the listeners' curve. This may be the effect of the more variable sounds heard by the "speakers" as compared with the "listeners". The listeners always heard the same recording of only one speaker, whilst the speakers may well have varied their pronounciation as their learning progressed, producing a feed back type oscillation in their results. Also – perish the thought – some equipment failure cannot be disregarded. The test equipment was basically a vocoder and vocoders' pitch detectors are notorious for making errors under certain combinations of speaker's voice, vowel quality and voice pitch. Whilst this could be controlled for the one prerecorded voice, pitch detector errors were observed for some of the other voices.

Further analysis of the recognition scores showed that almost three times as many errors were due to place-of-articulation than to manner-of-articulation confusions. This, of course, is quite consistent with the very rough quantization of the spectrum by the analyzing filters: the second formant region – so important for identifying the place of articulation of consonants – was divided into just four bands, each about 450 cps wide, so that only the really strong formant transitions were at all noticeable in the spectrum of the output. It may, in fact, have been more reasonable to have used a *formant* vocoder based system rather than the channel vocoder actually used. In this way, a much finer grain formant search at the analyzing end, would have allowed the compression of more formant information at the synthesizing end into the same narrow band as was used in the present experiment. By hindsight, this approach would have produced test signals that were still unlike normal speech yet richer in distinctive information about articulatory activity, making the learning of articulatory – auditory associations easier.

Was the output of the present system, in fact, distinctive enough to be learnable? On re-examination, the data showed definite evidence of the learnability of our nonspeechlike test sounds. The recognition scores obtained for certain key words after the first three learning periods were compared with the scores for the same words after the 9th, 10th and 11th learning periods. The confusions in recognizing word pairs such as for example *mirth* and *nurse* were examined. It was found that whilst at the start of the learning period confusions, were as high as 50% or more, by the end of the 11th learning period they were down to a very small value. And so on, for numerous other examples. The evident learnability of these processed "speech" signals is important not only theoretically but it also furnishes a guide to the promise of certain frequency-compression hearing aids. Unfortunately we have no time to discuss these hearing aids and the associated problems of relearning in this paper.

In conclusion then, it can be said that the tests have produced no firm evidence to support the motor theory of speech perception. The results have, however, shown some differences in the learning behavior of "listeners" and of "speakers", indicating that longer learning periods or perhaps more learnable signals might confirm

the motor theory after all. The results have also shown that some, at least, of the frequency transposed signals *were* learnable: an interesting result in itself. And finally, for those of you, who like myself, like the motor theory – because of the supporting evidence of psycho-acoustic results from Haskins and because of the unifying simplicity it promises for much of what we know about speech perception – there is a further point which I should perhaps have mentioned at the beginning of my talk: perhaps the kind of speech learning performed by adults – such as the subjects in my experiment – is different from that which takes place when a child learns speech.

Author's address: Dr. Peter B. Denes, Bell Telephone Laboratories Inc., *Murray Hill, N.J.* (USA)

Discussion

Jassem (Poznań): 1. I am not sure that the attempt has been made to construct a mechanical speech recognizer which possesses all the knowledge of language that has been collected by linguists.

2. Although Mr. *Denes* is not committing himself on this point, his paper appears to support the opposition against the motor theory of speech perception. There are many questions which would have to be answered before the theory can be unreservedly accepted. I will raise two:

(A) What is the delay time of the feedback system postulated by the theory? It might turn out to be so long that it might be an obstacle rather than a help in speech perception.

(B) What is the explanation of the fact that some speakers of a language (e.g. immigrants) have no difficulty in understanding its spoken form although they do not pronounce the language correctly?

Answer *Denes:* Mr. *Jassem* indicated that existing models of the human speech perception process could be more successful if only engineers would include all that is known about speech perception in the design of their automatic speech recognizers. Although it is true that in the past specialized engineering training, as well as knowledge of phonetics and linguistics was required for designing automatic speech recognizers, the availability of modern digital computers has changed this. Any phonetician or linguist, even if he has no knowledge of engineering, can put his theories on speech recognition to the test by computer simulation. All that is needed is a flow chart which specifies the sequence of logical operations which he considers are needed to implement the recognition process. If Mr. *Jassem* thinks that engineers have not utilized all that linguists and phoneticians know about speech recognition, I would ask him to specify the principles which we ignored. Any qualified computer programmer will be able to include his ideas into a suitable program and put them to the test.

Tillmann (Bonn): I want to ask you: Did you investigate whether the best of the trained subjects were able to identify even quite new not trained words presented to them after frequency compression?

Harms (Lawrence): 1. In preparing a learning program for phonetic transcription, students were observed to repeat the word they heard before they attempted to transcribe it. How might this be interpreted within the framework of the Motor Theory of speech perception?

2. What is a next experiment to follow the one you have completed?

Fry (London): I should like to make one general remark and also one specific comment on the experiments reported by Mr. *Denes*. The general point is one already mentioned by Mr. *Jassem*. Let me put it in this way: every individual spends a good deal of time acting as a speaker and a good deal as a listener. We are bound to assume that much of the brain mechanism is common to both processes. It would be entirely against what we know of the economy of the human being to imagine that there are completely separate mechanisms for the generation and for the reception of speech. On the other hand, there *are* differences between reception and generation and there are no doubt some mechanisms that are specific to the one or the other. For this reason I suggest we should not refer to "the motor theory of speech perception", nor even to "a motor theory" because speech can be perceived in a number of ways. Perception of speech at any time will be the result of a number of different factors, and what we are trying to find out is what weight should be given to the effect of motor memories in given circumstances. I have said before, for example, that I believe this factor has considerable weight in the perception of stress patterns.

The specific point concerns the task that subjects were asked to do in these experiments. It seems to me that the severity of the task in the case of the "speakers" has been rather under-rated. They are being asked not simply to learn associations between speech movements and sounds; they have first to dissolve existing associations which link articulatory movements with normal sounds before they can establish the new ones and this is clearly a very difficult thing to do. It would probably take quite a long time to achieve this and it is not surprising that the effect should not show up after the relatively short training period allowed in the experiments. It might be worthwhile to get both "speakers" and "listeners" up to some fairly high level of performance and then see how rapidly both groups could learn a fresh set of words.

Answer *Denes:* I agree completely that the motor theory can only represent *a* process of speech perception, rather than *the* process. Like with so many other aspects of the human speech communication process, or other human activities for that matter, the human being utilizes a variety of ways for achieving his aims. In speech perception, the motor theory probably accounts for only one for several ways in which perception is established.

Fourcin (London): Although the average performance for each of the two groups of subjects, those who spoke and listened and those who only listened, is almost the same, the difference in detailed shape of the two curves could be of importance. The speakers may, because of their prior practice, learn at a greater rate than the listeners during an experiment. In between sessions however they may at first continue to learn from their accumulated experience and then, if the interval is of sufficient length, start to forget owing to the disturbing influence of normal speech. Thus if the temporal spacing of experiments is short-long, the speakers' zig-zag performance curve could be expected. If these assumptions are correct, the zig-zag curve would be modified by having more experimental sessions in a working week. Then, if the inhibiting effects of fatigue are not pronounced, the speakers would have a greater average score than the listeners.

From this it seems possible that the basic experiment may have been more successful than at first appeared; it is not, however, of crucial consequence to the motor hypothesis, too many unknown factors are involved.

Proc. 5th int. Congr. phon. Sci., Münster 1964, pp. 259–260
(S. Karger, Basel/New York 1965).

Communication Sciences Laboratory, University of Florida

Perception of Inflected Vowels*

By DONALD DEW, Gainesville, Fla.

The commonly accepted technique for measuring the response of a resonator is to record the amplitude variations of a single pure tone as it is swept through the frequency range. Naturally, if measurements were made at only a few discrete frequencies the response would be less clearly indicated. In a similar manner this may also be true for the perception of the complex resonating characteristics associated with vowels. Specifically, it might appear that the intelligibility scores for vowels would be higher when formant frequency patterns were defined by the sweeping partials associated with an inflected fundamental than when only a few points of this pattern were defined by the fixed partials associated with a periodic fundamental. Indeed, synthesized vowels have been judged more natural when an inflection was used. Moreover, some evidence has been reported which indicates that an inflection does make a significant difference in the recognition of certain vowels.

The purpose of this investigation was to test the hypothesis that there is no significant difference between the intelligibility scores for inflected and for non-inflected spoken vowels.

To test this hypothesis 20 male and 20 female subjects read ten words (composed of different vowels in a "h – – d" context) until each word could be spoken with (a) a rising, (b) a falling, and (c) no inflection. A high fidelity system was used to tape record these 1,200 stimuli and to play them back in a random sequence to observers listening through a pair of earphones. In all, 17 observers, each having demonstrated prescribed phonetic transcription ability, listened independently to each stimuli until confident of the

* This research was supported by National Institutes of Health grant NB–04244.

phoneme identification. Then they recorded their judgment using the symbols of the International Phonetic Alphabet. The data were collated, and an analysis of variance was computed for the main effects and interactions of vowels, sex, and inflection.

The only main effect found significant at the one percent level of confidence was that for vowels. Neither the effects of inflection nor of sex were found significant at this level, and hence they were attributed to chance variations. A closer inspection of the data in terms of intelligibility scores for each vowel-inflection condition showed the vowel differences to be similar to those found in previous studies. In addition, the effects of inflection for certain vowels appeared large but inconsistent in direction. Thus, the null hypothesis that inflection has no significant effect upon vowel intelligibility could not be rejected.

Because so little is known about the identification of such complex stimuli, further generalizations must be quite speculative. It is possible that inflection is unrelated to the recognition of vowels; it is also possible that the quasi-periodic nature of laryngeal vibration defines the resonance characteristics as well as inflection. More research is needed before these phenomena can be clearly understood. In particular, the investigator plans to extend this research using synthesized vowels as stimuli. Thus, the difficulties of measuring and controlling the physical parameters of spoken vowels can be avoided.

Author's address: Dr. D. Dew, Communication Sciences Laboratory, University of Florida, *Grainesville*, Fla. (USA).

Discussion

Weingartner (Hamburg): The fact that Chinese vowels can be recognized easily even if they make up morphemes by themselves and are inflected in four tones appears as a corroboration of the paper read.

Proc. 5th int. Congr. phon. Sci., Münster 1964, pp. 261–263
(S. Karger, Basel/New York 1965).

From Georgetown University

The Phonemic Status of Juncture in Italian

By Robert J. Di Pietro

The key distributional unit in this discussion of Italian phonology is the *macrosegment*, which Professor *Charles F. Hockett* defines as the stretch of phonation that occurs between successive pauses[1]. Both vocoid and contoid allophones may be described in terms of their occurrence within or at the borders of the macrosegment. The pauses separating macrosegments are said to be instances of external juncture[2]. The binary opposite of external juncture is internal juncture, or the mode of transition from one sound to the next within the macrosegment. Our supposition is that all human languages have both types of juncture. Furthermore, no language consists wholly of mono-phonemic utterances. If such were the case, there would be no need to speak about juncture at all. It is quite possible, however, for a language to have two types of internal juncture – provided that features observed at locations of the external type are also found to be present within the macrosegment[3]. In such cases we would subdivide internal juncture into 'close' and 'plus'[4]. There are certain phonetic data in Italian which lead us to consider the postulation of a plus-juncture as a preferred alternative to the establishment of three additional phonemes.

The situation involving intervocalic phones [s] and [z] has led *Robert A. Hall* to speak of a "semi-componential transcription"[5]. In

[1] See Lit. 3, p. 44.

[2] Another term used to describe this type is 'open juncture'. See Lit. 3, p. 60.

[3] In addition to English, plus-junctures have been established notably for German (by *William G. Moulton*, Lit. 5) and Spanish (*R. P. Stockwell*, *J. D. Bowen* and *I. Silva-Fuenzalida*, Lit. 6). Both articles are reprinted in Readings in Linguistics, Lit. 4).

[4] The appropriateness of this terminology is not under discussion here. We have followed customary usage in the description of this contrast.

[5] See Lit. 1.

our eyes *Hall's* convention of indicating the component of voicing by a dot under /s/ (/ṣ/), as in /riṣorĝiménto/ (contrasted with /risalíre/), amounts to a phonemic statement in spite of its low-yield phonemic contrast.

In regard to semi-vowels, there is a possible contrast at one level of discourse between [ˌal–le–'vi̯a:–mo] 'we raise' and [al–le–vi–'a:–mo] 'we alleviate', or between ['pi̯a:–to] 'lawsuit' and [pi–'a:–to] 'peeped' (from the verb [pi–'a:–re] 'to peep'), in which comparable high front unrounded vocoid phones function as both semi-vowels and full vowels. Similar pairs with [u̯] and [u] are more difficult to find; but it is clear that in macrosegments like ['ku̯an–do], a semi-vocalic [u̯] occurs whereas either a full vowel *or* a semi-vowel is found in other cases, e.g., [at–te–'nu̯a:–re] or [at–te–nu–'a:–re], [e–'ku̯a:–re] or [e–ku–'a:–re].

In view of the phonetic data presented, it would seem necessary to establish phonemic semi-vowels /i̯/ and /u̯/ contrasting with /i/ and /u/, together with phonemic /z/ in contrast with /s/. However, the postulation of an internal transition of the 'plus' type allows us to account for all three variations in the same terms. We identify semi-vowels [i̯] and [u̯] as members of the phoneme classes /i/ and /u/, respectively. Their distribution would indicate that they never occur between consonant and plus-juncture. Our phonemic transcription of [ˌal–le–vi–'a:–mo] would be /allevi⁺ámo/; [ˌal–le–'vi̯a:–mo] – /alleviámo/; [pi–'a:–to] – /pi⁺áto/; ['pi̯a:–to] – /piáto/ and so on. As for [s] and [z], our statement of distribution would read that [z] never occurs following /⁺/ and preceding vowels. Phonemically [ˌri–sa–'li:–re] would be retranscribed /ri⁺salíre/, [ˌri–zor–ĝi–'men–to] – /risorĝiménto/, and so on.

The solution of phonemic plus-juncture is reinforced by consideration of features of consonant distribution. Analysts[6] usually state that sequences of [np nb nm] do not occur in Italian. The morphophonemic replacements of such sequences by [mp mb mm], respectively, are said to be automatic. Nevertheless, for many native speakers of Italian [np nb nm] are possible in cases where the preceding vowel has a degree of stress greater than weak, e.g., ['kon–pi̯a–'če:–re] vs. [kom–pi̯a–'če:–re]; [ˌin–'baŋᵑ–ka] vs. [im–'baŋᵑ–ka]; [ˌsan–'mar–ko] vs. [sam–'mar–ko]. In such cases, it would seem more advantageous to state that both the secondary stress and the occurrence of /n/ are conditioned by plus-juncture, rather than to change our statements of distribution within the macrosegment. We would then retranscribe [ˌkon–pi̯a–'če:–re] as /kon⁺pi̯ačére/; [ˌin–'baŋᵑ–ka] as /in⁺bánka/; and [ˌsan–'mar–ko] as /san⁺márko/. The form [kom–pi̯a–'če:–re], however, would have to be phonemicized /kompiačére/, because of the phonemic status of /m/.

[6] See especially *R. A. Hall*, Lit. 2, p. 12.

The status of plus-juncture as a phoneme is further enhanced by an important socio-linguistic factor. The use of standard Italian is expanding rapidly in regions where other Romance dialects are or were recently prevalent. As a result, each new idiolect added to the complex of 'standard' speakers incorporates a slightly different inventory of phones. That is to say, any given speaker of standard Italian may have one or more of the following in his idiolect: (1) [np nb nm] in contrast with [mp mb mm], (2) medial [s] vs. medial [z], (3) [i̯] vs. [i] and [u̯] vs. [u]. The convention of a plus-juncture would account for all three features. The other solution would be to alter the rules of distribution and add three new phonemes to the inventory with an indication that they are not part of the overall pattern. The preferred solution of plus-juncture would clearly obviate the positing of phonemes with marginal status while not obscuring the overall pattern.

References

1. *Hall, R. A.:* Italian /z/ and the converse of the archiphoneme. Lingua *9:* 2 (1960).
2. *Hall, R. A.:* Descriptive Italian grammar (Ithaca, N.Y. 1948).
3. *Hockett, C. F.:* A manual of phonology (Baltimore 1955).
4. *Joos, M.* (Ed.): Readings in linguistics (Washington, D.C. 1957).
5. *Moulton, W. G.:* Juncture in modern standard German. Language *22:* 200–248 (1946).
6. *Stockwell, R. P.; Bowen, J. D.* and *Silva-Fuenzalida, I.:* Spanish juncture and intonation. Language *32:* 641–665 (1956).

Author's address: Dr. Robert J. Di Pietro, Calle Ramón de la Cruz, 91 (4° B), *Madrid 6* (Spain).

Verh. 5. int. Kongr. Phon. Wiss., Münster 1964, pp. 264–266
(S. Karger, Basel/New York 1965).

Über einen [w] ~ [m] – Wechsel in den uralischen und türkischen Sprachen

Von J. Erdődi, Budapest

In manchen finnisch-ugrischen Sprachen (den sogenannten wolga-finnischen und den ob-ugrischen) und in den mit ihnen verwandten samojedischen Sprachen ist ein Wechsel [*w*] ∾ [*m*] zu beobachten. Derselbe Wechsel findet sich auch in gewissen türkischen Sprachen.

Den Wert dieses allophonischen Wechsels wollen wir zuerst an Beispielen aus der tscheremissischen Sprache klarlegen. Die Vertreter der Finno-Ugristik halten das /*w*/ und /*m*/ im Tscheremissischen für besondere Phoneme. So z.B. im Lehrbuch von *Thomas A. Sebeok* und *Frances J. Ingemann*, wo [*w*] auf folgende Weise charakterisiert wird: «bilabial (usually)... voiced fricative» ([2], S. 5). Als durch die Umgebung bestimmte Allophone werden [*b*] («bilabial voiced lenis stop») und [*v*] («labiodental voiced fricative») angegeben. Gemeinsames Phonemzeichen /*b*/, z.B. *bara* 'nachher' (eig. *wara*), *jeŋblak* 'Leute' (eig. *jeŋvlak*), und *kombo* 'Gans'. Der Mitlaut *m* wird als ein besonderes Phonem behandelt («bilabial nasal continuant» – [2], S. 6).

Nach unserer Beobachtung hat das [*w*] im Tscheremissischen auch ein Allophon [*m*], was eben durch den gemeinsamen bilabialen Charakter der beiden Laute zu erklären ist. Ich führe aus dem ungefähr 100 Beispiele enthaltenden Material nur einige an.

a) Anlautend: bergtscherem. *mərzi* 'Schneewiesel', *wərzi* 'Uferschwalbe' – osttscherem. *wem* 'Mark, Knochenmark', bergtscherem. *wim*, westtscherem. (J.) *wime* ∾ *wimə*, westtscherem. (V.) *mìmə*, im Waldtscherem. (M.) *mem;* – mitteltscherem. *woktem* und *moktem* 'Eigelb' (daneben auch *oktem* und *loktem*). Der Wechsel kommt auch in Lehnwörtern vor: westtscherem. *məχatir* 'Held' neben *pəχatir* (aus russ. bogatyr).

b) Inlautend: mitteltscherem. (Č.) *kowəzem* 'ich mahle': mitteltscherem. (C.) *koməzem* dss. – westtscherem. (K.) *jəwəžä* 'fein, weich', (J.) *ᵈjəwəža*, mitteltscherem. (CK. Č.) *ᵈjəwəža* dss., mitteltscherem. (UJ.) *jəməža*, (MK.) *dńimı̣ža* dss. Kommt ebenfalls in Lehnwörtern vor: *tawak* ∾ *tamak* 'Tabak'.

Dieser Wechsel ist in anderen oben schon erwähnten finnisch-ugrischen Sprachen und im Samojedischen zu finden. Wir führen bloß einige Beispiele an:

Mordwinisch (Ersä): *eŕma* – ∾ *eŕva* – 'warten'. Samojedisch (Č. NP.) *wandže*, (Tas. Kar.) *wuendž*, (B.) *muendž* 'Njelma-Fisch'.

G. N. Prokofjev schreibt in seiner Grammatik ([1], S. 22), daß in der Selkup-Sprache ein Wechsel *w* ∾ *B* ∾ *m* ∾ *p* besteht, der teilweise durch die Lautumgebung bedingt ist, teilweise individuell auftritt, z.B. *wəçi* 'Fleisch' neben *məçi*. Auch im Auslaut: *åtaep* ∾ *åtaem* 'das Rentier, Acc.' (eig. wahrscheinlich *åtaew*).

Ein derartiger Wechsel kommt auch in den türkisch-tatarischen Sprachen vor, wo inzwischen ein Lautwandel *w* > *p* vor sich gegangen sein muß, vgl. tat. (Kasan) *piräpäč* 'Pfannkuchen', tat. *pirämäč* dss., oder kas.-tat. *mamyk* 'Baumwolle, Daumen', dschag. *mamuk*, jedoch osm. *pamuk*, *pambuk*.

Ein anderes Beispiel aus den türkischen Sprachen: osm. *čulpan* 'Venus-Stern', baschkir. *sulpan*, kas.-tat. *čulpan* – teleut. *čolmon*, kalmük. *tsolwn̥*, *tsolmn̥*, chalca-mong. *čolmon*.

Der ursprüngliche bilabiale Charakter des alttürkischen bzw. uralischen *w* wird eben durch diese Wechselmöglichkeit bewiesen, wenn sie auch heutzutage in den meisten der zu diesen Sprachfamilien gehörenden Sprachen nicht mehr allophon ist. Die einstige phonetische Eigentümlichkeit, das bilabiale Wesen des [*w*] und sein allophoner Wert wird auch durch die verschiedenartige Vertretung im Finnischen und im Ungarischen bestätigt, was wir durch manche Beispiele illustrieren können:

	finnisch *nimi* 'Name'	ungar. *név* id.
	finnisch *liemi* 'Saft, Suppe'	ungar. arch. *lév* id.
dagegen	finnisch *lumi* 'Schnee'	ungar. dial. *lomos* 'kotig, naß durch Schnee'
	finnisch *silmä* 'Auge'	ungar. *szem* id.
	kolme 'drei'	ungar. *három* id.

Als aber das ursprüngliche *w* zu *v* (labiodental) wurde, ist die

Möglichkeit der Allophonie [*w*] $\backsim$ [*m*] verschwunden, daher u.a. ungar. *ver* 'er schlägt', *mer* 'er wagt'; finn. *veri* 'Blut', *meri* 'Meer'.

Literatur

1. *Prokofjev, G. N.:* Selkupische (ostjakisch-samojedische) Grammatik (Leningrad 1935), in russ. Spr.
2. *Sebeok, T. A.* and *Ingemann, F. J.:* An eastern Cheremis manual (Bloomington, The Hague 1961).

Adresse des Autors: Professor Dr. J. Erdödi, Fillér u. 50/A, *Budapest II* (Ungarn).

Discussion

Gyula Décsy (Hamburg): Herr *Erdödi* hat ungarisch *v* durch die Annahme von ursprünglich bilabialer Aussprache (w) dem *m* eine Stufe näher gebracht und dadurch ein neues Erklärungsprinzip für die vergleichende finnisch-ugrische Lautforschung geliefert (vgl. Fälle wie ung. *név* 'Name' < **neve* $\backsim$ finn. *nimi* 'dss.').

Verh. 5. int. Kongr. Phon. Wiss., Münster 1964, pp. 267–270
(S. Karger, Basel/New York 1965).

Psychische Gestaltungsfreiheit und sprachliche Verbindlichkeit in der Melodik des Ausspruchs

Von O. von Essen, Hamburg

Will man den melodischen Gestaltungsformen des Ausspruchs auf den Grund kommen, so hat man einerseits mit den situationsbedingten psychischen Motivierungen, also mit einer *expressiven* Komponente, andererseits mit der sprachlichen Gebundenheit, also einer *phonologischen* Komponente, zu rechnen. Auf diesen Sachverhalt hat zuletzt und am ausführlichsten *Wodarz* hingewiesen.

Wenn die Sprechmelodik Ausdruck seelischer Vorgänge ist, dann ist auch die Vielfalt der Formen selbstverständlich. Freilich haben sich einige Grundformen der Sprechmelodie stabilisiert, sind zu sprachlich relevanten Identifizierungs- bzw. Differenzierungsmerkmalen geworden und haben die Funktion der Kennzeichnung von Ausspruchskategorien (Aussage und Aufforderung, Entscheidungsfrage, unvollendeter Ausspruch) erlangt. Wenn man aber andererseits weiß, welch eine beherrschende Macht die seelischen Vorgänge auf die Melodiegestaltung ausüben, wird man sich fragen, ob der logisch gesteuerte Kommunikationsprozeß die emotionell gesteuerte Ausdrucksgestaltung zur Einhaltung der phonologisch gültigen Verständigungsgesetze zwingt oder sich von dieser verwischen oder gar unerkennbar machen läßt. Diese Frage hat zu folgender Untersuchung Anlaß gegeben.

Den Gewährspersonen wurden Texte ausgehändigt, die eine ausgeprägte emotionelle Färbung verlangen, z. B. ein Abschnitt aus Hebbels «Siegfrieds Tod», Storms «Abseits», Münchhausens «Todspieler» u. a. Die Sprechtexte wurden auf Tonband aufgenommen, auf Oscillomink in Bildkurven übertragen, in den Klarphasen der Silbenträger wurden die Frequenzen gemessen und deren logarithmische Werte in Tonstufenskalen aufgezeichnet. Aus diesen Aufzeichnungen sind je 8 Beispiele von zwei Personen ausgewählt (Bildvorführung).

Auffallend ist bei jedem Sprecher die unterschiedliche Höhenlage der Sprechtonbereiche: Ruhe, Beschaulichkeit, Stimmung der Trauer nötigten die Sprecher zur tiefen Stimmlage; Erregung wie im Ärger und Zorn – übrigens auch Angst – trieb in die Höhe; nur der unheimliche Haß, wie er Krimhild an der Bahre Siegfrieds befällt, drückte die Sprecher in drohende Tiefe hinab (männl. Vp. cis–A, weibl. Vp. g–H!). – Auffällig ist ferner die geringe melodische Breite in den Aussprüchen der beschaulichen Ruhe und stillen Trauer (Es ist so still...; Da war es einsam, einsam um mich her u. a.). Die Melodiebreite umfaßt hier in einem Falle eine Quart, sonst höchstens eine große Terz. Bei den aus seelischer Erregung quellenden Aussprüchen der Heiterkeit und des Zornes springt die Stimme über den Bereich einer Sext.

In manchen Einzelheiten sind die Bilder der verschiedenen Vpn. unterschiedlich. Es ist natürlich nicht zu erwarten, daß dieselben Texte bei verschiedenen Menschen gleiche Gefühlsregungen auslösen. Wenn man aber die *Ausspruchsschlüsse* vergleicht, so erkennt man eine fast völlige Übereinstimmung. Es ist das gewohnte Bild, wie wir es aus emotionell farblosen Aussprüchen kennen: Aussagetypus mit dem Absinken in die *spannungslösende* Tiefe, Entscheidungsfragen mit ihrem *spannungssteigernden* Anstieg, unvollendete Aussprüche mit *spannungserhaltendem* Ausklang. Durch diese phonologisch relevanten «Melodeme» kennzeichnen sich die Ausspruchskategorien. Wir haben es hier mit einer vom Sprachgebilde her bestimmten Funktion zu tun, die sich überall, auch – wie in unseren Beispielen zu erkennen – bei emotionell stark gefärbter Rede durchsetzt. Nur der (zornige) Ausspruch: Kannst du denn das überhaupt nicht begreifen? scheint einmal aus dem Rahmen zu fallen: Während die weibliche Vp. regelrecht das interrogative Melodem anwendet, bringt die männliche Vp. das terminale. Aber hier geht es um Interpretationsfragen: Der erste Sprecher hat diesen Ausspruch gar nicht als Frage gemeint, sondern als Befehl: Jetzt begreif das endlich!

Nachdem *Trojan* nachgewiesen hat, daß die Farbe der Sprechstimme von den psychisch-somatischen Reaktionen abhängig ist, liegt es nahe, eine solche Erklärung auch für die Modulation der Stimme in Anspruch zu nehmen. Es scheint mir berechtigt, von *offener*, *sperrender* und *abwehrender* Intonation zu reden. In den vier ersten der vorgeführten Beispiele war nichts abzuwehren; Eindrücke angenehmer Art, ein Sicherschließen, andererseits ein Sichabschließen, eine Introversion sind die steuernden Kräfte, die keine Lautheit, keine abwehrende Sperrspannungen der Rachen- und Stimmuskulatur, keine imperativische, akklamative oder aggressive Sprunghaftigkeit der Muskelaktionen aufkommen lassen. Deutlich aber wird die instinktive Sperrung und Abwehr des Körper- und Lebensfeindlichen in den Aussprüchen zorniger Erregung. Hingegen braucht das Herausdrängende, Überschäumende, das Sichhinwenden zur Mitwelt in dem Ausruf: Ungeheure Heiterkeit ist meines Lebens Regel! wieder den freien Stimmweg und die Gelöstheit bei ausholender Bewegung des Muskelspiels. Die melodische Gestalt schließt sich zwar derjenigen der emotionell farblosen Aufforderung an, aber mit stärkerer Profilierung – so auch in Aussprüchen der Freude und des Schmeichelhaften. Hohe Stimmlage und lebhafte Sprunghaftigkeit der Intervalle zeigte auch ein Ausspruch der Angst (Hast du gehört? Da ist jemand im Keller!).

Freilich ist die Sprechmelodik nur eine Komponente der expressiven Redegestaltung. Tempo, Stärke und Farbe spielen ebenfalls eine entscheidende Rolle. Aber auch in der Melodik zeigen sich deutlich die Symptome der Offenheit, der Sperrung und der Abschreckung. Dabei erstreckt sich der emotionale Bereich nicht nur bis an den phonologisch relevanten Bereich, sondern übergreift diesen, allerdings ohne dessen informatorische Funktion zu stören.

Reichhaltige Literaturangaben bei *H. W. Wodarz:* Satzphonetik des Westlachischen (Böhlau Verlag, Köln/Graz 1963).

Adresse des Autors: Prof. Dr. O. von Essen, Höpen 59, *2 Hamburg* (Deutschland).

Discussion

von Raffler Engel (Florenz): Dürfte ich Herrn *von Essen* fragen, ob man die zwei Ebenen unterscheiden kann, auf denen ich z. B. mit dem von mir erwarteten Pathos mechanisch Siegfrieds Tod vorlesen könnte, während ja meine Gedanken ganz woanders sind und mir in Wirklichkeit auch das, was ich vorlese, völlig gleichgültig ist.

Bock (Braunschweig): I. Der Referent hat zwei Möglichkeiten der Beschaffung von Sprechunterlagen aufgezeigt, an denen die Emotionen der Sprecher in der Melodie des Ausdrucks nachgewiesen werden können.

1. imaginierte, im Versuch künstlich herbeigeführte Sprechsituationen,
2. die Nachgestaltung von poetischen Texten.

Es fehlt die Heranziehung «echter» Sprechsituationen aus dem Leben, die allerdings schwer, oft nur durch Zufall beschafft werden können. Hierbei wäre auch an eine Auswertung der «Lautbibliothek der deutschen Mundarten» zu denken.

II. Bei der Bewertung der Sprechmelodie in nachgestalteten Dichtungen darf die Bedeutung der «Sprechmoden», also unbewußter Vorbilder, nicht unterschätzt werden; als Beispiel dafür sei die Vorliebe für ausdrucksstarke Gestaltung etwa zur Zeit eines *Moissi* der Ausdruckskargheit des heutigen Sprechstils gegenübergestellt.

Höffe (Dortmund): Im Anschluß an die Ausführungen von Herrn *Bock* möchte ich Sie bitten, Ihre Aufmerksamkeit einigen Punkten zuzuwenden:

1. Sprachaufnahmen, die nicht aus echten Sprechsituationen gewonnen wurden, sind auf ihre «Gültigkeit» hin zu prüfen, denn nur dann können die Ergebnisse und die aus ihnen gezogenen Folgerungen als «bedeutsam» bezeichnet werden (*Zwirner*).

2. Um sicherzustellen, daß absichtlich gestaltete Sprechaussagen sich innerhalb der zu fordernden Richtigkeitsbreite bewegen, ist ihr Ausdruckscharakter in einem «Siebverfahren» zu bestimmen bzw. zu bestätigen (*Höffe*).

3. Es genügt nicht, hierbei allein den Grundcharakter zu bestimmen, sondern es ist auch nach den «Färbungen» zu fragen (z. B. ärgerlich, dabei etwas ungeduldig). Diese Ausdrucksschattierungen finden bekanntlich ihren Niederschlag in der akustischen Struktur (vgl. *Höffe:* Phonetica *5:* 129–159, 1960).

Daneš (Prag): Manche Fragen, die hier von Herrn *von Essen* so interessant behandelt worden sind, habe auch ich in einigen meiner Arbeiten berührt. Ich habe dabei von Herrn *von Essens* Arbeiten profitiert; aber wie es jetzt aus dem heutigen Vortrag hervorgeht, sind wiederum meine Ansichten und Resultate, gemeinsam mit anderen theoretischen Punkten der Prager linguistischen Schule (*Mathesius*, *Trost*) offensichtlich die Hauptquelle der theoretischen Erwägungen, mit denen Herr *Wodarz* in der letzten Zeit auftritt. Es handelt sich z. B. um die Fragen der Beziehungen zwischen der expressiven und kommunikativen Funktion der Sprachmelodie oder zwischen dem spontanen (instinktiven) und konventionellen Gebrauch der Intonation. Erlauben Sie mir hier, ein Zitat aus meinem Aufsatz, der vor vier Jahren in «Word» erschien*, als eine Illustration anzuführen: 'From a genetic point of view, intonation formulas may indeed have developed from such instinctive signals: hence the relative similarity of intonational schemes in many languages. The degree of arbitrariness is proportional to the degree of intellectuality of the intonational function. Therefore, in its least intellectual, most "spontaneous" or instinctive uses, intonation is intelligible across language boundaries: in foreign languages it is sometimes easier to recognize and to render certain emotions than to distinguish a question from a statement. – *P. Trost*, in his article "O problémech větné intonace", Slovo a slovesnost *III*, 226 (1937), distinguishes three ranges of intonation use: (1) intonational mimicry in which intonation is spontaneous, natural, and psycho-physiologically stimulated; (2) intonational formulas, in which such natural values are intentionally utilized; (3) intonational oppositions, which enter into systems of form and meaning peculiar to a language, and for which the psycho-physiological value of intonation is irrelevant.'

Ich muß Herrn *von Essen* völlig zugeben, daß die expressive (oder) emotionale Färbung der Intonation das intonatorische Grundgebilde der Äußerung nur soweit modifizieren (deformieren) darf, um die Grundfunktion dadurch nicht zu stören. (Ich habe dies in meinem oben zitierten Aufsatz folgendermaßen formuliert: "Expressive intonation may operate only in such a way as not to interfere with the basic communicative intonation.") Man kann sich jetzt die Frage stellen, in welcher Weise sich die expressive Intonation geltend machen kann. Für das Tschechische habe ich diese drei Möglichkeiten festgestellt: (1) Die zulässige Variabilität in der phonetischen Realisation der intonatorischen kommunikativen Grundstrukturen (Konturen) ermöglicht, daß verschiedene Varianten dieser Strukturen zum Ausdruck der Expressivität ausgenützt werden können. (2) Es werden speziell (merkmalhaftige) expressive Konturen genutzt. (3) Eine spezielle Expressivität entsteht durch eine funktionelle Transposition der intonatorischen Konturen. – Mir scheint, daß sehr ähnliche Verhältnisse auch in der deutschen Schriftsprache vorkommen.

---

* *Daneš, Fr.:* Sentence Intonation from a Functional Point of View. Word *16:* 34–54 (1960).

Proc. 5th int. Congr. phon. Sci., Münster 1964, pp. 271–273
(S. Karger, Basel/New York 1965).

The Assimilation of ə in Present English

By S. S. Eustace, London

In rapid English speech, assimilations, dissimilations and elisions take place, most of which are scarcely considered in the textbooks. Evidence for some of them goes back for several centuries. They occur mostly in informal situations and are therefore difficult to observe. Not all individuals make them, but some do regularly; thus although, with few exceptions, they are not obligatory, the student should know of their existence. As we would expect, the phonemes most affected are those occurring frequently and in unstressed position, particularly t, d, n, ð, l, ɪ and ə. It is the behaviour of the last of these that concerns us today.

As a general principle we may say that ə not standing before a pause is assimilated to any neighbouring continuant phoneme. The ə disappears; the continuant is lengthened; if it is a consonant it changes from a margin to a peak; and the onset of the next syllable is transferred to the following phoneme. For instance *get along* gɛtə'lɒŋ > gɛtl·'ɒŋ [1]. The process often coincides with other assimilations, etc., and the resulting syllabics include sounds, such as *Φ* or *γ*, which are not generally reckoned as being part of the English sound system. Here are some examples:

m < vəm:	'nɛvm·'aɪnd *Never mind*
m < *məmpp < məntp:	wan'məʊm·ʔ'pliz *One moment, please* – not RP
m < vən:	'hɑ'pastl·'ɛbm *half past eleven*
n < tən:	batn· *button;* 'aftn·'un *afternoon*
n < dən < dɪn:	kɑdn·l· *cardinal*
n < tʃn·ət < tʃənət:	an'fɔtʃn·n·tlɪ *unfortunately*
n < aɪən:	laɪn· *lion;*/laɪn *line* – Midland, not RP

[1] The high point · indicates that the preceding sound is syllabic.

ŋ < *səŋgk<səŋdk:	'hɔsŋ·'kat *horse and cart* (but 'hæʔn·'kəʋt *hat and coat*)
Φ < səvp:	'lvtsΦ·'pipl· *lots of people*
Φ < *Φəb:	'wεɪspεɪΦbaskɪt *wastepaperbasket* Φ < p by dissimilation. Φ not syllabic here
β < ləvw:	'fʋlβ'wɔtə *full of water*. l remains clear as if before a vowel
f < vfəg:	af·'gɒʔn· *I've forgotten* – said by a child
f < təf:	'bjutf·l· *beautiful*
v < ɜvət:	kn·'sɜv·tɪv *conservative*. Stress on v level not diminishing
θ < nəθ:	'εnθ·ɪŋ *anything*
ð < εðə:	wεð·ðə'rεntəbbɪmpε:d *whether the rent had been paid*
ð < təð:	wɒtð·εgən'duðεə *What are they going to do there?*
s < məs:	bətams·'dʒεs *But I must just...*
s < vəs:	vs·ɪnətɪ *vicinity*
s < sðət:	'wɒtss·'taɪm *What's the time?*
z < sðət:	'wɒtsz·'taɪm
z < məz:	æmz·n· *Amazon;*/dæmzn· *damson*
z < ðəz:	ðz·ə'mænəʋvð·'εə *There's a man over there*
r < θrəp:	'ɑθr·pɒd *arthropod*
r < rən:	fɒr·nə *foreigner*. r scarcely syllabic
γ < *χəg < kəg:	'bæχγ·gεn *back again;* 'rεχγ·gnaɪz *recognize*
l < pəl:	prɒpl·ɪ *properly*
l < dəl:	'bɔdl·aɪn *borderline*
l < təl:	'kætl·ɒg *catalogue*
i̥ < təhɪ:	ti̥'hɪə *to hear*
ɪ < təj:	'mɪstɪ'ustəs *Mr Eustace*
ɪ < jəl:	ɑ'tɪkjɪlətrɪ *articulatory*
ʋ < təwəd:	'ɑftʋwʋdz *afterwards*. Here w remains, to keep the distinction, as tʋ'wɪŋk *to wink;* tʋ'ɪŋk *to ink*
*? < əʋ:	Does not occur. 1. This ə is of course often stressed. 2. The recent fashion for əʋ > ə“i”, ε“i” is parallelled by aʋ > a“i” and is probably an independent sound-change
ɪe < ɪə:	fɪe *fear* – Cockney, not RP
æ: < εə:	hɪ'læ: *Hilaire* – said by a child. Perhaps only on a rising or falling tone. Combined with post-war tendency to lower the front vowels

ɜ < ʋə:	kjɜrɪ-əs *curious*. Only after velars? 'djʋərɪŋ *during*
a < aə < aɪə, aʋə:	'faarɪŋ *firing;* 'ʃaabɑθ *shower-bath*

(After plosives and affricates the ə is simply elided. The elision is often marked by release of the voiceless plosives, as 'æpʰtaɪt *appetite*, 'ɑftʰ'nun *afternoon*, and the full voicing of the voiced ones, as nεɪbhʋd *neighbourhood*. In ædʒtənt *adjutant* the effect is very slight.)

Some of these assimilations occur only in rapid speech; but others, for instance n, ɫ, l ('mεtl·ɪŋ'gwɪstɪks), (tɪ'ju *to you*) and a ('kaarəʋ *Cairo*), are heard even in formal discourse.

They occur mostly in the posttonic, and the ə most often precedes the assimilating consonant. They also tend not to occur across morpheme boundaries: laa *lyre*/laɪə *liar*. Unfortunately time does not permit us to enquire further into the conditions of occurrence.

It is interesting to note how many places in the system are filled.

In deciding the phonemic status of the syllabics we are at first tempted to treat them as realizations of the sequences əC or Cə. But it is unnatural to analyse a single phone as a phoneme sequence; moreover in some cases the single phone can be contrasted with the sequence: bɪtn· *bitten*/bɪtən *bittern*, æpl·ɪ *apply* “apple-like”/æpəlɪ *Apperley*. A possible solution would be to bring in a phoneme of syllabicness [·], like phonemes of length or of juncture. Even in this rapid, informal and you may say slovenly style of speech, the distribution of the syllabics corresponds close enough to that of ə in a more formal style.

These facts also serve to illustrate a general rule, that the faster the speech, the more complicated its phonetic structure.

Owing to typographical difficulties it is not possible to show examples of the syllabic labio-dental and dental voiced nasals, labialised z, retroflex s and z, or unvoiced l. It is noteworthy that there are no examples of the voiceless correlates of the syllabic nasals, ɹ, l, ɫ, ʋ, and labialised z; nor of ʃ and ʒ; nor of χ at the time of going to press.

Author's address: S. S. Eustace, 11 First Street, *London SW 3* (England).

Proc. 5th int. Congr. phon. Sci., Münster 1964, pp. 274
(S. Karger, Basel/New York 1965).

From the Bell Telephone Laboratories, Murray Hill, N.J.

Digital Computer Calculation of Glottal Volume Velocity

By James L. Flanagan

Abstract

The acoustic volume flow through the vocal cords during voiced speech is described by non-linear differential equations. The equations are of second order and have time-varying coefficients which represent the acoustic resistance and inductance of the glottis. The resistance is approximated by a function of glottal area and glottal flow. The inductance is only a function of glottal area. Difference equation approximations are made of the continuous relations, and the difference equations are programmed for solution in a digital computer. The results show the relative influence of viscous and kinetic factors upon the form of the glottal wave. In addition, the formulation permits analysis of the effects of air density, with glottal pressure and acoustic load of the vocal tract.

Author's address: J. L. Flanagan, Bell Telephone Laboratories, Incorporated, *Murray Hill*, N. J. (USA).

Discussion

Denes (Murray Hill): In comparing the volume velocities at the vocal tract's driving point when the glottis is loaded with the tract and when it is short-circuited, you only showed volume velocity wave shapes. Quite often, the spectrum changes are more instructive than the wave shape changes, and I am wondering whether you have also calculated the spectrums of the waveshapes you showed in your slides.

C.R. 5e Congr. int. Sci. phon., Münster 1964, pp. 275–280
(S. Karger, Basel/New York 1965).

L'intonation interrogative et l'inversion, membres d'un paradigme hétérogène?

Par L. Flydal, Oslo

«La langue est un système où tout se tient», a-t-on dit, et d'autres ont ajouté: «Où tout se tient, oui; mais pas toujours également bien!»

Pour se convaincre de la justification de cette dernière remarque, on n'a qu'à prendre un exemple comme celui que fournit le paradigme du présent indicatif du verbe *aller: je vais, tu vas, il va, nous allons, vous allez, ils vont.* Du côté du contenu «ça se tient», il n'y a pas de doute, mais du côté de l'expression la systématisation inachevée et imparfaite que trahit la coexistence des deux radicaux hétérogènes saute aux yeux.

Il existe des paradigmes qui offrent un caractère d'hétérogénéité tel qu'on doit se demander si, réellement, les grandeurs qu'ils réunissent sont à considérer comme un tout systématisé ou bien comme des pièces disjointes qui ne se suppléent – et cela plus ou moins bien – que par l'effet d'un hasard.

Tel est le cas des moyens par lesquels, dans plusieurs langues, on exprime l'interrogation.

Dans l'espoir que les résultats auxquels nous ont conduit nos observations auront une validité plus générale que celle qui concerne les deux ou trois langues bien connues d'où nous tirerons nos exemples, nous allons essayer de démontrer l'existence de certains rapports paradigmatiques entre des grandeurs qui, souvent, ne sont examinées qu'isolément.

Le polonais est une langue qui fournit l'exemple d'un système d'«interrogatifs» – ce terme pris dans le sens très large de «moyens d'interrogation» – à la fois simple et clair. Si on en dresse l'inventaire, on trouvera que toutes les expressions utilisées pour indiquer qu'un énoncé est une question, se ramènent aux quatre moyens dont on s'est servi dans les phrases suivantes:

1° (On) był w domu?	?	*synonymes:* «A-t-il été à la maison?» «est-ce-que...»
2° Był on w domu?	? (+ inversion)	
3° Czy (on) był w domu?	(?+) czy	
4° a) (Kiedy, Gdzie, Jak, etc., (on) był w domu?)		a) («Quand, Où, Comment, etc.,...»)
b) Kto był w domu? (Który (pan), etc., był w domu?)	(?+) kto	b) «qui...» («Quel (homme)...»)

Etant donné que les trois premières formules sont synonymes, il y a redondance et partant synonymie dans l'exemple 2° entre d'une part le point d'interrogation, ?, qui, dans la manifestation graphique, représente l'intonation interrogative de la manifestation phonique, et d'autre part l'inversion; de même qu'il y a également synonymie et redondance dans l'exemple 3° entre l'intonation interrogative et le mot *czy*. Vu qu'on peut se passer de l'intonation interrogative dans les cas 3° et 4°, tandis qu'elle s'avère indispensable dans le cas 2°, on peut réduire le système au paradigme suivant composé de deux membres commutables, dont l'un se traduit par 3 variantes libres, l'une desquelles trouve son expression dans la substance phonémique*, l'autre dans la substance prosodique seule et la troisième à la fois dans cette même substance et dans celle de l'ordre des mots, que, pour plus de simplicité, on pourrait appeler la substance *séquentielle:*

$$\left\{\frac{\textit{czy}\ /\ ?\ /\ ?(+\text{inversion})}{\textit{kto (który..., kiedy, etc.)}}\right\}$$

Par quel trait essentiel le système de l'allemand actuel, diffère-t-il de ce système slave? Ayant perdu la possibilité d'exprimer pour le premier membre de ce paradigme le contenu interrogatif par des moyens appartenant à la substance phonémique (moyens que connaissait le gothique: *wileiz-u* «veux-tu?», v. *A. Meillet*, Caractères généraux des langues germaniques, Libr. Hachette, Paris 1926, p. 187), l'allemand moderne ne donne statut d'interrogatif qu'à l'inversion qui exige d'être accompagnée de l'intonation interrogative, ou bien à cette interrogation seule (*W. K. Jude*, Deutsche Grammatik, Berlin 1959, p. 218):

$$\left\{\frac{?\ +\ \text{inversion}\ /\ ?}{\text{Wer (Welcher, Wie, etc.)}}\right\}$$

Nous avons hésité un peu à appliquer le terme de *paradigme* interrogatif à cette liste de deux membres, comme nous avons hésité aussi à l'appliquer aux faits correspondants du polonais. C'est que non seulement les substances dans lesquelles se manifestent, avec leurs variantes, les deux membres de cet ensemble, sont-elles si différentes dans leur nature physique que nous éprouvons une certaine gêne à devoir constater que leur coopération intime à l'intérieur du langage pénètre jusque dans des systèmes aussi fondamentaux, mais on doit aussi se demander si les grandeurs que nous avons réunies dans cette perspective paradigmatique occupent, au point de vue structurel, c'est-à-dire au point de vue de leurs fonctions et de leurs contextes, des rôles qui, omissis omittendis, les rendent interchangeables.

Pour trouver jusqu'à quel point ce sont là des grandeurs comparables et paradigmatiquement systématisées, il va falloir procéder à une petite analyse de ces «interrogatifs» au point de vue des contextes dans lesquels ils fonctionnent.

Il apparaîtra alors que le rôle de l'intonation interrogative comme le rôle de celle avec laquelle elle fait paire d'opposition: l'intonation assertive, est de faire fonction de flexif dans le syntagme dont elle fait partie. Ce syntagme est toujours une nexie (c'est-à-dire une période consistant en une ou plusieurs propositions – ou: nexus). L'intonation interrogative et celle assertive fournissent donc ce qu'on peut appeler la flexion de la nexie (voir *Louis Hjelmslev*, Accent, Intonation, Quantité. Studi baltici VI, p. 17; 1936–37); ce sont des *flexifs de nexie.* Nous appellerons la partie fléchie de la nexie sa base (en pensant à son contenu) ou son thème (en pensant à son expression), respectivement pseudobase ou pseudothème pour des grandeurs inanalysables en unités plus petites de caractère linguistique:

* La phonématie *czy*, qui représente ici «lo variable en una manifestacion», est à considérer dans ce contexte comme une grandeur de substance conformément à la conception de *Eugenio Coseriu* dans *Forma y sustancia*, Montevideo 1954, pp. 185, 186.

Il est arrivé (Base) $\left\{\frac{.}{?}\right\}$

$\bar{\varepsilon}$ «Hein» (Pseudobase) $\left\{\ ?\ \right\}$

En français comme en allemand et en polonais la caractéristique phonique la plus fréquente du membre marqué de l'opposition que comprend le petit *paradigme prosodique* qui correspond à celui graphique du point et du point d'interrogation, est une hauteur musicale et surtout une intensité (force) plus grandes que celle des nexies assertives correspondantes. Ce trait phonique, uni à la même fonction, se trouve être plus répandu, dans les différentes langues du monde, qu'aucun autre élément morphologique jusqu'ici observé (voir *Eduard Hermann*, Probleme der Frage, 2. Teil, p. 363, Göttingen 1942), ce qui, en plus de son inanalysabilité en éléments plus petits, nous autorise à y voir non pas un signe linguistique (au sens saussurien et hjelmslevien du mot), mais un symbole motivé (c'est-à-dire une grandeur construite selon le principe de l'isomorphisme figuratif entre la face expressive et celle du contenu). Ce symbole, solidement conventionalisé et intégré au système des flexifs dans les trois langues auxquelles nous avons emprunté nos exemples, possède, dans une de celles-ci, le polonais, un synonyme, qui, sous la forme d'un mot, *czy*, peut faire corps avec le thème, c'est-à-dire la partie fléchie du syntagme que constitue la nexie: *Czy był w domu?* Ce flexif thématique rend redondant son synonyme, l'intonation interrogative, dont on peut, dans ce cas, se passer en polonais. Le fait intéressant de cet exemple polonais est qu'il permet de voir qu'on a, dans cette langue, le choix entre deux flexifs, dont l'un est thématique et manifesté dans de la substance phonémique, tandis que l'autre est extrathématique et réalisé en une substance prosodique. Par le fait même qu'elle est extrathématique, l'intonation interrogative s'inscrit dans un paradigme différent de celui des interrogatifs pronominaux. Dans ceux-ci, par exemple *kto*, *gdzie*, *jak*, etc., un contenu synonyme de celui de l'intonation interrogative se joint au contenu de quelque thème pronominal faisant partie du thème de la nexie. Un flexif interrogatif incorporé à la base d'un nexus se distingue d'ailleurs aussi d'un flexif extrathématique en ceci qu'il ne demande pas que le nexus en question fasse fonction de nexie, mais lui permet de passer à l'état de nexus subordonné:

Nie wiem, czy (on) był w domu.
Nie wiem, kto był w domu.

(«Je ne sais pas *s*'il a été à la maison; ... *qui* a été ...»). Cette règle de transformation s'applique tout aussi bien en allemand en ce qui concerne les interrogatifs pronominaux de cette langue. Elle se complique un peu en ce qui concerne le flexif thématique *ob*, qui ne s'emploie que dans les nexus subordonnés et, par la transformation de ceux-ci en nexies, se convertit normalement en substance séquentielle: inversion, flexif thématique moins explicite que *czy*, et qui présuppose l'appui que lui fournit le flexif extrathématique de l'intonation interrogative:

(Ich weiß nicht,) ob er zu Hause gewesen ist.
> Ist er zu Hause gewesen?

Les moyens par lesquels s'exprime l'interrogation en une langue slave comme le polonais, en une langue germanique comme l'allemand et en une langue romane comme le français, sont donc empruntés à trois substances différentes: la substance prosodique fournit les flexifs de nexies, qui, par leur constitution, sont des symboles et, en syntagme, sont extrathématiques; les flexifs thématiques, recherchant la substance phonémique, s'intègrent à des signes; la substance séquentielle, enfin, fournit un flexif thématique ayant constitution de symbole, mais se convertissant, avec le symbole prosodique qu'il régit, en signe à expression phonémique par le passage en nexus subordonné (inversion + intonation interrogative) > ob/si/czy). Les oppositions de contenu – et par là de contexte – qui, en polonais, font de *czy* une paire paradigmatique avec les autres interrogatifs thématiques, adverbes (kiedy, gdzie, etc.), substantifs (kto) et adjectifs (który) pronominaux, correspondent – abstraction faite de la sélection du prosodème – à celles qui existent en allemand et en français entre d'une part l'interrogatif thématique qu'est l'inversion (présupposant l'intonation interrogative) et d'autre part les interrogatifs pronominaux, adverbes, substantifs ou adjectifs.

Peuvent donc être membres d'un même paradigme à la fois des symboles et des signes, s'exprimant à la fois dans des substances aussi hétérogènes que le sont celle phonémique, celle prosodique et celle séquentielle.

Adresse de l'auteur: Professeur L. Flydal, Orionveien 9, *Oslo 4* (Norvège).

Discussion

Buyssens (Bruxelles): M. *Flydal* a rassemblé des faits bien connus pour les présenter sous un nouveau jour. En ce qui concerne les faits, je dois objecter qu'une question n'a pas toujours l'intonation montante. En ce qui concerne la théorie, je ne vois pas le moyen de considérer comme éléments de paradigme l'intonation interrogative et l'inversion, prèsqu'un paradigme ne réunit jamais des éléments ayant une même signification.

Rudnyckyj (Winnipeg) emphasized the "logical stress" in one of the phrases viz. *Byl on w domu?*

Meriggi (Pavia) bemerkt, daß im Deutschen Fragesätze mit oder ohne Inversion funktionell verschieden sind. Die echte einfache Frage erfordert die Umstellung. Wo diese unterbleibt, handelt es sich um eine rhetorische Frage oder eine solche mit besonderem Gefühlsausdruck (Staunen, Vorwurf usw.). Er fragt, ob im Polnischen dasselbe der Fall sei (und erhält eine negative Antwort).

Schubiger (Zürich): Während in der deutschen Entscheidungsfrage die Inversion nicht fehlen darf, wenn es sich um eine reine Frage handelt, ist in der französischen Umgangssprache die direkte Wortfolge viel üblicher als die Inversion.

Verh. 5. int. Kongr. Phon. Wiss., Münster 1964, pp. 281–286
(S. Karger, Basel/New York 1965).

Zur Gliederung der Satzmelodie

Von Ivan Fónagy

0. Wir studierten in den vergangenen Jahren im Sprachwissenschaftlichen Institut der Ung. Ak. Wiss. mit meiner Kollegin Dr. *Klara Magdics* auf Grund eines mannigfaltigen Materials (Theaterstücke, spontane Gespräche, Vorlesungen usw.) Form und Funktion der ungarischen Satzmelodie. Statt Teilergebnisse vorzuführen, möchte ich jetzt die seltene Gelegenheit ergreifen, einige Streitfragen in Gegenwart von hervorragenden Vertretern der verschiedensten Gebiete der Phonetik zur Diskussion zu stellen. Die syntagmatische und paradigmatische *Gliederung der Satzmelodie* liegt im Brennpunkt dieser Fragen.

1.1 In bezug der horizontalen, *syntagmatischen* Teilung der Sprechmelodie herrscht Unstimmigkeit. Einige behaupten, der Sprechtakt (*Palmer*, S. 13; *Kuhlman*, S. 11; *Pike*, S. 24) bzw. eine durch zwei Pausen begrenzte Lautreihe (*Jassem*, S. 46; *Hockett*, S. 44) sei die kleinste melische Einheit, andere erblicken sie im einfachen Satz (z.B. *Schubiger*, S. 9) oder im konstanten Tonschritt am Satzende (*Buning*, S. 91), wieder andere im «rhetorischen Syntagma» das oft länger als der Satz ist (*v. Essen*, S. 29). Manche behaupten, die Sprechmelodie sei unzerlegbar (*Buyssens*, S. 426).

Die gegensätzlichen Ansichten lassen sich gleichermaßen rechtfertigen. Denn: wäre das «rhetorische Syntagma», z.B. die Melodie einer Tirade, nicht einheitlich, wie könnte sie die Tirade als Einheit kennzeichnen? Wäre jedoch das «rhetorische Syntagma» eine homogene Einheit, wie könnte die Sprechmelodie die inbegriffenen Sätze und Wortgruppen begrenzen und kennzeichnen? Diese Betrachtungen führen uns zur Folgerung, daß die melische Einheit weitere Einheiten enthalte, daß sie daher eine *gegliederte Einheit* sei.

Die Schwierigkeiten und Widersprüche entstehen meines Erachtens auf Grund einer falschen Analogie zwischen phonematischer und melodischer Gliederung. Die konkreten Sprechlaute vertreten im wesentlichen einzelne Phoneme und tragen höchstens zur Kennzeichnung der Nachbarlaute bei. Eine Phonemreihe realisiert sich in einer entsprechenden Reihe von Sprechlauten. Im vollen Gegensatz zu dieser relativ klaren *Neben*ordnung steht die durchaus verwickelte *Hierarchie der Melodieformen* im Redeakt. Die konkrete Melodie einer Redeeinheit entsteht aus der Summierung von *ineinandergreifenden* und *überlagerten Melodieformen*. Die Gliederung der konkreten Sprechmelodie hat eine Art «Fourier-Analyse», die Auflösung der scheinbar einheitlichen Melodie in einzelne Komponenten zur Voraussetzung. Das Ergebnis dieser Analyse kann nur mit Hilfe eines *mehrschichtigen* Modells dargestellt werden (Abb. 1 und 2). Die einzelnen Melodieformen können sich auf eine Silbe beschränken oder

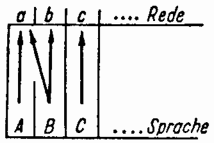


Abb. 1. Teilung der Lautreihe auf Grund einer verhältnismäßig einfachen Entsprechung der Sprechlaute und der Sprachlaute.

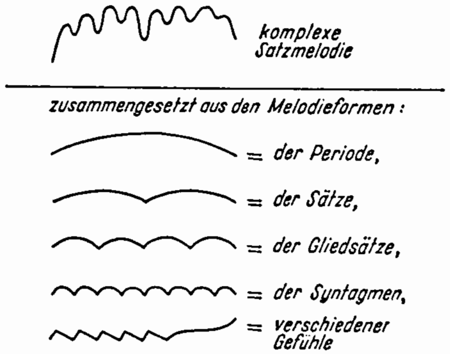


Abb. 2. Die konkrete Satzmelodie als ein Komplex von aufeinandergeprägten Melodieformen. Der hierarchische Aufbau der Redemelodie verhindert ihre unmittelbare Gliederung.

die ganze Redeeinheit erfassen. Ein Abschnitt dieser Redeeinheit ist infolgedessen Repräsentant einer einzelnen Melodieform und zugleich ein Teilvertreter anderer Melodieformen, die über diesen Abschnitt hinausreichen.

1.2 Die Melodie einer Redeeinheit (oder eines Abschnittes dieser Einheit) entsteht oft durch *Interferenz* von inhaltlich verschiedenen (modalen oder emotionellen) Melodieformen. Die konkrete Sprechmelodie des ungarischen Satzes: *Ott voltatok tegnap?* 'Wart

ihr gestern dort?' das als Frage und zugleich als ein Ausrufssatz empfunden wird, entsteht aus der Kreuzung der Frageform und der Melodieform des erstaunten Ausrufs (Abb. 3). Die Interferenz verschiedener Melodieformen ist durchaus kein Ausnahmefall. Das *Aufeinanderprägen* verschiedener Formen ist besonders für den künstlerischen Vortrag kennzeichnend (Abb. 4) (vgl. *Fónagy-Magdics*, S. 27–39)[1].

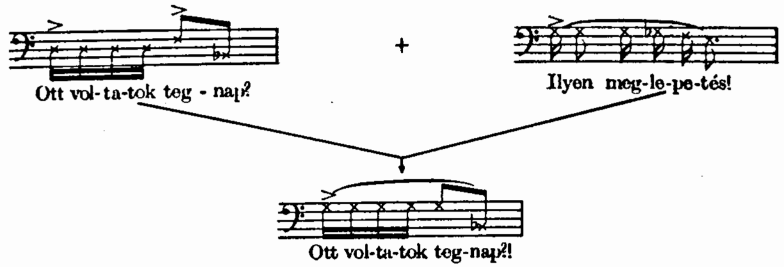


Abb. 3. Interferenz von einfachen Melodieformen: a) *Ott voltatok tegnap?* 'Wart ihr gestern dort?' (einfache Frage), b) *Ilyen meglepetés!* 'Solch eine Überraschung!' (Ausruf). Resultante: *Ott voltatok tegnap?!* 'Wart ihr gestern dort?!'.

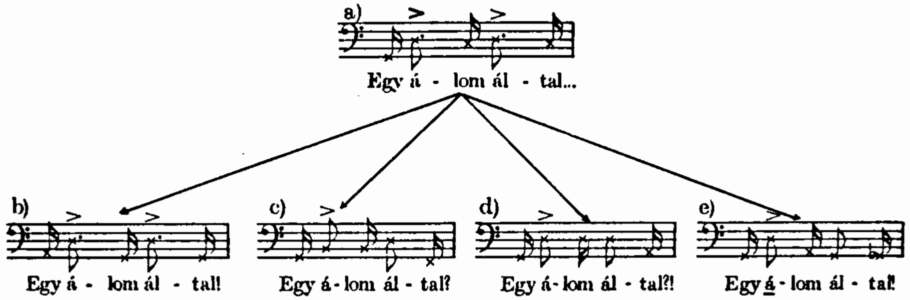


Abb. 4. Zerlegung einer kunstmäßigen Komplexform aus dem Hamlet-Monolog: a) «*Egy álom által...*» («und zu wissen, daß ein Schlaf...»); b) «Aber das ist ja reiner Unsinn!» Zurede; c) «Wäre es nur ein Traum?» Frage; d) «Ist das denn möglich?!» erstaunte Nachfrage; e) «Es ist bloß ein *Traum*» Hervorhebung.

1.3 Der Informationswert der Satzmelodie ist anderseits durch die *melischen Metaphern*, d.h. mittels der Vertretung einer Melodie-

[1] 20 Versuchspersonen verglichen unabhängig voneinander Satz für Satz den Hamlet-Monolog in der Interpretation von drei berühmten ungarischen Schauspielern und deuteten semantisch die als verschieden empfundenen Ausdrucksformen. Sie versuchten dabei den gehörten Satz formgetreu zu wiederholen, dies gelang aber nur im Fall von relativ einfachen, typischen Satzmelodien. An die Stelle der komplexen, künstlerischen Intonation traten stets die entsprechenden einfachen Formen. Die komplexe Satzmelodie wurde dadurch unwillkürlich in ihre Bestandteile zerlegt.

form durch eine andere erhöht. In Budapest verbreiteten sich neuerdings mit Fragemelodie gesprochene Aufrufssätze.

Einen gewissen Zunftcharakter verleiht dieser Metapher der Umstand, daß sie besonders für das Verkehrswesen typisch ist. So im Aufruf: *Tessék beljebb fáradni!* 'Bitte nur einwärts!' (Abb. 5). Die melische Metapher ist stets komplex: die konkrete Satzmelodie vertritt in solchen Fällen außer der tatsächlich realisierten Melodieform auch die erwartete, fehlende. Der Wert einer Metapher wird auf jeder sprachlichen Ebene durch den implizierten Zusammenhang des ersetzten und des ersetzenden Zeichens bestimmt (vgl. Abb. 6). In unserem Fall ergibt sich der Dynamik entsprechend die Bedeutung: «Bitte, ich frage ja nur eben» oder vielmehr: «Na wird's denn oder nicht?».

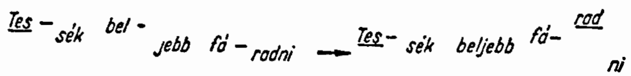


Abb. 5. Tessék beljebb fáradni! 'Bitte nur einwärts!'. Ein Aufrufssatz als Frage intoniert. Melische Metapher, die sich besonders im ungarischen Verkehrswesen allmählich verbreitet.

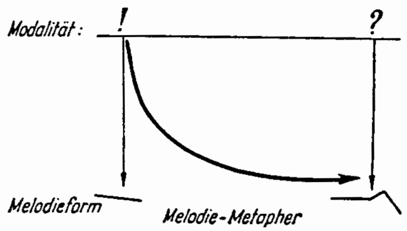


Abb. 6. Modell einer melischen Metapher (vgl. Abb. 5). Die Aufrufsform wird durch die Frageform vertreten.

2.1 Die Gliederung der Sprechmelodie ist in monotonischen Sprachen zugleich durch die unvollkommene *paradigmatische Gliederung* der Melodieformen erschwert. Das Phonemsystem besteht aus diskreten Einheiten, dies ermöglicht eben das Zerlegen der Redekette in Sprechlaute. Herrscht Unsicherheit in der Frage, ob eine gegebene Sprache Diphthonge hat, so wird auch die Gliederung der Lautreihe [auto] unbestimmt.

2.2 Der Melodieschatz wurde wiederholt als ein *Kontinuum* aufgefaßt, die Möglichkeit der paradigmatischen Gliederung bezweifelt (*Bolinger; Ebeling*, S. 54; *Martinet*, S. 30ff.). Die Frage, ob der ungarische Melodieschatz aus diskreten Einheiten bestehe (d.h. ob überhaupt von einem Melodiesystem die Rede sein kann), läßt sich nur mit einem *ja–nein* beantworten. Das Steigen der Melodie am Satzende verleiht der ungarischen Aussage ein affektives Gepräge. Es spiegelt je nach dem Grad des Aufstieges verschiedene Gefühle

(Staunen, Selbstrechtfertigung, Empörung usw.). Jedem wahrnehmbaren Stufenunterschied entspricht ein verschiedener emotioneller Gehalt. Insofern bestätigen auch unsere Ergebnisse *Bolingers* Behauptungen (S. 249) [2].

2.3 Hingegen ist der Gegensatz der morphematisch nicht gekennzeichneten Fragesätze einerseits, der übrigen Sätze anderseits, ganz eindeutig. Ähnliche Dichotomie der Melodieformen begegnet uns auch außerhalb des Gebietes der Modalität (*Fónagy-Magdics*, S. 17–22; vgl. *Bally*, S. 35ff.; *Lee*, S. 346; *Schubiger*, S. 90ff.).

Ja selbst im Kontinuum der emotionellen Melodieformen lassen sich Verdichtungspunkte, *Ansätze zur Gliederung* des Kontinuums aufweisen. Die statistische Verteilung der Intervalle der emotionellen Entscheidungsfragen ist höchst ungleich. Es lassen sich seltene und häufige, typische Intervalle unterscheiden.

So ist für die erstaunte Frage der Quint-Sprung besonders kennzeichnend, für die Verantwortlichmachung die verminderte Terz (vgl. Abb. 7a und b). Die mittleren Intervalle sind möglich, doch weniger wahrscheinlich. Dies ermöglicht eine beiläufige Unterscheidung von Typen und Varianten (*A. Martinet*, S. 37).

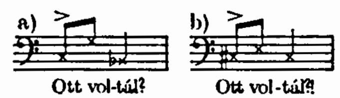


Abb. 7. a) *Ott voltál?* 'Warst du dort?' Verwunderte Frage. b) Verantwortlichmachende Frage.

3.1 Diese Ambivalenz der Melodieformen läßt sich auf einen grundlegenden Widerspruch zurückführen. Die Melodieformen der monotonischen Sprachen sind *konventionell* jedoch *nicht beliebig* (*Fónagy* 1956), sie sind sprachlich kodiert und beruhen gleichzeitig auf einem symptomatisch-symbolischen «natürlichen» Kode. Als «Lautgebärde» (*Martinet* «Gesturing with the glottis», S. 28) trägt die emotionelle Satzmelodie *ihre Bedeutung in sich.* Jede Melodieform ist auf Grund des kausalen Zusammenhanges von Ausdruck und Inhalt sinnvoll. Die natürlichen Anzeichen (Symptome und Symbole) bilden notwendigerweise ein Kontinuum, im Gegensatz zu den willkürlichen Zeichen, die ihre Bedeutung einzig und allein einer Übereinkunft verdanken (vgl. *Fónagy* 1956, *Meyer-Eppler*, *Ebeling*). Die willkürlichen Melodieformen der polytonischen Sprachen sind be-

[2] There is nothing in the nature of one pitch that distinguishes it from another, in the manner in which *m* is distinguished from *n* (p. 249).

grenzt an Zahl, und in den monotonischen Sprachen sind es stets die mehr willkürlichen Melodieformen, die zur Dichotomie neigen. Dem binären Gegensatz der Frage und Aussage steht die bunte Mannigfaltigkeit der emotionellen Ausrufssätze gegenüber (*Kaiser*, S. 293).

3.2 Das System der Melodieformen ist ein *System im Werden*, es steht auf einer Mittelstufe zwischen einer natürlichen Ausdruckssprache und einer beliebigen Zeichensprache («far periphery of the field of language», *Martinet*, S. 28). Sie eignet sich eben durch ihre Unvollkommenheit zur Mitteilung von seelischen Inhalten, die das Niveau des begrifflichen Denkens nicht oder kaum erreichen. Die menschliche Rede verdankt ihre Plastizität der Symbiose so verschiedener Kommunikationsmittel, wie das Phonemsystem und die Sprachmelodie.

Literatur

Bally, Ch.: Intonation et syntaxe. Cahiers Saussure *1* (1941).

Bolinger, D. L.: Intonation and analysis. Word *5* (1949).

Buning, E. J. and *Schooneveld, C. H. van:* The sentence intonation of contemporary standard Russian, as a linguistic structure (s'Gravenhage 1961).

Buyssens, E.: Development of speech in mankind. Manual of Phonetics, ed. *L. Kaiser* (Amsterdam 1957).

Ebeling, C. L.: Linguistic units (s'Gravenhage 1960).

Essen, O. von: Grundzüge der hochdeutschen Satzintonation (Ratingen 1956).

Fónagy, I.: Über die Eigenart des sprachlichen Zeichens. Lingua *6:* 67–88 (1956).

Fónagy, I. und *Magdics, Kl.:* Das Paradoxon der Sprechmelodie. Uralaltaische Jahrbücher *35* (1963).

Hockett, Ch. F.: A manual of phonology (Baltimore 1955).

Jassem, W.: Intonation of conversational English (Wrocław 1952).

Kaiser, L.: Psychologic and linguistic value of melody. Acta Psychologica *9:* 288–293 (1953).

Kuhlmann, W.: Tonhöhenbewegung im Englischen. Z. Phon. *5:* 203–224 (1951).

Lee, W. R.: English intonation: A new approach. Lingua *5:* 345–371 (1955–56).

Martinet, A.: Functional view of language (Oxford 1962).

Meyer-Eppler, W.: Grundlagen und Anwendungen der Informationstheorie (Berlin 1959).

Palmer, E.: A grammar of spoken English (Cambridge 1929).

Pike, K.: Intonation of American English (Ann Arbor 1953).

Schubiger, M.: English intonation, its form and function (Tübingen 1958).

Adresse des Autors: Ivan Fónagy, II Gábor Áron u. 1/b, *Budapest* (Ungarn).

Proc. 5th int. Congr. phon. Sci., Münster 1964, pp. 287–291
(S. Karger, Basel/New York 1965).

From the Department of Phonetics, University College London

A Note on the Spectral Analysis of Unvoiced Sounds

By A. J. Fourcin

The aim here is to describe a particular frequency spectrum analysis technique which can gather information for a closer approximation to a representative sample of an unvoiced sound than is normally available with a sound spectrograph (the Kay Sonagraph) and to illustrate some of the initial results obtained from the application of this analysis technique to the assessment of the sounds produced from a self excited acoustic model of the vocal tract.

Possibly a primary difficulty in the analysis of speech arises from our relative ignorance of the characteristics of the source of excitation applied to the vocal tract. For voiced sounds this difficulty is somewhat tempered by the quasi-periodic nature of the excitation, which leads to a useful degree of spectral coherence, and the well defined physical position of the source. For unvoiced sounds however the excitation is temporally random, which hampers spectral analysis, and often at least partially distributed in position along the vocal tract, which hinders structural interpretation.

The analysis of random signals has received a great deal of attention in other fields and the theoretical basis for their adequate spectral analysis is well established. The result to be expected in the simple case, for example, of a noise of gaussian amplitude distribution applied to a rectangular filter driving a linear rectifier followed by a finite integration has been given succinct expression by *Bennett* (1960, p. 149) who shows how the analysis tends to become more consistent as the filter bandwidth and the integration time are increased. For speech analysis *Fant, Fintoft, Liljencrants, Lindblom* and *Mártony* (1963) have considered the properties of a practical bandpass filter the effective output of which is averaged by a low pass filter. In the Kay Sonagraph however there is essentially no post filtering averaging either by integrator or low pass filter. In consequence the spectrum, or section, obtained from the frequency analysis of any unvoiced sound whose steady state duration is greater than the response time of the analyzing filter is less representative than need be the case and the spectral envelope of the analysis made by the machine is more disturbed than is theoretically necessary. An improvement can be obtained simply by averaging the filter output, before employing it in the sectioning circuit, for the maximum time permitted by the stability of the sound to be analyzed.

A simple arrangement for making this averaging possible is shown in figure 1. Once in every drum revolution an integrator is reset to zero by a short circuiting contact which can be placed at any point around the periphery of the drum base to discharge the capacity, C. After this resetting, integration of the analyzing filter output commences and the instant at which the normal sectioning microswitch is actuated determines the end of the integration interval. In this way an integration time between less than 0.1 sec and up to about 2.3 sec can readily be obtained. The only additions to the basic Sonagraph are the resistance R and the capacity short circuiting arrangement. It is an advantage, however, to use a more sophisticated integrator having a gain which may be adjusted to correspond to the integrating period and this has been employed to derive the lower spectra of figure 1, the integration intervals for which are shown below the spectrogram of the parent sound sequence. A more extreme indication of the improvement in spectral delineation which this post filtering integration affords is given in figure 2 where the ordinary narrow band sections for two fricatives are shown above and the corresponding long term spectra (2 sec integration intervals) are shown below. Narrow band analysis has been used for all the examples in the present note and this sets a limit to the degree of spectral smoothing which may be obtained, unless the speech sound analyzed contains spectral features of comparable or smaller bandwidth.

It might appear from the clarity of these spectra that this technique could profitably be employed for the determination of vowel formant frequencies; since the smooth spectral form which may be associated with a breathed excitation could overcome the disadvantage accompanying the discrete frequency spacing of voiced harmonics. This is correct for the case in which the breathed excita-

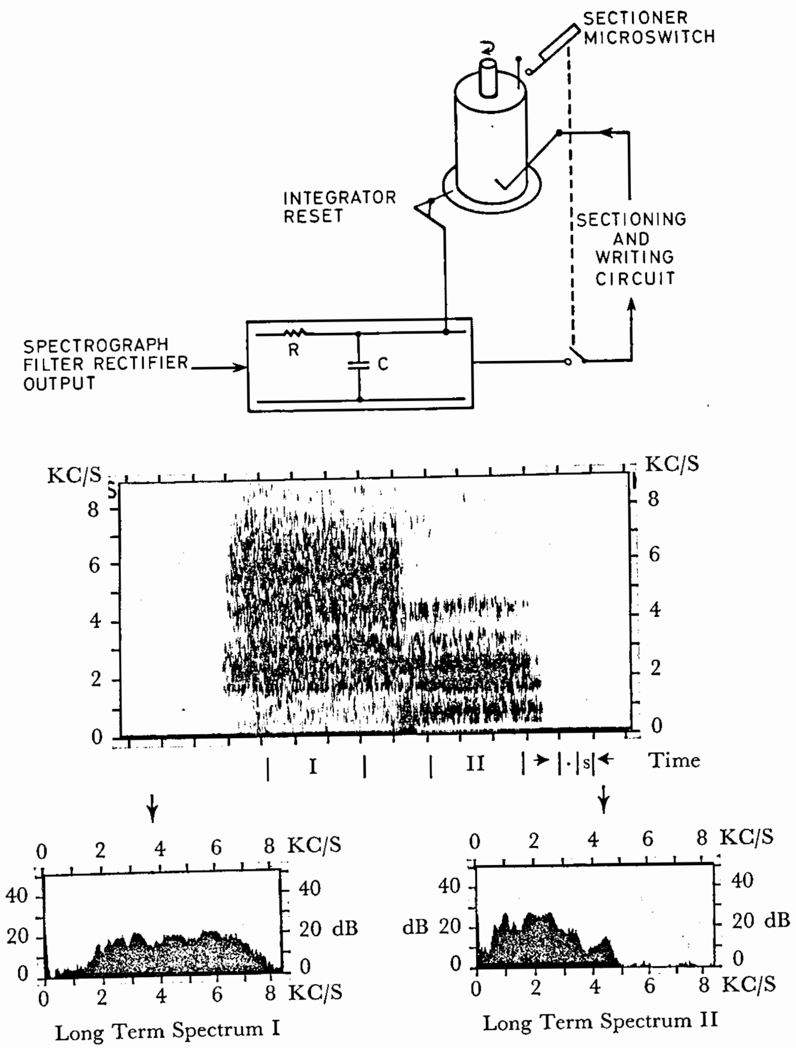


Fig. 1

tion is located at the glottis but when, as is often the case, friction is present along an appreciable length of the vocal tract the vowel spectrum will be seriously deformed. Results which are of more immediate interest can be derived from the analysis of unvoiced sounds which are associated with a single well defined source of turbulence in the vocal tract, such as those whose analyses are shown in figure 2. These two sounds were each produced with the same degree of lip rounding and an attempt was made to keep the tongue-alveolar

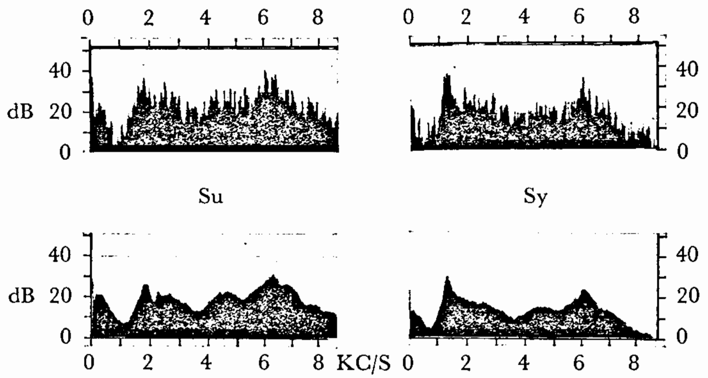


Fig. 2. Normal and Long Term Spectra.

constriction the same although the position of the back of the tongue was altered. It is evident from the long term spectra that the alteration of the pharyngeal cavity has an appreciable influence on the sound emerging from the lips. If the excitation of the vocal tract is entirely due to the sound produced by turbulence at the alveolar constriction it follows that the pharyngeal cavity is acting to absorb acoustic energy which would otherwise appear at the mouth.

The situation may be explored in a more controlled fashion by the use of acoustic models of the articulatory configuration involved in which the passage of an air stream through an appropriate internal constriction produces the acoustic excitation. In figure 3 (a) is shown a lateral medial view of an actual vocal tract in the configuration corresponding to a palatalized palato-alveolar fricative (taken from *Fant*, 1960, p. 170) and in figure 3 (b) is the outline of a corresponding planar acoustic model. This model has parallel transparent rigid cheeks spaced by 2 cm, so that the separation between the walls – which are curved in only one plane and, here, made of plasticene – is directly proportional to the cross sectional area of the corresponding section in the human original. The constriction was 1 cm long and of 0.1 cm² cross sectional area. The long term spectral form of the output from the lips when an air stream is applied from a simulated glottis is shown in figure 3 (c); an integration time of 2 sec was employed. In figure 3 (e) the same model has been modified to diminish the volume of the pharyngeal cavity whilst keeping the remainder of the artificial vocal tract unchanged. Once again the long term spectrum of the output, figure 3 (f), integration time of 2 sec, has been altered and a distinct spectral trough has been

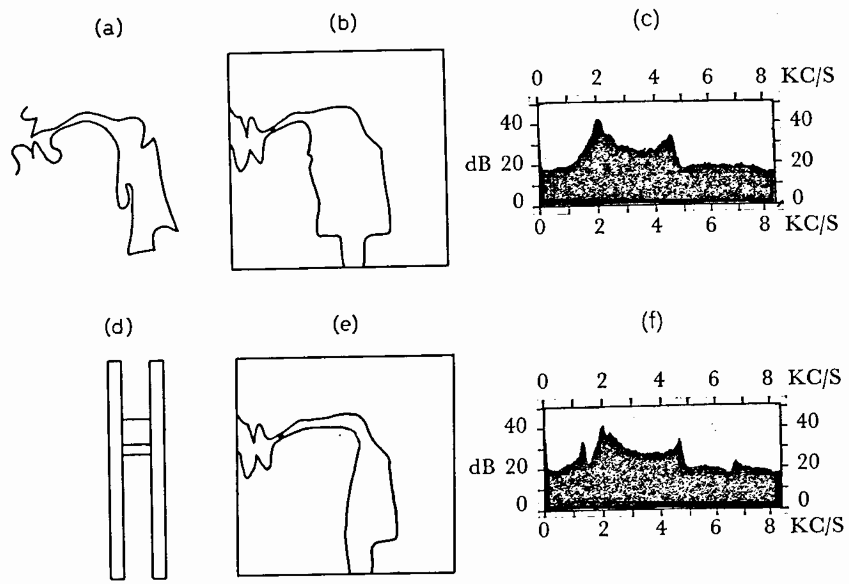


Fig. 3

introduced. Since one can now be sure that the constriction and the cavity arrangement which follows it are unchanged, this change in spectral form must be related to pharyngeal modification.

These spectral results could have been obtained by other means but it is of some practical convenience even with models to employ the same means of analysis that is used for speech, and with speech it is essential to be able to choose the integration time interval.

Acknowledgement. I have derived benefit in the preparation of this summary from discussion with Mr. *D. C. Bennett* and other colleagues in the Department of Phonetics and have been very much helped in the experiments by Mr. *J. E. West.*

References

Bennett, W. R.: Electrical noise (McGraw-Hill, New York 1960).
Fant, G.; Fintoft, K.; Liljencrants, J.; Lindblom, B. and *Mártony, J.:* Formant amplitude measurements. J. acoust. Soc. Amer. *11:* 1753–1761 (1963).
Fant, G.: Acoustic theory of speech production (Mouton, 's-Gravenhague 1960).

Author's address: A. J. Fourcin, University College London, Department of Phonetics, Gower Street, *London W.C. 1* (England).

Discussion

O'Connor (London): Would the material of which the teeth of the model are made have any effect upon the resulting friction for, say, an [s]?

(Mr. *Fourcin* thought not; and that the constriction itself would be responsible.)

C.R. 5e Congr. int. Sci. phon., Münster 1964, pp. 292–295
(S. Karger, Basel/New York 1965).

Perception et reproduction des traits pertinents

Par J. FOURQUET

L'association du magnétophone à plusieurs pistes et de la cabine insonorisée a suscité des méthodes d'enseignement des langues étrangères fondées sur le principe suivant: l'élève, après avoir entendu une phrase modèle, essaie de la reproduire; cet essai est enregistré sur une seconde piste, et l'élève peut entendre alternativement le modèle et son essai, et corriger sa prononciation jusqu'à reproduction parfaite.

Cette méthode repose sur le postulat que la seule cause d'erreur est dans le fait que l'élève n'a pas encore les habitudes *motrices* propres à la langue en question. On ne se pose pas le problème de la *perception.*

Plusieurs observations nous ont donné à réfléchir: il semble bien que l'habitude d'une langue crée chez l'adulte une sensibilité particulière pour les caractères sonores qui, dans cette langue, ont une fonction distinctive, et une sorte d'indifférence pour les autres. Le débutant n'*entend* pas, dans une langue étrangère, ce qui n'a pas une valeur distinctive dans la sienne, et n'essaie donc pas de le reproduire.

La regrettée *Marguerite Durand* a fait entendre à des Français, lors d'une séance de la Société de Linguistique, des mots géorgiens contenant les trois séries d'occlusives de cette langue; les différences qui séparent ces trois séries leur échappaient à l'audition, alors que les sonagrammes font apparaître à la vue des différences sensibles, qui assurent largement la distinction pour un membre de la communauté géorgienne.

D'après les observations de Mlle *M. Philipp*, un sujet alsacien perçoit dans le couple de mots français *vide: vite* une différence de durée de la voyelle, mais ne perçoit pas de différence entre *d* et *t*. C'est que sa langue I a une opposition de quantité vocalique tandis qu'elle ne possède qu'une seule série d'occlusives, des douces sourdes. Le francophone perçoit la différence entre *d* et *t*, mais n'a pas

conscience d'une différence de durée de la voyelle *i:* il faut la lui montrer sur un kymogramme.

Il semble bien qu'il s'agisse ici de faits de perception, qui relèvent de la «Gestaltpsychologie». L'aptitude bien connue de l'enfant à acquérir la phonétique d'une langue étrangère s'expliquerait dès lors par une sensibilité intacte, non encore sélective, pour tous les caractères sonores; les adultes qui conservent le mieux cette sensibilité auraient le plus de facilité pour acquérir une bonne prononciation de la langue étrangère.

S'il en est ainsi, on ne peut attendre du débutant adulte qu'il se corrige lui-même; il n'entendra pas de différence entre sa prononciation et le modèle, si cette différence porte sur un *trait pertinent* qui n'existe pas dans sa langue. Il faudra, pour l'avertir de cette différence un moniteur, qui, lui, la perçoit, parce qu'il a acquis le système phonologique étranger, ou le possède d'enfance.

Pour aider l'élève à surmonter la difficulté, le moniteur lui répétera le modèle en insistant sur le trait manquant, en l'exagérant; ou bien il essaiera d'amener l'élève à prendre la position articulatoire correcte, susceptible de produire l'effet acoustique voulu: cette seconde méthode est analogue à celle qu'on emploie pour faire parler un sourd-muet. Il s'agit de faire émettre quelque chose que le sujet ne perçoit pas.

On entrevoit un moyen d'accroître l'efficacité du self-teaching par magnétophone: il consiste à amplifier artificiellement les traits que le débutant ne perçoit pas, parce qu'ils n'existent pas dans sa langue I comme éléments distinctifs, c'est-à-dire à compenser la sensibilité inégale selon que le trait distinctif existe ou n'existe pas dans la langue I.

Le progrès de la synthèse de la parole devrait permettre bientôt des expériences d'amplification des traits non perçus ou imparfaitement perçus; on donnerait ainsi au débutant des modèles artificiellement *compensés*, par une amplification sélective, pour l'amener peu à peu au modèle normal. La parole synthétique a des possibilités d'exagération que n'a pas le moniteur humain: certains traits se laissent mal amplifier, du fait de la nature des organes de la parole. La mise au point d'une telle méthode suppose évidemment une analyse phonologique très approfondie des deux langues en cause, langue de départ et langue à enseigner.

En France, certaines expériences qui touchent dans une certaine mesure à ces problèmes ont été faites par le Dr *Tomatis*. Spécialiste de la pathologie de l'oreille, le Dr *Tomatis* a d'abord étudié les altérations de la parole corrélatives à un trouble de l'audition ou à une surdité partielle; l'émission est constamment contrôlée par l'audition: si ce contrôle est défaillant, l'émission est affectée.

Les expériences consistent à faire entendre au sujet sa propre voix par un casque d'écoute, en modifiant le niveau d'intensité respectif du grave, du médium et de l'aigu: pour l'observateur, le sujet change d'«accent»; il parle autrement, pour reproduire l'effet acoustique qu'il considère comme normal, et en croyant parler normalement d'après ce que lui fait entendre le casque. Le Dr *Tomatis* propose son appareil comme un moyen d'acquérir «l'accent» dans une langue étrangère.

Il aboutit à une théorie d'après laquelle à chaque communauté linguistique correspondrait un audiogramme spécifique, une répartition spécifique de la sensibilité aux divers niveaux de fréquence: nous ne serions plus ici dans le domaine de la *perception*, mais dans celui de la sensibilité physiologique. Les idées du Dr *Tomatis* sur l'origine de «l'oreille nationale» nous semblent contestables. Mais on ne peut exclure a priori l'hypothèse que la partie des organes auditifs qui «travaille» le plus dans le cas d'une langue donnée ne devienne plus sensible. Il serait assez facile d'établir à partir d'un corpus enregistré la quantité d'énergie émise sur chaque bande de fréquence. Il se peut d'autre part que la simple amplification relative d'un niveau de fréquence donne une première approximation grossière de l'amplification compensatrice dont il a été question plus haut: une zone où la langue se trouve avoir des traits distinctifs spécifiques se trouve renforcée.

Cependant il nous semble que la recherche devrait être abordée pour commencer au niveau de la *perception*, et au niveau des traits distinctifs de chaque *phonème*. Ceux-ci ne correspondent pas seulement à une répartition de l'énergie, mais à des faits de durée, de vitesse de variation, très complexes. C'est de la perception de ces traits que dépend la formation des habitudes motrices, et le contrôle de l'émission par l'audition.

Adresse de l'auteur: Prof. J. Fourquet, Université de Paris, Faculté des Lettres et Sciences Humaines, 17, rue de la Sorbonne, *Paris* (France).

Discussion

Nickel (Kiel): In Ergänzung zum mundartlichen Beispiel im Französischen, wo bei *vite* – *vide* die Phonemopposition t–d zugunsten von d aufgegeben wird und die Differenzierung in der Form antizipiert wurde, daß das i einmal lang und einmal kurz artikuliert wird, brachte ich das Beispiel *writer* – *rider*, wo in amerikanischen Dialekten ähnliches vor sich geht. Auch dort wird die Quantität des Diphthongs zur 'Ersatzdifferenzierung' benutzt. Dadurch wird auch klar, daß die Stabilität der Phoneme verschiedenen Grades ist: d–t scheinen besonders häufig gefährdet.

Proc. 5th int. Congr. phon. Sci., Münster 1964, pp. 296–299
(S. Karger, Basel/New York 1965).

Realization of Phonemes and Linguistic Norm

By G. Francescato, Utrecht

According to a paper by two distinguished scholars, "the central problem which confronts the science of linguistics in its attempt to describe the phonic aspect of language... is the nature of the relationship between phonological entities and sound" (*Mol* and *Uhlenbeck*, 1959[12], p. 161). If we discard – as the authors rightly do – the attempts to solve the problem either by a purely phonetic, or by a purely mathematical (in *Hjelmslev's* sense) approach, we are obliged to account in some other way for the difficulties involved in the identification of phonemes, on the one hand, and of phones, on the other. Of course, we admit the possibility of arriving at a satisfactory analysis of speech sounds by means of various techniques, starting from auditory, and ending (with the aid of modern technology) with acoustic analysis. Phonetic science has long been satisfied with the articulatory approach; only in relatively recent times, with the development of the necessary technical means, has a need for deeper knowledge been felt. As for phonemicists, extending their descriptions to more and more languages, they seem to imply that practical results are relatively easy to achieve. In both cases, however, one has the feeling of working in some sort of laboratory situation, where conditions are to be stated in advance (*Mol* 1954[12], p. 169).

What then will occur, when – as the authors quoted add a little later – "instead of deducing a distinctive function from speech situations which rarely occur, we prefer to start from what happens in everyday speech" (*Mol* and *Uhlenbeck*, 1959[12], p. 165)? They end up by suggesting that "recognition and identification of words... is not guaranteed by a certain phonic invariance of phonemes, but rests... upon the interpretative faculty of the listener (*Mol* and *Uhlenbeck*, 1959[12], p. 161).

In the present report I will try to investigate to what extent this "interpretative faculty" of the listener can be expected to work. It hardly needs to be said that we are rather scantily informed about the psycho-physiological operations of the ear and of the parts of the brain connected with it, when we try to consider in detail the way a listener "hears" the spoken utterances. The study of the listener, with the cooperation of physiologists, psychologists and phoneticians, has only just begun (*French*[8], p. 65 ss.), while researches in the past have been mainly concerned with the speaker.

We can however discover many cases in which the criterion of "phonematic interpretation" is insufficient to explain how the relation between sound and phoneme works. Usually speech sounds are thought of as variants, as against the phonemes, which are defined as invariants (*Dieth*[3], p. 336; *Grammont*[9], p. 9, and *Jakobson*[11], p. 231), but it is possible to list some situations, where phonemic invariance cannot explain by itself the identification of the phonetic entities involved.

1. Different sounds which have no phonematic value, can be accepted as realizations of the same phoneme and still be kept apart as symptoms of different "styles" (in which speed of utterance, emotional state of the speaker, etc., play a role) (e.g. the realization of various english phonemes in the pronunciation of 'seven' as registered by *Lane* (*Dieth*[3], pp. 334–335).

2. Different sounds, occurring as positional variants of the same phoneme, can be sharply identified (e.g. the German phoneme /x/ realized either in the *ich* or in the *ach* type; confusion of the two types is unacceptable* (*Heffner*[10], p. 66).

3. Different realizations, with phonemic relevance, can occur in the same place in certain word-forms, according to regional or social habits (e.g. a certain number of Italian words can be pronounced either with stressed *o* or *ɔ*; the difference between these vowels is supposed to be phonemic (*Francescato*[4], pp. 118, 121–122), but in practice both pronunciations are accepted, and act as a criterion for social or regional identification of the speaker).

4. Differing realizations of different phonemes can occur in the same total distributions ('diasystem'), according to regional or social habits [e.g. Friulian *fûk*, *fôk*, *fóuk*, *fúek*; these realizations are

* An analogous example is represented by the lowering of the vowels in Danish before *r*; but the lack of the positional variant seems not so hurting for the natives' feeling in this case, as in the other.

considered local idiosyncrasies (*Francescato*[5])*; the same holds for Friulian *s*, *ś*, *ŝ* in their various distributions (*Francescato*[7])].

5. The realization of certain phonemes is acceptable only when close enough to a certain phonetic type, even if there is no phonematic motive for this limitations [e.g. modern Serbo-Croatian, with a five-vowel system, admits as normal the realization of the mid vowels only in the open range, although there is no contrast with a close range (*Schmaus*[14], p. 8)**].

These situations are as a rule dealt with under the label of 'norm'. In the usual acceptation 'norm' means "what is in accord with a certain standard". Linguists use it to mean the "constant forms which are partially non-functional", as against the "constant forms which are always functional", viz. the phonemes (*Coseriu*[2], p. 208 ss.). We have seen that even the functional forms cannot always be identified as constants. It seems unavoidable that, to identify correctly the phonic entities involved in the situations listed, some sort of phonetic interpretation has to be recognized.

On the other hand, we are well aware of the fact that many important clues about social or emotional factors, interfering with linguistic usage, are gained by the listener just by comparing what he hears with some sort of ideal 'norm' he has in mind. Our everyday experience suggests that the commonest form of phonetic teaching consists precisely in offering 'normal forms', to which the speaker is supposed to adapt himself. Furthermore, the acquisition of the native phonologic system could not be understood, if we would not admit that there is some sort of 'norm', to which adult speech is accommodated, and to which children look, in order to establish their own linguistic habits***.

Communication by means of language is, to a certain extent, an interplay of guesswork between speaker and listener (*Mol*[13], p. 23), but the clues offered by phonetic elements probably play a more important part than recent research seems inclined to attribute them (*v. d. Berg*[1], p. 50 ss.).

* An interpretation in terms of 'diasystem' is suggested in my paper *Dialect borders and linguistic systems*, s. Lit. 6.

** Cf. *R. Jakobson*, TCLC *V:* 209 (1949); *Hodge*, Language *22:* 112 (1946).

*** Up to a certain age, children do all by themselves to learn their mother tongue and only on later stages adults can play actively upon the speech of children by means, of correction.

References

1. *Berg, B. van den:* Fonetiek en phonologie; in Taalonderzoek in onze tijd (Servire, Den Haag 1962).
2. *Coseriu, E.:* Forma y sustancia en los sonidos de lenguaje (Revista de la Facultad de Humanidades y Ciencias, Montevideo 1954).
3. *Dieth, E.:* Vademecum der Phonetik (Francke, Bern 1950).
4. *Francescato, G.:* Notes on relevant features with low functional yield. Lingua *11:* 118–127 (1962).
5. *Francescato, G.:* La dittongazione friulana. Ital. dialett. *23:* 43–54 (1959).
6. *Francescato, G.:* Dialect borders and linguistic systems. Proc. IXth int. Congr. Linguistics, Cambridge 1963, pp. 529–534.
7. *Francescato, G.:* Il sistema dei suoni sibilanti nel dominio friulano. Rev. linguist. rom. *26:* 51–70 (1962).
8. *French, N. R.:* Auditory considerations, in *L. Kaiser*, Manual of phonetics (North Holland, Amsterdam 1957).
9. *Grammont, M.:* Traité de phonétique (Delagrave, Paris 1956).
10. *Heffner, R. M. S.:* General phonetics (The Univ. of Wisconsin Press, Madison 1949).
11. *Jakobson, R.:* Selected writings (Mouton, Den Haag 1962).
12. *Mol, H.* and *Uhlenbeck, E. M.:* Hearing and the concept of the phoneme. Lingua *8:* 161–185 (1959). – The analysis of the phoneme in distinctive features. Lingua *4:* 167–193 (1954).
13. *Mol, H.:* Geloof en bijgeloof in de fonetiek (Mouton, 's-Gravenhage 1960).
14. *Schmaus, A.:* Lehrbuch der serbischen Sprache (Belgrad 1944).

Author's address: Prof. Dr. G. Francescato, Courbetstr. 20 I, *Amsterdam* (Holland).

Verh. 5. int. Kongr. Phon. Wiss., Münster 1964, pp. 300–305
(S. Karger, Basel/New York 1965).

Opposition und Distribution von oberdt. *st/št* in diachronischer Sicht: Zur Entwicklungsgeschichte von deutschmundartlich *fešt* ‚fest'*

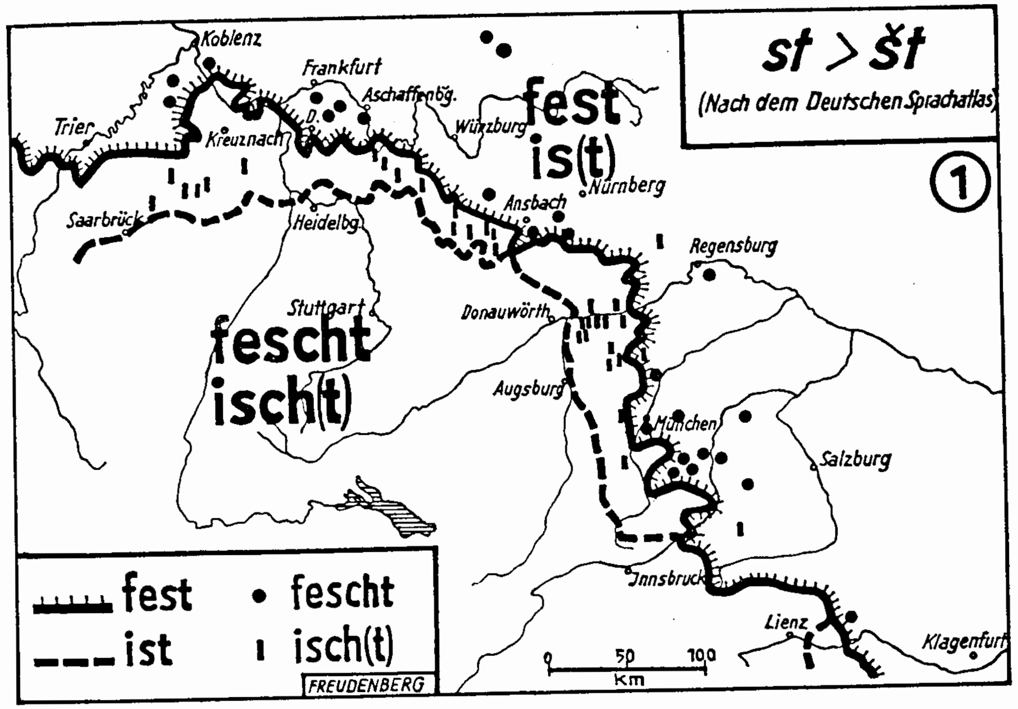
Von R. Freudenberg, Marburg a. d. Lahn

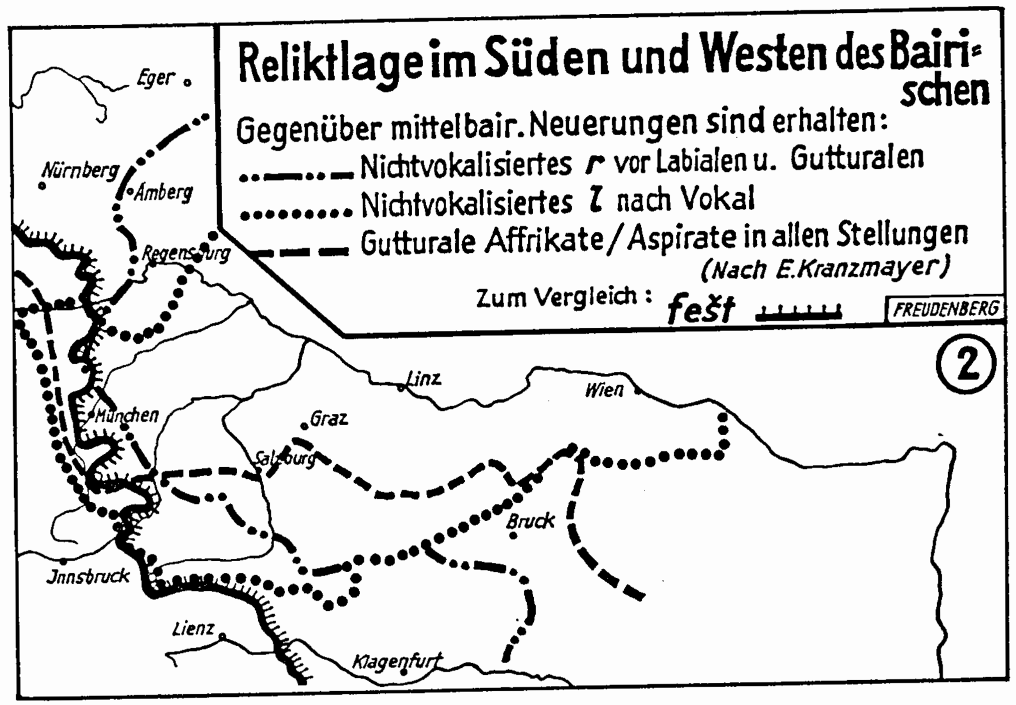
Ein auffälliges Merkmal der Mundarten im Südwesten und Süden des kontinentalgermanischen Sprachbereichs ist die Aussprache [*št*] für hochsprachliches [*st*] im In- und Auslaut: So wie *Schiller* nach dem Bericht des Weimarer Schauspielers *Genast* «meischterhaft» und «so ischt's recht» gesprochen hat[1], ist es selbst in der gehobenen Sprachstufe des sogenannten «Honoratiorenschwäbisch» noch heute der Brauch. In der dialektologischen Fachliteratur wird diese Erscheinung gewöhnlich mit dem Paradigma «*fescht*» der Sprachatlas-Karte 'fest' (DSA 23.43.75) bezeichnet und dem Alemannischen als Kennform zugeschrieben, was der dialektgeographische Befund auf den ersten Blick zu bestätigen scheint (Karte 1 zeigt den Verlauf der Isoglossen nach dem Material des Deutschen Sprachatlas). Bei genauer Betrachtung fällt jedoch auf, daß [*št*] schon auf reichsdeutschem Gebiet weiter ins Bairische hineinreicht als die sonstigen uns bekannten ostlechischen Alemannismen; und in der Tat haben die bisher am Lechrain vorgenommenen regionalen Untersuchungen [‿*št*] stets aus dem Problembereich der schwäb.-bair. Übergangserscheinungen ausgeklammert[2]. Zieht man die DSA-Ergänzungskarte für Österreich zu (was bisweilen bei der Grenzbeschreibung versäumt wird), so findet man [*fešt*] nicht nur

* Aus dem Forschungsinstitut für deutsche Sprache (Deutscher Sprachatlas). Marburg/Lahn (Direktor: Prof. Dr. *L. E. Schmitt*).

[1] ZdU *8:* 548 (1894).

[2] *Kranzmayer, E.:* Die Schwäb.-Bair. Mundarten am Lechrain. München 1927; *Bohnenberger, K.:* Über die Ostgrenze des Alemannischen. In: PBB *52:* 217ff. (1928).





in Westtirol, wo alem. Formen noch heute in größerer Zahl belegt sind, sondern noch wesentlich weiter östlich bis nach Kärnten hinein. Diese räumliche Lagerung macht die Annahme eines alem. Vorstoßes vollends unwahrscheinlich und legt für das Bairische eine andere Deutung nahe: Wir haben es hier mit einer für das Südbairische und den östlichen Lechrain typischen Reliktlage zu tun, die, abseits der mittelbair. Verkehrslandschaft im Städtedreieck Wien–Regensburg–München, ältere bair. Sprachzustände bewahrt hat: Man vergleiche etwa die heutige Randlage bei der Bewahrung der gutturalen Affrikate und der Liquiden (Karte 2). Diese Vermutung wird durch historische Zeugnisse bestätigt: *E. Kranzmayer* verdanken wir eine Anzahl wertvoller direkter und indirekter Belege für einstiges in- und auslautendes [*št*] im Mittelbairischen[3] (Karte 3). Damit ist nachgewiesen: [*št*] *im In- und Auslaut ist eine ursprünglich gesamtoberdeutsche Erscheinung.* Von hier aus verliert übrigens die von romanistischer Seite[4] gemachte Beobachtung, daß die [*fešt*]-Grenze in ihrem heutigen Verlauf in den Ostalpen genau auf die roman.-slaw. Sprachgrenze trifft, doch an Gewicht, und die daran geknüpften, an sich höchst anregenden Vermutungen von dt.-roman. Wechselbeziehungen werden fragwürdig.

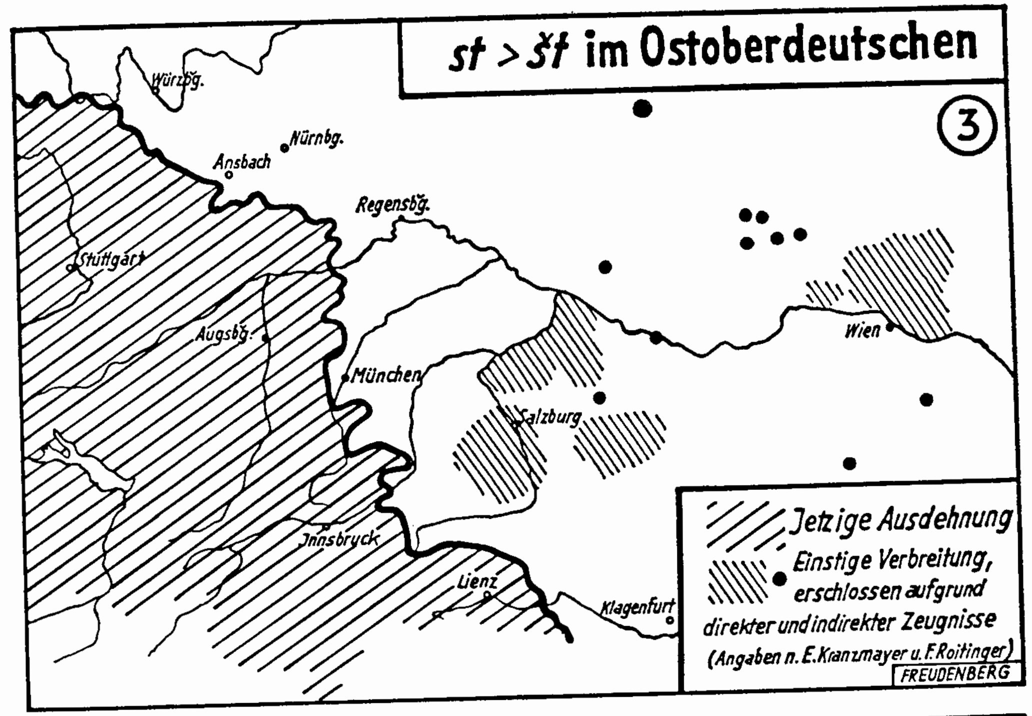
Wesentlich schwerer zu beantworten ist die Frage nach dem Alter dieser oberdt. Lautung. Die ältesten *scht*-Schreibungen reichen kaum über das 15. Jahrhundert zurück, setzen also erst ein, als das Mittelbair. schon weitgehend zu [*st*] geneuert hatte – übrigens eine plausible Begründung für das Fehlen direkter schriftlicher Zeugnisse aus diesem Gebiet. Dennoch hat es nicht an Versuchen gefehlt, die Entstehung von [*št*] im In- und Auslaut zeitlich hoch hinauf bis in die (alem.) Stammesgeschichte zurückzudatieren[5], und in der Tat legen die eben besprochenen bair. Vorgänge ein höheres Alter nahe. Immerhin bleibt zu fragen, warum diese für uns heute so ohrenfällige Erscheinung erst im 15. Jahrhundert ihren graphischen Niederschlag gefunden hat[6]. Im folgenden soll eine knapp andeu-

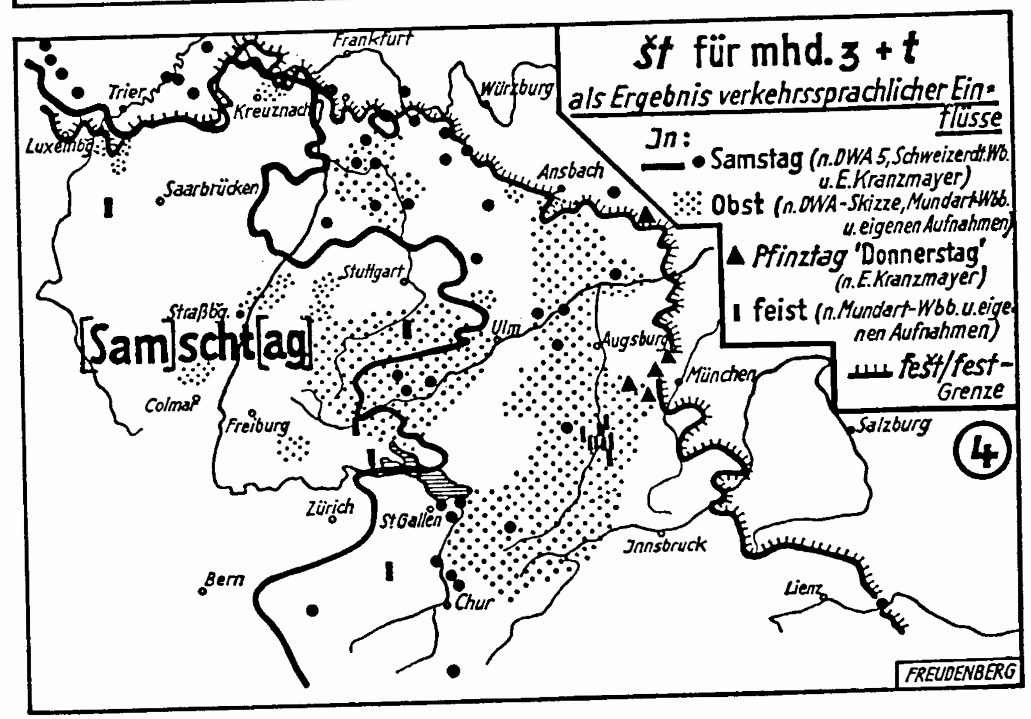
[3] *Kranzmayer, E.:* Historische Lautgeographie des gesamtbair. Dialektraumes (Wien 1956), S. 90. Vorher schon *Roitinger, F.* in: ZMaf *22:* 199 ff. (1954).

[4] *Schmid, H.: sp st sk > šp št šk* in roman. und dt. Mundarten. In: Vox Romanica *15, 2:* 30 ff. (1956).

[5] *Mitzka, W.* in: Dt. Phil. im Aufriß I², 1608. Anders: *Frings, Th.:* Sprache und Geschichte I, S. 69 (Halle 1956).

[6] Daß *scht*-Schreibungen in allen Positionen in der Überlieferung immer Ausnahmen darstellen, ist ein Problem für sich.





endet Begründung aus dem phonologischen System der Anblase- oder Zischlaute heraus versucht werden[7].

Die Sprachstufe des Vorahd. begnügt sich noch mit den Anblase-Phonemen /*s*/ und /*ss*/. Ein Merkmal-Gegensatz «spitz: breit» fehlt noch ebenso wie schon im Germanischen und Indoeuropäischen, so daß vorahd. /*s*/ ohne weiteres über diesen Gegensatz hinweg realisiert werden konnte.

Die ahd. Verhältnisse werden durch den Lautverschiebungsakt *t* > *ʒʒ* bestimmt. Daß ahd. *ʒ* dorsal gesprochen wurde ([*ʂ*][8]), ist für diese Periode durch die Lehnwortstudien von *E. Schwarz* und *P. Lessiak* gesichert. Gleichzeitig wurde das zuvor artikulatorisch «freie» /*s*/ in seiner Realisation auf eine apikale Zungenstellung ([*ś̯*]) beschränkt und erhielt dadurch einen «*š*-ähnlichen» Klangcharakter. Wir nehmen nun an, daß im Oberdt. der ahd. Zeit jedes /*s*/ vor Verschlußlauten nicht nur als [*ś*], sondern geradezu als [*š*] realisiert wurde. Das phonologische System war von dieser Artikulationseigenheit nicht betroffen, und eine lautgerechte Schreibung war deshalb nicht nötig (und übrigens auch nicht möglich).

Als nächste Stufe folgt der Lautwandel /*sk*/ > /*š*/. Phonetisch muß nach dem oben Gesagten für das Oberdt. [*šk*] > [*š*] angesetzt werden, wofür insbesondere die bair. Forschung einige überzeugende Beweise geliefert hat. Da die Schreibung diesen Lautwandel nur zögernd und mit den unzulänglichen Mitteln des lat. Alphabets wiedergibt (bekanntlich ist noch unser nhd. Graphem *s-c-h* ein Reflex der alten Aussprache), kann der (zweifellos frühmhd.) Vorgang weder zeitlich noch geographisch genauer festgelegt werden. Zwar fällt, strukturell betrachtet, /*sk*/ nunmehr aus der Reihe /*s* + Verschlußlaut/ heraus, aber bei jüngeren Entlehnungen (*riskieren*, *Maskerade* usw.) wird im Oberdt. [*sk*] durch [*šk*] ersetzt und so die Übereinstimmung mit [*št*] und [*šp*] (letzteres gilt heute auch ostfrk.) wiederhergestellt. Allerdings müssen wir nunmehr den Wandel /*sk*/ > /*š*/ im Norden von der bair.-alem. Entwicklung trennen, was sich aber vertreten läßt, wenn wir berücksichtigen, daß auch die gleichsinnige Entwicklung im Altengl. ohne Polygenese bzw.

[7] Eine ausführlichere Darstellung des Zischlaut-Systems in seiner Entwicklung vom Germ. aus wird an anderer Stelle gegeben werden. Dann werden insbesondere auch die anlautenden /*sp*, *st*/ [*šp*, *št*] mit einbezogen.

[8] Diese Umschrift in Anlehnung an *Joos, M.*: The Medieval Sibilants. In: Lang *28*: 222 ff. (1952). Das dort verwendete Diakritikum «untersetzter Strich» mußte aus drucktechnischen Gründen geändert werden.

«Entfaltung» (im Sinne *O. Höflers*) nicht zufriedenstellend erklärt werden kann.

Wie verhält sich nun dieses /*š*/ < /*sk*/ gegenüber dem älteren [*š*], das nur in den Verbindungen in- und auslautend /*st*, *sp*/ positionsgebunden vorkommt? Nachdem bereits in ahd. Zeit */*skp*, *skt*, *skk*/ durch Erleichterung schwerer Konsonanz in Wortbildungsfugen zu /*sp*, *st*, *sk*/ vereinfacht worden waren, gab es nunmehr keinen Fall, in dem das neue /*š*/ < /*sk*/ vor Verschlußlaut erscheinen könnte. Die alten Lautgruppen [*št*, *šp*] waren mithin nach wie vor ausschließlich Realisationen von /*st*, *sp*/ und als solche einer verdeutschließlich Schreibung ebenso wenig bedürftig wie unsere nhd. Anlaute in *Stein*, *Spiel* usw.

Das änderte sich natürlich, wenn in einer mhd. Wortbildungsfuge /*st*/ vor Konsonant zu stehen kam und die Dreierkonsonanz dabei um /*t*/ erleichtert wurde: Die so entstehende Verbindung [*š*] + Konsonant konnte nun graphisch ausgedrückt werden. Und in der Tat lautet der bei weitem älteste direkte Beleg (Reimzeugnisse müssen aus verschiedenen Gründen außer Betracht bleiben) *gaischlichen* (Mitte 13. Jahrhundert)[9]!

Ganz neue Verhältnisse schafft erst das Eintreten der *e*-Synkope. Vor allem bei der Verbalflexion kommt es jetzt zu einem phonemisch relevanten Gegensatz von in- und auslautendem /*st*/: /*št*/, z.B. in *reist*, *reiste*, *gereist* gegenüber *fischt*, *fischte*, *gefischt* (statt älterem *reiset*, *reisete*, *gereiset* gegenüber *fischet*, *fischete*, *gefischet*). Dieser Gegensatz wird graphisch bezeichnet. Da aber das alte [*št*] lautlich mit dem durch Synkope entstandenen /*št*/ zusammenfällt, wird es wie dieses behandelt und lautgerecht geschrieben. Eine genaue Überprüfung der chronologischen und geographischen Koinzidenz von Synkope und *scht*-Graphie scheitert allerdings daran, daß die *e*-Ausstoßung in den Denkmälern nicht zuverlässig wiedergegeben wird.

Abschließend sei noch erwähnt, daß im 13./14. Jahrhundert /*ʒ*/ und /*s*/ zusammenfallen, so daß die Verbindungen /*ʒet*, *ʒt*/ wie /*set*/ behandelt und normalerweise als [*st*] realisiert werden. Karte 4 zeigt an Einzelbeispielen die irreguläre Entwicklung zu [*št*].

Adresse des Autors: Dr. R. Freudenberg, Kaffweg 3, *Marburg a. d. Lahn* (Deutschland).

[9] *Kauffmann, F.*: Geschichte der schwäb. Mundart, S. 195 (Straßburg 1890).

Proc. 5th int. Congr. phon. Sci., Münster 1964, pp. 306–311
(S. Karger, Basel/New York 1965).

From the Department of Phonetics, University College London

The Dependence of Stress Judgments on Vowel Formant Structure

By D. B. Fry, London

The experiments reported in this paper attempted to explore the part played by vowel quality in stress judgments obtained from English listeners. Versions of the word-pairs *object*, *contract*, *subject* and *digest* were synthesized in which there was systematic variation of the frequency of the first and second formants in the first syllable of *object*, *contract* and *digest* and the second syllable of *object* and *subject*. Variations in vowel duration ratio were introduced in the same stimuli in order to provide a means of estimating the weight to be assigned to the changes in formant structure. The specification of the test stimuli is contained in the information shown in Table I. The fundamental frequency of the periodic sounds was kept constant at 120 cps. throughout. The overall intensity of syllables was regulated so that the maximum intensity in the two syllables of a test word was equal and a constant difference of 6 db. between formant 1 and formant 2 was maintained throughout. In Table I, f(1), f(2) and f(3) refer to frequencies used in the first syllable of *object*, *contract* and *digest*, f(4), f(5) and f(6) to those in the second syllable of *object* and *subject;* f(1) and f(4) are the formant arrangements likely to give the lowest number of 'noun' judgments. Duration ratios are labelled d(1), d(2) and d(3); d(1) indicates the duration ratio likely to give the lowest number of 'noun' judgments.

The stimuli were made into a listening test in which each stimulus occurred once. They were in random order and were preceded by five practice items. Stimuli succeeded each other at intervals of 2 sec with a gap of 10 sec after each set of 10 items. No carrier sentence or number anouncement was used. Stress judgments were

Table I

Experimental Duration Ratios (V1/V2)

	d(1)	d(2)	d(3)
object	0.45	0.66	1.33
contract	0.2	0.53	0.8
digest	0.63	1.0	1.3
subject	0.25	0.6	1.0

Experimental F1/F2 Values (cps.)

		First Syllable				Second Syllable		
		F1	F2	F1	F2	F1	F2	F3
object	f(1)	570	1380			570	1980	
	f(2)	600	1260			570	1980	
	f(3)	600	1020			570	1980	
contract	f(1)	570	1380			720	1800	2520
	f(2)	600	1260			720	1800	2520
	f(3)	600	1020			720	1800	2520
digest	f(1)	480	2040			570	1980	
	f(2)	840	1440 ➤	570	1800	570	1980	
	f(3)	840	1440 ➤	480	2040	570	1980	
object	f(4)	600	1020			570	1980	
	f(5)	600	1020			540	2160	
	f(6)	600	1020			400	2280	
subject	f(4)	720	1320			570	1980	
	f(5)	720	1320			540	2160	
	f(6)	720	1320			400	2280	

obtained from one hundred subjects who were all young speakers of Southern English, nearly all brought up in the south or in the Midlands.

In presenting the results of these experiments, a recurring difficulty is the problem of comparing the scale of variation in the two dimensions. In the present case, it can be said that the range of variation in F1/F2 was as wide as it could be without introducing a marked unnaturalness in the vowels which provided the end-point of each range; that is to say that in the first syllable of *object* and *contract*, for example, the F1/F2 values at the ends of the range gave as clear an [o] and as clear an [ə] as could be obtained. It has been shown in an earlier paper that context must be expected to play a considerable part in any judgments based on vowel quality. In previous experiments, it has been found consistently that this context effect works in the direction of contrast, that is to say that the perceptual difference between vowels is increased by their juxtaposition (*Fry et al.*, 1962). If we assume that the same tendency holds good with regard to the stimuli in the present experiments, this would mean that an [o], f(3), would sound more back and more open when it followed the corresponding f(2) or f(1), and an [ə], f(1), would sound more central when it followed the corresponding f(2) or f(3). The effect would therefore be to force the judgments at one end of the scale further in the direction of 'noun' and at the other end, further in the direction of 'verb', thereby increasing the apparent effect of the variation in F 1/F 2. That some influence of this kind was at work is suggested by the rather anomalous results obtained with the middle F 1/F 2 value in some word-pairs, where in one case f(2) produced slightly more 'nouns' than f(3) and in two other cases the same number of 'nouns' as f(1).

The relative effect of the formant changes and the duration changes can be judged from the range of percentage 'noun' judgments given in Table II. In every case the increase in 'noun' judgments produced by the change in duration ratio is greatly in excess of that produced by the change in formant frequencies. There is no

Table II

	Percentage 'nouns'	
	f (1)–(3)	d (1)–(3)
object	45–62	32–72
contract	35–52	12–72
digest	33–56	14–76
	f (4)–(6)	d (1)–(3)
object	60–65	39–84
subject	57–63	31–82

doubt that in the conditions of this experiment, the weight of the duration cue is very considerably greater than that of the formant structure cue. This fact is expressed in Figure 1 where the 'noun' scores for all word-pairs are pooled and the effect of the three steps of duration change and three steps of formant change are abstracted. The difficulty of comparing or equating the two scales has already been pointed out; here the steps of change are set off arbitrarily on the horizontal scale as though they were equal, with the duration

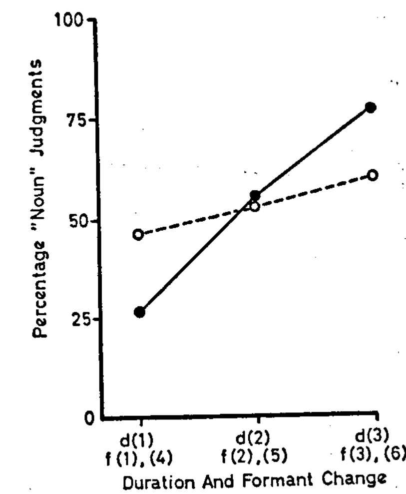


Fig. 1. The effect of changes in duration ratio and in F1/F2 plot on the percentage of "noun" judgments, scores for all word-pairs and all subjects pooled. The continuous curve refers to duration ratio and the dotted curve to formant structure.

and formant values giving the greatest number of 'verb' judgments nearest the origin. In the case of the formant values, f(1) and (4), f(2) and (5), and f(3) and (6) are pooled.

It seemed clear from preliminary experiments that there was a certain asymmetricality with respect to the effect of formant change in the first and the second syllables of the words; vowel change in the first syllable had a greater effect on the judgments than change in the second syllable. This fact is shown in Figure 2 where the effects of f(1)–(3) and of f(4)–(6) are plotted separately. This difference might be explained simply by the fact that in the particular words employed in the test the formant change in the second syllables was smaller than that in the first, judged either phonetically or by the crude test of the difference between F 1 and F 2 in cycles per second. The change from [o] to [ə] would probably strike many people as being perceptually greater than the change from [e] to [i]; in the stimuli used in the test, the frequency difference between F 1 and F 2 changed in the first progression from 420 to 810 cps. and in the second from 1410 to 1880 cps., a much smaller proportional change. When listening to the synthetic stimuli, however, it is difficult to resist the impression that there is something in the situation which renders vowel change in the second syllable really less effective than change in the first syllable. It is not easy to see how this can be check-

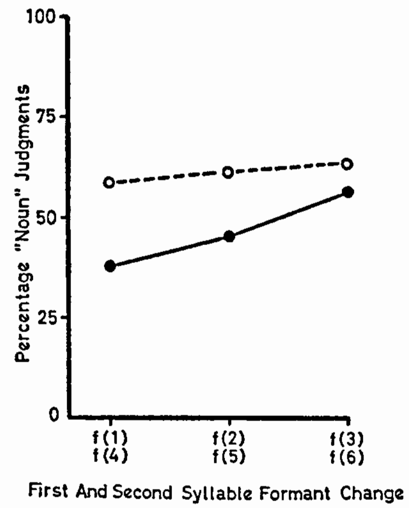


Fig. 2. The effect of changes in formant structure in the first and second syllable. The continuous curve refers to change in the first syllable, and the dotted curve to change in the second.

ed further by experiment except perhaps by synthesizing a series in which the vowel quality in the second syllable of such word-pairs as *subject* and *object* varied from [e] to [ə]. The latter pronunciation has a certain regional flavour but it might be possible to find out whether the effect could be enhanced in this way.

The Relative Weight of Formant Structure and Duration Ratio

The difficulty of arriving at any quantitative estimate of the relative weights of stress cues is aggravated in the case of formant changes by the fundamental problem of formulating a reasonable method of quantizing distances in the F1/F2 space. Any arrangement of scales chosen for this purpose is necessarily somewhat arbitrary and, in the absence of much fundamental work on the perceptual scales relevant to speech, is not likely to be very closely related to impressions of quality. In the present experiments, an attempt to arrive at some notion of the relations between the duration and the formant cues has been made by using the method outlined in a previous paper (see *Fry*, 1958) of plotting the logit values for percentages resulting from changes in duration ratio and in F1/F2 plot. In this way it is possible to find out approximately what change in duration ratio has an effect on the stress judgments equal

to that produced by the total change of formants used experimentally in any word-pair. These computations show that the swing in listeners' judgments due to the whole range of formant shift in the first syllable of *object* and of *digest* could be effected by a change in duration ratio of about 0.4 in the same words; for the first syllable of *contract*, the equivalent duration ratio change is about 0.16 and for the second syllable of both *object* and *subject*, about 0.1. If these ratios are compared with the ranges shown in Table I, it will be seen that they constitute a small proportion of the total ranges used experimentally. Values given previously as a basis for comparing the effects of duration and intensity in the same word-pairs were as follows: a difference of 20 db. between the two syllables of the stimulus was about equivalent to a difference in duration ratio of 0.4 for *object*, 0.16 for *digest*, 0.35 for *contract* and 0.6 for *subject*. A difference of 20 db. is rather large, and also it must be remembered that the stimuli in the two experiments were not identical, but a comparison of the two sets of figures suggests at least that the formant structure cue for stress may in fact be less effective than the intensity cue. A firmer conclusion on this point should, however, await the results of a direct experimental comparison of the two kinds of cue in action.

References

Fry, D. B.: Experiments in the perception of stress. Language and Speech *1:* 126–152 (1958).

Fry, D. B.; Abramson, A. S.; Eimas, P. D. and *Liberman, A. M.:* The identification and discrimination of synthetic vowels. Language and Speech *5:* 171–189 (1962).

Author's address: Prof. D. B. Fry. Department of Phonetics, University College London, *London* (England).

Verh. 5. int. Kongr. Phon. Wiss., Münster 1964, pp. 312–315
(S. Karger, Basel/New York 1965).

Untersuchungen über akustische Akzentmerkmale des Ungarischen

Von I. Futaky, Göttingen

Bei der Frage nach relevanten Akzentmerkmalen einer Sprache postulierten die meisten Forscher mit Recht, daß diese Merkmale im Lautprodukt des Sprechers enthalten sein müssen und daß deshalb an den Lautkurven betonter und unbetonter Redeteile deutliche Unterschiede auftreten sollten. Aber weder die Annahme höherer Lautdruckgipfel für die Akzentsilben noch das Vorherrschen hoher oder extrem tiefer Tonlage bei akzentuierten Silben noch die Auffassung vom Akzent als eigentümlicher Melodieverlaufsform konnten in den bisherigen Arbeiten mit hinreichender Sicherheit bewiesen werden.

In meiner Untersuchung über die Akzentmerkmale der ungarischen Mundart von Oberwart in Österreich habe ich deshalb sowohl nach neuen Fragestellungen hinsichtlich der Lokalisation der Akzentmerkmale als auch nach anderen Methoden zur Bearbeitung und Auswertung des Materials gesucht. Durch Messungen an möglichst vielen charakteristischen Punkten der Lautkurven sollte ein Zugang zur 'Feinstruktur' der Silben geschaffen werden. Um eine Auswertung nach verschiedenen Gesichtspunkten zu ermöglichen, sollten die Werte in den beiden Betonungsklassen nach Silbentypen getrennt registriert werden. Schließlich sollten die gefundenen Unterschiede mittels statistischer Verfahren – ich verwendete den Mediantest nach *Mood* – auf ihre Signifikanz überprüft werden.

Im Rahmen der Untersuchungen[1] haben drei zuverlässige ungarische Abhörer Tonbandaufnahmen aus der Mundart hinsichtlich der Akzentverhältnisse beurteilt. Zu 846 übereinstimmend beurteil-

[1] Ausführlicher Bericht: Untersuchungen über die Akzentmerkmale des Ungarischen – durchgeführt an der Mundart der Sprachinsel Oberwart, Göttingen 1964 (maschinenschriftliche Dissertation).

ten Silben wurden dann an der Pegelkurve und der Tonhöhenaufzeichnung des *Grützmacher-Lottermoser-Kallenbach*schen Tonhöhenschreibers je 10 Werte bestimmt. Die Meßpunkte an der Pegelkurve ergaben sich aus der *Sovijärvi*schen Aufteilung (vgl. Die sonagraphische Auswertung der Bandaufnahmen, Phonetica *6:* 34–36 [1961]). Demnach wurden der Aufbau und der Abbau des zum Vokal gehörigen Lautdrucks in je zwei Phasen erfaßt. Nach Qualität ihrer Anlautkonsonanten ordnete ich die Silben in drei Prüfgruppen. Bei der Tonhöhe registrierte ich die Anzahl der vom Vokaleinsatz an nach oben oder nach unten durchschrittenen Tonstufen. Zur Untersuchung der Lautdauer wurden die Gesamtdauer des Vokals sowie die Dauer der einzelnen Phasen der Pegelkurve gemessen und nach Vokalquantität bzw. -qualität vier Prüfgruppen gebildet.

Hypothesen und Ergebnisse

Im Lautdruckbereich stellte ich drei Hypothesen auf.

1. Die Lautdruckgipfel liegen bei betonten Silben höher als bei unbetonten Silben. Ein signifikanter Unterschied ergab sich nur bei Silben mit stimmlosen Anlautkonsonanten, während sich betonte und unbetonte Silben mit stimmhaften Anlautkonsonanten nicht auf dem festgesetzten Signifikanzniveau (1 Promille) unterschieden.

Die 2. Hypothese: Der Übergangspunkt zwischen Anlautkonsonant und Vokal liegt bei betonten Silben niedriger als bei unbetonten Silben – ließ sich nur für zwei Prüfgruppen (stimmlose bzw. stimmhafte Verschluß- und Engelaute sowie Affrikaten) verifizieren. Meiner Ansicht nach kann dieses Ergebnis mit der 'Präzision der Lautbildung' als Begleitmerkmal des Akzents in Verbindung gebracht werden.

Die 3. Hypothese lautete: Der Schalldruckzuwachs in der Stoßphase der Vokale ist bei betonten Silben größer als bei unbetonten Silben. Im Gegensatz zu den zuvor untersuchten Merkmalen ließ sich hier der erwartete Unterschied in allen drei Prüfgruppen mit hoher Signifikanz nachweisen. Das scheint ein Beweis dafür zu sein, daß – entsprechend der Ansicht von *Sovijärvi* – in der Stoßphase des Sonanten ein relevantes akustisches Akzentmerkmal zu sehen ist.

Da aber die Stoßphase – wie meine Ergebnisse ebenfalls zeigen – zugleich ein differenzierendes Merkmal bestimmter konsonantischer Anlautgruppen ist, kann nicht der absolute Lautdruckzuwachs als Akzentmerkmal fungieren, sondern der Stoßphase kommt offenbar nur innerhalb verschiedener Bezugssysteme, die von der Qualität der Anlaute bestimmt sind, eine akzentsignalisierende Funktion zu.

Hinter dem Druckgipfel wiesen betonte und unbetonte Silben keine Unterschiede auf.

Im Bereich der Tonhöhe untersuchte ich erstens die Verteilung der Bewegungstypen. Bei Akzentsilben dominierte die Steigung, bei nicht akzentuierten Silben das Fallen. Gleichbleibende Tonhöhe kam nur bei nicht akzentuierten Silben und hier relativ häufig vor. Zweitens verglich ich – in der Annahme, daß es weniger auf die Art als auf den Umfang der Bewegung ankommt – betonte und unbetonte Silben nach dem Ausmaß des Steigens und des Fallens der Tonhöhe. In beiden Fällen waren die Ergebnisse hochsignifikant, so daß man in dem größeren Maß der Tonhöhenbewegung, gleichgültig ob nach oben oder nach unten, das relevante Akzentmerkmal im Tonhöhenbereich erblicken kann. Die Bewegung konzentriert sich bei den betonten Silben fast immer auf die Stoßphase, während sie bei unbetonten Silben, insbesondere beim fallenden Typ, meist hinter dem Druckgipfel beginnt.

Außerdem wiesen betonte stimmhaft anlautende Silben eine größere Tonhöhensteigung auf als betonte stimmlos anlautende Silben. Angesichts der Tatsache, daß der Lautdruck als Akzentfaktor umgekehrt bei stimmhaft anlautenden Silben schwächer ausgeprägt war, scheint die Annahme nicht unbegründet, es handle sich hier um eine Interdependenz des Lautdrucks und der Tonhöhe und zwar in Form eines umgekehrten Verhältnisses, deren Ursache vielleicht darin liegt, daß zwischen den Hervorbringungsorganen eine Art Kompensation in der Beteiligung an der Akzentuierung stattfindet.

Die Untersuchung der Lautdauer zeigte, daß sowohl die Gesamtdauer der Vokale als auch die Dauer der Stoßphase bei betonten Silben signifikant länger ist als bei unbetonten Silben, während sich die Betonungsklassen hinter dem Druckgipfel nicht unterscheiden. Die Längung der Vokale betonter Silben ist also auf eine längere Dauer ihrer Stoßphase zurückzuführen.

In allen drei Bereichen – Lautdruck, Tonhöhe und Lautdauer – ließen sich also die für die Akzentwahrnehmung wichtigen Änderungen in der Stoßphase des Vokals lokalisieren.

Es wäre eine Aufgabe späterer Untersuchungen, das Verhalten der gefundenen Akzentmerkmale in zusammenhängender Rede zu

klären. Einige Stichproben haben bereits gezeigt, daß die Merkmale auch dort feststellbar sind, wo es mit den bisherigen Methoden in manchen Fällen nicht möglich war, als akzentuiert empfundene Silben auch auf akustischer Ebene zu identifizieren.

Adresse des Autors: Dr. István Futaky, Finnisch-ugrisches Seminar, Prinzenstraße 21, *Göttingen* (Deutschland).

Discussion

Sovijärvi (Helsinki): Im Vortrag wurde erwähnt, daß sich die Meßpunkte an der Pegelkurve aus der *Sovijärvi*schen Aufteilung ergaben. Es scheint jedoch möglich zu sein, daß Herr *Futaky* den Begriff «Stoßphase» nicht ganz richtig verstanden hat, weil er in seinem Vortrag über den Druckgipfel statt über den Endpunkt der Stoßphase spricht. Ich habe ausdrücklich betont, daß sich der Druckgipfel hinter dem Endpunkt der Stoßphase befinden kann; besonders oft geschieht das in den langen betonten Vokalen. Der unveränderliche Zuwachs der Intensität im Verhältnis zu der Zeitachse ist nach meiner Definition für die Stoßphase charakteristisch.

Nicht nur die Qualität der *Anlaute*, sondern auch die der nachfolgenden *Vokale* bestimmen die verschiedenen Bezugssysteme, die der Stoßphase als einer akzentsignalisierenden Funktionsgruppe zukommen.

Proc. 5th int. Congr. phon. Sci., Münster 1964, pp. 316–320
(S. Karger, Basel/New York 1965).

Accent, a Chief Factor in Linguistic Change

By Herbert Galton, Lawrence

The connection between the topic of my paper and the general theme of this Congress may not be immediately obvious, but will, I hope, emerge from my remarks. When I say "linguistic change", I should really make it clear that it is only phonological change I have in mind and take the latter as the generic term covering both allophonic variation and phonemic change. Phonological change may start with allophonic variation due to a specific environment and remain on the allophonic level as long as the conditioning factors remain unchanged; as the conditioning factors lapse, the mode of realization acquires phonemic status, if it does not lapse together with them. On the other hand, phonological change may from the beginning be irrespective of the phonetic environment and offer us no tangible allophonic variations as evidence of a transitional stage. If we now take up the problem of causation, the following avenues of approach suggest themselves from a purely theoretical point of view:

1. The cause of the sound change is to be sought in the nature of the speech sound alone, in its "Eigenart", as *Brugmann* called it when he formulated this view.
2. We must study the totality of the phonemes and their phonetic realization.
3. The cause at least of conditioned sound change lies in the phonetic environment.
4. The ultimate cause of phonological change lies beyond the phonemic pattern.

We will quite briefly discuss these various possibilities. *Brugmann's* point of view embodies the old "atomistic" approach, which we even find hard to understand nowadays, with all due respect to our predecessors. It becomes obvious from the study of any group

of genetically related languages that *Brugmann's* proposition must be inadequate, because these languages, with their different phonological structures, proceed from what were originally identical phonemes.

Approach 2. endeavors to overcome the atomistic approach. The phonemes are arranged in certain patterns based on the classification of their physiological or, more recently, acoustic qualities. In this way, our urge to study facts not in their isolation, but in their interrelation with a Gestalt character seems to be satisfied. Various symmetries and assymetries naturally emerge from these arrangements; somewhere, gaps emerge in the pattern, or else some phonemes do not fit into them at all and remain on the outside. It then becomes natural to look for the causes of instability in these disharmonies within the system. Despite all the evidence suggested by entropy in the surrounding universe, it is tacitly assumed that there is inherent in this system a tendency to arrive at some sort of pre-established harmony, to abolish dissymetries and fill in gaps; altogether, the pattern, which is really nothing but the result of our abstractions, is being endowed with some kind of immanent force. Please do not believe that I wish to belittle the validity or usefulness of the phonemic pattern – at the synchronic level; but that it presides over the drift of phonological change I do not believe. We will, therefore, regard this approach as nothing but an interesting attempt to take a structuralist view of diachronic change, based on a notion of linguistic structure that differs from ours.

Now for 3., the phonetic environment. Here it is time to give a concrete example of what I have in mind. Both Germanic and Slavic started out from one common form with their word for the "lie", say **lugi*. The phonemes with their "Eigenart" as well as the phonetic environment were identical. What about the phonetic realization? To take Germanic first, the realization of the consonantal phonemes was not significantly affected by the ensuing vowels, neither the *l* got velarized under the impact of the following back vowel nor the *g* significantly palatalized by the following front vowel. What allophonic variation there may have been in the *g* did not lead to the establishment of a new phoneme. But we get something else in Germanic instead; the first vowel developed into *ü*, at first only a fronted allophone of *u*, under the impact of the front vowel following in the next syllable, and this acquired phonemic status as soon as the original front vowel changed its nature in the history of

German. Thus we had in this particular stage of Germanic no significant changes due to the phonetic environment within the same syllable, but on the other hand a marked transsyllabic effect (not due to a palatalization of the intervening consonant).

Things were different in Slavic. Here, the *l* was velarized by the following *u*, giving rise to a velar variant that eventually acquired phonemic status where it became independent of the back versus front character of the ensuing vowel due to subsequent phonological changes. In the second syllable, the *g* was shifted to the prepalatal region by the action of the ensuing *i*: *gi* > *ži*, and subsequently, a new phoneme *ž* came about. Also in Slavic, the vowels changed, but irrespective of each other, that is to say, without any transsyllabic effect, resulting in Slavic [*lŭžĭ*] as against N.H.G. [*ly:gə*]. I have no time to go into the details of this development, but I view it against the background of the Slavic "musical" accentuation.

And here we come up against approach 4., factors beyond the phonemes and their patterns. We subsume these factors under the term "accentuation" which, in an adaptation of *A. Schmitt's* excellent definition, is nothing but the statement of the relationship between the syllables of the word or word-group in terms of intensity. He calls the accent of Germanic "strongly centralizing", that is to say, the accented syllable strongly dominates the unaccented ones, closely knitting the syllables together into the phonetic word. Such a definition is perfectly in keeping with the transsyllabic developments in Germanic of which I could only give you one example. From the autonomous, as I have called it, development in the Slavic syllables, however, we may infer a more even character of the accentuation allowing more scope for the interplay of vowels and consonants within the same syllable.

I cannot, in the brief time at my disposal, give more than a few examples from two more or less diametrically opposed accentual types. There is, then, on the one hand, the Old Slavic with its even accentuation, with its consonants and vowels evolving regardless of the place of the accent, but in close conjunction with each other within the syllable. The phonemes develop allophonic variaties and eventually change phonemic status in accordance with the following 3 factors:

a) along the lines laid down by their nature (Eigenart), i.e. their physiological or auditory criteria;

b) the framework established by the neighboring phonemes, in the special case of Slavic, in the great majority of cases within the same syllable only, and

c) in connection with b), under the effect of a particular type of accentuation that treated, generally speaking, all syllables alike. When I call accentuation something "beyond" the phonemes, this should be understood as exhibiting the same relationship as that of the pattern to its constituent parts. It is something over and above the elements and yet pervading them. In the case of speech, this overall pattern is, in accordance with the one dimension of speech, the succession and gradation, in terms of intensity, of the phonetic units of speech – the syllables.

In fact, in Germanic we find the same principle at work not only as between the successive syllables of the phonetic word, strongly dominated by one accent, but also within the syllabic nucleus itself. West Germanic provides many good examples of one part of a long and accented vowel differing greatly in its tension from another part, resulting in the establishment of a diphthong. Thus, the perfectly symmetrical pattern of the inherited six long vowels was upset in O.H.G. by the diphthongization *ô*, *ê* > *oa*, *ea* > *ua*, *ia*, etc., the forerunner of the N.H.G. diphthongization. Slavic, which we have taken as the opposite, in many ways, of the Germanic development, contracted on the contrary all the I.-E. diphthongs and, moreover, coordinated the resulting monophthong with the character of the preceding tautosyllabic consonant. Diphthongization also characterizes the development of the Romance languages out of local Latin dialects that had evolved a marked accent of intensity (cf. French *poire* < *pera*, Spanish *pierdo* – *perdemos*, etc.).

This is how we imagine the accentual principle to work: the accentual pattern of the word or word-group precedes in the speech centers the realization of the individual phoneme and their sequence, imposing certain features on them, in accordance with the proposition that the whole precedes the part. The accent, by no means a mysterious entity, is nothing but the organizing principle of the phonetic whole (not a delimitative sign, etc.). This principle asserts itself in the mutual relationship of the syllables in terms of intensity. In itself, "accentuation" is nothing but a generic term, whose concrete content varies from one language to another. By postulating the dominance of this factor in phonological change, we are only trying to bring linguistics into line with contemporary science, many of whose concepts defie material definitions and are seen to be mere expressions of relationship. Accent, as a suprasegmental feature, is not on a line with the phonemes, but hierarchically dominating them, so that both synchronic realization and diachronic change alike take their cue from it.

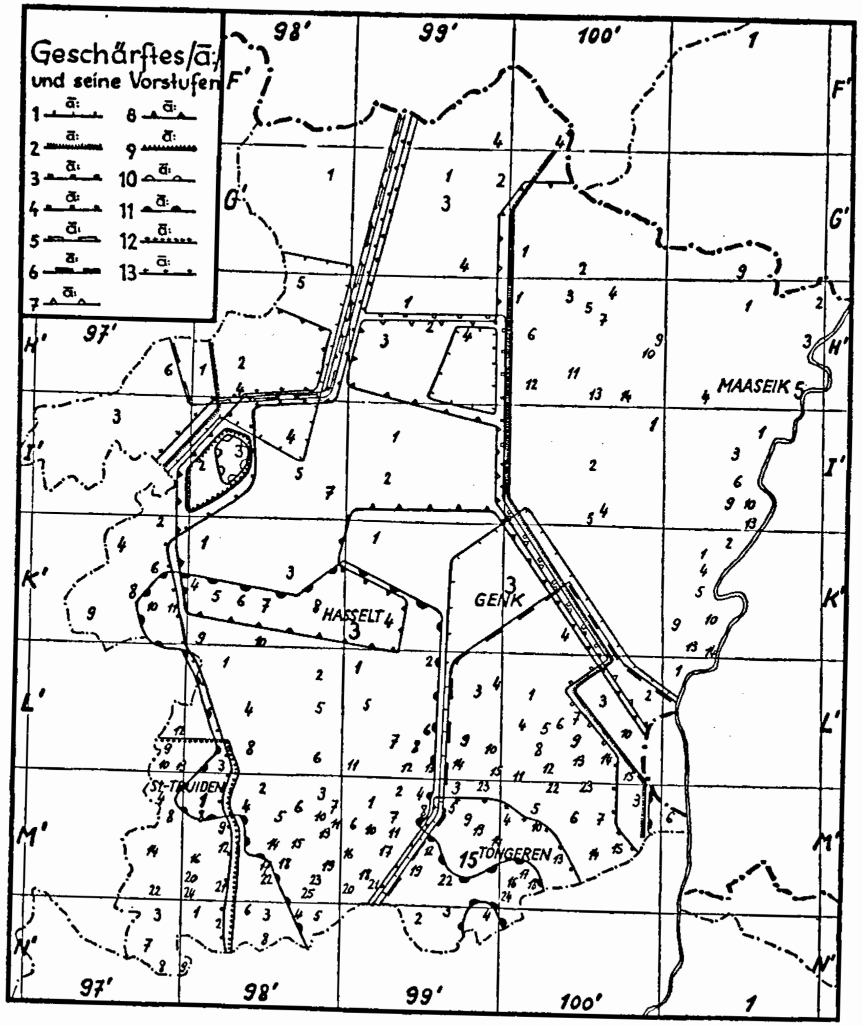
Author's address: Prof. Herbert Galton, Slavic Dept., University of Kansas, *Lawrence, Kansas* (USA).

Verh. 5. int. Kongr. Phon. Wiss., Münster 1964, pp. 321–324
(S. Karger, Basel/New York 1965).

Geschärftes /ā:/ und seine Vorstufen in Belgisch-Limburg

Von J. Goossens, Marburg a. d. Lahn

Die zu besprechende Karte wurde unter Mitwirkung des limburgischen Dialektologen *A. Stevens* gezeichnet. Das Verfahren ist der strukturellen Sprachgeographie schon bekannt. Es ist das umgekehrte des in der traditionellen Lautgeographie üblichen. Den Ausgangspunkt bildete ein Phonem, das in den verschiedenen Mundarten des Untersuchungsgebietes vorkommt. Es wurde untersucht, welche in den einzelnen Mundarten dieses Gebietes die Vorstufen dieses Phonems sind. Die Gebiete der einzelnen Vorstufen wurden abgegrenzt und die Isoglossen alle auf eine Karte eingetragen. Absichtlich wurde ein Phonem gewählt, das mit Sicherheit eine große Variation von Vorstufen ergeben mußte. Auf diese Weise war es möglich, der Frage nachzugehen, ob die einzelnen Isophonen Funktionen voneinander sind oder nicht. Wenn die Isophonen zweier Gebiete, die über dasselbe Phonem verfügen, zusammenfallen, wobei das Gebiet mit der ersten Vorstufe links und das mit der zweiten rechts von der gemeinschaftlichen Grenze zu finden ist, liegt die Vermutung nahe, daß die Verbreitung des Phonems aus der ersten Vorstufe an die Verbreitung desselben Phonems aus der zweiten Vorstufe gebunden ist. Der Umfang, den das erste Gebiet angenommen hat, scheint von der Verbreitung des zweiten abhängig zu sein und umgekehrt. Wenn zahlreiche Isoglossen zusammenfallen, wird diese Vermutung zu einer Gewißheit. Wenn anderseits die Isoglossen der einzelnen Gebiete überhaupt nicht zusammenfallen, muß die Erklärung des Umfanges und der Form der Gebiete in einer anderen (z.B. einer expansiologischen) Richtung gesucht werden.



Die eingezeichneten Isophonen umgrenzen die folgenden Gebiete[1]:

1. Drei Gebiete mit /ā:/ < *e* oder *a* vor geminiertem *r*. Beispiele: /sxā:rə/, šā:rə/ «scharren», /kā:r/ «Karre».

2. Ein Gebiet mit /ā:/ < *a* vor intervokalischer Verbindung *-nd-* (*d* = ð, *þ*). Beispiel: /hā:n/ «Hände».

[1] Es handelt sich bei jeder der dreizehn Nummern um Reihen von Wörtern. Bei einzelnen zu diesen Reihen gehörigen Wörtern kann die Grenze der /ā:/-Gebiete von der eingezeichneten «Idealgrenze» abweichen. Es war gar nicht daran zu denken, die Abweichungen ebenfalls einzuzeichnen.

3. Ein Gebiet mit /ā:/ < *a* vor intervokalischer Verbindung *-ld-* (*d* = ð, *þ*). Beispiel: /kā:/ «kalte, Kälte».

4. Ein Gebiet mit /ā:/ < *a* vor auslautender Verbindung *-ld* und vor *-lt-*. Beispiele: /kā:t/ «kalt», /smā:t/ «Schmalz».

5. Ein Gebiet mit /ā:/ < *o*, *u* vor der Verbindung *-ld-*, *-lt-*. Beispiele: /γā:t/ «Gold», /hā:t/ «Holz».

6. Ein Gebiet mit /ā:/ < *a* vor intervokalischer Verbindung *-l-* + beliebigem stimmhaftem Konsonanten. Beispiel: /hā:və/ «halber».

7. Ein Gebiet mit /ā:/ < *a* vor intervokalischem stimmhaftem Dental. Beispiel: /γlā:zə/ «gläsern».

8. Ein Gebiet mit /ā:/ < *a* vor intervokalischem stimmhaftem Labial oder Velar. Beispiele: /hā:m/ «Hamen» (= Kumt), /jā:γə/ «jagen».

9. Ein Gebiet mit /ā:/ < *a* vor intervokalischem stimmlosem Labial oder Velar. Beispiele: /ā:p/ «Affe», /krā:kə/ «krachen».

10. Ein Gebiet mit /ā:/ < *a* vor intervokalischem stimmlosem Dental. Beispiel: /kā:tər/ «Kater».

11. Zwei oder drei Gebiete mit /ā:/ als zweiter Komponente eines Diphthongs /jā:/ < *e* vor intervokalischer Verbindung *-r-* + stimmhaftem Dental. Beispiel: /pjā:t/ «Pferde».

Anmerkung: Es sind zwei Gebiete, wenn man die Verbindung von M' 99', 22 und N' 99', 3 zu M' 100', 16, 17, 18, 24 herstellt. Es sind drei Gebiete, wenn man N' 99', 4 mit M' 99', 15 verbindet.

12. Ein Gebiet mit /ā:/ < nicht umlautsfähigem *û* oder auslautendem *-ô*. Beispiele: /hā:s/ «Haus», /kā:/ «Kuh».

13. Zwei Gebiete mit /ā:/ < *â* vor *w*. Beispiel: /blā:/ (im S.O.), /blā:t/ (im N.W.) «blau».

Aus der Karte geht hervor, daß die einzelnen Gebiete in der Regel nicht nur von ihren eigenen Isophonen, sondern auch von denen anderer Gebiete begrenzt werden. Die Gebiete schließen sich gegenseitig aus. Wo sich ein Phonem /ā:/ aus einer bestimmten Vorstufe entwickelt hat, wurde die gleiche Entwicklung aus anderen Vorstufen ausgeschlossen und umgekehrt. Die einzelnen Isophonen sind Funktionen voneinander: die Verbreitung eines Phonems /ā:/ aus einer Vorstufe A ist abhängig von der Verbreitung desselben Phonems aus einer Vorstufe B. Die geographische Verbreitung eines Lautgesetzes kann also durch die Verbreitung anderer Lautgesetze in anderen Gebieten determiniert werden. Dieser zwar nicht völlig neue Gedanke wurde u. W. noch nie so deutlich mittels einer Sprachkarte dargelegt.

Adresse des Autors: Dr. Jan Goossens, Forschungsinstitut Deutscher Sprachatlas, Kaffweg 3, *Marburg a. d. Lahn* (Deutschland).

Proc. 5th int. Congr. phon. Sci., Münster 1964, p. 325
(S. Karger, Basel/New York 1965).

From the Hindi Department, University of Jodhpur (India)

Some Undecided Hindi Phonemes*

By Moti Lal Gupta, Jodhpur

I have taken for my study the standard form of Hindi, tipped as the national and official language of India. This form of Hindi is an adaptation of the Khari Boli, belonging to the Western Hindi group of languages, according to *Grierson*. It is spoken over a vast area of Northern India and is followed in almost all the parts of the country, leaving some far off rural areas. Sometimes there is a controversy over its status, and one of the arguments advanced against it is the nature of some of its phonemes which I call 'undecided' – may also be called 'ambiguous'.

The commonly accepted vowel phonemes are 12, and the consonant phonemes 33. Out of these the following ones are still undecided –

Vowel /*ai*/ equivalent to [*ai*] and [*ɛ*]
/*au*/ equivalent to [*au*] and [*ɔ*].

The phonetic symbols are the same but they represent different phonemes from place to place and situation to situation, as also from person to person.

/*ṛ*/ being reduced now to /*ri*/ and /*ru*/.

The phonetic symbol continues but its importance at its phonetic level no longer exists.

Consonant /*ŋ*/, /*ñ*/.

The symbols are there but they are neither spoken nor written having yielded their places to /n/ and for all purposes reduced to [*ŋ*], [*ñ*] ∾ /*n*/.

/*ṣ*/ having become one with /*ś'*/.

New sounds have crept in but symbols have not been devised. They are having a double phonemic representation. /*n͜h*/, /*m͜h*/, /*r͜h*/, /*l͜h*/ and the like.

* The paper in full will be published in Phonetica.

Author's address: Dr. M. L. Gupta, Hindi Department, University of Jodhpur, *Jodhpur* (India).

Proc. 5th int. Congr. phon. Sci., Münster 1964, pp. 326–331
(S. Karger, Basel/New York 1965).

From the Haskins Laboratories, New York

Intonation Contours Evaluated by American and Swedish Test Subjects

By K. Hadding-Koch* and M. Studdert-Kennedy**, New York, N.Y.

Both in Swedish and in American English the pitch contours of yes-no questions tend to display a final rise, pitch contours of statements a final fall. But naturalistic observations suggest that earlier portions of the contour are also relevant. Such questions tend to be spoken on a comparatively high pitch in Swedish[2], and with a continuously rising contour in American English[3,6]. The two languages agree in displaying a statement contour with a moderate rise followed by a fall. The present study simulated Swedish intonation contours, by controlled variation of the fundamental frequency, and examined the distributions of responses to these contours made by Swedish and American subjects. The two groups were expected to agree more in their statement than in their question responses.

The words, "*Fő̈r Jane*" [foe'jein] = *for Jane* (duration: 640 msecs), so spoken as to be acceptable as English by Americans, as Swedish by Swedes, were recorded on magnetic tape. The tape was passed through a special synthesizer at Haskins Laboratories, New York, which permits the fundamental frequency of the utterance to be varied without otherwise changing the recorded speech[1]. The output of this synthesizer was recorded on magnetic tape.

Forty-two intonation contours were used. The fundamental frequency values were based on detailed spectrographic analyses of the Swedish speaker's natural speech. All started at 250 cps (sustained over "Fő̈r" for 140 msecs), rose to a peak of either 370 cps or 310 cps,

* Also at Institute of Phonetics, University of Lund, Sweden.
** Also at Barnard College, Columbia University, New York.

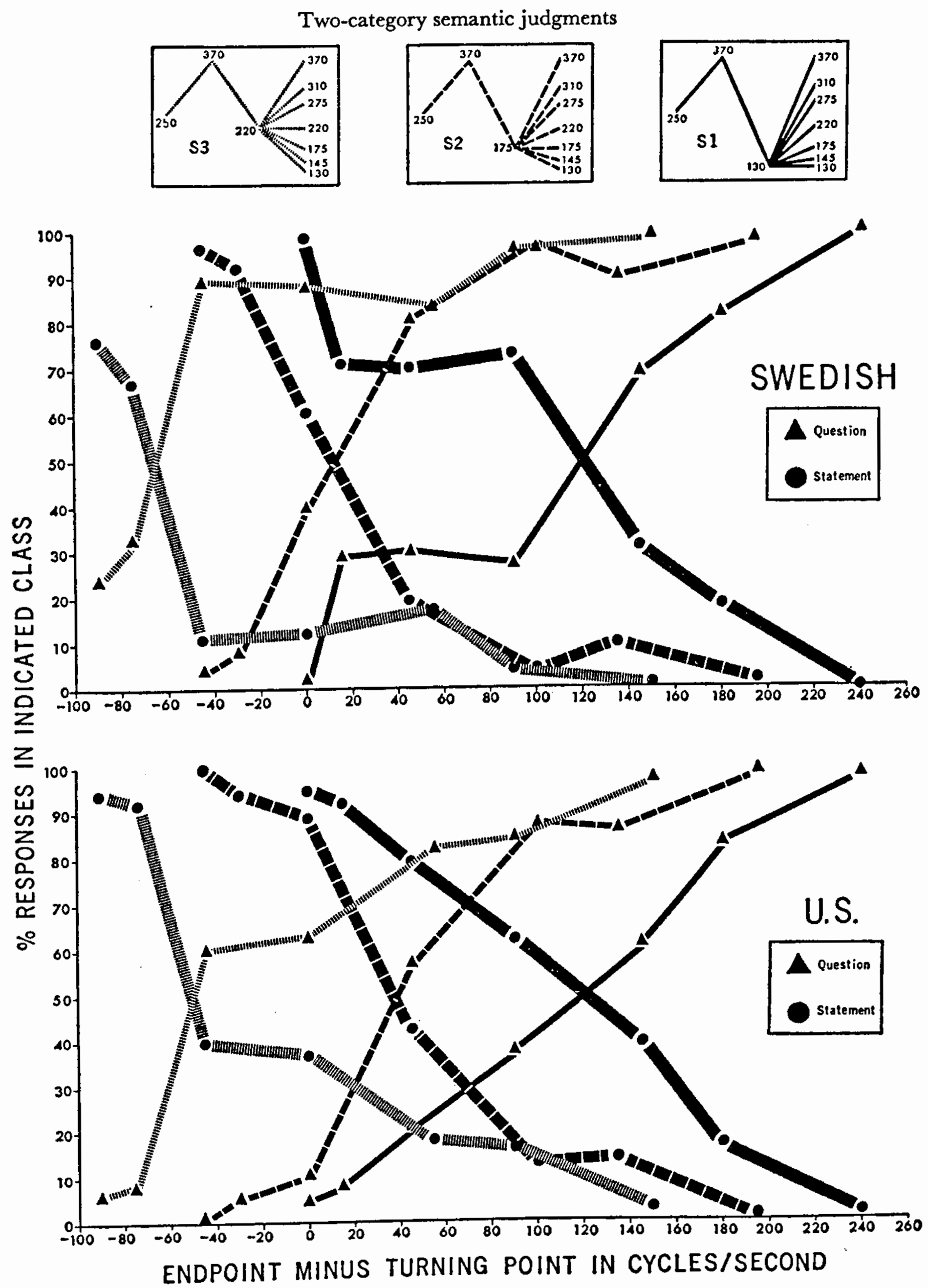


Fig. 1 and *2.* Two-category semantic judgments with peak fundamental frequency at 370 cps: percentage of statement and question responses as a function of the terminal rise (positive) or fall (negative) in cycles/second of fundamental frequency (end point minus turning point). Parameters of the curves are turning point fundamental frequency: 130 cps *(S1)*, 175 cps *(S2)*, and 220 cps *(S3)*. The Swedish data are plotted above, the U.S. data below.

dropped to one of three turning points: 130 cps, 175 cps, or 220 cps (300 msecs), and then proceeded to one of seven end points between 130 cps and 370 cps (200 msecs).

The forty-two stimuli were arranged into five different random orders and presented in an appropriately counterbalanced series to 25 Swedish and 24 American undergraduates. In four separate sessions, subjects were instructed to indicate for each stimulus: (1) whether it would be better characterized as a statement or a question; (2) whether it ended with a rise or a fall; (3) whether it would best be characterized as a question, a statement or a non-communicative utterance, that is, a private "reflection", spoken by the speaker to herself; (4) whether it ended with a rise, a fall or a level pitch. The third category of (3) and (4) was suggested by previous analyses of natural speech in which "reflections" appeared to display more or less level final contours, and was introduced here to see whether it might not collect the responses to stimuli that were found to be ambiguous when only the statement-question categories were used.

In the two-category semantic test, both Swedish and American responses to a stimulus with a given final rise or fall in frequency were found to vary with the frequency values of peak and turning-point: the higher the frequency of either, the more questions were heard. Thus, if we consider the value in cycles of the final rise at which Swedish responses crossed over from predominantly statements to predominantly questions, we find that: with the peak at 310 cps, the crossover is at 160 cycles for a turning point of 130 cps, at 130 cycles for a turning point of 175 cps, at 80 cycles for a turning point of 220 cps; with the peak at 370 cps, the crossovers for the three turning-points are at a rise of 120 cycles, a rise of 12 cycles and a fall of 65 cycles, respectively. From this last figure (crossover from statement to question at a final fall of 65 cycles), it is evident that the effect of the peak and turning point frequencies overrode the effect of the final fall. See Figs 1 and 2.

Turning to the results of the psychophysical sessions in which listeners were asked simply to indicate whether the final movement of the contour was a rise or a fall, we find more overall uncertainty – more disagreement between subjects – but, at the same time, a remarkable degree of agreement with the semantic judgements. That is to say, stimuli heard as questions tended to be heard also as having a final rise (even if, in fact, the contour displayed a fall), while stimuli heard as statements tended to be heard as having a fall

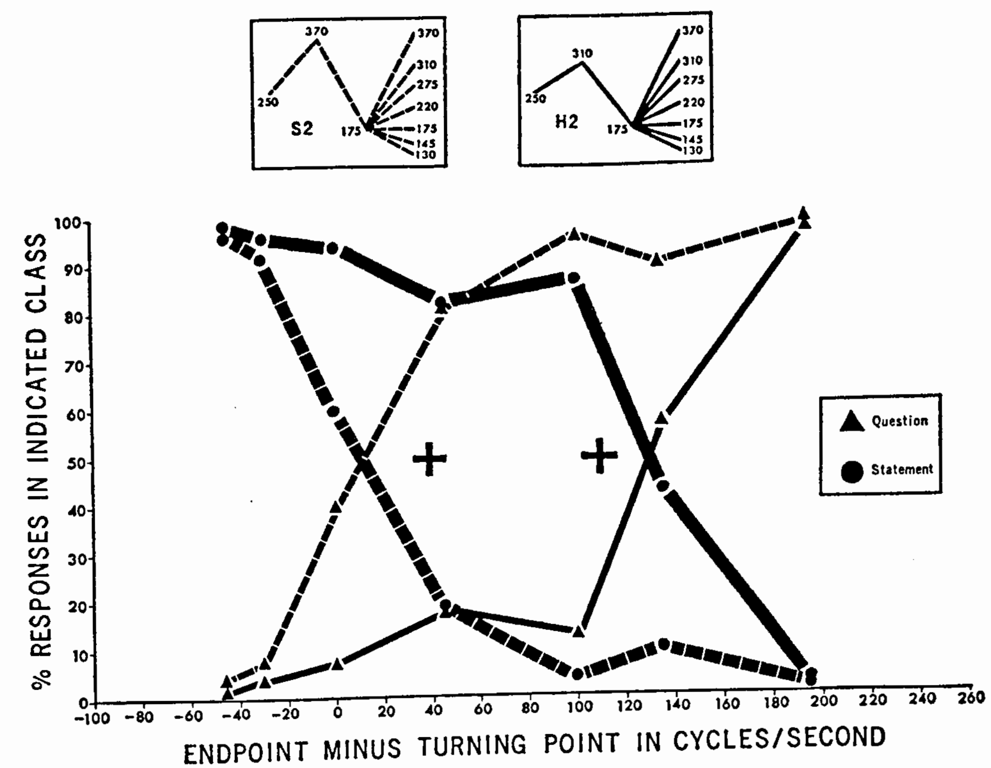


Fig. 3. Two-category semantic judgements of Swedish subjects on the *S2* (peak frequency: 370 cps) and *H2* (peak frequency: 310 cps) series. Percentage of responses in indicated class is plotted against terminal rise or fall in cycles/second of fundamental frequency. The crosses indicate the crossover values for the U.S. subjects on the *S2* (left) and *H2* (right) series.

(even if, in fact, the contour displayed a rise). Thus the discrimination of pitch within these speech signals was markedly less fine than might have been predicted from pure tone pitch discrimination[4].

The results of the two sessions in which three category judgements were called for were somewhat unsatisfactory – although once again there was remarkably good agreement between semantic and psychophysical judgements. For the most part, the third category served only to confuse the picture. The response curves were, on the whole, very similar to those of the two category responses, except that there was less agreement, the third category "nibbling" pieces from the other two. But in one group of stimuli, for which the peak

U.S. two-category semantic and psychophysical judgments.

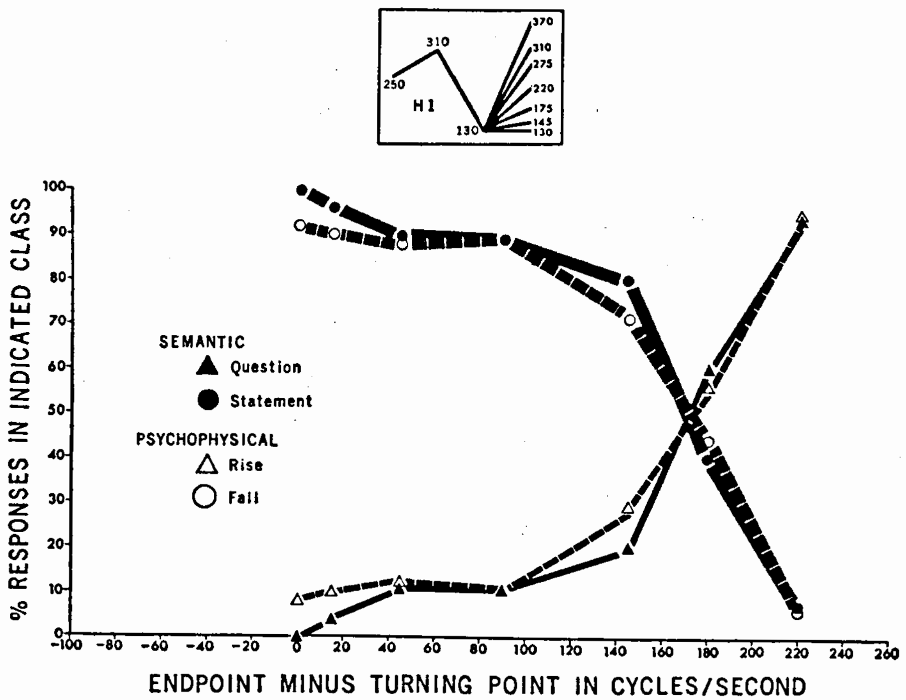


Fig. 4. Two-category semantic (solid line) and psychophysical (hatched line) data for U.S. subjects on the *H1* stimulus series. Percentage of responses in indicated class is plotted against terminal rise or fall in cycles/second of fundamental frequency.

was relatively low (310 cps) and the turning point relatively high (220 cps), the third category did show an appreciable percentage of responses: 58 % and 72 % in the semantic test for the Swedish and American groups respectively, 70 % and 66 % in the psychophysical test. As was expected, these maximum figures were reached on stimuli with a level final contour. See Fig. 4.

To sum up, at least three variables were operating to determine listeners' semantic responses: peak frequency, turning point frequency and final contour. In general, the higher the peak and the turning point, the more likely were both Swedish and American listeners to hear a question, the lower the peak and the turning point, the more likely were they to hear a statement. Only marked rises and falls overrode the effects of peaks and turning points. The third category ("reflection") proved to be of use for a limited group of stimuli

having a level final contour and a relatively low precontour. All these effects were also present, though less marked, in the psychophysical data. For the most part, particularly in the semantic responses, these effects were more marked in the Swedish group than in the American. The American listeners showed greater disagreement among themselves – as was expected, since the contours used were derived from the spectrographic analysis of Swedish speech.

Further experiments, designed to clarify this and other points, are now in progress. Our new stimuli include typical American English, as well as typical Swedish, contours. For purposes of control we are collecting psychophysical data on pure tone contours identical with the fundamental frequency contours of the speech.

Acknowledgement. We should like to acknowledge that the research reported here was supported in part by a grant from the National Science Foundation, Washington, D.C.

References

1. *Borst, J. M.* and *Cooper, F. S.:* Speech research devices based on a channel vocoder (abstract). J. acoust. Soc. Amer. *29:* 777 (1957).
2. *Hadding-Koch, K.:* Acoustico-phonetic studies in the intonation of Southern Swedish (Lund 1961).
3. *Pike, K. L.:* The intonation of American English (Ann Arbor 1945).
4. *Stevens, S. S.* and *Davis, H.:* Hearing, its psychology and physiology (New York 1938).
5. *Uldall, E. T.:* Ambiguity: question or statement? or "Are you asking me or telling me?", Proc. IVth Int. Congr. Phonet. Sci., Helsinki, 1961, pp. 779–783 (The Hague 1962).

Authors' addresses: Docent Dr. Kerstin Hadding-Koch, Fonetiska Institutionen, Kävlingevägen 20, *Lund* (Sweden) and Dr. Michael Studdert-Kennedy, Haskins Laboratories, 305 East 43rd Street, *New York, N.Y.* (USA).

Discussion

Gårding (Lund): If the generated stimuli that have been described are part of experimental work in juncture I do not think some of them are very realistic. A prejunctural fall of fundamental frequency is most often accompanied by a loss of intensity. Four years ago, at Haskins Laboratories, *Arthur Abramson* and I varied the fundamental frequency in some American English intonation contours. One of the experiments was to pit a falling statement contour against a rising question contour. By means of listening tests we found that we needed a considerable drop of frequency to make listeners place the generated contour in the statement category. A contour with a drop of ten cycles for instance was rather put in the question category. I think this is due to the fact that our carrier phrase did not have the intensity curve that normally goes with a falling contour.

Answer *Hadding-Koch:* Yes, I believe it is quite probable at least that we would have got more statement responses if we had used falling intensity with falling intonation. It was not possible to vary the intensity feature with the device used but we have done it in our new sets of stimuli: we have one set with sustained intensity and one with falling intensity. This could be done simultaneously on two identical spectra stored in the Digital Spectrum Manipulator at Haskins Laboratories, New York.

C.R. 5e Congr. int. Sci. phon., Münster 1964, pp. 332–335
(S. Karger, Basel/New York 1965).

Sur la hiérarchie des éléments phoniques des sons du langage

Par B. HÁLA, Prague

Dans les recherches portant sur la phonétique historique, il est très utile de faire la distinction, en ce qui concerne la nature acoustique des sons du langage, entre, d'une part les éléments ou traits phoniques primaires réalisés intentionnellement et indispensables pour l'identification auditive de ces sons, et d'autre part les éléments ou traits phoniques secondaires dont la naissance est due aux hasards de l'articulation et qui sont, de façon générale, à considérer comme superflus ou redondants.

L'évolution historique des sons du langage nous apprend que les deux classes d'éléments phoniques qui viennent d'être signalées, assujetties à la hiérarchie que nous venons de préciser, ne maintiennent pas leurs positions respectives sans présenter parfois d'exceptions. Il peut se produire en effet, au cours de l'évolution historique d'une langue, des changements dans l'appréciation auditive de ces éléments, dont la conséquence est que des éléments originellement primaires perdent quelquefois du terrain pour faire place à des éléments secondaires qui, presqu'imperceptiblement, ont gagné en importance pour finir par dépouiller tels éléments primaires de leur supériorité sur le champ acoustique de la parole.

Le bien-fondé de ce que nous venons d'avancer pourrait être appuyé sur un grand nombre d'observations phonétiques, mais le peu de temps dont nous disposons nous contraint à n'en citer ici qu'une seule. Elle est puisée dans le domaine de la langue tchèque, où l'on rencontre une consonne aussi curieuse que rare, à savoir le fameux *ř* tchèque (comme par exemple dans le mot *moře* «la mer»). Les étapes successives de l'évolution de cette consonne, issue de l'ancien *r*, feront l'objet de ce bref exposé.

Stade primitif. Comme ailleurs, l'*r* proto-slave était un *r* apico-alvéolaire, dont l'effet auditif rappelle le son de «roulement» pro-

voqué par des amortissements périodiques du son laryngien dûs aux battements de la pointe de la langue contre la zône supradentale du palais dur. Ce roulement est, dans l'*r*, le seul élément acoustique produit consciemment, mais il est suffisant pour distinguer cette consonne de toutes les autres. Superposé à ce son, il existe un second élément acoustique: c'est le «son propre» de la cavité buccale. Mais ce dernier, prenant naissance en dehors de l'intention du locuteur, ne constitue qu'un élément secondaire; il est perçu par l'auditeur dans l'ensemble acoustique de la consonne sans jouer aucun rôle dans l'identification de l'*r* comme son du langage.

1re étape de transformation: r se modifie en ŕ (r palatalisé)

Dans certaines conditions, l'*r* primitif du slave a été palatalisé (mouillé), c'est-à-dire transformé en *ŕ*. La palatalisation se manifeste, d'une façon générale, par un fort soulèvement du dôme de la langue vers le palais dur, de sorte que la cavité buccale se trouve quasi-comblée par elle. Son volume diminue; sa résonance devient plus aiguë, par opposition à l'*r* pour laquelle elle est relativement assez grave. Cette acuité est, au surplus, accrue par la forme de l'ouverture buccale qui s'allonge des deux côtés en se rétrécissant en même temps. Ainsi, la fréquence de la cavité buccale, dépourvue d'intérêt dans l'*r* tout seul, devient un élément distinctif là où il existe un *ŕ* mouillé à côté d'un *r* non-mouillé ou «dur».

Le mouvement de la langue vers le haut de la bouche est accompagné d'un soulèvement de la mandibule qui se rapproche du maxillaire supérieur. L'air expiratoire, en s'échappant à travers la fente étroite constituée entre les deux rangées d'incisives, ajoute au son de l'*ŕ* un léger bruit fricatif. Celui-ci, négligé d'habitude par l'auditeur, ne représente qu'un élément secondaire (qui apparaît, d'ailleurs, dans toutes les consonnes palatalisées).

2e étape de transformation: ŕ mouillé se modifie en ř', c'est-à-dire en ŕ mouillé fricatif

Sur de vastes territoires du slave occidental, et principalement en tchèque et en polonais, le rapprochement des incisives est plus prononcé qu'ailleurs, et dès lors le bruit expiratoire revêt un caractère sibilant. L'auditeur a pu fixer ce phénomène dans son impression auditive, lui attribuer peu à peu la valeur d'un élément indispensable pour l'identification de cette consonne, et, à la fin, le reproduire intentionnellement. De ce fait, la sibilation est devenue un élément primaire. En revanche, le «son propre» de la cavité buccale a perdu progressivement son importance pour devenir ultérieurement un élément tout-à-fait secondaire.

3e étape: ř' palatalisé se modifie en ř non-palatalisé

A une époque suivante, ř' a subi une dépalatalisation. Il est assez difficile de déterminer les causes de cette transformation nouvelle de l'ancien ŕ. Y a-t-il eu une dépalatalisation générale du système consonnantique tout entier? Celà pourrait avoir été possible pour le tchèque, mais nullement pour le polonais où la coexistence des deux séries, palatale et non-palatale, s'est maintenue jusqu'à présent. Il faut donc chercher ailleurs la cause véritable de cette évolution, que nous allons tenter de découvrir. Comme on sait, l'ř' palatalisé devait disposer de deux éléments phoniques le différenciant de l'*r:* le bruit sibilant et la fréquence aiguë de la cavité buccale: c'était plus qu'il n'en fallait. Donc, rien de plus opportun que d'en supprimer un. Le choix se décida en faveur du bruit sibilant, car toute palatalisation exige un accroissement du travail articulatoire.

Nous assistons ainsi, dans l'évolution de *r* vers ŕ, à un échange total des rôles des éléments phoniques composant les deux consonnes: l'élément primaire, à savoir la fréquence de la cavité buccale, cède la place à un élément secondaire, le bruit sibilant. En revanche, l'élément primaire de groupe, le «roulement» qui distingue les consonnes dites «roulées» de toutes celles appartenant aux autres groupes consonantiques, demeure intact et maintient sa place pendant toutes les transformations qui se sont succédées*.

Voilà la situation finale en tchèque, où l'évolution s'est arrêtée au point que nous venons de préciser. Il en est tout autrement en ce qui concerne le polonais. Dans cette langue, l'évolution a continué jusqu'à la suppression totale de la vibration de la pointe de la langue, le bruit sibilant finissant par dominer tous les autres éléments phoniques, y compris le trait distinctif de groupe, le «roulement». C'est ainsi que l'ř a fini par aboutir à une fricative pure identique à la fricative chuintante *ž* [ʒ]**. Cette évolution s'est

* Il est curieux que la forme de l'ouverture buccale pour l'ŕ se soit maintenu même après la transformation de cette consonne en ř.

** Il est à noter que l'ř, originellement sonore, est devenue quelquefois sourd par un effet d'assimilation (comparer à ce sujet le tchèque *moře* avec ř sonore et *tři* avec ř sourd, le polonais *morze* [moʒe] et *trzeba* [tʃeba]).

accomplie malgré les homonymies qui en résultent quelquefois, comme par exemple dans [moʒe], signifiant, soit «la mer» (orthographiquement *morze*), soit «il peut» (orthographiquement *może*).

Adresse de l'auteur: Prof. B. Hála, Foneticky Ustav, Nám. Krasnoarmejcü 2, *Prague I* (ČSSR).

Discussion

Meriggi (Pavia) hebt mit einer Danksagung die Bedeutung der Mitteilung hervor und stellt einige Fragen über Einzelheiten, indem er vor allem anzweifelt, daß man vom Willen des Sprechenden, einen Laut so oder so auszusprechen, reden kann.

Martinet (Paris): On rend hommage à la clarté de la représentation d'un problème délicat de phonétique évolutive. On se demande si le trait noté par un sigma suscrit peut résulter d'une friction entre les dents.

(Après l'intervention de M. *Meriggi* qui reproche à M. *Hála* de faire intervenir la volonté du locateur.)

Traiter un trait distinctif comme un trait volontaire résulte d'une hypothèse qui n'ajoute rien à la clarté de l'exposition.

Sovijärvi (Helsinki): Während seiner Antwort an Herrn *Martinet* hat der Vortragende darauf hingewiesen, daß der distinktive artikulatorische Unterschied zwischen den Phonemen s und š (= ʃ) darin besteht, daß es einen Kesselraum zwischen den Vorderzähnen und der Zungenspitze beim *š* gibt, daß dieser Raum aber beim *s* fehlt. Ich möchte in diesem Zusammenhang kurz erwähnen, daß dies nicht in allen Sprachen unbedingt der Fall sein muß. Zum Beispiel im Finnischen sind die beiden Phoneme medioalveolare Zungenspitzenlaute, aber in dem *š*-Laut (*šakki* 'das Schach') gibt es eine viel längere und breitere Zungenrille als in dem *s*-Laut (*sakki* 'Gruppe, Menge'). Das finnische Phonem *š* wird nur in den Fremdwörtern angewendet. (Vgl. *Sovijärvi:* Suomen hielen äännekuvasto, Jyväskylä 1963.)

Verh. 5. int. Kongr. Phon. Wiss., Münster 1964, pp. 336–339
(S. Karger, Basel/New York 1965).

Können Lautveränderungen erklärt werden?

Von G. Hammarström, Uppsala

Der Sprecher übermittelt dem Hörer Informationen durch Lautmerkmale, die mit Rücksicht auf ihre Funktion auf drei Niveaus verteilt werden können:

1. *Das α-Niveau* mit den Lautmerkmalen, die die intellektuellen «Bedeutungen» unterscheiden: Phone–Phoneme, «Prosode»–Prosodeme usw.

2. *Das β-Niveau* mit den Lautmerkmalen, die berichten, wie etwas gesagt wird (mit Bewunderung, Ironie usw.).

3. *Das γ-Niveau* mit den Lautmerkmalen, die Erläuterungen über die Person des Sprechers geben: a) idiolektale, b) «soziolektale» und c) dialektale Merkmale[1].

Die «Erklärungen» der Lautveränderungen gründen sich oft (wie bei *A. Martinet*) auf Verhältnisse des α-Niveaus. Wenn man annimmt, daß eine «Erklärung» einen ausreichenden Grund implizieren muß, kann sie jedoch nicht auf dem α-Niveau ihre Begründung finden. Weder in «paradigmatischen» Verhältnissen (Druck im System, mangelnde Ökonomie des Systems usw.)[2] noch in «syntagmatischen» Verhältnissen (Assimilation, Dissimilation usw.) können hinreichende Gründe gefunden werden. Die in der sprachlichen Kommunikation verwendeten Laute scheinen zu jeder Zeit ungefähr gleich gut funktioniert zu haben. Folglich sind die Veränderungen in dieser Hinsicht nie notwendig gewesen. Tatsächlich hat man ja nie erklären können, warum eine Lautveränderung zu dieser und nicht zu jener Zeit eingetreten ist. Das alles heißt, daß die bisherigen «Erklärungen» nicht ausreichend gewesen sind.

[1] Vgl. S. 12 meines Aufsatzes *Réflexions...*.

[2] Vgl. *K. Togeby* und Fußnote 48 meines Aufsatzes *Inquéritos linguísticos*. Wie *Togeby* bin ich der Meinung, daß diejenigen Teile der diachronischen Phonologie, die sich auf die Beschreibung der Veränderungen (ohne sie zu «erklären») beziehen, relevant und wertvoll sind.

Gründe (die für die Linguistik als hinreichend betrachtet werden können) scheinen aber in Verhältnissen des β- und besonders des γ-Niveaus vorzuliegen. Aussprachegewohnheiten – wie andere sprachliche Tatsachen – unterscheiden Gruppen von Individuen gemäß ihren sozialen (γ, b) und geographischen (γ, c) Bedingungen. Die Sprecher ändern ihre sprachlichen Gewohnheiten, um soziolektale und dialektale Unterschiede zu vergrößern oder zu verkleinern[3]. Zu erklären, wie das zugrunde liegende Gefühl des sozialen und geographischen Gruppierens sich ändert, ist hingegen keine linguistische Angelegenheit.

Der Versuch scheint nutzlos, zu erklären, warum ein Laut *a* von den Sprechern in *b* und nicht in *c* geändert wird, weil die Änderung als arbiträr anzusehen ist. Es ist ja überhaupt nur ausnahmsweise möglich, denjenigen zu finden, der eine neue Aussprache als erster verwendet hat. Und auch wenn man ihn finden würde, wäre es im allgemeinen unmöglich festzustellen, warum er *b* und nicht *c* gewählt hat.

Geben wir ein Beispiel (unter den Tausenden von möglichen Beispielen). In Stockholm wird der Langvokal */a:/* meistens dunkler (velarer) ausgesprochen als in anderen Teilen Schwedens. Die dunkelsten Varianten findet man bei «ungebildeten», oft jungen Individuen. Es gibt aber auch sehr helle (palatale) Varianten, die einer kleinen Gruppe von hauptsächlich Intellektuellen eigen sind. Der Unterschied zwischen den zwei Vokalvarianten hat soziolektale Funktion. Beide Varianten zusammen, gegen Varianten anderer Mundarten gestellt, haben dialektale Funktion.

Wenn das Bedürfnis der Sprecher, sich, so wie es eben erklärt wurde, gemäß sozialer und geographischer Zugehörigkeit zu unterscheiden, verschwindet, dann wird sich auch das erwähnte sprachliche Verhältnis ändern: alle Schweden werden das */a:/* ungefähr gleich aussprechen.

Daß bestimmte Sprecher die erwähnte helle Vokalvariante und andere die dunkle Variante verwenden, ist arbiträr (im Sinne *F. de Saussures*). Die Verhältnisse hätten ebensowohl umgekehrt sein können, denn es liegt nicht in der «Natur» eines hellen Vokals, «gebildeter» als ein dunkler Vokal zu klingen.

In der Welt von heute, wo die Gruppen politisch, kommerziell, kulturell usw. immer größer werden, verschwinden, wie man er-

[3] Diese Auffassung gründet sich auf den Finalitätsbegriff, wie er von *E. Coseriu* entwickelt wurde.

warten muß, nicht nur soziolektale und dialektale Unterschiede, sondern sogar ganze Sprachen[4].

Literatur

Coseriu, E.: Sincronía, diacronía e historia (Univers., Montevideo 1958).
Hammarström, G.: Inquéritos linguísticos II. Revista de Portugal, Série A: 9–32 (1961).
Hammarström, G.: Réflexions sur la linguistique structurale et la phonétique expérimentale. Phonetica *9:* 11–16 (1963).
Togeby, K.: Les explications phonologiques historiques sont-elles possibles? Romance Philology *13:* 401–413 (1959/60).

Adresse des Autors: Prof. Dr. G. Hammarström, Fonetiska Institutionen, Thunbergsvägen 3, *Uppsala* (Schweden).

Discussion

Martinet (Paris): La phonologie diachronique n'exclue aucun des trois plans de l'analyse bühlerienne. Il y a interaction naturelle, comme le montre l'exemple même choisi par *M. Hammarström:* les prononciations populaires sont celles qui sont le plus directement affectées par les pressions internes, les prononciations des gens cultivés tendent naturellement à la conservation. Dans le cas du suédois *gata* le *a* long accentué tendra très naturellement à passer vers (ɔ) dans une langue où *u* est devenu une sorte de [y], *o* un [u] *a* presque un [o]; en danois, où *u* est resté [u], l'asymétrie des organes tend, au contraire, à faire passer le *a* de *gade* (= suédois *gata*) à [ε] et ceci se manifeste surtout dans le parler populaire.

Sovijärvi (Helsinki): Herr *Hammarström* hat in seinem Vortrag u. a. die sogenannten «soziolektalen» Faktoren der Lautveränderungen erwähnt. Ich möchte ihn fragen, ob auch diejenigen Lautveränderungen, die wahrscheinlich auf Grund der Einwirkung der speziellen phonetischen Eigenschaften der *Nachbarsprache* erklärt werden können, zu den erwähnten Faktoren gehören. Im Wotischen z. B., das zu den ostseefinnischen Sprachen gehört, hat sich das [k] vor den Vordervokalen im 14. Jahrhundert (laut *Lauri Posti*) durch Einwirkung der russischen Sprache in eine Dorsalaffrikate [tʃ] verändert: z. B. im Wort [tʃeŋŋæt] lautete der Anfang früher [ke–] (vgl. Fi. *kengät* 'Schuhen'), aber gemäß der russischen Phonetik wurde der Anlaut stark palatalisiert [k'é–], wonach er sich in eine Affrikate verändert hat.

Tillmann (Bonn): Sie sprechen von einer Absicht der Kommunikationspartner, sich mit Hilfe der Merkmale der γ-Ebenen voneinander und von andern zu unterscheiden. Ich weiß nicht, ob man unbedingt eine Absicht (abgesehen sei von möglichen «Bewußtseinsgraden» derselben) annehmen muß. Auch die Rolle der γ-Merkmale, und zwar aller drei Kategorien, dürfte als eine soziale kulturell geprägt und von Kultur zu Kultur unterschiedlich ausgeprägt sein.

Tatsächlich aber werden die Merkmale der γ-Ebenen erst interessant, sobald (verschiedene) Individuen, (verschiedene) Sozio- und (verschiedene) Dialekte aufeinandertreffen. Man muß deshalb annehmen, daß, wenn wir das Individuelle einmal ausklammern, für den typischen Vertreter eines bestimmten Soziolektes selbst das

[4] Vgl. S. 16 (insbesondere Fußnote 26) des oben erwähnten Aufsatzes *Inquéritos linguísticos, II.*

Soziolektale schon durch die α- und β-Ebenen restlos gegeben ist. Andererseits ist ein und derselbe Soziolekt für den Vertreter verschiedener anderer Soziolekte mit Sicherheit jeweils verschieden. Wenn man nicht von einem abstrakt dialektologischen Gesichtspunkt aus die sozio- und dialektalen Merkmale als geradezu geographisches Neben- und Miteinander aufzeichnen, sondern die *kommunikativen* Gegebenheiten berücksichtigen will, muß man die Kommunikatoren zum Maßstab nehmen.

Der einfachste Fall, daß zwei Vertreter verschiedener Sozio- und Dialekte A und A′ miteinander in Kommunikation stehen, wäre graphisch folgendermaßen zu skizzieren:

Vertreter von A		Vertreter von A′
$\alpha\beta+\gamma \leftarrow$ selbst	$\longleftarrow$	nicht selbst $\leftarrow \alpha'\beta'$
$\alpha\beta \rightarrow$ nicht selbst	$\longrightarrow$	selbst $\rightarrow \alpha'\beta'+\gamma$

Wenn mehr als zwei Soziolekte, gegeben durch typische Vertreter, berücksichtigt werden sollen, so ergibt sich ein recht komplexes Netz, insofern jeder Soziolekt zu jedem anderen in zwei Relationen gesehen werden muß.

Ich glaube, daß auch diese Überlegungen die Fruchtbarkeit der *Hammarström*schen Klassifikation phonetischer Merkmale demonstrieren können.

Proc. 5th int. Congr. phon. Sci., Münster 1964, pp. 340–343
(S. Karger, Basel/New York 1965).

Programmed Learning in a Two-Person Speech System*

By L. S. Harms, Lawrence, Kansas

1. In a manner not anticipated a few years ago, the sequence and control techniques of programmed learning may be utilized in the manipulation of a two-person conversation. My purpose in this talk is to discuss programmed learning in terms of a speaker-listener pair, or more exactly, in a two-person speech system. The important inputs and outputs of this two-person learning operation may be viewed from a system orientation. One basic problem is to identify the relevant variables of the tutor's "information" and teaching technique, and to provide a corresponding account for the activities of the student.

2. The first two-person learning form is the natural learning situation and it is found in any family. With neither special training nor electronic apparatus, it is a 90% safe bet that through an extended conversation the mother will succeed in teaching the child to speak her dialect before the child reaches five years of age. Such a two-person instructional relationship, which is often called the Socratic dialogue, assumes the tutor has both information and teaching technique. The student is viewed as cooperating through conversational interaction with the tutor. The student attempts to acquire information, attitude or skill, and in the end, to approximate the speech behavior of the tutor.

3. A learning program provides a carefully sequenced arrangement of all the materials a good tutor might present to the student. Material is included or excluded on the basis of its relevance to a specified educational objective. The student has the programmed material he is to learn presented to him in "conversational fashion"

* Supported, in part, by University of Kansas General Research Fund Grant 3871–5038

by a teaching machine; he makes frequent responses to this material and immediately receives confirmation or correction of each response. This presentation, response, confirmation event typically requires on the order of one minute. In the body of the learning program, these events must be carefully graded, considerably more detailed than textbooks or other educational materials, and so arranged that the student can independently proceed from the first event in the program to the last one without difficulty. For example, to meet an educational objective of 95% accuracy in phonetic transcription of isolated English syllables, a learning program was prepared which caused a student to hear, transcribe, and immediately check his transcription of 1600 or more syllables[1]. On the completion of a program, a student must be able to demonstrate he can meet a specified objective. If he cannot, the program must be revised.

4. The automated teaching situation permits a student to work through a learning program at his own best rate with the aid of a mechanical tutor, the teaching machine. In this instance, a teaching machine presents a learning program to the student, and thereby performs the functions of the tutor. The responses each student makes while completing the program provide a detailed record of his learning attempts which may be used to revise and refine both the learning program and the teaching machine.

With a simple teaching machine and learning program, our students have achieved 95% accuracy in transcription of isolated English syllables in 4 to 12 hours[2]. The mean time was 6–7 hours – when the students had the same dialect as the voice being trancribed. In learning speech production of selected consonant sounds, foreign students have been able to reduce the number of deficient sounds 63%[3]. Fifteen to 90 minutes per sound was the time range. The mean was 53 minutes per sound. The sounds were English; the learners were foreign students from several countries. While the results in speech production are better than chance they are still considerably below the target of 90% accuracy. Work is underway to increase the efficiency of the system and extend it to larger units in connected speech.

5. Teaching machines for speech learning cost money. Good tutors are scarce and expensive. Students are plentiful. The presentation function of the teaching machine is performed by a student in a two-person learning program. For this approach, we have gained

useful information from programmed learning theory, game theory, psychiatric work on dyads, system analysis and other areas.

While materials are being developed for several speech learning tasks, because of its simplicity, only one example from speech intelligibility will be presented in this brief time[4]. Two students, *A* and *B*, are each given a complete but complementary set of materials; each has one-half of a complete learning program and sufficient instructions to play the tutor role for the other person. After working through the set of instructions *A* and *B* begin saying "words" to each other. In the beginning a matching pattern is employed. *A* looks at his list and sees: 1. latter, 2. ladder. If number 2 is the test word, *A* will say: "latter, ladder, (pause), ladder." If he receives the word correctly, *B* replies: "ladder two." *A* replies "correct". *B* looks at his list and pronounces his test material while *A* responds. And so on. When an error occurs, the word is marked, and both *A* and *B* work to correct it. From the example, if *B* had replied: "latter one", *A* would record the error and then repeat; if repetition was not sufficient as a clue, context would be added: "A step ladder is used for climbing." Other error correction procedures are also employed. When 12 words are correctly identified in sequence, *A* and *B* move to the next condition, and by degrees, to larger units.

Our initial work in two-person learning programs centered on speech intelligibility. The technique may be used for speech learning of units both larger and smaller than words. Person *A* and person *B* may be, but need not be similar in background. The data collected from the error and time records of each pair are used to guide the refinement of the two-person system. Error data can be entered into a confusion matrix; frequency of occurrence of particular materials in the program may be regulated on the basis of error information. When used in this way, programmed learning serves as a means of instruction for the student, and an instrument for collection of data on the speech learning process for the researcher.

6. Three possibilities for acquiring speech control of various language units are considered here. First, the natural learning situation, and second, the automated instruction situation are discussed. Third, in the two-person system, the listener controls the speaker in a way that ensures that the speaker quickly adapts to the changing requirements of the listener. Thus, the social aspect of the natural learning situation and the control techniques of automated instruction are combined to provide a highly flexible speech centered learning operation.

References

1. *Harms, L.S.:* Phonetic transcription: a programmed introduction (Scott, Foresman and Co., Chicago 1964).
2. *Harms, L. S.:* Programmed learning for phonetic transcription. Phonetica *10:* 42–50 (1963).
3. *Scheib, M.E.* and *Harms, L.S.:* Toward a generalized self-instructional system for speech sound acquisition. KU/CRC/63/10/R/1.0. Mimeo Report, 1963, pp. 1–25.
4. *Harms, L.S.* and *Beisecker, Th. D.:* A two-person self-instructional system for speech intelligibility (report in preparation).

Author's address: Prof. L. S. Harms, Communication Research Center of the Department of Speech and Drama, 116 Strong Hall, University of Kansas, *Lawrence, Kansas* (USA).

Verh. 5. int. Kongr. Phon. Wiss., Münster 1964, pp. 344–347
(S. Karger, Basel/New York 1965).

Die 'distinctive features'-Theorie und Probleme der automatischen Erkennung gesprochener Sprache

Von Georg Heike, Marburg a. d. Lahn

Wir möchten auf einige grundsätzliche Probleme eingehen, die sich unter anderem bei unserer Beschäftigung mit der sogenannten automatischen Erkennung gesprochener Sprache stellen. Diese Bezeichnung wie auch der noch ungenauere Terminus 'Spracherkennung' sind offenbar entsprechend zu 'speech recognition' gewählt. *G. Ungeheuer* schlägt korrekterweise die dem Stand der Forschung angemessenere Bezeichnung 'Identifizierung von Sprachschallmerkmalen' vor. Das Ziel dieser Forschung läßt sich formulieren als die apparative Feststellung von invarianten Sprachschallmerkmalen, die die Signifikanten zu identifizieren und zu unterscheiden gestatten. Insofern gilt es bis zu einem gewissen Grade, nämlich unter Ausschluß von Kontexthilfen, den menschlichen Identifizierungsprozeß durch ein Modell nachzubilden. Das Auffinden und Definieren solcher invarianter Merkmale ist, wie es sich in der Praxis zeigt, weitaus schwieriger, als es die bereits vorliegenden akustisch-phonetischen Untersuchungen erwarten lassen. Zunächst scheint sich die 'distinctive features'-Theorie und die darin formulierten Binärmerkmale in akustischer Definition als weitgehende Näherung an die Lösung des Invarianzproblems anzubieten. Es sind bereits Versuche zur 'speech recognition' zu verzeichnen, die auf diesen Definitionen aufbauen. An dieser Stelle ist es leider notwendig, ausdrücklich darauf hinzuweisen, daß es sich bei der von uns formulierten Aufgabe nicht darum handeln kann, eine apparative Phonem*analyse* anzustreben. Noch viel weniger kann beispielsweise der Sonagraph als ein automatischer Phonemdetektor angesehen werden, wie *Isačenko* in seiner Arbeit über den velaren Nasal im Deutschen ja sehr richtig sagt. Insofern wird der Ansatz von 'distinctive features' zur Beschreibung und weiteren Durchleuchtung eines be-

reits elaborierten Phonemsystems nicht von dieser Feststellung betroffen, nur weil man bei diesem Vorgehen unter anderem auch mit dem Sonagraphen arbeitet. Augenscheinlich basieren solche und andere Mißverständnisse auf der scheinbaren Unversöhnlichkeit von abstrahierender Phonem*definition* und der Untersuchung von Phonem*unterscheidungsmitteln* im Kommunikationsakt. Der vermeintlich strenge Strukturalist fühlt sich zu einer Ablehnung phonetischer Ansätze verpflichtet und der substanzorientierte Phonetiker und Physiker glaubt, eine apparative Phonemerkennung oder gar Phonemanalyse erreichen zu können. Vielleicht ist auch die übersehene Zweideutigkeit des Terminus 'Phonemanalyse' der Grund für solche Mißverständnisse, da einerseits darunter die Bestimmung des Phoneminventars einer Sprache verstanden wird und andererseits die Zerlegung eines konkreten Textes in phonematische Einheiten, wie sie beispielsweise in einer phonemischen Transkription zum Ausdruck kommt. Selbstverständlich ist der zweite Vorgang nicht möglich ohne den ersten und *beide* nicht möglich ohne die Basis phonetischer Daten. Ebenso einleuchtend ist es, daß das weitgesteckte Ziel eines sogenannten Phonemerkenners oder -detektors nur durch eine Bewertung apparativ extrahierter Signalmerkmale mit den für die betreffende Sprache maßgebenden distributionellen Regeln möglich sein könnte. Es sind bereits in einigen Arbeiten Ansätze hinsichtlich der Einbeziehung von – wie man sagt – linguistischer Information zu verzeichnen. Für einen Automaten, der in der Lage sein soll, ein einlaufendes, optimal produziertes Sprachsignal in eine Folge von Phonemsymbolen zu codieren, möchten wir den Terminus 'Phonemcoder' vorschlagen. Die Termini 'Erkenner' oder 'Detektor' sind nicht angemessen, da es sich nicht nur um eine Feststellung objektiv vorhandener Signaleigenschaften handeln kann, sondern letztlich um das Ergebnis einer Bewertung solcher Signaleigenschaften nach Maßgabe eines vorgegebenen Codes. Wenn wir die folgende Frage positiv beantworten können, so lassen sich einige Bedingungen für das Modell eines Phonemcoders formulieren. Die Frage lautet: «Ist ein phonetischer Experte in der Lage, ein optimal produziertes Sprachsignal einer ihm unbekannten Sprache phonemisch zu transkribieren, wenn ihm das Phonemsystem in expliziter Formulierung des Phoneminventars, der phonetischen Beschreibung der zugehörigen Varianten und von Distributionsregeln vorgegeben ist?» Bejahen wir diese Frage – wie wir es gemäß der beanspruchten Geltung phonologischer Analysiererergebnisse tun müßten – so könnten folgende Stufen eines Phonemcoders grob formuliert werden: 1. Zerlegung eines einlaufenden Schallsignals in simultane Schichten observabler Merkmale, die für die Beschreibung aller Signale der betreffenden Sprache notwendig sind, 2. Codierung von Merkmalskombinationen oder -komplexen in ein begrenztes Allophoninventar, das in akustischer Definition vorliegt, 3. Codierung so definierter Allophone in ein kleineres Repertoire von Phonemen auf Grund geltender distributioneller Regeln. Die Verwirklichung der Stufe 2 setzt eine Lösung des Invarianzproblems voraus und stellt die Hauptaufgabe der gegenwärtigen Forschung dar. Die Stufe 3 wäre im Gegensatz zur Analyse in 'distinctive features' notwendig, da Fälle möglich sind, wo *ein* Merkmalsbündel nicht für alle positionellen Allophone eines Phonems gilt, zum Beispiel die Allophone des /r/-Phonems im Deutschen und weiter die Codierung des velaren Nasals vor dem Plosiv [k] in das Phonem /n/ und in allen anderen Positionen in die Phonemfolge /ng/, falls man die entsprechende Analyse von *Isačenko* akzeptiert. Der Ansatz von 'distinctive features' ist von großer Bedeutung als Bindeglied zwischen der Phonemdefinition als distinktiver Einheit und den Voraussetzungen, unter denen die Distinktion im Sprechakt garantiert wird. Insofern stellt diese Theorie eine wichtige Basis für die Stufen 1 und 2 unseres Modells dar. Mehrere Probleme, von denen wir im folgenden einige nennen wollen, bedürfen jedoch einer Klärung. Unter anderem ist zu nennen: 1. die relationelle Definition einiger Merkmale, die wohl mit gewissen Bedenken zur Beschreibung eines Phonem*systems* ausreicht, wobei alle Phoneme gleichzeitig betrachtet werden, jedoch nicht zur Identifizierung einer einzelnen Phonemrealisation geeignet ist. Ohne Berücksichtigung dieses Problems ergibt sich ein weiteres daraus, daß bei der analytischen Transkription nach erfolgter auditiver Segmentation eine Merkmalsbeschreibung der Segmente vorgenommen wird, während in unserem Fall die Segmentation gerade aus der Merkmalsbeschreibung hervorgehen müßte. Die dichotomische 'distinctive feature'-Definition auf paradigmatischer Ebene findet im kontinuierlichen Schallstrom keine eindeutige Entsprechung etwa in Form eines abrupten Merkmalssprunges. 2. das Invarianzproblem kontextueller und interindividueller Varianten; daraus ergibt sich unter anderem die Unmöglichkeit, die Invarianz von Qualitäten verschiedener individueller Bezugssysteme durch fixe, absolut definierte Merkmale zu erfassen; 3. die nicht ausreichende Entsprechung der Merkmalsdefinitionen zu Wahrnehmungsqualitäten. Die Notwendigkeit einer solchen Entsprechung ist von *R. Jakobson* und *M. Halle* mehrmals ausgesprochen worden. Fassen wir die 'distinctive features' mit *G. Ungeheuer* primär als Wahrnehmungsqualitäten auf, so bleibt dennoch die Notwendigkeit, sie nur in ihren genetischen und akustischen Korrelaten quantitativ bestimmen zu können. Aus dem Gesagten ergibt sich die Forderung nach einer umfassenden psychophonetischen Grundlagenuntersuchung, die eine Zuordnung von Signaldaten zu Wahrnehmungsqualitäten im Sinne auditiv-essentieller Sprachschallmerkmale zum Inhalt hat.

Adresse des Autors: Dr. Georg Heike, Forschungsinstitut für deutsche Sprache, Kaffweg 3, *Marburg a. d. Lahn* (Deutschland).

Proc. 5th int. Congr. phon. Sci., Münster 1964, pp. 348–352
(S. Karger, Basel/New York 1965).

The Articulation of Final '–nh' and '–ch' in Vietnamese

By Eugénie J. A. Henderson, London

The official romanized writing system of Vietnamese has four conventions for the spelling of nasal consonants, namely: 'm', 'n', 'ng'[1] and 'nh'. There is general agreement that in syllable initial position these represent bilabial, alveolar, velar and palatal nasals respectively. Most scholars from the seventeenth century onwards have assumed that in Northern Vietnamese the same four-fold distinction is also valid in syllable final position, both for nasals and for stops, the latter being spelt 'p', 't', 'c' and 'ch'. In Southern Vietnamese the position is quite different in that finally non-labial articulations are regularly post-alveolar after front vowels, and velar after central and back vowels.

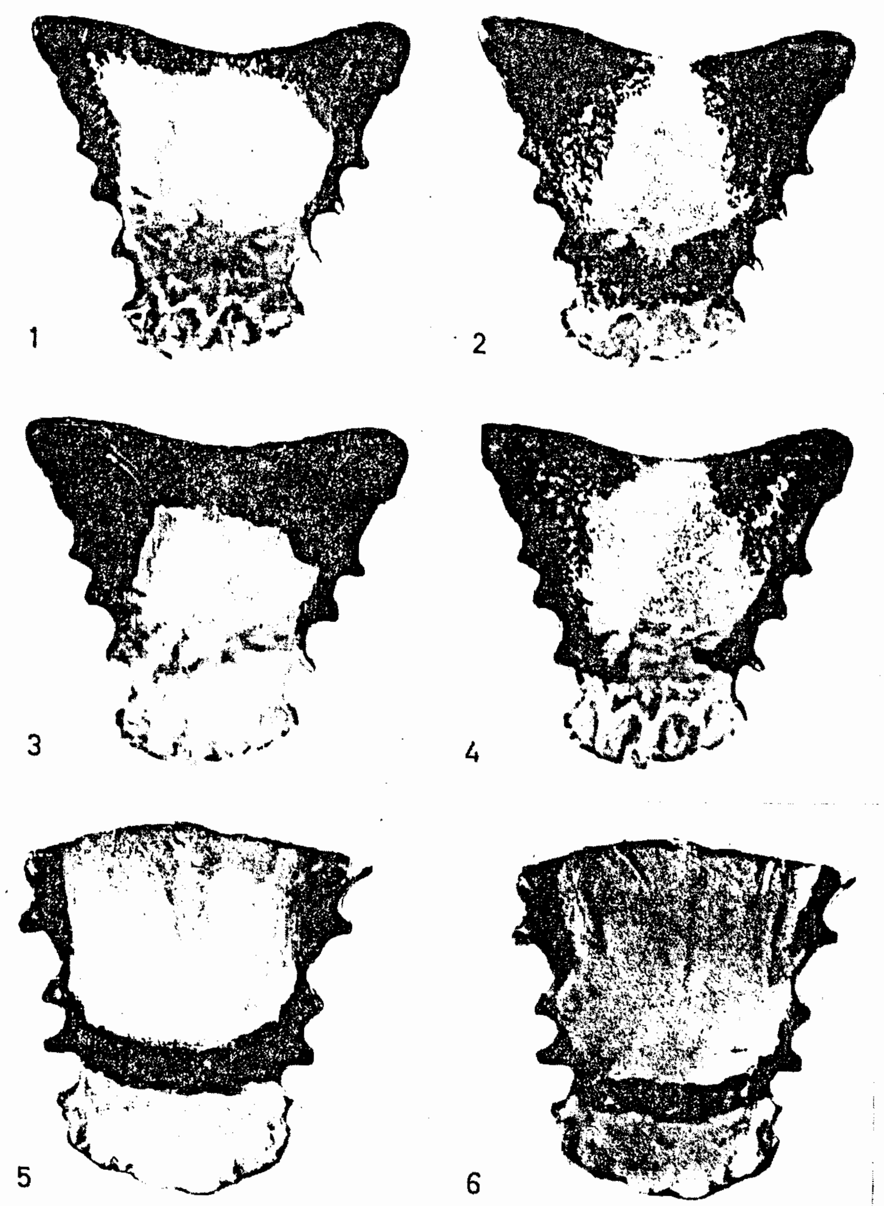
Writing of Tonkinese pronunciation in 1651[2], *Alexandre de Rhodes* described the pronunciation of 'nhà' *house* as being the same as that of the Italian sequence 'gna', and that of 'manh' *strong* as being like Portuguese 'manha' pronounced without the final vowel, thus implying palatal articulations both initially and finally. In similar fashion he equated the initial sound of Vietnamese 'cha' *father* with that of Italian 'cia', Portuguese 'cha', and the pronunciation of Vietnamese 'cách' *manner* with that of Portuguese 'cacha' with the final vowel omitted. Subsequent lexicographers, grammarians and phonologists[3] down to this day have, almost without exception[4], followed *de Rhodes* in assuming identical or very similar articulations of 'nh' and 'ch' initially and finally in Northern

[1] 'ngh' before front vowels.
[2] In the short description of the language appended to the *Dictionarium Annamiticum*.
[3] E.g. *Bonet*, *Génibrel*, and *Legrand de la Liraye* among lexicographors; *Barbier*, *Bulteau*, *Chéon*, *Diguet*, *Dirr*, *Emeneau*, *Gouzien*, *Julien* and *Vallot* among grammarians; *A.C. Day* and *Le Van Ly* among phonologists.
[4] A notable exception is the Vietnamese scholar *Nguyen Bat Tuy*.

Vietnamese, although some[5] have recognized that finally (certain onomatopes and loans apart) they are in complementary distribution with the velars. The spelling has undoubtedly played an important part in the perpetuation of this assumption, as have certain peculiarities in the pronunciation of syllables containing final 'nh' and 'ch'. There is, for example, in the NV pronunciation of such syllables, a clearly perceptible palatal on-glide to the final consonant, which inclines the listener to expect a following 'palatal' articulation. Moreover, the three vowels (of three degrees of openness) that may occur before 'nh' or 'ch' are always pronounced very short and are markedly centralized, which makes their identification with vowels in other contexts difficult. The seventeenth century solution was to write these vowels 'i', 'ê' and 'a', and the coexistence of written forms 'ang' and 'ac' beside 'anh' and 'ach' has seemed to many scholars since then to support, and indeed to require, the hypothesis of four-fold systems of final nasals and of final stops. Misled by such considerations, the present writer[6] also at first concluded that syllables spelt with final 'nh' and 'ch' were closed by palatal articulations not materially different from those found in syllable initial position. Subsequent experience both in listening to a variety of NV speakers and in teaching NV pronunciation to English students gave rise to doubts as the accuracy of this conclusion which were confirmed by palatographic investigation of the speech of a typical NV speaker, a native of Hanoi. In the course of the investigation scores of palatograms were made over a period of some months, and to encourage naturalness of pronunciation whole phrases and sentences, as well as isolated words, were used. Subsequently a similar investigation of SV pronunciation was carried out. Fig. 1–4 on page 350 are photographs of characteristic NV utterances of phrases which may be freely translated as follows: 1. There's no pulse 2. Your grandson 3. Your younger brother 4. He is at home. Fig. 5 and 6 show photographs of characteristics SV utterances of the first and third phrases.

Palatograms of initial 'nh' and 'ch' in both dialects regularly indicated raising of the sides of the tongue in the alveolar or post-alveolar region so as to narrow the air passage, and sometimes even to block it completely, together with some bunching and narrowing of the sides of the rear part of the tongue towards the back of the

[5] E.g. *Emeneau* and *Day*.
[6] See *Le Maître Phonétique*, No. 79, 1943.



1. NV Không có mach

2. NV Cháu của ông

3. NV Em của anh

4. NV Ông ở nhà

5. SV Không có mạch

6. SV Em của anh

hard palate (see fig. 2 and 4). NV final 'nh' and 'ch', on the other hand, never showed the slightest sign of contact or narrowing in the alveolar region, but regularly indicated firm contact of the back of the tongue extending over quite a considerable area of the back of the hard palate (see fig. 1 and 3). SV final 'nh' and 'ch' showed, as expected, a firm post-alveolar contact, (see fig. 5 and 6) which contrasted sharply with that shown for NV utterances of the same phrases (see fig. 1 and 3)[7]. It is clear that whatever articulatory labels we may attach to initial 'nh' and 'ch', the same labels cannot properly be applied to the NV final sounds also. If it is decided to label the NV finals 'fronted velars' or 'palatalized velars'[8] the label 'palatal' may without ambiguity be retained for the initials, although the articulations so designated are not of the dorso-palatal type commonly adduced for the palatal column in the IPA consonant chart. If, on the other hand, the finals are considered as displaying sufficient contact with the hard palate to justify the label 'palatal', some other term must be found for the initials. Since final 'nh' and 'ch' are, in phonemic terms, clearly the allophones after front vowels of phonemes written 'ng' and 'c' in other contexts, the label 'fronted velar' has the advantage both of suggesting the congruence of phonetic substance and phonemic interpretation and of emphasizing the fact that the final consonantal alternance, whether of nasals or of stops, is one of three terms, not of four.

Author's address: Professor E. J. A. Henderson, School of Oriental and African Studies, Department of South East Asia and the Islands, University of London, *London, W. C. 1.* (England).

Discussion

Fujimura (Stockholm): The NV "fronted velars" in final position are here described as phonetically different from the corresponding consonants in initial position. It would be good if the shift in the phonetic value could be acounted for from a general point of view, e.g. by considering the physiological constraints in articulation or general trends of the pertinent language (dialect) in contrasting finals to initials, or more specifically, by giving a systematic set of rules that would prescribe the differences depending on the position of the sound.

Annan (Leeds): The Vietnamese recognize diphthongs before certain nasals in final position, but is this a question of loan-words and/or linguistic sophistication?

[7] It is emphasized that the statements made in this paper about articulation are valid for a variety of contemporary Vietnamese only. It is, of course, impossible to say whether or not a similar articulatory pattern existed in the Tonkinese of de Rhodes' day.

[8] This is the term used by the writer for these sounds in an article that appeared in the Transactions of the Philological Society of Gt Britain for 1951.

Mangold (Saarbrücken): How would you transcribe the finals in questions in narrow transcription for practical touching purposes?

Gage (Washington): First, I should like to ask what the action of the tongue tip is for the sounds you are discussing, since for initial 'nh-' and 'ch-', at least for Southern Vietnamese, the tongue tip can easily be seen to be pressed quite firmly against the back of the lower front teeth.

As my personal feeling, I would say that the existence of the 'eng' and 'éc' types of words, and similarly words with 'oong', has gone beyond the point where they can easily be regarded as really outside the system.

And you somewhere overstate the case when you say that "phonemically nobody has ever been in doubt it was an allophone of ε" in 'ahh' or that "final 'nh' and 'ch' are, in phonemic terms, clearly the allophones after front vowels of phonemes written 'ng' and 'c' in other contexts". Most authors who have treated these sounds as /ŋ/ and /k/ have written /ay/ in 'anh', as with *Jones* and *Thông*, or *Hòa*. And *Emmenneau*, including 'eng' words in his system, phonemicised 'anh' as /a+palatal n/, with the same vowel as 'ǎn'. And, probably prejudiced by listening mostly to Southern Vietnamese, and by my ideas about its phonemic system, I usually hear 'anh' as starting with a vowel sound closest to that usually spelled 'â'. It is of course quite clear historically and from internal reconstruction that these sounds were the allophones of /ŋ/ and /k/ after front vowels at some earlier time, but their present phonemic status is not quite such a straightforward matter.

C.R. 5e Congr. int. Sci. phon., Münster 1964, pp. 353–355
(S. Karger, Basel/New York 1965).

Plan phonique et plan significatif, phonétique et phonologie – Les fondements de la distinction

Par J. G. C. Herculano de Carvalho, Coimbra

Le besoin de distinguer d'une part entre un plan phonique et un plan significatif (plan de l'expression et plan du contenu selon la terminologie de la glossématique, les deux articulations du langage d'après *M. Martinet*), d'autre part entre une discipline phonétique et une discipline phonologique, est sans doute un des postulats les plus solidement établis de la linguistique moderne. Nous ne nous proposerons donc pas d'apporter de nouvelles preuves ou de nouveaux arguments à l'appui de ces distinctions, mais uniquement de les reformuler et d'établir d'une façon peut-être plus explicite le principe général sur quoi elles se fondent.

Ce principe réside premièrement dans le caractère double du signe linguistique – qui n'est pas à proprement parler *un* signe, mais une association de *deux* signes, le signifiant et le signifié – et ensuite et surtout dans la nature même du signifiant.

En effet, ce que nous appelons *signifiant* est en réalité ce que, dans une théorie générale des signes, on nommerait un *signe instrumental*. Par *signe instrumental*, *extérieur* – opposé à *signe formel* (*signe intérieur*, auquel appartient la signifié) – nous comprenons un signe qui, étant constitué par un objet, c'est à dire par une chose physique, est d'abord connu en tant qu'objet, faisant ensuite connaître une autre chose différente de lui-même.

Comme exemples, on pourrait citer les empreintes sur le sable d'un animal déterminé, signe de cet animal; la fumée, signe du feu; le son d'une cloche, signe de la cérémonie religieuse qui va avoir lieu; la lumière verte, signe de passage libre. Tout cela, les empreintes, la fumée, le son de la cloche, la lumière verte, ce sont d'abord des choses, connues en tant que telles, mais des choses qui fonctionnent comme des signes, nous portant à la connaissance d'autres choses.

C'est exactement ce qui se passe avec le *signifiant*: il est une chose, un objet sonore, qui fonctionne comme signe, puisque, une fois connu en tant qu'objet sonore, il nous fait connaître autre chose, c'est-à-dire le *signifié* et, dans celui-ci, la réalité qui y est contenue. Il appartient d'ailleurs à cette catégorie spéciale de signes instrumentaux qui n'existent que pour servir de signes, dont l'existence s'épuise, pour ainsi dire, dans leur fonction de signes.

Voilà donc établi le fondement même de la distinction entre un plan phonique et un plan significatif. C'est en effet cette nature du signifiant qui nous oblige à distinguer dans le langage deux plans totalement différents: d'un côté un plan phonique, sur lequel se situe le signifiant en tant que chose physique, objet sonore, et de l'autre côté un plan significatif, sur lequel se placent non seulement le signifié mais le signifiant en tant que signe.

Dans sa qualité d'objet, le signifiant, considéré indépendamment de sa valeur significative, est analysable en syllabes, en phonèmes (ou en phones), en propriétés distinctives (ou simplement phoniques) – toutes entités qui sont antérieures au signifiant en tant que signe, n'étant en somme que des parties ou des pièces dont se compose l'objet signifiant.

Dans sa qualité de signe, le signifiant ne se décomposera (s'il est décomposable) qu'en d'autres signifiants, de telle façon que les deux analyses ne coïncident pas, si ce n'est que par hasard.

Comme discipline du plan phonique nous ne considérerons qu'une seule, que nous nommerons *Phonétique* «lato sensu». Mais, selon la perspective que l'on adopte pour considérer l'objet signifiant, il y aura lieu de distinguer, non pas *deux*, mais *trois* sous-disciplines phoniques différentes.

Le signifiant pourra donc être envisagé comme une réalisation momentanée, instantanée; comme un objet matériel unique, donné une fois pour toutes. Il sera alors analysé dans sa physicité même, dans sa matérialité immédiate, au même titre (presque...) que le cri d'un animal. La discipline phonique placée dans cette perspective s'appellera *Phonétique matérielle* ou *Phonétique de la parole*.

Le même signifiant pourra ensuite être considéré comme un schéma normal de réalisation. C'est-à-dire, dans cet objet matériel unique on fera abstraction de ce qui y est justement unique et accidentel, pour n'y fixer que ce qui y est constant, se trouvant identique (au moins intentionnellement) dans d'autres objets matériels, également uniques dans leurs apparitions phénoméniques, comme

autant de manifestations d'une seule configuration phonique, toujours la même. Nous désignerons cette perspective comme *Phonétique normale* ou *Phonétique de la norme (du système normal)*.

Jusqu'ici, on a volontairement laissé de côté toute référence à la valeur significative du signifiant (qui est pourtant toujours sous-entendue). Dans la troisième perspective, au contraire, il faudra envisager cette configuration phonique, déterminée dans la perspective précédente, en tant que porteur d'une signification. Il n'importe pourtant pas d'établir *quelle* est la valeur significative – le signifié – de ce signifiant particulier, mais uniquement de constater qu'il possède une valeur significative et que celle-ci est distincte de celle portée par d'autres signifiants. C'est-à-dire que cet objet signifiant x est un signe distinct de ces autres objets signifiants y, z, etc. Ensuite – et voilà le but de l'analyse entreprise sous ce troisième point de vue – il importera de déterminer ce qui fait que l'objet x soit distinct, comme objet-signe, des objets y, z, etc.; ce qui fait que chacun d'eux soit toujours identique à soi-même dans chacune de leurs réalisations concrètes et en même temps distinct de tous les autres. La discipline placée dans cette perspective sera donc, bien entendu, la *Phonétique fonctionnelle* ou *Phonétique du système fonctionnel (du schéma*, dans la terminologie que je préfère) ou simplement *Phonologie*.

Adresse de l'auteur: Prof. José G. Herculano de Carvalho, Rua Dr. Augusto Rocha, 7 *Coimbra* (Portugal)

Verh. 5. int. Kongr. Phon. Wiss., Münster 1964, pp. 356–361
(S. Karger, Basel/New York 1965).

Über Sprechdauer und Sprechtempo bei thematisch ähnlichen Aufgaben mündlicher Sprachgestaltung

Von WILHELM L. HÖFFE, Dortmund

Vorbemerkung: An einer Dortmunder Gemeinschaftsschule wurden über 3 Jahre Versuchsreihen durchgeführt, deren Ziel es war, die gesamtsprachliche Entwicklung von Kindern einer Schulklasse über einen längeren Zeitraum zu verfolgen. Im Rahmen dieser Aufgabe wurde unter anderem auch nach der mündlichen Aussageweise und Gestaltungskraft der Kinder gefragt. Im Anschluß an diese Untersuchungen habe ich die Daten einer Arbeit zur Grundlage von weiteren Fragen genommen.

A. Die Fragen

1. Welches sind die absoluten Gesamtwerte von *Sprechdauer* und *Sprechtempo* bei zwei thematisch ähnlichen Aufgaben mündlicher Gestaltung, welches die entsprechenden Durchschnittswerte?
2. Wie stehen die Durchschnittswerte der beiden Versuchsreihen zueinander?
3. Wie stellen sich die Ergebnisse dar, wenn man die gefundenen Daten einzelner Versuchspersonen miteinander vergleicht?
4. Welches sind die Endergebnisse und welche Folgerungen können aus ihnen gezogen werden?

B. Das Untersuchungsverfahren[2]

Sprecher waren 11 Kinder, 7 Jungen und 4 Mädchen, einer 4. Grundschulklasse. In ihrem eigenen Klassenraum saßen sie ihren Mitschülern gegenüber und sprachen in ein Mikrophon, eine Form, die ihnen nicht unbekannt war. Der *Versuchsleiter*, eine Dame, saß zwischen Sprecher und Hörer am Rande des Klassenraumes, um so gegebenenfalls helfend eingreifen und das Tonbandgerät möglichst

unauffällig bedienen zu können. Vor jeder Aufgabe führte der Versuchsleiter mit der ganzen Klasse ein kurzes Gespräch (etwa 5 min), um so die Kinder, Sprecher und Hörer, für das einzelne Thema zu öffnen. Die *Aufnahmen* der beiden Versuchsreihen wurden im Abstand von mehreren Wochen durchgeführt, im Sommer Versuchsreihe A und kurz nach Weihnachten Versuchsreihe B. Als Aufnahmezeiten wurden die 2. und 3. Unterrichtsstunde gewählt; die Aufnahmen nahmen nie mehr als 45 Minuten in Anspruch; das heißt: die Kinder konnten wohl kaum ermüdet sein.

C. Die Aufgaben

1. Erzählt einmal, was ihr alles auf der Straße gesehen und erlebt habt! (Versuch A).
2. Erzählt ein besonderes Ereignis oder Erlebnis von Weihnachten oder Silvester! (Versuch B).

Beide Aufgaben können entweder mehr als *Sachbericht* oder mehr als *Erlebnisbericht* gelöst werden, wenngleich das 2. Thema zweifellos stärker zum Erlebnisbericht hinzielt (vgl. *Winkler*[7]). Der breitere Rahmen des 1. Themas kann sich – theoretisch gesehen – sowohl negativ wie positiv auf *Sprechdauer* und *Sprechtempo* auswirken (vgl. *O. von Essen*[1]).

D. Die Ergebnisse

I.

Zu Frage 1 (Sprechdauer)

Die Gesamtwerte betragen für Versuchsreihe A = 659 sec, für B = 1246 sec. Als Durchschnittswerte wurden errechnet: 59,9 sec bzw. 113,3 sec.

Zu Frage 2 (Sprechdauer)

Setzt man den Durchschnittswert von Versuchsreihe A mit 100% an, so ergeben sich für Versuchsreihe B = 188%. In Worten: bei Thema 2 werden durchschnittlich beinahe doppelt so lange Zeiten der Sprechdauer gemessen.

Zu Frage 3 (Sprechdauer)

Zwei Versuchspersonen (4, 11) benötigten für Versuch A etwas mehr Zeit als für B; die Differenz beträgt – grob gerechnet – aber nur etwa 10%. Beide Versuchspersonen gehörten übrigens in beiden Versuchen zu den Sprechern, deren Sprechleistung die geringsten Meßwerte aufweisen:

Versuchsperson 4, Versuch A = 70 sec, Versuch B = 65 sec
Versuchsperson 11, Versuch A = 40 sec, Versuch B = 35 sec

Annähernd gleiche Werte für beide Themen sind bei Versuchsperson 1 abzulesen: Versuch A = 58 sec, Versuch B = 61 sec. Auch Versuchsperson 1 benötigt zu beiden Aufgaben nur eine kurze Sprechzeit. Bei den übrigen 8 Versuchspersonen zeigen die Dauerwerte ein deutliches Gefälle. Die extremsten Beispiele liefern die Versuchspersonen 2 und 8:

Versuchsperson 2, Versuch A = 34 sec, Versuch B = 188 sec
Versuchsperson 8, Versuch A = 60 sec, Versuch B = 227 sec

Anders ausgedrückt: Versuchsperson 8 spricht zu Thema 2 fast viermal so lange wie zu Thema 1, Versuchsperson 2 sogar mehr als fünfmal so lange.

Zu Frage 4 (Sprechdauer)

Beinahe 75 % (72,7 %) der Kinder (8) verwenden mehr als die doppelte Zeit für Thema 2 (100 %:221 %).

II.

Zu Frage 1 (Sprechtempo)

Als Gesamtwerte sind errechnet worden für Versuchsreihe A = 1359 W/min, für Versuchsreihe B = 1607 W/min.

Die Durchschnittswerte betragen für Versuchsreihe A = 123,5 W/min, für Versuchsreihe B = 146,1 W/min.

Zu Frage 2 (Sprechtempo)

Wird der Durchschnittswert von Versuchsreihe A mit 100 % gewertet, dann beträgt der Zuwachs für Versuchsreihe B = 18,4 %. Das bedeutet: das Sprechtempo hat sich für Thema 2 beinahe um 20 % erhöht.

Zu Frage 3 (Sprechtempo)

Bei 2 Versuchspersonen zeigten die Ergebnisse annähernd gleiche Zahlen:

Versuchsperson 1, Versuch A = 121 W/min, Versuch B = 118 W/min,
Versuchsperson 3, Versuch A = 89 W/min, Versuch B = 84 W/min,

genauer gesagt: die beiden Versuchspersonen sprechen zu Thema 2 etwas langsamer als zu Thema 1. Die Unterschiede liegen jedoch bei etwa 5 % und darunter. Bei Versuchsperson 9 ist das Gefälle in gleicher Richtung indes wesentlich größer: Versuch A = 163 W/min, Versuch B = 131 W/min. Hier beträgt die Differenz etwa 25 %. Auch beim *Sprechtempo* – wie bei der *Sprechdauer* – sind es 8 Versuchspersonen, bei denen das Zeitmaß in Versuch B größer ist als in Versuch A. Während es bei Versuchsperson 5 nur um 5 Einheiten differiert (137:142), zeigen die Zahlen der übrigen 7 Versuchspersonen deutliche Unterschiede. Stellen wir die Durchschnittswerte dieser Sprecher gegenüber, so ergeben sich folgende Zahlen: 839 zu 1132; das heißt: betrug vorher bei allen 11 Versuchspersonen die Differenz 18,4 %, so wächst sie bei 7 Versuchspersonen im Durchschnitt auf 33,5 % an. Die größten Extremwerte sind bei Versuchsperson 10 festzustellen: Versuch A = 95 W/min, Versuch B = 176 W/min. Hier beträgt der Unterschied somit fast 100 %. Diese Versuchsperson sprach also zum Thema 2 beinahe doppelt so schnell wie zum Thema 1 (übrigens auch mehr als doppelt so lange!).

Zu Frage 4 (Sprechtempo)

Fast 75 % (72,7 %) der Sprecher (8) haben ein um 30 % größeres Sprechtempo bei Thema 2 als bei Thema 1 eingeschlagen.

E. Zusammenfassung und Folgerungen

Fast 75 % aller Kinder sprachen zu Thema 2 wesentlich länger (Sprechdauer 221 %) und schneller (Sprechtempo 130 %) als zu Thema 1 (Sprechdauer und Sprechtempo je 100 %). Dies war grundsätzlich zu erwarten, nicht vermuten konnte man indes den Umfang der Veränderung: a) die Größe der verlängerten Sprechdauer, b) die Größe des veränderten Sprechtempos (vgl. *Kainz*[5];

Tabelle

Versuchsperson Nr.	Name	Sprechdauer Versuch A	Sprechdauer Versuch B	Sprechtempo Versuch A	Sprechtempo Versuch B
1	Almut	58 sec	61 sec	121 W/min	118 W/min
2	Detlev	34 sec	188 sec	155 W/min	165 W/min
3	Dietmar	90 sec	102 sec	89 W/min	84 W/min
4	Heinz	70 sec	65 sec	114 W/min	133 W/min
5	Joachim	52 sec	81 sec	137 W/min	142 W/min
6	Jürgen	84 sec	147 sec	159 W/min	204 W/min
7	Klaus-Jürgen	67 sec	113 sec	107 W/min	155 W/min
8	Matthias	60 sec	227 sec	136 W/min	169 W/min
9	Monika	38 sec	69 sec	163 W/min	131 W/min
10	Ulrike	66 sec	158 sec	95 W/min	176 W/min
11	Ursula	40 sec	35 sec	83 W/min	130 W/min

Höffe[4]). Durch unsere Versuche werden unseres Wissens hier erstmalig exakte Werte vorgelegt. Darüber hinaus ist erneut folgendes nachgewiesen worden: Die gezielte, das persönliche Erlebnis anrufende Aufgabe weckt im Kinde ein größeres Ausdrucksverlangen (vgl. *Trojan*[6]; *Höffe*[3]), das sich sowohl in der Dauer wie in der «sprecherischen Dichte» mündlicher Sprachgestaltung widerspiegelt. Ob und in welcher Weise den gefundenen *quantitativen* Ergebnissen *qualitative* entsprechen, geht über eine phonetische Fragestellung hinaus. Auf dieses Problem sollte in einer gesonderten Abhandlung näher eingegangen werden.

Literaturverzeichnis

1. *Essen, O. von:* Sprechtempo als Ausdruck psychischen Geschehens. Z. f. Phon. u. allg. Sprachwiss. *3:* 317–341 (1949).
2. *Höffe, W. L.:* Zum Experiment in der Sprechwissenschaft. Wiss. Z. der M. Luther-Univ. Halle/Wittenberg *1956:* 397–400.
3. *Höffe, W. L.:* Sprachlicher Ausdrucksgehalt und akustische Struktur. Habil. Schrift. Wiss. Z. d. Fr. Schiller-Univ. Jena, pp. 793–830 (1956/57).
4. *Höffe, W. L.:* Über Beziehungen von Sprachmelodie und Lautstärke. Phonetica *5:* 129–159 (1960).
5. *Kainz, Fr.:* Psychologie der Sprache. Bd. 3 (F. Enke, Stuttgart 1954).
6. *Trojan, F.:* Der Ausdruck der Sprechstimme (Maudrich, Wien 1952).
7. *Winkler, Chr.:* Peter und der Wolf, zwei Sechsjährige erzählen nach. Wirkendes Wort *11:* 345–360 (1961).

Adresse des Autors: Professor Dr. W. Höffe, Bergstraße 18, *Ergste über Schwerte/Ruhr* (Deutschland).

Discussion

Meriggi (Pavia): Wäre es nicht günstiger, der Berechnung des Sprechtempos die Laute und nicht die Wörter zugrunde zu legen?

Antwort *Höffe:* Die Frage ist sehr berechtigt. Bei den weiteren Untersuchungen mit dem bisher verwendeten Material werden wir durch Stichproben feststellen, ob die Auszählung der Laute die bisher genannten Ergebnisse bestätigt oder in Frage stellt. Im Verlaufe dieser weiterführenden Untersuchungen werden wir auch den Problemen «Pause» und «Sprecheinheiten» nachgehen. Wenn wir uns zunächst mit den Angaben: Wörter je Minute begnügten, dann aus zwei Gründen: 1. Unseren Versuchen liegen Themen zugrunde, die sich weder im Hinblick auf das seelische Engagement noch auf die sprachliche Gestaltungsaufgabe voneinander unterscheiden; das heißt: wir können auch bei dem Maß W/min echte Vergleiche anstellen. 2. Der in der sprachpädagogischen Literatur bekannte und häufig zum Vergleich herangezogene Aufsatz von *Christian Winkler:* «Peter und der Wolf, zwei Sechsjährige erzählen nach», Wirkendes Wort, (S. 345 ff. 1961), fußt gleichfalls auf dem Maß W/min.

Sovijärvi (Helsinki): Nach meiner Erfahrung kann man das Tempo des laufenden Sprechens auch unter dem Gesichtspunkt der Frequenz der rhythmischen Perioden betrachten. Ich habe die Tonbandaufnahmen des gesprochenen Finnischen seit 1946 u. a. so analysiert, daß die Abhörtexte in solche Sprecheinheiten zergliedert wurden,

deren Dauer vom rhythmischen Standpunkt aus verhältnismäßig wenig (höchstens $\pm$ 50%, meistens aber zwischen 0 und $\pm$ 25%) um den Durchschnittswert schwankt. Diese Sprecheinheiten sind oft, aber nicht immer, von einer Pause begrenzt, die meistens als Schlußteil der betreffenden rhythmischen Periode betrachtet wird. Jede Periode hat gewöhnlich 2–3 Sprechtakte, die innerhalb der betreffenden Periode ungefähr als gleich lang zu betrachten sind. Es gibt auch solche Perioden, die sogar 5 Sprechtakte besitzen, und oft bildet die Schlußpause einen eigenen Paustakt, dessen Dauer der Länge der anderen Takte derselben Periode – rhythmisch genommen – entspricht.

Das Tempo des Sprechens wird nun dadurch bestimmt, daß jeweils eine Gaußsche Kurve aus den einzelnen Periodendauerwerten der zu messenden Abschnitte der Abhörtexte errechnet wird. Die Meßresultate machen es möglich, die Durchschnittswerte der rhythmischen Perioden der einzelnen Abschnitte und die Breiten der entsprechenden Schwankungsgebiete derselben Abschnitte miteinander zu vergleichen. In meinem Material schwanken die Durchschnittswerte der rhythmischen Perioden zwischen 1 und 3 Sekunden.

Proc. 5th int. Congr. phon. Sci., Münster 1964, pp. 362–364
(S. Karger, Basel/New York 1965).

Communication Sciences Laboratory, University of Florida

Stroboscopic Laminagraphy of the Vocal Folds*

By HARRY HOLLIEN, Gainesville, Fla.

To the present time application of conventional laminagraphy for the study of laryngeal structure and physiology has suffered from the basic limitation that it has provided only a single exposure of rapid vibratory motion. The result has been an x-ray photograph in which the shadow outlining the vocal folds results from the blurring of many complete vibratory cycles. Because of this complexity, it is impossible to know the nature of the time function summation of vocal fold motion which is represented by the x-ray shadow. It is obvious, however, that the resultant image does not represent any single position of the vocal folds during a vibratory cycle. Moreover, both the shape and size of the shadow obtained are possibly affected by dynamic factors such as amplitude of motion, relative duration of the closed phase, etc. Since some of these factors could vary differentially for certain vibratory conditions, measurements made on conventional laminagrams may be contaminated by the inadequacy of the observational technique. Thus, the validity of vocal fold cross sectional area and thickness measurements can be questioned.

A stroboscopic laminagraph (STROL) was developed in order to resolve these problems. It provides a series of laminagraphic x-ray photographs which show coronal views of the vocal folds at each of several phases distributed equally throughout the vibratory cycle. Although each picture is a composite of several short exposures obtained from a number of vibratory cycles, each x-ray firing occurs when the folds are in the same position. Thus, the folds are seen in

* This program of research is being supported by National Institutes of Health grant NB-04407.

several different positions throughout a cycle and accurate measurements of area and thickness can be performed.

The availability of STROL also permits the study of laryngeal vibration. Until this device was constructed, the only satisfactory way of investigating the vibratory patterns of the vocal folds during phonation was by means of indirect laryngoscopy and associated ultra-high speed photography. Virtually all the basic information relative to laryngeal vibration has resulted from this and allied methodologies. While a substantial research tool, ultra-high speed photography of the vocal folds has two limitations. First, since many individuals are unable to expose their vocal folds due to anatomical unsuitability or severe gag reflex, only a relatively few individuals can be used as subjects. Moreover, since the image photographed is of the superior surface of the folds, activity below this level is visualized imperfectly. On the other hand, stroboscopic laminagraphy can be used with virtually any subject and clear images of vocal fold cross section may be obtained during actual phonation. Of course, this technique also has limitations, the most serious being that the resulting images do not portray successive cycles but rather are a composite image of numerous cycles. Moreover, only one plane along the anteroposterior extent of the vocal folds can be visualized at any given time. Nevertheless, STROL provides important information relative to the dynamics of vocal fold vibration as well as data on the mass and thickness of the folds.

It was the purpose of this presentation to briefly describe STROL primarily by means of photographs of the prototype model being used at the Communication Sciences Laboratory. Finally, photographs of the vocal folds during phonation were presented.

Author's address: Dr. H. Hollien, Communication Sciences Laboratory, University of Florida, *Gainesville, Fla.* (USA).

Discussion

Snidecor (Santa Barbara): The previous commentator showed a distrust of singers in saying that their vibrato would make them less usable in a stroboscopic study. Some years ago, working with *Tiffin* and *Saidveit*, and with relatively crude equipment, we needed to have subjects who could phonate within one or one and one half cycles at 130 cycles per second. Singers proved to be the best subjects although it was necessary to train them to use a steady state tone rather than a fluctuating one. Their precision in holding a given frequency more than compensated for any difficulty with vibrato. Our best subject could not only hold a steady tone, but we could also photograph the total length of the vocal cords at various pitches and intensities. Your method does not, of course require more than a steady tone.

Falc'hun (Rennes): La technique dont M. *Hollien* vient de nous montrer un si beau spécimen nous laisse espérer une solution satisfaisante et définitive d'un grand problème de phonétique historique: celui du passage des occlusives sourdes aux aspirées. On a autrefois expliqué l'aspiration des occlusives germaniques par une articulation à glotte ouverte pendant la tenue, sans doute parce qu'on pensait ne pouvoir expliquer l'aspiration que par un retard dans la fermeture de la glotte après l'explosion marquant la fin de la tenue.

A cette époque, aucune technique ne permettait d'observer ni surtout d'enregistrer la position de la glotte pendant la tenue d'une occlusive. La laminographie stroboscopique, synchronisée avec l'enregistrement de la parole, permettra d'observer la position de la glotte à chaque phase de la tenue et de l'aspiration d'une occlusive. Il est à souhaiter que M. *Hollien* applique bientôt sa nouvelle technique à l'étude de ce problème, et fournisse aux linguistes une série de «laminogrammes stroboscopiques» synchronisés avec des oscillogrammes et des sonagrammes représentant des occlusives, aspirées et non aspirées. Cette documentation sera des plus précieuses pour les linguistes qui essaient de mettre sur pied une théorie générale de la mutation consonantique, et surtout de l'altération des occlusives. Elle leur évitera du moins d'asseoir leur argumentation sur un mécanisme du larynx qui ne serait qu'une vue de l'esprit.

Smith (Hellerup): Your interesting technique has shown – as far as I can see – some gross mobility in the heavy structure of the vocal cords not realized before.

What cannot be observed on your pictures is a longer closure phase. This may partly be dependent on the voice of the subjects at just the moment of take (husky, or breathy voice?), partly on the fallacy of X-ray to give a picture of the movements of the thin mucosa.

It seems as if a movement in the vertical plane of ventricular bands may be observed in your film – unobservable from above.

Answer *Hollien:* I wish to thank Mr. *Smith* for his kind comments concerning my film on Stroboscopic Laminagraphy. He is most correct in suggesting that my technique can be used to study 1. the mass and thickness of the vocal folds under a variety of phonatory conditions and 2. the dynamics of vocal fold vibration in a manner not previously realized.

I should like to differ with Mr. *Smith*, however, relative to his second comment, i.e., that the longer closed phase associated with certain phonatory events, cannot be seen in *Strol* pictures. In this regard, it is possible that the vocal fold vibratory action of the two subjects seen in my movie is somewhat misleading. I should like to point out that in both instances, the individuals were producing a relatively low fundamental frequency (100 cps – 125 cps respectively) at *very soft vocal intensities*. Thus, the closed phase for one of the subjects is a little over 10% of the total cycle and for the other about 20%. Quite obviously, in other modes of vibration, a much longer closed phase would be produced and seen on the films. Such materials are being developed currently and will be discussed in later reports.

I agree with Mr. *Smith* that a possible movement in the vertical plane of the ventricular vocal folds seemed to be the case but I do not know if it resulted from mispositioning the x-rays for motion picture photography or whether such action is universally true for these modes of phonation. In any event, subsequent research may clarify these observations.

Verh. 5. int. Kongr. Phon. Wiss., Münster 1964, pp. 365–366
(S. Karger, Basel/New York 1965).

Zur Theorie der unterscheidenden Eigenschaften
('Distinctive Features')

Von K. HORÁLEK, Prag

Die Ausbildung der Theorie der unterscheidenden Eigenschaften (weiter UE) hängt eng mit der Tätigkeit der Prager linguistischen Schule (Cercle linguistique de Prague) zusammen. Es zeigte sich bald, daß in der allgemeinen Theorie der phonologischen Oppositionen in gewissen Fällen (besonders bei den sogenannten Korrelationen) der Begriff der UE unentbehrlich ist. Den entscheidenden Wendepunkt stellt dabei der konsequente Übergang auf eine akustische Grundlage dar. Dieser Übergang ermöglichte es *R. Jakobson*, eine neue Theorie der konsonantischen Opposition aufzubauen, in der schon der Begriff der UE eine große Rolle spielt (1938). Die Theorie der UE wurde später unter *Jakobsons* Leitung in den USA weiter ausgebildet.

Jakobson unterschied im Jahre 1938 und auch später zwei Arten von phonologischen Gegensätzen: kontradiktorischen und konträren (Kontrastoppositionen). Das Prinzip der Binarität ist in dieser Auffassung beibehalten, die Struktur der binären Oppositionen ist hier aber nicht einheitlich. Bei der Beschreibung der vokalischen Phoneme führte man sogar den Begriff der graduellen Oppositionen (im Einklang mit der Theorie *Trubetzkoys*) ein.

Daraus entstehen verschiedene Schwierigkeiten, besonders bei der Systematisierung der UE in den Matrizentabellen. Man versuchte in der letzten Zeit, diese Schwierigkeiten dadurch wegzuräumen, daß die konträren Gegensätze in kontradiktorische übergeführt wurden. Man machte z.B. aus der Opposition kompakt – diffus zwei Oppositionen: kompakt – nichtkompakt und diffus – nichtdiffus. Diese Lösung ist aber ganz formell und führt zu neuen Schwierigkeiten, besonders bei der phonetischen Analyse dieser Oppositionen.

Wenn man verschiedene Ergebnisse bei der Anwendung der Theorie der UE auf eine und dieselbe Sprache bei verschiedenen Forschern vor Augen hat – ein schönes Beispiel dafür bietet die phonologische Analyse des russischen Phonemsystems, die in der letzten Zeit von einem amerikanischen und von einem tschechoslowakischen Forscher ausgearbeitet wurde –, muß man notwendig daraus den Schluß ziehen, daß die Theorie der UE noch nicht in solchem Maß ausgebildet wurde, wie es für eine strenge Wissenschaft unentbehrlich ist. Man kann sogar heute von einer Krise der Theorie der UE sprechen. Dabei bleibt noch die Frage offen, inwieweit bei der Lösung dieser Krise die akustische Analyse behilflich sein kann. Die phonologischen Gegensätze (und somit auch die UE) sind eine Angelegenheit des Sprachsystems und nicht der phonetischen Realisation. Ein und derselbe Laut kann in einer Sprache eine phonematische Einheit bilden, in einer anderen Sprache aber nicht. Schon daraus folgt, daß die akustische Analyse für die phonologische Beschreibung nur als eine Hilfsmethode ihre Gültigkeit haben kann.

Adresse des Autors: Prof. Dr. Karel Horálek, Komornická 7, *Praha 6 – Dejvice* (ČSR).

Discussion

Rudnyckyj (Winnipeg) hat die Frage gestellt, ob die akzentuierten und nichtakzentuierten Vokale (z. B. im Slavischen) zur Kategorie der Oppositionen oder Korrelationen gehören.

Antwort *Horálek:* Der russische Akzent hat zwei Funktionen, die phonemischunterscheidende (distinktive) und die kulminative. Beide sind untrennbar aneinander gebunden und kommen nicht einzeln vor. Der tschechische Akzent ist gebunden, aber nicht funktionslos (er hat die delimitative Funktion).

Proc. 5th int. Congr. phon. Sci., Münster 1964, pp. 367–370
(S. Karger, Basel/New York 1965).

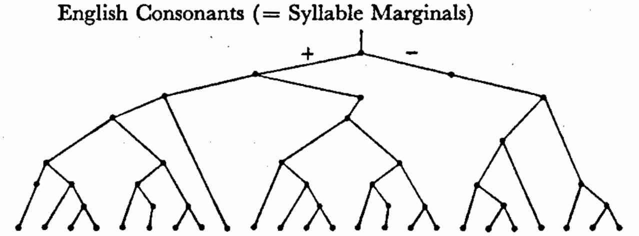
From the University of Illinois

Degrees of Difference of English Consonants

By Lee S. Hultzén, Urbana, Ill.

The ordering of features in the tree shown below was fixed for another purpose; here the tree serves to explain the assignment of features in the accompanying table. In the table the vocalic feature is reversed in sign and consonantal put next to vocalic to agree with the usual practice, and redundant features are marked ·, o, and × to specify certain operations in calculating degrees of difference. In accord with the scheme used by *Sol Saporta* in 1955 (Language *31:* 25–30), which was based on the analysis by *Jakobson, Fant,* and *Halle* of 1952 (Preliminaries to Speech Analysis), the difference between

English Consonants (= Syllable Marginals)



	č	k	t	p	ʃ	s	θ	f	h	ǰ*	g	d	b	ʒ	z	ð	v	ŋ	n	m	l	r	j	w
vocalic	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	+	+	+	+	+	+
consonantal	+	+	+	+	+	+	+	+	−	+	+	+	+	+	+	+	+	+	+	+	+	−	−	−
tense (fortis)	+	+	+	+	+	+	+	+	+	−	−	−	−	−	−	−	−	o	o	o	o	o	o	o
interrupted	+	+	+	+	−	−	−	−		+	+	+	+	−	−	−	−							
nasal																		+	+	+	−			
strident	+	−	−	−	+	+	−	−		+	−	−	−	+	+	−	−							
compact	+	+	·	·	+	−	·	·		+	+	·	·	+	−	·	·	+	·	·		+	·	·
acute	·	·	+	−	·	×	+	−		·	·	+	−	·	×	+	−	·	+	−		·	+	−

* This symbol is used because the standard symbol for the lax-interrupted-strident (affricate) is not available.

+ and — for any feature may be counted as 2 and that bet een + or — and · counted as 1. The o, which makes allowance for the nondistinctive phonetic–tense of sonorants and semivowels, is counted as 1 different from + and not different from —. The ×, for the phonetic acuteness of /s/ and /z/, is counted as 1 different from — and not different from · or +. It is obvious that I am assuming that the distinctive features have phonetic reference, i.e. are observable without or with the aid of instruments, which it seems to me they must be to be of any significance beyond mere coding.

In the main table, on this page, the degrees of difference between consonants are shown below and to the left of the diagonal line of symbols. The analysis is shown to the right and above in blocks of the eight features, ordered:

vocalic	consonantal	tense	interrupted
nasal	strident	compact	acute

The marks ", ', and · indicate two, one, and no difference counts for each feature.

Degrees of Difference

č	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
2	k	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
4	2	t	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
4	2	2	p	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
2	4	6	6	ʃ	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
4	6	5	6	2	s	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
6	4	2	4	4	3	θ	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
6	4	4	2	4	4	2	f	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
2	2	2	2	2	2	2	2	h	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
2	4	6	6	4	6	8	2	4	ǰ	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
4	2	4	4	6	8	6	6	4	2	g	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
6	4	2	4	8	7	4	6	4	4	2	d	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
6	4	4	2	8	8	6	4	4	4	2	2	b	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
4	6	8	8	2	4	6	6	4	2	4	6	6	ʒ	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
6	8	7	8	4	2	5	6	4	4	6	5	6	2	z	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
8	6	4	6	6	5	2	4	4	6	4	2	4	4	3	ð	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
8	6	6	4	6	6	4	2	4	6	4	4	2	4	4	2	v	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
3	3	5	5	3	5	5	5	5	2	2	4	4	2	4	4	4	ŋ	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
5	5	3	5	5	4	3	5	5	4	4	2	4	4	3	2	4	2	n	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
5	5	5	3	5	5	5	3	5	4	4	4	2	4	4	4	2	2	2	m	[illegible]	[illegible]	[illegible]	[illegible]
3	3	3	3	3	3	3	3	5	2	2	2	2	2	2	2	2	2	2	2	l	[illegible]	[illegible]	[illegible]
5	5	7	7	5	7	7	7	3	4	4	6	6	4	6	6	6	2	4	4	2	r	[illegible]	[illegible]
7	7	5	7	7	6	5	7	3	6	6	4	6	6	5	4	6	4	2	4	2	2	j	[illegible]
7	7	7	5	7	7	7	5	3	6	6	6	4	6	6	6	4	4	4	2	2	2	2	w

While the analysis and the numbers are disputable, I do not intend to discuss them now. I want to comment on some situations in which the use of degrees of difference calls for the application of phonetic or distributional considerations which cannot get into the overall systematic array.

My scheme shows differences of 6, 5, and 2 between /t/ and /ʃ/, /s/, and /θ/ respectively. *Saporta* has 7, 4, and 2. Either set of numbers applies well enough for what might be a useful application to contrasts of *tie* with *shy*, *sigh*, or *thigh* in neutral context. Neither scheme will work when the question is on successive segments in a text, as in *Saporta's* article. A /t/ is as completely homorganic with an adjacent /ʃ/ or /s/ as it is with an adjacent /θ/, and there can be no difference among the degrees of difference except for stridency, where /ʃ/ and /s/ are each two degrees farther from /t/ than is /θ/. The counts must be 4, 4, and 2 here.

A similar adjustment must sometimes be made when the contrast between words is in question, perhaps the most generally useful application of the scheme. The feature of voicing is now almost universally considered to be nondistinctive in English, but one cannot deny that in most cases it plays a part in the distinction in perception, and perception is as much a factor in evaluating the usefulness of degrees of difference as is economy of production. In *tray* the /r/ is voiceless, in *dray* voiced. The difference between *tray* and *dray* may be no greater than that between *Tay* and *day*, where the equivalent supplementary differentiation in the aspiration of the /t/ is also not taken care of in the table. It is, however, surely greater than that between *rate* and *raid* before pause with /t/ not aspirated, and /d/ mostly devoiced, where the difference in the length of the vowel is the conspicuous realization of the 2-point fortis-lenis opposition. In the notorious *writing* versus *riding*, the voicing of the /t/ greatly reduces the reliability of perception of the /t/ – /d/ contrast. The degrees of difference should certainly be reduced from the 2 shown in the table to 1.

For *tray* or *dray* versus *ray*, it will hardly do to count up the number of distinctive features in /t/ or /d/ and call that the difference. In this case I think the best count is the difference between /t/ or /d/ and /r/, *tray* no more different from *ray* than is *Tay*, and *dray* – *ray* equal to *day* – *ray*. Note from the table that the first difference is greater than the second. In not quite the same way, *stray* cannot be more than 2 degrees of tension different from *dray* because the /s/

and the phonetic structure of the syllable are predictable; all one needs to perceive is that there is more to the former. Some such operation will, I think, do for all cases where the –l consonant, that next to the vowel, is the same for both words, and perhaps also for contrasts of the shape CV– versus V–.

For such a prevocalic contrast as *black* versus *track*, from an actual list (given by *J. W. Black* in Journal of Speech and Hearing Disorders *28:* 81), the difference can be no more than that between /b/ and /t/ for another reason: in /ˈb–æk/ only /l/ occurs in an available English word and in /ˈt–æk/ only /r/. For *stubble* versus *trouble*, from the same list, the orthographically apparent difference in tension for the –l consonants is of course cancelled by the phonetic structure and the other possible differences by occurrence, leaving only the /s/ – /t/ difference effective.

I make only two comments on postvocalic clusters. The difference between *things* /ˈθɪŋz/ and *thinks* /ˈθɪŋks/ is only that between /z/ and /s/, because the /k/ is predictable. No two clusters can have more than 2 degrees of difference in tension, i.e. one +—, because such difference must be fixed by the first or second consonants and beyond that is invariable, exx.: /–vdz/ – /–fts/ (words?), /–ndz/ – /–lts/.

Author's address: Mr. Lee S. Hultzén, 1102 So. Orchard, *Urbana, Illinois* (USA).

Discussion

Black (Columbus): May I elaborate on the 4-word group Mr. *Hultzén* has put on the board: fair, bare, care, and pair, and the respective scores 40, 76, 47, and 57. These words appear together because when one of them was used as a stimulus the other three occurred most frequently among the error responses. Later, and purely for experimental purposes, each of these three words was used as a stimulus and auditors were asked to identify from the 4-word group which one had been spoken. Thus, when *fair* was the stimulus it was reorganized correctly by 40% of the listeners; when *bare* was spoken it was reorganized by 76% of the listeners, etc.

Now with respect to predicting the foregoing outcomes, I would only call attention to the characteristically low scores that attend [f] and thus would expect *fair* to be the least identifyable of the 4 words, in keeping with the obtained results. I wonder whether or not Mr. *Hultzén* has tried other predictive formulas, for example ones that give weightings to sound pressure level.

Verh. 5. int. Kongr. Phon. Wiss., Münster 1964, pp. 371–373
(S. Karger, Basel/New York 1965).

Deutsche Satzintonationen als Konfigurationen zweier diskreter Tonebenen (mit Demonstrationen von Tonbändern)

Von A. V. Isačenko, Berlin

1. *Hypothese.* Es wurde von der Voraussetzung ausgegangen, daß sich die Variationen der Grundfrequenz, die die syntaktisch relevanten deutschen Intonationsmuster ergeben, als Abfolge von nur zwei diskreten Tonstufen darstellen lassen. Unter syntaktisch relevant werden konventionell «neutrale Aussage», Entscheidungsfrage, Kontrasthervorhebung und weiterweisende Intonation verstanden. Intonationen affektbetonter Äußerungen wurden nicht untersucht.

2. *Versuch I.* Es wurden natürliche Sätze jeweils mit zwei verschiedenen Grundfrequenzen an einem Vocodergerät monotonisiert und auf Tonband aufgenommen. Sodann wurden aus jeweils zwei Tonbändern Intonationsmuster nachgebildet, indem Segmente mit tieferer Grundfrequenz und solche mit höherer Grundfrequenz zu einem Band zusammengefügt wurden. Diese simulierten Intonationsmuster mit diskreten Grundfrequenzen wurden von deutschen Versuchspersonen zu 81,3% richtig identifiziert.

3. *Versuch II.* Zwischen Tief- und Hochton wurden verschiedene Intervalle gewählt. Als charakteristisch für die natürliche Intonation werden in der Literatur Intervalle angegeben, die zwischen zwei Ganztönen und einer Sext liegen. Es zeigte sich, daß simulierte Intonationsmuster auch dann einwandfrei identifiziert werden, wenn das Intervall zwischen den beiden Tonstufen nur *einen Halbton* beträgt. So wurde Frageintonation von Nichtfrageintonation bei einem Intervall von einem Halbton zu 100% richtig unterschieden. Für die Identifizierung syntaktisch relevanter Intonationsmuster ist einzig und allein erforderlich, daß sich Hoch- und Tiefton wahrnehmbar unterscheiden. Ein Halbton liegt aber gerade an der Schwelle

des in zusammenhängender Rede noch wahrnehmbaren Kontrasts. Größere Intervalle werden für die Vermittlung außersyntaktischer (emotionell-expressiver) Information verwendet.

4. *Versuch III.* In Sätzen, in denen die Grundfrequenz einer Silbe gegenüber der monotonisierten Grundfrequenz aller vorangehenden und folgenden Silben verschieden (prominent) ist, wird diese Silbe unabhängig von ihrer akustischen Intensität als hervorgehoben aufgefaßt. Darüber hinaus werden alle Silben, vor denen ein (steigender oder fallender) Tonbruch erfolgt, als hervorgehoben identifiziert. Auch hier wird die Perzeption der Hervorhebung von der akustischen Intensität der Silbe nicht beeinflußt.

5. *Ergebnisse.* Syntaktisch relevante Intonationsmuster können im Deutschen als Abfolge von nur zwei diskreten Tonebenen dargestellt werden, was vor allem die Beschreibung von komplizierteren Satzintonationen ganz wesentlich vereinfacht. Diese Erkenntnis kann im Sprachunterricht bei der Darstellung der deutschen Intonationsverhältnisse erfolgreich verwertet werden.

Das Minimalintervall beträgt bei deutschen Satzintonationen einen Halbton.

Die Hervorhebung einer Silbe wird im Deutschen unabhängig von ihrer akustischen Intensität durch vorangehenden Tonbruch erzielt.

Adresse des Autors: Prof. Dr. A. V. Isačenko, Karl-Marx-Allee 26, *Berlin C 2* (DDR).

Discussion

Trim (Cambridge): Die Versuche beweisen, daß zwei Tonstufen zu den hier erwähnten syntaktischen Funktionen genügen. In einem mehrgliedrigen Satz aber, in dem mehrere betonte Wörter vorkommen, wird oft das eine betonte Wort gegenüber den anderen durch Höherlegung hervorgehoben, was wohl auch als syntaktische Funktion zu bezeichnen wäre. Wenn in einem weiterweisenden Sprechtakt, der nicht vor einer Pause steht, das Kernwort ein zusammengesetztes Wort ist, wird das zweite Glied durch die Senkung des Tones um eine Stufe, nicht aber zur Grundstufe, gekennzeichnet. Auch hier handelt es sich um eine syntaktische Funktion. Natürlich sind diese beiden syntaktisch relevanten Tonabstufungen nicht so unerläßlich wie diejenige, die *Isačenko* beschrieben hat. Sie sind aber nicht zu übersehen, und sie werden wohl nicht die einzigen sein.

von Essen (Hamburg): Die Ausführungen behandeln lediglich die sprachlich-phonologische Funktion der Ausspruchschlüsse und sind für das Deutsche aufschlußreich und zweifellos richtig. Sie haben sicher nicht allgemeinsprachliche Gültigkeit; verschiedene Sprachen verfahren verschieden.

van Katwijk (Eindhoven): Ihrem inspirierenden Vortrag habe ich entnommen, daß zwei Tonhöhenstufen zur Beschreibung der syntaktisch relevanten Intonationserscheinungen genügen. Auch haben Sie uns deutlich gemacht, daß es hauptsächlich die Tonhöhenveränderung – aufwärts oder nach unten, schneller oder weniger schnell – ist, die die Prominenz eines Wortes oder Satztyps bestimmt.

Darf ich Sie fragen, ob Sie mit mir einverstanden sein können, daß die Beschreibung des Intonationsverfahrens vielleicht noch essentiell reduziert werden kann, wenn man nur die *Veränderungen* in Betracht zieht und die *Stufen* der Tonhöhe als sekundäre Gegebenheiten anerkennt, deren Zahl von dem Umstand abhängt, wie groß eine Serie von Tonhöheveränderungen in gleicher Richtung ist?

Pilch (Freiburg): Die Versuchsergebnisse halte ich für hochinteressant. Sie, Herr Kollege *Isačenko*, sprechen von Hypothesen über deutsche Satzintonation. Darf ich fragen, welche? Ich glaube nicht, daß solche Hypothesen sich an die vorliegenden Ergebnisse knüpfen lassen, weil synthetische Rede nicht wirkliche deutsche Rede ist.

Hammarström (Uppsala): Sie meinen, daß Ihnen Ihre durch Hören bewerteten, simulierten Intonationen «Fakten» über relevante Eigenschaften der «natürlichen Intonation» gäben. Ich glaube, es wäre richtiger, von «Hypothesen» anstatt von «Fakten» zu sprechen.

Sie meinen weiter, daß in den zitierten Sätzen die «Intonation» relevant wäre. Ich bezweifle nicht, daß Ihr Experiment möglich und richtig gewesen ist. Nur muß ich annehmen, daß man das, was Sie beim Hören erzielt haben, auch in anderer Weise (durch Pausen, Intensitätsänderungen usw.) erzielt werden kann.

Tillmann (Bonn): Ich hatte den Eindruck, daß durch die schlechte technische Qualität der Modelle der Mangel an Natürlichkeit ein wenig vertuscht wurde. Ich möchte aber gerade deswegen die mangelnde Qualität für testpsychologisch sehr glücklich halten.

Proc. 5th int. Congr. phon. Sci., Münster 1964, pp. 374–377
(S. Karger, Basel/New York 1965).

From the Instytut Podstawowych Problemów Techniki, Poznań

Measurements on the Distinctive Features of Polish Phonemes

By Wiktor Jassem, Poznań

Investigations into the nature of the phonemic distinctive features in Polish are being carried out mainly with a view to obtain data for the design of speech compression systems and mechanical speech recognition possibly leading to real-time translation of spoken language. The distinctive features as here defined are based on results of acoustic (mainly spectrographic) analysis of undistorted speech and experiments with frequency and time filtered speech. In the following definitions it is to be understood that if the feature is present the phoneme is classified as one to which the first of the two-term label applies (e.g. if the phoneme has more than 3 formants, with amplitude levels within a range of 15 db beginning from the second formant, below 3 kc, and/or any antiformants, it is consonantal, otherwise it is vocalic, etc.).

1. *Consonantal : vocalic.* More than 3 formants, with amplitude levels within 15 db beginning from the second formant, below 3 kc, and/or any antiformants.

2. *Supraglottal : glottal.* Noise-like aperiodic excitation (alone or superimposed on quasi-periodic excitation).

3. *Nasal : oral.* More than 3 formants below 3 kc, with amplitudes within a range of 15 db beginning from the second formant.

4. *Smooth-abrupt.* No appreciable pulse-like rapid event.

5. *Compact-diffuse.* [General description: concentration of energy in a middle range of frequencies.] Specific definitions: (A) vocalic phonemes: $F_1 > 450$ cps, (B) consonantal phonemes: (a) nasal: no formants in the range 1.2 to 2 kc, (b) other: most, or all, energy concentrated in the range 1 to 4 kc.

6. *Acute : grave.* [General description: relatively high level of higher frequencies]. Specific definitions: (A) vocalic: (a) compact: $\frac{F_2 + F_3}{2} - F_1 > 1.45$ kc, (b) diffuse: $\frac{F_2 + F_3}{2} - F_1 > 2.2$ kc, (B) consonantal phonemes: (a) nasal: (α) compact: the level of the formant just above 2 kc more than 10 db higher than level of other formants except the first, (β) diffuse: a formant between 1.4 and 1.8 kc, (b) other: almost all energy above 3 kc.

7. *Low-tone : high-tone.* [General description: Low frequency of all formants.] Specific definitions: (A) vocalic phonemes: $F_1 + F_2 < 1.7$ kc, (B) consonantal phonemes: $F_2 + F_3 + F_4 < 8$ kc.

8. *Short : long.* (A) vocalic phonemes: Position next to another vocalic phoneme and duration of steady F_1 segment less than 100 msec, (B) consonantal phonemes: duration of the noise segment less than 100 msec.

9. *Voiced : voiceless.* Quasi-periodic glottal source excitation.

It will be seen that some of the features have complementary variants. E.g., the feature acute:grave has one definition for vowels and another for consonants. Even among the consonants the definition of the acute:grave contrast depends on whether the contrasting phonemes are diffuse or compact. The phoneme /ʒ/, to take an example, is defined according to (1), (2), (4), (5Bb), (6Bb), (7B), (8B), (9).

Wherever, in the above definitions, the distribution of energy in wide frequency bands is mentioned, it is to be understood that a band below 0.7 kc (in which energy, if present, is due mainly to the glottal source) is not considered.

Using two elementary signs (e.g. + and —, or 0 and 1, etc.) each phoneme is uniquely coded by a definite combination of such signs. A binary distinctive feature code is naturally suitable for transmission over bistable systems. Any sequence of phonemes, coded according to the system here proposed, with the signs in temporal succession, may be uniquely decoded from a text message without any additional signs being necessary for the indication of phoneme boundaries*.

The order of the distinctive features is such that the redundancy of the code is reduced to a minimum. The code is constructed in such a way that of the two elementary signs one is always attached

* The present system is a revision of a first version, cf. QPSR, STL 1/1962, Stockholm, pp. 7–14.

Table I

	ī	u	w	i	j	a	o	e	r	l	m	n	ŋ	ɲ	p	b	t	d	k
Consonantal	−	−	−	−	−	−	−	−	+	+	+	+	+	+	+	+	+	+	+
Supraglottal									−	−	−	−	−	−	+	+	+	+	+
Nasal									−	−	+	+	+	+					
Smooth									−	+					−	−	−	−	−
Compact	−	−	−	−	−	+	+	+			−	−	+	+	−	−	−	−	+
Acute	−	−	−	+	+	−	−	+			−	+	−	+	−	−	+	+	−
Low-tone	−	+	+			−	+												
Short		−	+	−	+														
Voiced															−	+	−	+	−
N	4	5	5	4	4	4	4	3	4	4	5	5	5	5	6	6	6	6	6
10^3 p	53	35	20	43	45	90	92	91	36	27	34	44	9	18	27	17	48	24	22

	g	c	ɟ	f	v	s	z	ts	dz	x	c,	z,	tc,	dz,	ʃ	ʒ	tʃ	dʒ
Consonantal	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Supraglottal	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Nasal																		
Smooth	−	−	−	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Compact	+	+	+	−	−	−	−	−	−	+	+	+	+	+	+	+	+	+
Acute	−	+	+	−	−	+	+	+	+	−	+	+	+	+	+	+	+	+
Low-tone											−	−	−	−	+	+	+	+
Short						−	−	+	+		−	−	+	+	−	−	+	+
Voiced	+	−	+	−	+	−	+	−	+		−	+	−	+	−	+	−	+
N	6	6	6	6	6	7	7	7	7	5	8	8	8	8	8	8	8	8
10^3 p	14	8	1	20	32	34	19	14	3	11	15	1	9	5	19	13	13	0*

* $p < 0.0005$.

to the group with more stages of subdivision (thus compact phonemes receive a + because some of them are divided again at 4 stages whereas no diffuse phonemes are divided again more than 3 times. Similarly, acute phonemes may be divided again 3 times, but grave phonemes only twice, so acute is a «plus» feature, etc.). This results in the phonemes tending strongly to order themselves according to the number of features. The table shows this to be certainly true of the consonantal phonemes. Strangely enough, the tendency is almost reversed in the small group of vocalic phonemes.

Measurements leading to the most accurate definitions of the individual features are still in progress at the Acoustic Phonetics Laboratory of the Polish Academy of Sciences, and some details concerning female voices (the frequencies given above in the definitions refer to male voices) have not yet been solved.

The entropy of the information source of the Polish phonemes being $H_s = -\sum_{i=1}^{i=37} p_i \log_2 p_i = 4.744$ bits/phoneme, and the entropy of our natural code being $H_c = \sum_{i=1}^{i=37} p_i N_i = 5.046$ distinctive feat./phoneme (N – number of distinctive features for phoneme with occurrence frequency p), the redundancy of our code is 6%. This redundancy is chiefly due to the fact that N does not always tend to increase with decreasing p. Considering, however, that our features are entirely natural the code redundancy is considered very low. Redundancy due to transitional probabilities may be used for human or mechanical error correction.

Author's address: W. Jassem, Instytut Podstawowych Problemów Techniki, ul. Matejki 48/49, *Poznan* (Poland).

Discussion

Daneš (Praha): I am afraid that the surprising agreement between both rows of members (viz. the numbers of bits and of distinctive features, respectively) is only an incident. Without doubt, one can imagine such a vowel change by which the frequency of a phoneme increases and that of an other one proportionally decreases (e.g. in the case of an incomplete merging of two phonemes), while the phonemic structure of the given language, the structure of the distinctive features, does not change at all.

Hill (Harlow): I should like to raise two points (which concern the discussion we have just heard). First, when constructing a Shannon code the most frequently occurring items for transmission should be represented by the least number of binary choices (in this case distinctive features) for greatest efficiency. Thus the point at issue is resolved by saying that, if a certain phoneme, requiring a certain number of binary choices (distinctive features) for identification, becomes more frequently used, then the speakers of that language are using it less efficiently than they were.

I will ask Mr. *Jassem* to elucidate further the second point, which is how, precisely, he arrived at the series of divisions of the total phonemes to produce the tree illustrating the grouping of phonemes by frequency of occurrence.

Bluhme (Amsterdam): The F_1/F_2 diagram exposed by Mr. *Jassem* shows quite considerable overlapping of the Polish vowel phonemes. However, as we know from phonemic analyses that they are distinctively different from each other, we have serious doubts in the validity of this acoustic approach. Mr. *Mol* has been able to show that the significant differences come out much better if the formants of vowels of one and the same speaker are compared with each other. So it would have been more appropriate if Mr. *Jassem* had used the formular $\sum\left(\frac{F_1}{F_2}\right)$ instead of $\frac{\sum F_1}{\sum F_2}$.

Verh. 5. int. Kongr. Phon. Wiss., Münster 1964, pp. 378–380
(S. Karger, Basel/New York 1965).

Das reduzierte Vokal-Phonem im Wogulischen

Von B. KÁLMÁN, Debrecen

Das Wogulische ist eine finnisch-ugrische Sprache, die mit dem Ostjakischen und dem Ungarischen die sogenannte ugrische Sprachgruppe bildet. Das Wogulische wird heute nur von etwa dreieinhalbtausend Leuten in Westsibirien zwischen dem Uralgebirge und dem Ob gesprochen. Die in Dialekte zersplitterte Sprache bekam erst vor dreißig Jahren ihr Alphabet. Als Basis für die Literatursprache wurde der Norddialekt gewählt.

Mit dem Vokalismus des Wogulischen hat sich zuletzt *Wolfgang Steinitz* befaßt. Seiner Meinung nach besitzt das Nord-Wogulische folgendes Vokal-Phonemsystem:

In der ersten Silbe			In den folgenden Silben	
	Labiale	Illabiale	Volle Vokale	reduz. Vokal
(lange Vokale)	*ɔ̄*	*ā*	*a*	
	ū	*ē*	*e*	*ə*
(kurze Vokale)	*o*	*a*	*i*	
	u	*i*		

Wie wir sehen, ist die Zahl der Vokale auch in der ersten Silbe ziemlich beschränkt und symmetrisch: 4 lange und 4 kurze; 4 labiale und 4 illabiale. Die langen Vokale unterscheiden sich auch qualitativ von den kurzen, sie haben einen größeren Öffnungsgrad /*ɔ̄* – *o*, *ū*˯ – *u*˃, *ā* – *a*˰, *ē* – *i*/. In den nichtersten Silben kommen – unabhängig von der Zahl der Silben –, die Vokale *ə*, *a*, *e*, *i* vor. Labiale Vokalphoneme kommen in nichtersten Silben nicht vor (*Steinitz*, S. 37).

Wie wir sehen, gibt es in den nichtersten Silben nur vier Phoneme, halbsoviele wie in der ersten Silbe. Dabei tauchen zwei Phoneme auf, die in der ersten Silbe nicht vorkommen: *e* und *ə*. Die Klangfarbe des *e* ist dieselbe wie die des *ē* der ersten Silbe, doch ist es ein wenig kürzer.

Das reduzierte Vokalphonem [*ə*] hat nach *Steinitz* folgende Varianten: 1. ᵘ vor den bilabialen *m* und *p*: *kātᵘm* 'meine Hand', *jūntᵘp* 'Nadel'. 2. Vor und nach mouillierten Konsonanten, und nach *j* tritt eine palatalisierte Variante *ı̣* auf: *āmı̣s'* 'Rätsel', *sol'ı̣t* 'Nägel'. 3. Neben *χ* erscheint eine hintere velarisierte Variante *ə̂*, z.B. *n'oχə̂s* 'Zobel'. 4. Sonst tritt «gewöhnliches» *ə* auf, z.B. *lātəŋ* 'Sprache', *χāpət* 'Kähne'. 5. «In der Stellung vor *w* erscheint ein deutlicher *u*-Laut; da [*u*] in nichtersten Silben in anderer Stellung aber nicht vorkommt, ist es hier offenbar als kombinatorische Variante von [*ə*] aufzufassen» (a.a.O.).

Steinitz' Beobachtungen sind im allgemeinen richtig. Die velarisierten Varianten kommen aber bei meinen Aufzeichnungen nicht nur neben *χ* vor. Die Bemerkung aber, daß der Vollvokal *u* in den nichtersten Silben nur vor *w* vorkommt, ist nicht stichhaltig. Im Nordwogulischen (ausgenommen die Ob-Mundart) haben die Verba mit unpaarigen Stammsilben die Infinitivendung *–uŋkwe*, die mit paarigen *–aŋkwe;* z.B. *ajuŋkwe* 'trinken', *aluŋkwe* 'töten', *n'ūnsalaluŋkwe* 'ausbreiten', aber *ajalaŋkwe* 'verschlingen', *alis'laŋkwe* 'jagen', *χājtmi̯γtali̯γlaŋkwe* 'anlaufen'.

Die Literatursprache schreibt die fünf Hauptvarianten des [*ə*]-Phonems mit zwei Buchstaben. Für die Varianten *ə*, *ı̣*, *ə̂* schreibt man /y/, für ᵘ und *u* aber /u/. In diesem Fall hat die Rechtschreibung – meines Erachtens – recht, obwohl sie in anderer Hinsicht sehr mangelhaft und unpraktisch ist. Sie macht z.B. keinen Unterschied zwischen langen und kurzen Vokalen, obwohl diese Unterscheidung im Wogulischen sehr wichtig ist, und so entstehen Hunderte von Monographen, die gleich geschrieben, aber verschieden ausgesprochen werden, z.B. *tūjt* 'Schnee', *tujt* 'Schlitten', beide gleich geschrieben; so auch *sām* 'Gegend' und *sam* 'Auge' usw.

Wenn aber ein Laut in den nichtersten Silben genauso ausgesprochen wird – wenn auch nur in bestimmter phonetischer Stellung – wie ein Laut, der in der ersten Silbe als *u* geschrieben wird, sehe ich nicht ein, warum er mit einem anderen Buchstaben geschrieben werden soll. Phonematisch würde ich sagen, daß das *u* auch in nichtersten Silben vorkommt, als ein Phonem mit beschränkter Funktion.

Literatur

Steinitz, W.: Geschichte des wogulischen Vokalismus (Akademie-Verlag, Berlin 1955).

Adresse des Autors: B. Kálmán. Universität, *Debrecen* (Ungarn).

Discussion

Sovijärvi (Helsinki): Ich wurde gefragt, ob es meines Erachtens zwischen dem überkurzen und reduzierten [ᵘ]-Vokal und dem «deutlichen» [u]-Vokal der nordwogulischen Mundarten nicht nur einen Dauerunterschied, sondern auch andere distinktive Merkmale geben könnte. Im vorigen Jahre habe ich einen Artikel veröffentlicht, in dem ich die Resultate der sonagraphischen Analysen der mokschamordwinischen reduzierten [ə]- und [ᵊ]-Vokalallophone und des Vollvokals [e] behandelt habe (Commentationes fenno-ugricae in honorem Paavo Ravila, Mém. de la Soc. Finno-ougrienne 125, Helsinki 1963). Das Verhältnis der Summen der Breiten der F1, F2 und F3 in den reduzierten Varianten und in dem Vollvokal ist 61,4:100, was also bedeutet, daß die Summe der Formantbreiten bei den reduzierten Varianten rund 40% kleiner ist als beim Vollvokal. Es ist deshalb sehr wohl möglich, daß spätere Messungen einen ähnlichen Unterschied zwischen den genannten [u]-Varianten erkennen lassen.

Verh. 5. int. Kongr. Phon. Wiss., Münster 1964, pp. 381–383
(S. Karger, Basel/New York 1965).

Abtönung von reduziertem _e_ im Urindogermanischen

Von WILLY KROGMANN, Hamburg

Ein Kennzeichen des Urindogermanischen ist die Abtönung von _e_ zu _o_. Sie ist jünger als die Abstufung, durch die neben der Vollstufe _e_ die Dehnstufe _ē_, die Reduktionsstufe _e_ und die Schwundstufe entstanden sind. Ich halte es daher für verfehlt, wenn _Trautmann_[1], _Walde_[2] u.a. innerhalb der _e_-Reihe mit reduziertem _o_ rechnen. Allgemein anerkannt ist, daß die Abtönung nicht nur die Vollstufe, sondern auch die Dehnstufe betroffen hat, die zu _ō_ werden konnte. Ich möchte zeigen, daß sie auch reduziertes _e_ verändert hat, das dann zu _o_ wurde.

Ein sicherer Beleg für aus _e_ umgefärbtes _o_ ist griech. _γυνή_ «Frau». _Walde_ irrt, wenn er als griechische Fortsetzung von _o_ im allgemeinen _o_, vor Nasalen aber _a_ annimmt. _o_ geht vielmehr auf idg. _o_, _a_ jedoch, soweit es einen reduzierten Vokal vertritt, auf _e_ zurück. Im Griechischen ist _o_ durchweg zu _υ_ geworden. Hier steht _γυνή_ neben böot. _βανᾱ'_ < _*gṷenā_ und _μνᾱ–_ < _*βνᾱ–_ < _*gṷnā–_ in _μνάομαι_ «freie». Reduziertes idg. _*gṷen–_ begegnet noch in arm. _kanai-k‘_ < _*gṷen-ai-k_ Plur., _kanani_ Koll.; air. _ban_ < _*gṷen-ōm_ Gen. Plur., _ban–_ in Zusammensetzungen wie _ban-chú_ «weiblicher Hund» und wohl auch in aisl. _kona_, kaum aber in mhd. _kone_, das vielmehr aus ahd. _quena_ entstanden sein wird. Schwundstufiges idg. _*gṷn–_ liegt noch in ai. _gnā_, Gen. _gnā's_, av. _γnā;_ air. _mná_ < _*gṷnās_ Gen. Sing.; gheg. _grue_, tosk. _grua_ < _*gṷn-ōn_ und toch. A _śnu_ Plur., toch. B _śno_ Obl. vor. Außerdem finden wir vollstufiges idg. _*gṷen–_ in ai. _jáni-ḥ_, _jánī_, av. _dž̌aini–_, npers. _zan;_ arm. _kin;_ air. _ben_, _bé_ N., kymr. _ben-yw_ «weiblich», korn. _ben-en_ «Gattin»; alb. _zonjë_ < _*gṷeniā;_ got. _qino_, ae. _cwene_, as. ahd. _quena_

[1] Baltisch-slavisches Wörterbuch, 1923. Slavia _2:_ 1–4 (1924).

[2] _o_-farbige Reduktionsvokale im Indogermanischen. In: Stand und Aufgaben der Sprachwissenschaft. Festschrift für Wilhelm Streitberg, S. 152–199 (1924).

mit aisl. *kvenna* Gen. Plur.; apreuß. *genna*, Vok. *genno*, aksl. *žena;* toch. A *śäm*, B *śana* und wohl auch messap. *benna* und lepont. *venia*, dehnstufiges *g$^{u̯}$ēn– in ai. *–jāni–*, av. *jāni–*; got. *qens*, aisl. *kvæn*, *kvān*, ae. *cwēn*, as. *quān* < *q$^{u̯}$ēni–. Die Annahme *Hirths*[3], daß griech. γυνή auf idg. *g$^{u̯}$n– beruhe, wird durch die Vertreter der Schwundstufe widerlegt. Seine Behauptung, daß diese auf eine *u̯*-lose Form zurückgingen oder in ihnen der reduzierte Vokal nachträglich geschwunden sei, entbehrt jeder Grundlage. Unhaltbar ist auch *Schwyzers*[4] Vermutung, daß böot. βανā idg. *g$^{u̯}$n̥ā fortsetze. Bei ihr bleibt das Verhältnis zwischen *g$^{u̯}$nā und *g^{u}n̥ā undurchsichtig. Auch ist es fraglich, ob wir überhaupt mit silbischen Liquiden und Nasalen vor Vokalen rechnen dürfen. *Schwyzers* Hinweis auf skr. *–ro–* = *r̥o–* nützt nichts, da *–r̥l–* zugrunde liegt. Zutreffend hat er dagegen sizil. γάνα als Umbildung von βανā' nach γυνή erklärt. Daß das Nebeneinander von griech. γυνή und böot. βανā', das sich auch in den Kasus obliqui wie γυναικός und γβανῆκας (= βανῆκας) Hes. spiegelt, nicht auf verschiedener mundartlicher Entwicklung beruht, lehrt böot. Γυνόππαστος «von einer Frau erworben». Es besteht daher nur die Möglichkeit, griech. γυνή aus idg. *g$^{u̯}$onā herzuleiten. Daß idg. *g$^{u̯}$onā zugrunde liegen könnte, ist mit Recht niemals erwogen worden, da idg. *o* im Griechischen nach Ausweis von βολή < *g olā «Wurf» u.dgl. als *o* auftritt. Zwar ist dieses mundartlich vereinzelt in *u (υ)* übergegangen, doch lassen sich Fälle wie ὑπωρυφιά (SGDI 3325,42) oder böot. inschr. Διύδοτος, lesb. ὑμοίως nicht mit γυνή vergleichen.

Wie dieses enthalten auch andere griechische Wörter υ < idg. *o*. Als verläßliche Beispiele nenne ich noch:

Griech. νύξ, νῦκτός «Nacht», in Zusammensetzungen νυκτι–, νυκτο– < *nok$^{u̯}$t– gegenüber *nok$^{u̯}$t– in ai. *nák* Nom. Sing., *naktā* Dual; toch. *naktim* «nächtlich»; alb. *natë;* lat. *nox*, Gen. Sing. *noctis;* air. *innocht*, akymr. *henoid*, korn. *haneth*, mbret. *henoz* «heute Nacht», kymr. korn. *nos*, bret. *noz;* got. *nahts*, aisl. *nōtt*, *nātt*, ae. *neaht*, *nieht*, afries. *nacht*, as. ahd. *naht;* aksl. *noštь*, lit. *naktìs*, lett. *nakts*, apreuß. *naktin* Akk. Sing. *e*-Vokalismus sichern hett. *nekuz* Gen. Sing. «Nacht», *neku–* «dämmern». Im Griechischen wird das υ vom Gen. Sing. νυκτός < idg. *noktós aus verallgemeinert worden sein. Schwundstufiges idg. *n̥k$^{u̯}$t– setzen ai. *aktā* «Nacht», *aktú* «Dunkel, Nacht, Licht, Strahl» und got. *ūhtwo*, aisl. *ōtta*, ae. *ūht(a)*, as. ahd. *ūhta* «Dämmerung, Frühe» fort.

Griech. κύκλος «Kreis», Plur. κύκλοι, κύκλα «Räder», wozu in diesem Falle toch. A *kukäl*, B *kokale* «Wagen», wahrscheinlich auch phryg. κίκλην «großer Bär < *Wagen» treten, gegenüber ai. *cakrá–*, Plur. *cakrā* «Wagenrad, Rad, Scheibe, Kreis», av. *čaxra–* «Rad»; aisl. *hvēl*, *hjōl*, ae. *hwēo(g)l*, *hweohl*, *hweowol* «Rad». Vgl. auch griech. κυλίνδω, –έω «rolle, wälze», κύλινδρος «Walze, Rolle, Zylinder» und mit idg. *o* πολέω «bewege mich herum».

Griech. κύρβις «drehbarer Pfeiler mit Gesetzestafeln» gegenüber got. *hwairban*, aisl. *huerfa*, ae. *hweorfan*, afries. *hwerva*, as. *hwerƀan*, ahd. *(h)werban* «sich drehen, wenden, kehren» und schwundstufigem idg. *q$^{u̯}$r– in ai. *kr̥mi–* «Wurm, Made»; ir. *cruim*, kymr. korn. *pryf*, bret. *prenv* «Wurm».

Schon toch. A *kukäl*, B *kokale* ließen erkennen, daß sich abgetöntes idg. *e* nicht nur im Griechischen zu *u* entwickelt hat. Wenn ich es auch vor allem mit Hilfe dieser Sprache nachzuweisen suchte, so will ich doch wenigstens noch kurz auf Entsprechungen in anderen Sprachen hinweisen. In Betracht zu ziehen ist vor allem das Baltisch-Slawische, aus dem viele der von *Trautmann* für reduziertes *o* angeführten Beispiele hierher gehören. Von ihnen nenne ich nur lit. *ugnìs*, lett. *uguns* «Feuer», serb. *vìganj* «Amboß, Schmiede», tschech. *výheň* «Esse, Ofen, Schmiede» < *vygnь < *zgnь gegenüber ai. *agní–*; lat. *ignis* < *egnis; aksl. *ognь*, *ognj* «Feuer»; aksl. *grznz* «Kessel», *grznilo* «Ofen» gegenüber ai. *ghr̥ná–* «Glut, Hitze», lat. *fornus*, *furnus* «Backofen», *fornax*, *furnax* «Ofen», *formus* «warm»; griech. θερμός «warm» und apreuß. *gulsennien* «Schmerz» gegenüber griech. δέλλιθες. σφῆκες, ἢ Ζῷον ὅμοιον μελλίσσῃ Hes., βελόνη «Spitze, Nadel»; air. *at-baill* «stirbt»; ae. *cwelan* «sterben», ahd. *quelan* «Schmerzen leiden»; lit. *gélti* «stechen (von Bienen); heftige Schmerzen verursachen», *gėlà* «Schmerz». Sonst scheinen mir namentlich noch im Altindischen Zeugnisse für den Lautwandel *o* > *u* vorzuliegen. So stellt sich ai. *kúla–* «Herde, Schwarm, Menge; Geschlecht, Familie» zu griech. τέλος N. «Schar»; aksl. *čeljadь* «Gesinde» und gehört ai. *gula–* «glans, penis; clitoris» zu griech. βάλανος; arm. *kalin;* lat. *glans*, Gen. Sing. *glandis;* lit. *gìlė*, *gylẽ*, lett. *dzīle*, apreuß. *gile* «Eichel». In welchem Umfang solche *u* auf abgetöntes *e* zurückzuführen sind und wo sich dieses auch sonst noch nachweisen läßt, müssen weitere Untersuchungen lehren.

Adresse des Autors: Dr. phil. Willy Krogmann, Isestraße 88, *Hamburg 13* (Deutschland).

[3] Indogermanische Grammatik; Teil II: Der indogermanische Vokalismus (Indogermanische Bibliothek. Erste Reihe: Grammatiken 13²), § 184–188 (1921).

[4] Griechische Grammatik; Bd. 1 (Handbuch der Altertumswissenschaft, II 1), Anm. 1, S. 342 (1934).

Proc. 5th int. Congr. phon. Sci., Münster 1964, pp. 384–386
(S. Karger, Basel/New York 1965).

Phonemics and Lexicon

By YVAN LEBRUN, Bruxelles

Most, if not all, phonological systems show combinatory, also called conditioned, or contextual, variants. In Received Southern English, /t/ is generally alveolar, but before /l/ it is lateral.

Again, most, if not all, phonological systems show free, also called stylistic, or optional, alternants. In RP [ʔ] may be substituted for [t] at the end of a syllable before a consonant.

The relative frequency of free alternants may depend on the phonological context. In RP [ʔ] compared with [t] is more frequent before /m/, /n/, and /r/ than before other consonants.

The relative frequency of free variants may also depend on the style of speech. In colloquial standard German an unaspirated variety of /t/ is often substituted after /ʃ/ for the slightly aspirated variety which is commonly used in formal speech.

Besides, a given sound variety may be both a free and a combinatory variant. At the end of a syllable before /r/ both a post-alveolar [t] and a glottal stop may be used in RP. But if the sequence /tr/ occurs as an onset, only a post-alveolar [t] can be used.

These phonological traits are parallelled rather closely by lexical traits. As an example we may quote the English auxiliary verbs 'can' and 'may'. In present-day English 'can' and 'may' are used to denote

(a) a physical possibility, e.g.

"As *may* be seen from figure 4, the buffer has a very large core array."
"As *can* be seen from the illustrations, this feature is automatic."

(b) a moral possibility, e.g.

"Under Decree Law 40,350, persons belonging to certain categories *can* be detained as a security measure for an indeterminate period from six months to three years, which *may* be extended by successive periods of three years as long as they continue to show themselves dangerous."

(c) a logical possibility, e.g.

"The x's in non-numerical work are, if not arbitrarily restricted, essentially an unbounded set from an infinite set which is everywhere dense. That is, one cannot tell, unless one in fact looks in the table, whether there is not another x which lies between x_1 and x_2, since x and the set of x's belong essentially to the continuum. For example, in a dictionary with the ordinary alphabetical order, there *may* be the words 'bee' and 'beef'. Between these *can* be 'beech', 'beeches', and any idiomatic phrase beginning with either of these words."*

In such instances as those quoted above, 'can' and 'may' are interchangeable. In other words, they are free lexical variants.

There are contexts, however, in which only one of the two words can be used: 'can' does not occur

(a) when it expresses a moral possibility
  - in rhetorical questions of the type
    "May I ask where you are from?"
  - in clauses with 'as well' immediately after the auxiliary

(b) when it expresses a logical possibility
  - in purely positive clauses** with a perfect infinitive after the auxiliary
  - in subordinate clauses of concession

'may' does not occur

(a) when it expresses a physical possibility
  - in negative, semi-negative, and positive interrogative clauses

(b) when it expresses a moral possibility
  - in interrogative clauses with a subject in the 2nd or 3rd person
  - in negative interrogative clauses with a subject in the 1st person

(c) when it expresses a logical possibility
  - in clauses with a semi-negative immediately after the auxiliary.

Thus, 'can' and 'may' are optional variants in some clauses and combinatory alternants in others.

When they are free variants, their relative frequency may depend on the formal context. In comparison with 'can', 'may' expressing a physical possibility is more frequent before passive than before active infinitives.

Their relative frequency may also depend on the style that is

* 'May' can also be used to mark a clause as optative, but this function is disregarded here because it can never be fulfilled by 'can'.

** I.e. without so much as a semi-negative.

used. Relatively to 'can', 'may' denoting a moral possibility is more frequent in formal than in colloquial English.

Previous studies have shown that the distinctive features that make up a phoneme and are actualized in each member of this phoneme together with adventitious or phonologically irrelevant traits, can in a way be compared with the distinctive features that make up a significatum and are designated as an entity by a significant, sometimes by several substitutable significants.

We may add that when a significatum has more than one significant, these may be free alternants in some contexts and complementary variants in others, in much the same way as the various members of a phoneme may be in free alternation in some positions and in complementary distribution in others. Moreover, the relative frequency of free lexical variants may, like the relative frequency of optional phonetic alternants, depend on style and context.

Author's address: Dr. Y. Lebrun, 8, av. G. Bergmann, *Bruxelles 5* (Belgique).

Discussion

Nickel (Kiel): Hinweis auf die unterschiedliche Rangordnung und Wichtigkeit der partiellen «free variants» auf den einzelnen Ebenen. Auf der lexikalischen Ebene (*can – may*) ist auch bei totalem Austausch in allen Positionen das Kommunikationsrisiko relativ gering. Gering ist es auch bei Allomorphen (puis – peux). Hier handelt es sich im wesentlichen um stilistische Risiken. Am größten ist das Risiko zweifellos auf der phonologischen Ebene, wobei freilich, wie so oft, der Kontext dieses Risiko wieder reduzieren kann.

Die Teilsubstituierbarkeit ist wohl auch als eine Warnung vor Verabsolutierungsversuchen bei Definitionen von Begriffen wie Phonem, Morphem, Synonym usw. zu betrachten.

Proc. 5th int. Congr. phon. Sci., Münster 1964, pp. 387–388
(S. Karger, Basel/New York 1965).

Speech Transmission Laboratory
Royal Institute of Technology, Stockholm, Sweden*

Dynamic Aspects of Vowel Articulation**

By BJÖRN LINDBLOM, Stockholm

In an earlier study a compact description of formant frequency data was obtained for a set of Swedish vowels pronounced by a male talker under varying timing conditions in symmetrical consonantal contexts. The extent to which formant frequencies in the vowels reach their target values as a function of vowel segment duration was found to be determined by the consonantal environment and was described by an exponential function. The target of a vowel was specified by the asymptotic values of the first two formant frequencies and was independent of consonantal context and duration and thus an invariant attribute of the vowel. These results suggested the interpretation that an increase in the rate of utterance does not primarily involve an increase in the rate of articulation but rather in the degree of coarticulation.

This hypothesis has been put to further tests which are described in the present paper. Thus a cineradiographic study was made of a Swedish talker producing vowels embedded in a C–C context at various rates. The results of a series of measurements taken from individual profile tracings afford further evidence in favor of the interpretation suggested. It can be shown that as the time scale of an utterance is compressed there does not seem to be

---

* At present at the Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA.

** The present research was supported in part by the Wenner-Gren Foundation for Anthropological Research, Research Institute of the National Swedish Defence, Swedish Technical Research Council, Telefonaktiebolaget LM Ericsson, Swedish Foundation for Scientific Research and Industrial Development, U. S. Air Force under Grant AF EOAR 63–55, U.S. Army under Contract DA–91–591–EUC–3080, National Institutes of Health under Research Grant NB–04003–01.

any marked increase in the speed at which articulatory components move and instead undershoot relative to the target positions for the vowel in question can be observed in the direction of the consonantal environment; the higher the rate of utterance the larger the amount of undershoot.

For the type of material considered so far it seems reasonable to conclude that the realization of a vowel phoneme is associated with a fixed articulatory target. The allophonic variation observed at the articulatory as well as acoustic levels appears to a large extent to be coarticulation effects attributable to the temporal overlap between neural events corresponding to consecutive phonemes and to the dynamic properties of the articulatory structures.

Author's address: Dr. Björn Lindblom, Speech Transmission Laboratory, Royal Institute of Technology, *Stockholm* (Sweden).

Proc. 5th int. Congr. phon. Sci., Münster 1964, pp. 389–391
(S. Karger, Basel/New York 1965).

Haskins Laboratories, New York, N.Y.

Stop Categorization and Voice Onset Time

By Leigh Lisker and Arthur S. Abramson, New York, N.Y.

It is a phonetic commonplace to say that in very many languages stop categories may be divided into sets that differ with respect to a feature of voicing: voiced stops are characterized by the presence of glottal buzz during the interval of occlusion, while absence of buzz during this interval is a mark of the voiceless stops. Spectrographically the two kinds of stops are in most cases easily distinguished: for voiced stops the formantless pattern segment corresponding to closure is traversed by a small number of low frequency harmonic traces, and for the voiceless stops this portion of the spectrographic pattern is blank. But while this difference is an adequate basis for the physical separation of stop categories in many languages, there are some, like English and German, for which it works only in part. In English, for example, the sets /bdg/ and /ptk/ cannot be neatly separated on such a basis: in medial position /bdg/ are voiced and /ptk/ voiceless, but initially both sets are commonly produced with silent closure intervals and ought therefore to be classified as voiceless. While phoneticians rarely call English initial /bdg/ out-and-out voiceless stops, they regularly cite a second phonetic attribute, that of aspiration, which distinguishes /ptk/ from /bdg/ in initial position. In many positions, then, /bdg/ are voiced and /ptk/ voiceless, but initially /ptk/ are released with an audible explosion and an interlude of noise, while /bdg/ are not. Spectrographically aspiration may be detected as noise largely in the mid and higher frequencies within the range important for speech perception. Thus the two phoneme sets may be said to differ in either (or both) of two ways: medially /bdg/ are distinctively marked by low frequency harmonics preceding the "burst" of the release, and initially /ptk/ are distinctively marked by an interval of higher frequency noise immediately following the burst. It is, however, also possible to extract a single difference, one in the timing of voice onset relative to release, which is common to both the voiced-voiceless and the aspirate-inaspirate contrasts. Now we can say that /bdg/ are everywhere characterized by earlier voice onset than are /ptk/, and at the same time it is true that in initial position each set shows a delay in voice onset as compared to other positions. Since this single measure of relative onset time effectively separates the two stop categories in many positions, one may be tempted not only to consider the aspiration difference linguistically redundant in English, but to regard the aspiration noise as no more than the automatic consequence of a large delay in voice onset. This is to say that one might view the features of voicing and aspiration as phonetic attributes which are not completely independent, although such a view is incompatible with phonetic accounts of certain languages in which these two features operate to distinguish two, three or even four categories of stops.

The fact that the measure of voice onset time provides an effective way of separating both voiced from voiceless and aspirated from inaspirated stops in English suggests the possibility that differences in relative voice onset time might prove to be important for separating stop categories generally in languages, whether these categories are called voiced and voiceless or aspirated and inaspirated or fortis and lenis or different in still other ways. Relevant data for measuring the effectiveness of voice onset time were obtained for some eleven languages which vary in the number and phonetic nature of their initial stop categories. Of the languages examined six are two-category languages, – that is, there are at most two stop categories for a given place of closure; these are Dutch, Spanish, Hungarian, Tamil, Cantonese and English. Three other languages have three categories each: Eastern Armenian, Thai and Korean; and the remaining languages have four categories each: Hindi and Marathi. The distribution of measured values for the eleven languages supports the view that in general this feature of relative onset time serves very effectively as a means of separating stop categories quite independently of whether they are said to be distinguished solely by voicing or solely by aspiration or by a combination of the two features; the only categories clearly not distinguishable on this basis are the so-called voiced aspirates and voiced inaspirates of Hindi and Marathi. Moreover, while there are

differences from language to language, they all appear to place their categories along the dimension of voice onset time in such a way that we may speak of three crosslanguage phonetic categories, one for which voice onset leads release by about 100 msec, a second where voice onset occurs just after release, and a third with voice onset lagging about 75 msec behind release. The fact that the distribution of voice onset time values is thus far from random suggests that the time of voice onset is not entirely a matter of how a particular linguistic community "decides" to exploit a phonetic dimension, for there may perhaps be physiological constraints operative as well.

Author's address: Prof. L. Lisker, Haskins Laboratories, 305 East 43rd Street, *New York 17, N.Y.* (USA).

Proc. 5th int. Congr. phon. Sci., Münster 1964, pp. 392–396
(S. Karger, Basel/New York 1965).

Acoustic Correlates of Some Hungarian Emotive Intonation Patterns

By K. Magdics, Budapest

1. Acoustic analyses and various experiments were carried out on the synthesized one-word Hungarian utterance /e:p/ "It is undamaged" pronounced with 20 different emotions or attitudes. 10 of the synthesized sentences were questions, 8 were statements and 2 exclamations. One of the questions was a neutral yes-or-no question, the others were coloured by astonishment, two different degrees of curiosity, impatience, suspicion, consternation, sarcasm, indignation and threat. One of the statements was neutral, the others implied consolation, astonishment, incomprehension, warning, sarcasm, argumentation and indignation. The exclamatory sentences expressed joy and surprise.

2. The changes of fundamental frequency, duration and formant amplitudes were measured on spectograms and sections. Fundamental frequencies were plotted against time, and these graphs automatically provided measures of the durations. Formant amplitudes in dB were plotted against frequency.

a) *Fundamental frequency* of both questions and statements ranged from 80–180 cps. (this was the full fundamental frequency range of the synthesizer), that of the exclamations from 140–180 cps.

The neutral question had a medium rising pitch pattern starting at 100 cps. and rising to 140 cps.; the slightly and strongly inquisitive, impatient, indignant and threatening questions had high rising pitch patterns ranging from 100–180 cps.; the questions coloured by surprise and consternation showed low rising patterns moving between 80–120 cps.; the suspicious and sarcastic questions were characterized by rising-falling patterns, the frequency changing between 80–150 cps.

The neutral statement had a medium falling pitch pattern

moving from 110–80 cps.; the sarcastic statement showed a low falling pattern changing from 90–80 cps.; consolation, incomprehension, astonishment and joyful exclamation had high falling patterns with frequency values between 180–80 cps.; warning, argumentation, indignation and exclamation of amazement had falling-rising patterns moving from 180–120 cps. with a frequency value of 120–140 cps. between the two.

b) There was a certain relationship to be found between pitch patterns and *duration*. The sentence with a medium rising pattern was medium long (0.42 sec). The sentences with high rising and medium falling patterns were rather short (0.32–0.36 sec). Both the low rising and the low falling patterns were characterized by a fairly long duration (0.44–0.46 sec). The rising-falling patterns were extremely slow in speed (0.55–0.65 sec), whereas the fall-rise occurred with both an increased (warning = 0.36 sec, indignant statement = 0.40 sec) and a slower speed (exclamation of amazement = 0.55 sec). High falling patterns had either shorter (joyful exclamation = 0.35 sec) or longer duration (consolation = 0.49, incomprehension = 0.46, astonished statement = 0.44 sec).

In the rising pitch patterns the first, lower line was longer (0.10–0.25 sec) than the following higher one (0.07–0.15 sec), whereas in the falling patterns the first, higher line was remarkably shorter (0.08–0.15 sec) than the second, lower one (0.10–0.25 sec). In the rising-falling and falling-rising patterns the first and the last melody lines were more or less the same in length and the middle one was rather short.

c) The *amplitude* generally decreased with a medium and a high rise[1] (A_1: 0 – —2.5 dB, A_2: —1 – —4 dB), but increased with a low rise (A_1: 0 – +2 dB, A_2: —2 – 0 dB). Medium falling and high falling patterns were generally characterized by a falling amplitude (A_1: +2.5 – —1 dB, A_2: 0 – —3 dB), whereas low fall showed a rising amplitude (A_1: 0 – +2 dB, A_2: —3 – 0 dB). Both rise-fall and fall-rise had a falling-rising amplitude (A_1: +1 – —2 – +1 dB, A_2: 0 – —4 – 0 dB).

3. 10 English and 10 Hungarian subjects were asked to define the emotions or attitudes the synthesized sentences implied. Out of the 200 English answers 78 (39%), of the 200 Hungarian answers 140 (70%) turned out to be correct.

[1] The dB amounts are related to the amplitude of the A_1 of the neutral yes-or-no question.

4. Modifications of the position, the pitch intervals, the duration and the amplitude were carried out one after the other in each of the sentences. Spectograms and tape recordings were made of each modification. Then 10 English and 10 Hungarian subjects were asked to define the emotions or attitudes the modified utterances implied.

a) *Position was raised or lowered* with 40 cps. This time the English subjects produced 50 (25%) and the Hungarian 104 (52%) correct answers. The following sentences were generally misjudged: neutral question and statement, surprised and dismayed question, statements coloured by sarcasm, incomprehension and surprise, joyful exclamation.

b) *Pitch intervals were increased or decreased* with 20 cps. In this case the percentage of the correct answers decreased to 10 (20 in number) with English subjects and to 32 (64 in number) with Hungarians. Correct judgements were given only in the case of consolation, sarcastic statement and questions coloured by astonishment, impatience, sarcasm and threat.

c) The *duration of the whole utterances were increased or decreased* by 0.1 sec. The number of correct answers was 30 (15%) with English subjects and 82 (41%) with Hungarians. Neutral question and statement, impatient and surprised questions as well as exclamation of surprise were generally not recognized.

The *durations of the first and the last pitch level* in the rising and the falling patterns *were inverted:* only 50 (38%) out of 130 answers given by Hungarian subjects proved to be correct in comparison with the 82 (63%) given by the same subjects listening to the original forms of the sentences.

d) The *amplitude* of F_1 or F_2 *was increased or reduced* with 4 dB in the beginning or at the and of the utterance. The percentage of the correct answers given by English subjects was 21 (42 in number), those given by Hungarians was 35 (70 in number). Nobody recognized the neutral question, the slightly and strongly inquisitive, surprised, dismayed, sarcastic and threatening questions nor the statements coloured by consolation, incomprehension astonishment and sarcasm.

5. *Conclusions:* Four acoustic cues: position, pitch intervals, duration and amplitude were determined in a synthesized one-word Hungarian utterance pronounced with 20 different emotions or attitudes. The synthesized sentences were presented as tests material to 10 English and 10 Hungarian subjects. The experiment resulted

in 39% correct answers by the English and 70% by the Hungarian subjects. The successive modifications of the acoustic components reduced the number of the correct answers and revealed the acoustic cues relevant in the recognition of the emotive patterns in question. It turned out that *each of the four cues* were *equally relevant* in the recognition of the neutral question. *Three of the cues:* position, duration and amplitude determined the astonished question; position, pitch intervals and amplitude made the dismayed question, the astonished and uncomprehended statement recognizable; position, pitch intervals and duration were important in the neutral statement. *Two of the cues*: pitch intervals and amplitude were necessary for the recognition of both kinds of inquisitive question; position and amplitude were relevant for the sarcastic statement, position and pitch intervals for the joyful exclamation, pitch intervals and duration for the exclamation of surprise. *One of the cues:* duration was enough for the recognition of the impatient question, the pitch intervals for the suspicious and indignant question, as well as the warning, the arguing and the indignant statement, the amplitude for the consolation, the sarcastic and threatening question.

The synthesis was produced by using the speech synthesizer constructed by *P. Denes* and *J. E. West*, and a Kay Sona-Graph was used for the analyses at the Phonetics Department of University College, London. The synthesis followed the author's pronunciation, who had previously examined Hungarian intonation including emotive patterns; see *I. Fónagy* and *K. Magdics:* Das Paradoxon der Sprechmelodie. Ural-Altaische Jb. *35:* 1–55 (1963). – Emotive Patterns in Intonation and Music. Z. Phon. *16:* 293–326 (1963).

Author's address: Klara Magdics, MTA Nyelvtudományi Intézete, Szalay u. 10–14, *Budapest V* (Hungary).

Discussion

Von Kürthy (Aachen): You can distinguish Hungarian from other languages without understanding because

a) the syllables are very short
b) the words are either intonated in a high or in a low cue

For example:
High intonation: Egynegyed egy
Low intonation: Bútorraktár

Gårding (Lund): I am surprised to hear that it is possible to make listeners identify as much as twenty different intonation contours differing only in the fundamental frequency of the carrier phrase. Together with *Arthur Abramson* at Haskins Laboratories I worked on a project similar to the one reported on by Miss *Magdics*. We then started with a great number of intonation contours that we had elicited from American English informants. These were put in a listening test. I cannot describe the test procedure here but it may be of interest to mention that the recorded sentences chosen for the test were put on cards and the card reader that Mr. *Cooper* described in his talk was used. Out of the original contours, twelve could be identified by listeners. Later, when we applied the pitch curves from these twelve contours to a common carrierphrase only five were identified by listeners.

Verh. 5. int. Kongr. Phon. Wiss., Münster 1964, pp. 397–402
(S. Karger, Basel/New York 1965).

Zur Problematik der strukturellen Relevanz prosodischer Erscheinungen

Von IRMGARD MAHNKEN, Göttingen

Die Frage der strukturellen Relevanz der prosodischen Erscheinungen steht seit einiger Zeit im Zentrum des Interesses der phonetischen und der phonologischen Forschung. Es sind vornehmlich zwei Problemkreise, auf die sich die bisherigen Untersuchungen und Ausführungen konzentrieren: einerseits hat man begonnen, sich eingehend der Problematik der Sprechmelodie zu widmen und dabei dem Gesichtspunkt ihrer linguistisch-strukturellen Relevanz gebührende Geltung zu verschaffen; andererseits ist in der letzten Zeit immer wieder von Neuem besonders eindringlich die Frage nach der Relevanz der einzelnen prosodischen Komponenten bei der Iktusbildung gestellt worden. Man darf wohl sagen, daß unsere Kenntnis von den prosodischen Erscheinungen der Rede inzwischen recht erhebliche Fortschritte gemacht hat. So darf hier z.B. das besondere Verdienst der Prager Linguisten und Phonetiker hervorgehoben werden, die Grundlage für ein phonologisch orientiertes Verständnis der satzmelodischen Kadenzen geschaffen und damit gezeigt zu haben, daß von einer adäquaten Fragestellung aus auch zu den so schwierigen Problemen der linguistischen Relevanz und des Systemcharakters der prosodischen Erscheinungen ein Zugang gefunden werden kann.

Es läßt sich aber nicht bestreiten, daß wir – trotz aller bedeutsamen Fortschritte hinsichtlich bestimmter Teilfragen – immer noch weit entfernt sind von einer wirklichen Bewältigung des umfangreichen Fragenkomplexes der suprasegmentalen Erscheinungen der Rede. Die Einsicht in die grundlegenden Voraussetzungen des sprachlich-kommunikativen Funktionierens der suprasegmentalen Erscheinungen hat bisher auch noch nicht annähernd jenen Erkenntnisgrad erreicht, der inzwischen bezüglich des Charakters und der linguistischen Struktur der Lautsysteme erarbeitet worden ist. Natürlich lassen sich eine ganze Reihe von Gründen dafür anführen, warum sich einer adäquaten Erfassung der prosodischen Erscheinungen solche Schwierigkeiten in den Weg stellen. In dem engen mir hier zur Verfügung stehenden Rahmen kann und soll nur auf einzelne dieser Gründe eingegangen werden, und zwar vornehmlich auf jene Umstände, die in den bisherigen Diskussionen noch nicht gebührende Beachtung gefunden haben.

Von sprachwissenschaftlichem Interesse sind bekanntlich speziell jene Erscheinungen an den sprachlichen Äußerungen, denen im Hinblick auf den Zeichencharakter der Sprache funktionelle Relevanz zukommt. Will man Klarheit über das Laut*system* einer Sprache gewinnen, so verschafft man sich zunächst einen Überblick über den Phonem*bestand* dieser Sprache. Die Methoden der Erfassung und Analyse des Phonembestandes einer Sprache basieren wesentlich auf dem Umstand, daß es sich bei den Phonemen um diskrete Einheiten handelt, bei deren Auswechselung die Bedeutung jener sprachlichen Zeichen eine Änderung erfährt, die aus bestimmten Phonemfolgen gebildet sind. Insofern in einzelnen Sprachen bestimmten «Tönen» oder Tonverlaufsoppositionen eine ähnliche, für die Bedeutung der Wörter relevante Funktion zukommt, lassen sich auch diese von der Bedeutungsdifferenzierung her in ihrer sprachlichen Funktion erfassen und dann auch einer weiteren Untersuchung zuführen. Bei den satzprosodischen suprasegmentalen Erscheinungen hingegen ist weder eine solche segmentarische Abgrenzung diskreter Einheiten noch die Bestimmung der funktionell relevanten Kriterien auf dieser Basis ohne weiteres möglich.

Ist schon das Lautsystem einer Sprache nur vom funktionellen Aspekt her adäquat zu erfassen, so wird man – darüber dürfte heute weitgehende Übereinstimmung bestehen – den suprasegmentalen Erscheinungen der Rede erst recht nicht ohne Berücksichtigung des funktionellen Aspekts beikommen können. Daß trotzdem auf diesem Wege bislang nur in Teilfragen gewisse Ergebnisse erzielt werden konnten (so z.B. bezüglich der satzmelodischen Kadenzen), hat seine tiefere Ursache in dem Umstand, daß – trotz der prinzipiellen Ablehnung eines Ausgehens von den akustisch-physikalischen Daten – im Grunde *auch* bei den sprachlich orientierten Analysen suprasegmentaler Erscheinungen in erster Linie mit *akustisch* fundierten oder beeinflußten Begriffen gearbeitet wird (und nicht mit primär linguistisch bestimmten Begriffen). Es wird nämlich i.a. nicht hinreichend beachtet, daß in der vorzugsweisen oder ausschließlichen Berücksichtigung der *auditiven* Qualitäten der Tonhöhe (bzw. Tonhöhenbewegung), der Lautstärke (bzw. dynamischen Modulation) und der Dauer (bzw. des Tempos) ungewollt doch eine Anlehnung an die drei *akustischen* Dimensionen der Frequenz, der Amplitude und der Zeit vollzogen wird. In der Tat aber wird die Aufgabe einer linguistisch orientierten Untersuchung der suprasegmentalen Gestaltungsmittel der Sprache meist darin gesehen, die sprachlichen Funktionen der eben genannten (auditiven) Qualitäten zu ermitteln. Würde man aber bezüglich der suprasegmentalen Gestaltung der Rede wirklich in gleicher Weise wie bezüglich der Kriterien der phonematischen Gestaltungsebene von den *sprachlich* relevanten auditiven Kriterien ausgehen, so würde das zwar gewiß nicht zu einer Vernachlässigung der erwähnten Qualitäten der Tonhöhenmodulation, der dynamischen Modulation und der Tempo- bzw. Quantitätsgestaltung führen, doch würde man dann zunächst einmal konstatieren müssen, daß in der Ebene der suprasegmentalen Erscheinungen mit und neben *diesen* Qualitäten *auch* noch *anderen Qualitäten* sprachliche Relevanz zukommt: so etwa den Qualitäten der Abgehobenheit und der Hervorgehobenheit (die nicht ohne weiteres mit «Akzentuierung» identifiziert werden dürfen) oder den Qualitäten des Insichgeschlossenseins und des Gegeneinanderabgesetztseins (auch dies erfolgt nicht nur vermittelst der Pausen und der «Grenzsignale» der phonematischen Ebene, sondern in besonderem Maße gerade vermittelst bestimmter Gestaltung der einzelnen suprasegmentalen Komponenten). Diesen verschiedenen Qualitäten kommt im sprachlichen System ihre bestimmte sprachliche Funktion zu, und *diese* Funktionen im einzelnen zu untersuchen, ist Aufgabe der Sprachwissenschaft. Ein vertieftes Verständnis ist allerdings nur möglich, wenn zugleich auch die phonetischen Voraussetzungen der Realisierung dieser Qualitäten ermittelt werden. Es sind z.B. Gestaltungskriterien dieser Art (bei denen die Tonbewegung demnach erst sekundär funktionell relevant wird), die in bestimmten Sprachen bei sonst gleichem Wortlaut einer Äußerung die unterschiedliche syntaktische Funktion bestimmter Redeteile bedingen und somit den Sinn einer Aussage verändern können. In solchen Fällen etwa die Tonbewegung (die in funktioneller Abhängigkeit von der jeweiligen Gliederung steht) *unmittelbar* als relevantes Kriterium für die «syntaktische Bedeutung» des betreffenden Wortes anzugeben, bedeutet eine gefährliche Verkennung der funktionellen Verhältnisse in der Struktur des betreffenden Sprachsystems.

Geht man von der Notwendigkeit einer adäquaten Berücksichtigung des funktionellen Aspekts aus, so stellt sich eine besonders wichtige Aufgabe also in der Ermittlung des *Inventars der prosodischen (suprasegmentalen) Qualitäten*, die der Sprache zur adäquaten Gestaltung des Sinns der Rede zur Verfügung stehen. In dieses Inventar gehören u.a. sicherlich auch die o.e., von Linguistik und Phonetik bisher nicht hinreichend beachteten Qualitäten, denen in der (nicht nur linearen, sondern auch hierarchischen) *Gliederung* der Rede funktionelle Relevanz zukommt. Ihre Gestaltung muß übrigens vom Sprache lernenden Kind ebenso erlernt werden wie etwa die Handhabung des Phoneminventars einer Sprache. Diese *der Gliederung dienenden auditiven prosodischen Qualitäten* erstehen weitgehend aus einer bestimmten, in ihren Gestaltungsbedingungen analysierbaren (und z.T. in den letzten Jahren bereits analysierten) Gestaltung der *akustischen* Komponenten der Frequenz- und Amplitudenmodulation sowie der Zeitrelationen (ohne daß sie *als* Tonhöhenmodulation usw. wahrgenommen werden müssen!). Die ungenügende Beachtung dieser Tatsache ist einer der Gründe für die scheinbaren Schwierigkeiten bei der Zuordnung akustischer Daten und auditiver Kriterien, und nur aus der Nichtbeachtung dieser grundlegenden Tatsache ist es zu verstehen, daß viele Untersuchungen prosodischer Erscheinungen der Versuchung erliegen, auch die suprasegmentalen prosodischen Erscheinungen in einer ausschließlich linear gerichteten segmentierenden Analyse erfassen zu wollen.

Adresse des Autors: Dr. I. Mahnken, Seminar für Slawische Philologie, Hospitalstraße 10, *34 Göttingen* (Deutschland).

Discussion

Ezawa (Köln): 1. Die linguistisch adäquate, nicht physikalisch-akustisch befangene Bestimmung bzw. Untersuchung, die die Referentin für prosodische Eigenschaften fordert, ist gerade das, was von der Zwirnerschen Phonometrie für lautliche Erscheinungen überhaupt initiiert und durchgeführt worden ist. Wie beurteilt die Referentin die inzwischen in größerem Umfang vorliegenden phonometrischen Untersuchungen prosodischer Eigenschaften?

2. Die Referentin führt als linguistisch relevante Qualitäten prosodischer Erscheinungen die «Hervorgehobenheit», das «Insichgeschlossensein», das «Gegeneinanderabgesetztsein» usw. an. Bei der Hervorhebung des «Vorrangs» des linguistischen Gesichtspunktes bei phonetischen Untersuchungen geht es aber nicht um die Verwendung solcher linguistisch plausibel klingenden Bezeichnungen, sondern um die Annahme, daß zwischen den faktisch-funktionellen Untersuchungen der Phonetik und den normativ systematischen Untersuchungen der Linguistik eine Kongruenz besteht und entsprechend erschlossen werden muß («Zuordnung»).

Krámský (Prag): Ich habe eine Bemerkung zum Problem, das zwar hier nicht erwähnt worden ist, doch für die Problematik der strukturellen Relevanz prosodischer Erscheinungen wichtig ist. Es ist das phonologische Gesetz über die Inkompatibilität freier Quantität und freier Betonung, das *Roman Jakobson* als erster formulierte. *Trubetzkoy* hat in seiner «Anleitung zu phonologischen Beschreibungen» die Koexistenz freier Betonung und freier Quantität in beschränktem Maße anerkannt. Ich möchte Sie darauf aufmerksam machen, daß dieses Gesetz nicht so ausnahmslos ist, wie man ursprünglich dachte. Wenn wir einige orientalische Sprachen in dieser Hinsicht überprüfen, können wir solche finden, die freie Quantität mit freier Betonung ohne die früher erwähnte Beschränkung kombinieren. *Trubetzkoy* kannte solche Sprachen, behauptete aber, daß es sich in diesen Fällen nicht um Silbensprachen, sondern um Morensprachen handelte. Ich bin der Meinung, daß der Unterschied zwischen Silben- und Morensprachen nicht phonologisch, sondern phonetisch ist, und daß es grundlos ist, ihn in einer prosodischen Typologie in Betracht zu ziehen. Es zeigt sich bei der Berücksichtigung des Japanischen, des Urdu und iranischer Mundarten, daß das Gesetz über freie Quantität und freie Betonung nicht ausnahmslos gilt, aber – andererseits – eine große Verbreitung aufweist.

Horálek (Prag): Bei den prosodischen Eigenschaften muß man die suprasegmentalen von den segmentalen unterscheiden. Zu den segmentalen gehört z.B. die vokalische Quantität, zu den suprasegmentalen der Wortakzent (sowohl mit deliminativer wie mit kulminativer Funktion). Der sogenannte freie Akzent (z.B. russischer Art) hat zwei Funktionen, die distinktive und die kulminative (gipfelbildende), beide sind verknüpft und können nicht selbständig auftreten. Deswegen kann man den Wortakzent bloß als Angelegenheit des phonemischen Planes behandeln, wie es die Vertreter der Distinctive-Features-Theorie machen. Die prosodischen Eigenschaften gehören zu verschiedenen Plänen der Sprache: dem phonemischen, dem silbischen (silbische Quantität und silbische Intonation), dem Wortbauplan und dem syntaktischen Plan.

Antwort *Mahnken* zu *Ungeheuer* (der die Anführung eines Beispiels für Umsetzung in auditive Qualitäten anderer Art wünschte): Leider ist Herr *Pollok* nicht anwesend, der in seinem Vortrag ein solches Beispiel an Hand der serbokroatischen Akzentopposition geben wollte. Ein weiteres Beispiel würde z.B. die gesamte Problematik der Phrasierungserscheinungen darstellen. – Wichtig ist die Bemerkung, daß die Bezeichnung «auditiv» nicht ganz eindeutig ist. Man sollte präziser, als dies bisher im allgemeinen der Fall gewesen ist, auseinanderhalten, daß die Perzeption in der sprachlichen Kommunikation einerseits und die Möglichkeiten, die das Abhören bei der Bearbeitung einer Sprachaufnahme bieten, andererseits zwei recht verschiedene Vorgänge sind. Beim Abhören können gewisse Erscheinungen durchaus erfaßt werden, die normalerweise in der Perzeption des Kommunikationsprozesses unterschwellig bleiben. Für die Linguistik spielen im allgemeinen (bei der Konstatierung der einzelsprachlichen Systeme) jene «auditiven» Qualitäten die entscheidende Rolle, die der «unbefangenen» Wahrnehmung entsprechen. Selbstverständlich müssen bei der Zuordnung zu den akustischen Daten sowohl diese, dem Angehörigen einer Sprachgemeinschaft geläufigen «auditiven Qualitäten» als auch die durch eingehendes Abhören überhaupt auditiv erfaßbaren auditiven Erscheinungen zugeordnet werden.

Zu *Ezawa:* Meine Stellungnahme zu den mein Arbeitsgebiet betreffenden Untersuchungen der Phonometrie habe ich seinerzeit in meiner Arbeit «Redegebilde oder Zufallsstreuung?» dargelegt. Ich habe im übrigen *stets* die Auffassung vertreten, daß Untersuchungsergebnisse, die die Antworten auf unterschiedliche Fragestellungen geben, sich gegenseitig im Hinblick auf die tiefere Erkenntnis des untersuchten Gegenstandes ergänzen müssen, soweit die Untersuchungen selber einwandfrei durchgeführt sind. Aus diesem Grunde bin ich auch der Auffassung, daß meine eigenen Untersuchungsergebnisse – bei allen Unterschieden in der Fragestellung – grundsätzlich *nicht* in einem Gegensatz zu den durch phonometrische Methoden zu erreichenden Ergebnissen stehen können. Sie antworten vielmehr nur auf andere Fragen und scheinen mir darüber hinaus sogar in erheblichem Maße geeignet, die statistischen Ergebnisse phonometrischer Untersuchungen erläuternd zu kommentieren.

Proc. 5th int. Congr. phon. Sci., Münster 1964, pp. 403–408
(S. Karger, Basel/New York 1965).

From the Institute of Phonetics, Lund

Stability and Instability of Syllabic Structures

By Bertil Malmberg, Lund

The syllable is a basic unit in any expression structure[1]. The only phonemic structures without syllables are those which only have vocoids, whose function is to make the pronunciation and identification of the phonemes ("consonants") possible. These non-phonemic elements are determined as to their quality by the surrounding sounds (as is supposed for Proto-Indoeuropean; *Borgström*, etc.)[2]. In such a case, there is identity between phonemes and syllables. A language like Proto-Indoeuropean according to the description referred to – the correctness of which is not in principle essential for our discussion – may be said to have only pseudo-syllables, which by definition are open. Consequently the first step in the direction of a phonemisation of vocoids – the introduction of a non predictable vowel colour – necessarily implies the creation of real, open syllables. We know that the open syllable is the most widely spread in the languages of the world. All languages have open syllables. Many languages have *only* open syllables. Thus the open syllable conditions the closed one, but not inversely. This is a consequence of Jakobson's law on the level of syntagmatic structure[3]. It would accordingly be natural to suppose the open syllable to come earlier in the child's phonemic development, and to be the more resistant one in aphasia. We know that this is so (*Jakobson*, *Durand*, *Ombredane*, *Grégoire*, etc.)[4]. It should also be added that

[1] Expression used in the Glossematic sense, opposed to content.

[2] See e.g. *Borgström* in: Norsk tidsskrift for sprogvidenskap XV, 1949, p. 137–187, and cf. *Malmberg*[2], p. 129–139; and cf. id., in Ref. 1, p. 81–97.

[3] See Ref. 1, chap. VII, particularly p. 129–130.

[4] See my Stabilité et instabilité des structures phonologiques, with references (in print).

consonant clusters are in the same way secondary to single consonants, syllable-initial as well as syllable-final, and that the syllable *CV*, according to our premises above, is the basic, general, and most primitive, syllabic type.

This basic syllabic structure has, however, another aspect too. In languages where closed syllables are possible it seems to be almost general that the number of oppositions is smaller, and/or the number of combinations more reduced (the distribution more limited at the ends of syllables). In any case, a system is never richer in syllable-final positions. I do not know of any language which admits more distinctive possibilities at the end of syllables than in initial positions[5]. A good example is the loss of the voiced ∼ voiceless distinction in word-final position (in Slavic and German), another the numerous cases of syncretism at the end of syllables in Spanish. Even if, in these examples and numerous others, the syllable in spite of the reductions mentioned remains equally closed, the phenomenon indicates a structural weakness, a poverty of distinctive possibilities which nevertheless can be looked upon as a tendency towards open syllables. In Spanish, where this reduction in vulgar and dialectal speech is often very far-going (with a distinction between a palatal and a labio-velar semi-vocalic element as the only remaining opposition), the final generalization of the syllabic tendency may sometimes be a fact. This is particularly the case in peripheric bilingual societies without any linguistic tradition[6]. Even the general weakness of the implosive elements in Spanish, as compared with other languages (and particularly the Germanic ones)[7], is an effect of this syllabic tendency.

The Germanic languages, and particularly German and Scandinavian, are well-known for their complex implosive clusters. Swedish admits e.g. *CCCVCCCCC* (initially *str–*, *spr–*, *skr–*, *spl–*, etc.; finally *–tskt*, *–lmskt* [in *skälmskt*], etc.); German has groups like *–lk(s)t*, *–rk(s)t*, etc. There is consequently a profound structural difference between languages like Swedish or German where groups like those mentioned are possible, and a language like Japanese where a syllable always contains only a consonant plus a vowel: *CV*, and

---

[5] See my articles on Spanish in Boletim de filologia IX. 1948, p. 99–120, and Orbis XI, 1962, p. 131–178, and my handbook on Spanish phonetics, Ref. 4, § 42, p. 87–90.

[6] Cf. hereto Ref. 5.

[7] See *Navarro Tomás*, Ref. 7, §§ 72, 132.

where in European loan-words consonant clusters are split up in as many syllables as there are consonants (*club* > *kurabu*, *film* > *hirumu*). The main difficulty for most African people to pronounce European languages consists in articulating the syllable-final consonants. But although Scandinavians, Germans, etc. have this peculiar habit of making numerous distinctions and of combining consonants at the end of syllables, a close examination of the way in which they pronounce them reveals far-going phonetic reductions of such groups in fluent speech. Thus Sw. *hemskt* normally becomes *hemst* (in the same way *löms(k)t*, *skälms(k)t*, *jor(d)gubbe*, etc.), a simplification which is immediately perceptible and admitted by any speaker whose attention is drawn to the reduction. Instrumental evidence is, however, needed to discover more subtle phenomena. Examples will easily be found in spectrograms, etc. There is no doubt that these far-going simplifications are due to the difficulty of realizing too complex structures in such an extremely weak position as is the postvocalic part of the syllable. This weakness is of course primarily an articulatory phenomenon, a successive weakening of the speech muscles involved, but also, in consequence hereof, an acoustic weakening, a blurring of the characteristic properties of the implosive elements resulting in difficulty for the listener to identify them. This confusion, according to Jakobson's law, first affects the minor, most subtle distinctions. It is a well-known fact that complex clusters come late in the speech development of the child and are early to disappear in aphasia. Certain Swedish children retain those simplifications for years; some linguistically weak individuals do so for ever. The same is true of assimilations, structurally speaking reductions of distinctive contrasts within the syntagm. Even though Swedish – as opposed e.g. to Spanish – admits three, and sometimes even four[8], nasals in final position and before a following consonant (e.g. *–m* + *p*, *–n* + *p*, *–ŋ* + *p*, and even *ɳ* + *p*), many individuals, as a mostly unnoticed speech defect, assimilate, so that, just as in Spanish, the nasal is adapted to the following phoneme. Children generally do so. The name of my native town *Hälsingborg* /ˌhɛlsiŋ-ˈbɔrj/, with /ŋ/ followed by /b/, is pronounced by some people, among them otherwise linguistically quite normal individuals, with /m/. The examples quoted all indicate that even in languages where complicated syllabic structures are current, the phonetic manifes-

---

[8] The alveolar [ɳ] is a realisation of the *r* + *n*-cluster (*järnbalk* [ˈjæːɳˌbalk] etc.).

tation of the complex syllables is often reduced as compared to the norm, and subject to individual simplifications, to a kind of infant speech which gives evidence for the particular efforts needed for their correct pronunciation. They are the first to be affected in any kind of defect or weakened speech.

The insufficient resistance of the complicated syllabic structures referred to here has also another aspect. The stability of a phonemic – or more generally a linguistic – element is also inversely proportional to its amount of information. The more predictable, the less resistant. The least predictable of the elements in a chain is of course its first phoneme, since nothing can make us guess at its character (unless the non-linguistic context strongly limits our choice). Already the second is to some extent signalled by the first, a vowel by the colour of the preceding consonant, a consonant by different combinatory phenomena (formant transitions, etc.), by distributional rules, etc. If it is true, on the one hand, that languages like Germanic or Slavic have very complex clusters, it is also true, on the other, that their combinatory possibilities are very much restricted and that only a limited number of possible orders is admitted. Since most of the complex types in e.g. German or Swedish are due to morphemic elements added to a less complicated stem, the nature of the cluster may also to a large extent be inferred from the morphological or syntactic structure of the sentence.

The idea of statistical probability as an element of phonetic or phonemic strength or resistance has recently been put forward by *Rebecca R. Posner*[9]. It is no doubt a very fruitful idea and contributes to a structurally more adequate conception of the syllabic tendency referred to here, and also of the kind of phonetic change connected with it[10].

References

1. *Catalán, D.* (Ed.): Miscelánea Homenaje a André Martinet, 3 vol. (La Laguna 1963).
2. *Malmberg, B.:* Structural Linguistics and Human Communication (Springer, Berlin 1963).
3. *Malmberg, B.:* Stabilité et instabilité des structures phonologiques (in print).
4. *Malmberg, B.:* Spansk fonetik (Gleerup, House of publication, Lund 1963).
5. *Malmberg, B.:* Linguistique ibérique et ibéro-romane (Lund 1961).
6. *Malmberg, B.:* Gémination, force et structure syllabique en latin et en roman. Orbis litterarum, Suppl. *3:* 106–112 (1963).

[9] Ref. 8, particularly p. 26 s.
[10] The same problems were touched upon from a somewhat different point of view in my article, see Ref. 6.

7. *Navarro Tomás, T.:* Manual de pronunciación española (Hafner, New York, 10th ed. 1961).
8. *Posner. R. R.:* Consonantal dissimilation in the Roman languages (Blackwell, House of publication, Oxford 1961).

Author's address: Prof. B. Malmberg, Fonetiska Institutionen, Kävlingervägen 20, *Lund* (Sweden).

Discussion

Westring-Christensen (Copenhague): M. *Malmberg* a dit qu'un système de consonnes est plus riche à l'initiale qu'à la finale, qu'il y a moins de distinctions possibles à la finale qu'à l'initiale. Je ne crois pas que cela soit vrai pour toute langue.

1° En français moderne, plus précisément le français de la bonne société parisienne, on trouve 16 consonnes à l'initiale:

/p t k b d g f s ʃ v z ʒ m n l r /;

à la finale on trouve les mêmes consonnes et en plus / ñ j /; c'est une condition qu'on interprète (j – ɥ – w –) comme variantes de / i– y – u – / comme M. *Martinet* l'a fait et qu'on exclue les quelques mots d'argot commençant par /ñ –/, par ex. /ñuf/ «gniouf».

2° Les combinaisons binaires de consonnes sont plus riches à la finale qu'à l'initiale.

A l'initiale l'ordre des consonnes est unique; on trouve par ex. /tr –/ mais non */rt –/: trãt/ «trente», /gl –/ mais non */lg –/: /glã/ «gland».

A la finale on trouve et l'ordre de l'initiale et l'ordre inverse, par ex. /mɛtr/ «mettre» /tart/ «tarte» /ɛgl/ «aigle» et /alg/ «algue».

Il est vrai que l'existence des groupes à la finale est assez précaire, mais elle est assurée par l'*ə* caduc qui peut les alléger; c'est-à-dire que souvent les syllabes sont phonétiquement ouvertes, mais fermées du point de vue phonématique: il n'y a pas commutation entre, par ex. [sartrə] et [sartr̥] qui sont tous les deux variantes de /sartr/ «Sartre».

Martens (Hamburg): Man sollte bei der Entgegnung auf Herrn *Malmbergs* Darlegungen nicht die Struktur der Silbe außer acht lassen.

Es ist kein Beweis gegen *Malmberg*, die Folgen /rt/ und /tr/ am Ende französischer Silben aufzuführen und dabei unberücksichtigt zu lassen, daß /rt/ in tarte dem Wort keine Silbe hinzufügt, während /tr/ in *mettre* ein silbisches /r/ als Träger einer weiteren Silbe hat.

Außerdem dürfte die Struktur von aigle (stimmhafter Explosivlaut + stimmhafter Dauerkonsonant) nicht in eine Kategorie geworfen werden mit Folgen von stimmlosen Explosivlauten und stimmlos realisiertem (-normaliter stimmhaftem) Dauerkonsonanten.

Auf diese Weise lassen sich *Malmbergs* Ausführungen wahrlich nicht widerlegen.

Krámský (Praha): The instructive and very interesting lecture of Mr. *Malmberg* has centred our attention to the fact that the basic syllabic type which is also most widely spread, is CV; another very important fact is that the number of oppositions is smaller in languages where closed syllables are possible.

I should like to add a remark which does not concern the instability of the syllable quite in the sense of the lecture, but rather the instability of syllabic boundary. It is a well known fact that the same complex of identical sounds in different languages can have a different syllabic boundary. Thus the word *Krve* (genitive singular) in Czech has two syllables, whereas the same word in Polish forms only one syllable. As one of the most important components of the syllable formation we can consider the muscular tension (according to *Grammont*), because this tension determines the syllable peak and the length of the optimum phase. Another very important factor of the syllable formation is different adhesion of the elements of the syllable, that is the looseness or tightness of the junction of particular phonemes of the same syllable or of the final phoneme of one syllable and initial phoneme of the next syllable. This looseness or tightness of syllable junction is different even in the same language, between different phonemic junctions. According to *Alena Skaličková* the sequence VC is not so tight as the sequence CV and the sequence CC is not so tight as the sequence CV or VC. This is in accordance with Mr. *Malmberg:* open syllables, that is the sequence CV, show greater adhesion of their elements, that is they are more stable than closed syllables. Let me add to this that the tightness of the junction of two phonemes is different even when sequences VC and CV or CC are composed of different kinds of the sounds in question. I shall quote only an English example: the sequence n + ə in the word *finer* is different from that in *finely* and *finally*. Examples from Czech would be still more instructive.

Proc. 5th int. Congr. phon. Sci., Münster 1964, pp. 409–411
(S. Karger, Basel/New York 1965).

Speech Transmission Laboratory Royal Institute of Technology (KTH), Stockholm

Information Bearing Aspects of Formant Amplitude

By JÁNOS MÁRTONY and GUNNAR FANT, Stockholm

A formant may be specified by its frequency, amplitude, and bandwidth. Most of the experimental studies of formant structure up-to-date have been concerned with formant frequencies. In recent years communication engineers have shown a rising interest in studying formant bandwidth. This is motivated by the role of the frequency and the bandwidth as the measurable counterparts of the imaginary and the real part of a complex number, the pole entering the theory of resonance phenomena. Formant amplitudes, however, have not been studied to the same extent. Data on formant amplitudes of English vowels have been given by *Peterson* and *Barney*[1] and corresponding data on Swedish vowels by *Fant*[2].

Formant amplitudes are related to formant bandwidth to the extent that an increase in formant bandwidth, everything else held constant, results in a reduction of the formant amplitude. Formant amplitudes are in addition dependent of the overall pattern of formant frequencies and possible antiresonances (antiformants) so that a formant amplitude increases, when it is shifted to a position closer to an adjacent formant and it decreases as it is made to approach the frequency location of an antiresonance. A shift up in the frequency of the first formant is associated with an increase in the intensity level of the entire spectrum above the first formant including all formant amplitudes. This is the natural mechanism whereby a more open (high F_1) sound is more intense than a more closed (low F_1) vowel phonated with the same vocal source effort.

In addition to these changes of formant amplitudes which are a function of the articulation, i.e. the vocal cavity configuration, and thus are fully predictable the main source of variations is the phonation, i.e. the particular spectrum of the raw material supplied by the vocal cord source or by any other primary source. The source characteristics changes with the speaker, his voice effort, pitch level, and the prosodic stress pattern within connected speech. They are to some extent also dependent on the articulation. Spontaneous variations in the glottal waveform and thus in the corresponding source spectrum occur frequently. Recent studies[3] have revealed considerable fluctuations of formant amplitude in connected speech which are related to the particular ratio of voice fundamental frequency F_0 to a formant frequency F_n, particularly F_1. These fluctuations involve nonlinear interaction between the source and the filter.

It is well known that a vowel retains its phonetic identity irrespective of formant amplitude changes providing these are within reasonable limits (10–20 dB) and the pattern of formant frequencies remains constant. Individual variations typical of a speaker's voice category may be studied.

In connected speech the transition from a vowel to a consonant is typically accompanied by a change in formant amplitudes. In case of a transition from a nasal consonant to a vowel the instant of oral opening is associated with a radical change of formant amplitudes. Typically the second formant is much weaker in the consonant than in the vowel. This is one of the cues of nasal consonants.

The object of the present study was to see if appropriate formant amplitude changes in synthetic speech introduced via formant bandwidths could create the effect of a transition from a nasal consonant to a vowel.

From our experience of speech analysis and synthesis and earlier investigations, e.g. *Fant*[4], *Fujimura*[5], it would be expected that a correct reproduction of formant amplitudes would enable a listener to perceive a synthetic syllable as lateral + vowel if the bandwidths were low and nasal + vowel if the bandwidths were larger, i.e. when formant amplitudes were accordingly decreased. A clear tendency in this direction was observed.

Our study was devoted to the nasals [m] and [n] in CV syllables with the vowels [i] [u] and [a]. The synthesis was made on OVE II, a synthesizer with the formant circuits in cascade. In addition to the normal control via conducting ink lines drawn on a function generator there were relay controls incorporated for introducing changes in formant bandwidths and thus in formant levels.

Our preliminary results may be summarized as follows.

[ni] A low level of the second formant in the nasal murmur is the primary cue. The nasalization cues in the form of bandwidth increase of the higher formants were not significant.

[na] The correct reproduction of formant amplitudes in the nasal murmur was not sufficient. The nasalization cue in the vowel in the form of a decreased first formant amplitude was necessary. Without this modification the syllable is heard as [da] or [la] and the perception of [l] is further enhanced by reducing the bandwidth of the second formant within the consonant interval.

[mi] The low L_2 in the nasal segment is important. If this level is raised the nasal shifts to [l] or [b]. The nasalization of the [i] was not important as a cue for the nasal consonant.

[ma] Both the nasal murmur and the vowel are of importance for the identification of the nasal.

[mu] The optimum synthesis required a low L_2 in both the nasal
[nu] and the vowel. The [mu] was, however, not quite natural.

The increase of formant bandwidth as a means of reducing formant amplitude has the unavoidable effect of spectrum flattening which does not allow the reproduction of a valley in the 700 c/s range of the nasal murmur.

References

1. *Peterson, G. E.* and *Barney, H. L.:* Control methods used in a study of the vowels. J. Acoust. Soc. Amer. *24:* 175–184 (1952).
2. *Fant, G.:* Acoustic analysis and synthesis of speech with applications to Swedish. Ericsson Technics *15,* 1: 3–108 (1959).
3. *Fant, G.; Fintoft, K.; Liljencrants, J.; Lindblom, B.* and *Mártony, J.:* Formant amplitude measurements. To appear in J. acoust. Soc. Amer. (November 1963).
4. *Fant, G.:* Acoustic theory of speech production (Mouton, 's-Gravenhage 1960).
5. *Fujimura, O.:* Analysis of nasal consonants. J. acoust. Soc. Amer. *34:* 1865–1874 (1962).

Authors' address: Dr. J. Mártony and Dr. G. Fant, Speech Transmission Laboratory, Royal Institute of Technology, *Stockholm 70* (Sweden).

Verh. 5. int. Kongr. Phon. Wiss., Münster 1964, pp. 412–415
(S. Karger, Basel/New York 1965).

Zum Lautwandel

Von Piero Meriggi, Pavia

Die ganze Sprachwissenschaft zerfällt in Laut- und Bedeutungslehre, und sofern sie als Geschichte betrieben wird, sind Lautwandel und Bedeutungswandel ihre Gebiete. Es ergibt sich daraus, welch großen und grundlegenden Teil der Sprachwissenschaft die Frage nach dem Lautwandel ausmacht, der immer noch ganz unaufgeklärt ist. Zwar liegen ein Dutzend Theorien zu seiner Erklärung vor, aber das ist, wie immer, der beste Beweis, daß das Problem ungelöst ist. Denn sonst gäbe es ja nur eine allgemein angenommene. Eine Prüfung zeigt denn auch, daß keine der vorgeschlagenen Lösungen auch nur der einfachsten Kritik standhält. Zwar könnte man sich mit der Erkenntnis trösten, daß die Sprachwissenschaft fast nie auf die Frage nach dem Warum der sprachlichen Erscheinungen, mögen sie synchronischer oder diachronischer Art (oder auch panchronischer Art, wie ich hinzufügen muß) sein, antworten kann. Sie vermag nur in günstigeren Fällen das Wie aufzuzeigen, d.h. die Entwicklung zu erzählen, die zur gegebenen Erscheinung geführt hat. Doch wird man sich nicht so leicht zu dieser resignierenden Haltung bequemen, indem man auf jeden Versuch der Erklärung verzichtet.

Nun habe ich schon 1952 beim Londoner Linguistenkongreß eine Betrachtung geltend gemacht, die zwar keine Lösung des alten Problems darstellt (es wäre reine Anmassung, so etwas zu denken), aber wohl zu einer neuen etwas aussichtsreicheren Problemstellung führen kann. Bei meinen experimentalphonetischen Untersuchungen über die Artikulation der italienischen Konsonanten hatte sich ergeben, daß sie ungeahnt starke Veränderungen je nach ihrer Stellung im Satze erleiden, insofern die normale Aussprache nur in den (natürlich aus Bedeutungsgründen) hervortretenden Satzteilen stattfindet. In den unwichtigeren Satzteilen erfährt die Artikulation eine Lockerung, die zu starken Veränderungen führt: z.B. wird ein

intervokalisches *t* zu aspiriertem *th* oder gar zur Frikativa *þ* bzw. *ð* und weiter über *h* sogar zu Null. Was dabei beeindruckt, ist die Feststellung, daß diese synchronisch möglichen Veränderungen ja die gleichen sind, die wir in der Diachronie von jeher feststellten, sofern die Urkunden es uns erlaubten, oder analogisch annahmen, wenn uns nur die Extreme dieser Entwicklungskette bezeugt waren: etwa *t* > *þ*/*ð* oder gar *t* > *0*.

Zu dieser ersten Beobachtung kam später eine weitere hinzu. Diese Phasen der Entwicklung liegen auch synchronisch in ganz anderer Weise vor: in der geographischen Verteilung nämlich. Man braucht nur eine Karte des AIS aufzuschlagen um zu sehen, daß die im Experiment bei einer und derselben Versuchsperson auftretenden Phasen, sich im Raum so verteilen, daß von einem Zentrum mit normaler Aussprache *–ato* ausgehend, in der zunächst nahen, dann ferneren Umgebung sich all die Entwicklungsphasen
–at^{h}o > *–aþo*
–ado > *–aðo* > *–aho*, *–ao*, schließlich einerseits *–ô*, andererseits *–â*
finden.

Um ein an sich zu stark schematisches Beispiel zu geben, das aber den Vorteil hat, Tatsachen zu verwerten, die jedem bekannt sind: wenn man die Entwicklung der latein. Velaren vor palatalen Vokalen (z.B. in *centum* oder *amici*) verfolgt, sieht man, daß die nächste Stufe die toskanische Affricata *č* (bzw. *j**) ist, dann kommt die norditalienischen (z.T. auch französischen) Fricativa *š*, dann im Französischen die Dentale *s* und schließlich im Spanischen *þ*. Mit dieser Interdentalen erreichen wir die weiteste Entfernung von der ursprünglichen Artikulationsart. Will man diese Verteilung im Raum, die der physiologisch bedingten Verschiebung der Artikulation entspricht, wozu auch die Verteilung in der Zeit (da wir normalerweise ein *h* nicht gut direkt von *t* oder *d* ohne Zwischenstufen ableiten können) hinzukommt, für ein Spiel des Zufalls erklären? Gibt man aber eine kausale Verbindung zwischen diesen drei Reihen zu, so heißt es nun, diese Betrachtungsweise bei der Behandlung des Lautwandels anzuwenden. Jedenfalls wird jeder Versuch, die Ursachen dieses Wandels zu erklären, mit diesen Feststellungen rechnen müssen. Steht der Versuch mit ihnen nicht im Einklang, so wird er sich als verfehlt erweisen wie alle bisherigen.

Kehren wir nun zu unserem Ausgangspunkt zurück, den Experimentaluntersuchungen, die mich zu diesen Betrachtungen führ-

* *j* (mangels einer genaueren Type) hat hier den englischen Lautwert.

ten, so wird nun klar werden, daß die alte am meisten angenommene Theorie, der Lautwandel lasse sich im Grunde auf die Bequemlichkeit zurückführen, zwar nicht ganz falsch, aber völlig ungenügend ist. Sie hat für sich, daß die häufigste und frappanteste Folge des Lautwandels der Lautschwund ist, und die meisten sonstigen Veränderungen sich mehr oder weniger leicht als Abschwächungen auffassen lassen. Doch gibt es auch klar entgegengesetzte Fälle, z.B. die Diphthongierungen oder die Verschärfung (bzw. Dehnung) der Konsonanten. Also beide Vorgänge, Schwächung und Verstärkung, sind nachzuweisen. Höchstens bleibt es zu erklären, warum der erstere weitaus überwiegt. Auch darauf scheint mir die neue Betrachtungsweise eine Antwort zu geben.

Wollen wir, was ja der normale Fall ist, im Satze etwas hervorheben, so müssen wir alles Übrige artikulatorisch herabsetzen (sonst wären wir gezwungen, bei dem Teil, den wir hervorheben wollen, höchste Energie anzuwenden). Bei dieser artikulatorischen Herabsetzung oder Dämpfung erleiden aber die Laute offenbar normaler- und begreiflicherweise die Lockerung und Veränderung, von der wir sprachen. Mengenmäßig sind diese schwächeren Lautstrecken im Satze überwiegend, da wir ja normalerweise nur ein Element hervorheben wollen, z.B. in: *ich wollte, daß du mir das Buch bringst!* Also die energische (eventuell übertriebene) Artikulation ist seltener als die schwächere. Kein Wunder also, daß diese dazu neigt, vorherrschend zu werden.

Diese Energieverteilung rührt also nicht von der Bequemlichkeit her, sondern von dem Streben nach Wirksamkeit des Gesamtausdrucks, das ja auch auf semantischem Gebiet die grundlegende Triebkraft ist (also wieder ein Fall des meines Erachtens in der ganzen Sprachwelt dominierenden Parallelismus zwischen Lautung und Bedeutung). Man muß die längsten Strecken im Satze (also phonetisch: in der Artikulationskette) dämpfen, damit der uns wesentliche Teil desto besser und leichter hervortritt. Alles gleichmäßig stark und sorgfältig zu artikulieren, würde den Hauptzweck unserer Mitteilung vereiteln. Daß wir die wirklichen Verhältnisse gewöhnlich so verkennen, ist natürlich, wie so viel anderes, dem verhängnisvollen Einfluß des Schriftbildes zuzuschreiben, von dem wir uns so selten freimachen. Deshalb ist die Experimentalphonetik für den Sprachforscher von so wesentlicher Bedeutung, daß ich mir ohne sie wie ein herumtappender Blinder vorkäme.

Adresse des Autors: Prof. Piero Meriggi, Facoltà di Lettere, Università di *Pavia* (Italien).

Discussion

Horálek (Prag): Eine Bemerkung zur Frage der Unterschiede zwischen der junggrammatischen und strukturalistischen Auffassung der Sprachentwicklung.

Den junggrammatischen Standpunkt kann man als sprachwissenschaftlichen Kausalismus charakterisieren, die Strukturalisten (wenigstens die Prager Schule) streben zur Teleologie, nicht aber im Sinne der Theorie des Fortschrittes (wie z. B. bei *Jespersen*), sondern im Sinne der zweckmäßigen Anpassung. (Was damit gemeint ist, wurde an dem Beispiel der Schicksale der konsonantischen Weichheitskorrelation und der vokalischen Quantitätskorrelation im Altpolnischen und im Alttschechischen gezeigt.)

Hoffmann (Ibadan): Ich möchte Herrn *Meriggi* zugeben, daß die von ihm angeführten Gründe der Lautveränderung sicher eine Erklärung sind, glaube aber nicht, daß sie ausreicht, alle Erscheinungen zu erklären. Ich hätte gern von Herrn *Meriggi* gehört, wie er sich die Unterschiede in der Entwicklung z. B. zweier so nahe verwandter Sprachen wie des Litauischen und des Lettischen erklärt, die sich bei engstem räumlichem Zusammenhang doch so unterschiedlich entwickelt haben.

Bluhme (Amsterdam): Für den Lautwandel, den Herr *Meriggi* zeigte, spielen gewiß die Akzentverhältnisse im Wort eine entscheidende Rolle. Die Wirkung des Satzakzents läßt sich im Deutschen an der phonologischen Quantität sehen: Unter dem Satzakzent ergibt sich für das Verhältnis der langen zu den kurzen Vokalphonemen ein höherer Quotient als bei den wortbetonten Vokalen.

Verh. 5. int. Kongr. Phon. Wiss., Münster 1964, pp. 416–419
(S. Karger, Basel/New York 1965).

Zur Theorie der Phonembildung

Von W. Merlingen, Wien

Eines der markantesten Beispiele dafür, daß Phonetik und Phonologie verschiedene Aspekte haben, bilden sicherlich die sibilantischen Affrikaten, die phonetisch mit *ts*, *dz*, *tš*, *dž* (usw.) wiedergegeben werden: «einheitliche» Laute und dergleichen für viele Sprachwissenschaftler, gesondert klassifiziert von Phonetikern, Einzelphoneme /c, ʒ, č, ǯ/ für die Phonologen (Explosive mit besonders hörbar spirantischer Explosion, bedingt durch eine besondere Zungenhaltung; die Zwangsläufigkeit einer solchen spirantischen Explosion kann man sich etwa bei den lateralen Vertretern besonders deutlich machen). Nun ergibt aber die phonologische Analyse (Trubetzkoyscher Prägung) einen Umstand, der mit der streng logischen Gesetzlichkeit des phonologischen Status unvereinbar erscheint: es gibt verschiedene *Grade* der Einphonemigkeit. Höchster Grad: eine Sprache kennt keinerlei selbständige Spiranten, wohl aber Lautverbindungen wie *tš*, *dž*, deren [š, ž] also keine Phoneme, sondern nur Modifikationen, d.h. Merkmale von Phonemen bilden können. Die meisten Sprachen mit [tš] usw. verfügen aber daneben durchaus über selbständige *š*, *ž*, *s*, *z*, und die Einphonemigkeit, derer wir in den meisten Sprachen gewiß sind, gibt sich zwar auch, aber nicht mehr so scharf zu erkennen. Unter den verschiedenen Kriterien ist am häufigsten das eines *allgemeinen Prinzips* in der betreffenden Sprache anwendbar (Analogie); je allgemeiner das Prinzip, desto klarer die Einordnung fraglicher Fälle. Der höchste Grad wäre hier etwa der, daß eine Sprache grundsätzlich keine Konsonantenverbindungen besitzt; in diesem Fall sind beschränkte Ausnahmen wie *tš* usw. ebenfalls «Einzelkonsonanten». Es gibt aber, wie z.B. die europäischen Sprachen auf den ersten Blick zeigen, weit schwächere Bedingungen oder «Grade» der Einphonemigkeit (für die Einzelheiten sei auf ZfPh *13:* 116ff., 133ff., verwiesen). Trotz

dieser verschiedenen Schärfe der Kriterien sprechen wir – und wir dürfen es – bei *allen* Fällen von frei und regelmäßig auftretenden *č* usw. von Einzelphonemen; wir hätten ja auch keine Handhaben, irgendwo Grenzen zu ziehen.

Aber nun haben wir zu fragen: Wer *entscheidet* denn (für den Sprachträger), ob ein [tš] usw. diesseits oder jenseits der Grenze der Ein- oder Zweiphonemigkeit liegt? Für den Sprachträger jedenfalls kann wohl die Zugehörigkeit immer nur eindeutig sein, so verschwommen sie auch etwa vom rein phonetischen Standpunkt aus sein mag. Es stellt sich heraus, daß die Entscheidung *im Sprachträger* getroffen wird, daß also in unseren Fällen die Ein- und Zweiphonemigkeit eine *subjektive* Angelegenheit ist. Es wird schon beim Erlernen der Muttersprache eine einfache Entscheidung getroffen, die dann für das Phonemsystem des ganzen Lebens gilt.

In diesen Jahrzehnten der mathematischen Linguistik und der vom Menschen losgelösten Sprachbetrachtung dürfen und müssen wir die Aufmerksamkeit darauf zurücklenken, daß die Sprache, obwohl allen möglichen vom Menschen unabhängigen Bedingungen unterworfen, doch in hohem Maße wieder an den *Menschen* gebunden ist. Und wenn wir die Aufgaben der Sprachwissenschaft in der Erforschung der Natur der Sprache und der Sprachen und ihrer Veränderungen erblicken, so erfahren wir das, was wir wissen wollen, nur dann, wenn wir auch nach den *subjektiven* Vorgängen bei der Sprache fragen. Und eine dieser Fragen lautet also: Worin besteht die Phonembildung für den Sprachträger, vom Sprachträger aus? Sie liegt tief unter dem Bewußtsein des Erwachsenen (von Schrift, Schule usw. überlagert) und kann, wie in unserem Fall von /č/ usw., durch gewisse Tricks der Phonologie bloßgelegt werden. Wichtige Beispiele dafür scheint es mir bei der Frage nach «merkmallosem und merkmalhaltigem Glied von Oppositionen» zu geben (wobei wir die akustische Betrachtung zu der vom Menschen losgelösten Sprachbetrachtung rechnen und uns an die Artikulation halten, in der wir die wirklichen Vorgänge am nächsten greifen können):

In slawischen und anderen Sprachen gibt es die bekannte Erscheinung der sog. *palatalisierten Konsonanten*, die sich von den anderen durch vorgedrückte Vorderzunge unterscheiden. Man kann aber auch sagen, daß sich die anderen von den pal. Kons. durch zurückgezogene Zunge unterscheiden. Was gilt? Das kann nicht von vornherein für jede Sprache feststehen. Wenn die Phonologen nur von palatalisierten Konsonanten sprechen, so sind sie der Suggestion einer Art Weltnorm unterlegen (die ja allerdings dem Standpunkt der reinen, nicht phonologisch interpretierenden Phonetik entspricht). Tatsächlich ist z.B. im Russischen und Polnischen das Umgekehrte zu erkennen, daß nämlich die palatalen Konsonanten die merkmallosen sind, die nicht-palatalisierten die merkmalhaltigen (Mermal: «Velarisierung»). Der Trick: «hartes *i*» (*y*) und «weiches *i*» (*i*) hängen von der Zungenhaltung des vorhergehenden Konsonanten ab; wenn eine der beiden Zungenhaltungen das zusätzliche Merkmal ausmacht, so ist das merkmallose Glied dort erkennbar, wo das *i* unbeeinflußt ist, also im Anlaut steht; das ist aber im Russischen und Polnischen [i], nicht [y]; [y] kann nur von vorhergehenden Konsonanten verursacht werden, die ihrerseits eben das zusätzliche Merkmal – Zurückziehung der Zunge – tragen (s. «Zur Phonologie...»).

Ein anderes Beispiel. Die Vokale *e/i* und *o/u* unterscheiden sich phonetisch durch den Öffnungsgrad bzw. durch die Zungenhebung. Zugleich ist aber die Zunge beim *i* weiter vorn als beim *e*, beim *u* weiter hinten als beim *o*. Was gilt als zusätzliches Merkmal? Zur Unterscheidung kann nur eines von beiden dienen (die an sich nichtssagende Gruppierung der Vokale zu Dreiecken und Vierecken führt irre). Es muß aber nicht eine – auch hier von den Phonologen bevorzugte – Weltnorm gelten, wie z.B. wieder das Russische zeigt. Kriterium: /e/ und /o/ gibt es nur in betonter Stellung, unbetont verlieren sie ihr zusätzliches Merkmal; aber /e/ wechselt zwar mit /i/, /o/ jedoch nicht mit /u/ – was sich auf der Bahn Öffnungsgrad/Hebung abspielen würde –, sondern mit /a/; Gemeinsamkeit: /e/ und /o/ verlieren unbetont die *Zurückziehung der Zunge*, und dies ist das Unterscheidungsmerkmal, übrigens vergleichbar mit dem bei den Konsonanten. (Russ. [o] steht also in Relation zu /a/, von dem es sich phonologisch durch Zurückziehung der Zunge unterscheidet.)

Ähnlich im Französischen. Welche Merkmale die vermeintlich offenen und geschlossenen Vokale haben, zeigt die keinen Einwirkungen ausgesetzte Stellung, nämlich die in offenen Silben an nichtletzter Stelle: «geschlossenes» *e* und *a*, aber «offenes» *o* und *ö*, ein Widerspruch, wenn es sich um Öffnungsgrad/Hebung handeln würde; das gemeinsame zusätzliche Merkmal ist wieder die Zurückziehung der Zunge, wenn auch in engerem Bereich als im Russischen.

Wir sehen, es gibt nicht nur die Frage, *ob* ein lautliches Merkmal distinktiv ist, sondern auch *welche Seite* der Artikulation es ist. Ein wichtiger Teil der inneren, subjektiven Vorgänge bei der Phonembildung. Die Sprachen der Welt bieten dafür noch unabsehbares Material.

Literatur

Merlingen, W.: Zur Phonologie der sogenannten palatalisierten Konsonanten. Sudia linguistica in hon. acad. Stephani Mladenov, 493ff. (Sofija 1957).
Merlingen, W.: Über Ein- und Zweiphonemigkeit. Z. f. Phon. *13:* 98–176 (1960 [1962]).

Adresse des Autors: Dozent Dr. W. Merlingen, Rudolf-Bärenhart-Gasse 15, *Wien 17* (Österreich).

C.R. 5ᵉ Congr. int. Sci. phon., Münster 1964 pp. 420–421
(S. Karger, Basel/New York 1965).

Les fondements de la phonétique quantique

Par R. F. Mikuš, Zadar

Les beaux et très encourageants résultats de la phonétique instrumentale, surtout acoustique, me semblent dans un certain sens compromis par l'obstination avec laquelle se maintiennent en phoné(ma)tique et en linguistique les notions traditionnelles de «son», «voyelle», «consonne», «phonème (consonantique, vocalique)» et semblables qui témoignent d'un décalage entre un ordre de choses dûment analysé par les sciences et une nomenclature inadéquate qu'une phonétique théorique aurait dû remplacer il y a bien longtemps par une autre, plus rationnelle.

Certes, derrière ces notions certaines réalités se cachent, tout comme derrière l'expression telle que «le soleil se lève à 6 heures» en est escamotée une. Bien que nous sachions depuis *Galilée* que le soleil «ne se lève» ni ne «se couche», nous continuons par inertie linguistique à parler du «lever (coucher)» du soleil. En est-il de même de nos notions? Evidemment non, car on a pas encore posé sérieusement la question de savoir ce qu'elles signifient en fait.

Dans un article qui paraîtra prochainement dans la revue *Phonetica*, j'ai essayé de montrer que «son», «voyelle», «consonne» ne sont que des noms donnés par l'empirisme primitif et le réalisme naïf de l'antiquité aux qualités («timbres» ou «couleurs») de l'expiration rendue audible grâce aux certains mouvements et aux certaines positions des organes phonatoires. Pour décrire le phonétisme d'une langue, tout autre terme est aberrant sauf celui de *qualité expiratoire* dont les variétés sont définies par les variations topographiques et par différentes phases (implosion, tenue, explosion) de l'acte phonatoire.

Parallèlement à l'abolition de ces termes néfastes, dont l'imprécision tenait la phonétique pendant deux milliers d'années dans une impasse, il a été nécessaire de définir d'une manière prégnante aussi l'unité phoné(ma)tique fondamentale du langage.

Etant donné que par sa nature la parole n'est que l'expiration «colorée» et que celle-ci est linéaire, l'unité phoné(ma)tique fondamentale ne saurait être que linéaire elle aussi. Comment la chaîne parlée pourrait-elle s'étendre dans la linéarité s'il n'en était pas ainsi? Voilà pourquoi l'unité fondamentale n'est ni «voyelle», ne «phonème», mais un segment infinitésimal de l'expiration que j'ai appelé *quantum phonétique*, et qui est un invariant dont les propriétés: quantité (longueur) relative, «couleur», intensité et hauteur musicale relatives constituent les «traits distinctifs»; la «syllabe» est la notion empiriste primitive de quantum, mais ne s'identifie pas avec lui, car il y a des quanta que la tradition ne connaît pas comme «syllabes».

En analysant différents jeux des phases phonatoires, j'ai trouvé, dans l'application de la théorie, que le français possède environ 27 types quantiques, fondés sur 8 types fondamentaux pour lesquels la phonétique française ne connaît que «syllabe fermée» et «syllabe ouverte».

Adresse de l'auteur: Prof. Dr. Radivoj Franciscus Mikuš, Beogradska 16, *Zadar* (Yougoslavie).

Discussion

Váchek (Praha): Es ist sehr fraglich, ob die Quantisierung der qualitativen Tatsachen einen positiven Wert bringen kann. In der Sprache sind es gerade die qualitativen Tatsachen, die einen funktionellen Wert haben.

Proc. 5th int. Congr. phon. Sci., Münster 1964, pp. 422–425
(S. Karger, Basel/New York 1965).

Some Problems around the Growth of the Vocal Tract

By H. Mol, Oegstgeest

In phonemic theory the anatomic differences between the articulatory organs of men, women and children have never been taken seriously into account. The same can be said of the differences within one of these three groups. The level of abstraction on which the phoneme is defined is so high above, or in any case so distant from the actual mechanisms involved in the process of speaking and listening that the indifference of the phonemicists towards these anatomic variations can be easily imagined: the underlying doctrine of phonemics is the statement that isolated words are 100% intelligible, no matter what.

As long as the acoustic theory of the vocal tract and reliable measurements of the vowel formants were still in the cradle much faith could be attached to compensatory mechanisms that were supposed to alleviate many of the anatomic differences between the speakers of the same language.

In the light of modern phonetics two questions can be asked:

1. Are compensatory mechanisms powerful enough to alleviate the acoustic results of the gross anatomic differences between men, women and children?

2. If they are, are they actually brought into play by the speakers? Both questions must be answered in the negative.

By applying, for instance, *Webster's* horn equation it can be proved that the frequencies of *all* formants of the vocal tract are inversely proportional to the length of that tract as measured between the lips and the vocal cords. From this fact a very interesting conclusion may be directly drawn.

Suppose a young speaker has mastered to produce the inventory of vowels characteristic of his language and, during the growth of

his vocal tract, he does nothing to compensate for the acoustic results of this axial growth, all his 'adult' formants will be lower by the same factor than the corresponding 'juvenile' formants. It means that he then keeps on giving the same commands to his articulators. Measurements prove that this situation is the normal.

Though the authors themselves do not stress the point, this proporti nality factor can indeed be found in the well-known formant measurements of *Peterson* and *Barney*[1], who presented separate averages for the formants of men, women and children in carefully pronounced words containing one vowel in medial position.

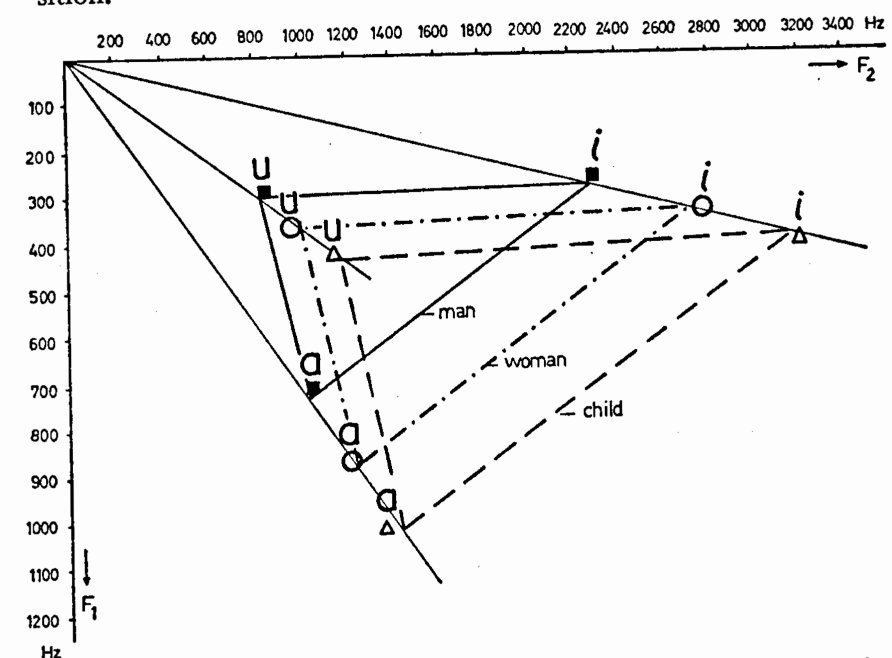


Fig. 1. Illustration of the fact that there are constant ratios between the formants of men, women and children.

In Fig. 1, based on the measurements of *Peterson* and *Barney*, for the sake of clarity, only the three vowels forming the familiar vowel triangle of *Hellwag* have been brought to the fore, though the properties of proportionality also hold for the other vowels. Within reasonable limits of accuracy the triangles can be derived from each other by simple multiplication from the zero point of the frequency scales. The corresponding vowels of men, women and children are on straight lines through that zero point, the so-called origin, provided linear scales are chosen for both formants. The ratio between the formants of women and men equals 1.16. The same factor relates children and women, whereas between the formants of children and men the ratio 1.35 obtains.

While learning to speak the child does not (and indeed cannot) imitate the formants of the adult speakers in its environment. Unlike a parrot, it does not give exact imitations on an acoustic basis. In the long run the child learns to master the muscular activity necessary for producing the *same number* of perceptively different vowel sounds used for coding purposes by adult speakers of the same language in situations where the need for utmost clarity is greatest. Gradually the child learns to choose the, in his language, normalized number of perceptually different tongue positions realizable within the limiting boundaries of the teeth, the back, the roof and the floor of the mouth. He uses and needs his ear in this process but certainly does not adapt the absolute values of the formants to the example.

Once settled in a certain period of life the neurological programmes for controlling the articulatory muscles do not change materially as the child grows up. Apparently the tongue keeps the same relative position in the vocal tract for each vowel. Lengthening of the vocal tract makes the formants shift downwards and there are no compensatory attempts on the part of the speaker to make them stay at their 'juvenile' positions. Obviously the role of the often postulated auditory feed-back mechanism for regulating the articulatory movements is an extremely minor one.

Every speaker has his own vowel system. The range of the frequencies of his formants is influenced by the dimensions of his vocal tract. The listener has to cope with the difficulty that the absolute values of the formants of the vowels vary from speaker to speaker[2]. He is helped, among other things, by the fact that in a given language all speakers have the same *number* of vowels. A listener is able to adapt himself in a surprisingly short time to the formant positions of a speaker. When he is not able to do so, for instance when isolated words of unknown speakers are presented to him, his identifications of the intended vowels become inaccurate[1].

If we admit that the exact acoustic imitation of the vowels is not the aim of learning to speak, we can understand why the inherited shapes of parts of the vocal tract, like the hard palate, can withstand

'compensatory' actions. It is a well-known fact that quite often the voice of a speaker resembles that of his father as soon as his larynx and his vocal tract have reached their final dimensions.

In general it can be said that the encouragement for an articulatory programme to 'settle' is not the result of successfully matching an example. This encouragement must stem from the dawning of the idea that the sounds produced really function. The parrot is in a different position: he is a real imitator. We have established the fact that a parrot learns to speak without training. He is able to store a word or a sentence heard only once at a certain time, often under conditions of psychological stress, in his memory and to reproduce this utterance suddenly at a considerably later moment. Apparently he possesses a built-in inventory of acoustically 'labelled' articulatory programmes; as soon as he hears a sound he immediately selects the corresponding articulatory programme. Let us call this gift 'parroting'. Every now and then one hears about children who seem to have the gift of parroting. They are very silent, so long that the parents become uneasy about it, until suddenly they produce 'complete' words.

Besides permanent growth the vocal tract is subject to temporary length-variations controlled by the will of the speaker.

The axis of the vocal tract can be lengthened by moving the larynx downwards, moving the hump of the tongue upwards and backwards and by pouting the lips. There even are formant shifts that can only be explained by a length-variation of the vocal tract, for instance the shift from [a] to [ɑ] in Dutch.

References

1. *Peterson, G.E.* and *Barney, H.L.:* Control methods used in a study of the vowels, JASA Volume 24, Number 2, p. 175–184, March 1952.
2. *Ladefoged, P.* and *Broadbent, D.E.:* Information conveyed by vowels, JASA 29, p. 98–104 (1957).

Author's address: H. Mol, Prinses Beatrixlaan 28, *Oegstgeest* (Netherlands).

Discussion

Sovijärvi (Helsinki): Herr *Mol* hat darauf hingewiesen, daß die Relation des ersten bzw. zweiten Vokalformanten bei den Männern und Kindern 1:1,35 sein würde (vgl. Fig. 1). Nach meinen Messungen, die ich von der laufenden finnischen Rede eines Mannes und eines 10jährigen Jungen gemacht habe, schwanken die Formantgipfelhöhen des F1 in demselben Bereich, welche Tatsache dadurch erklärt werden kann, daß jeweils die phonematisch am besten passenden F1-Höhen auf sehr verschiedene Weise wegen der wechselnden Grundtonhöhen erreicht werden. Dagegen ist die erwähnte Relation 1:1,35 betreffs der F2-Höhen auch nach meinem Material stichhaltig (vgl. meinen Artikel in «Phonetica, Symposium Trubetzkoy», 1958).

Proc. 5th int. Congr. phon. Sci., Münster 1964, pp. 426–430
(S. Karger, Basel/New York 1965).

Are Phonemes Really Realized?

By H. Mol, Oegstgeest

It has become increasingly clear since the 4th Congress that only in isolated words a speaker brings into full play the complete system of perceptual differences characteristic of his language.

For the correct identification of isolated words the listener is solely dependent on the acoustic information contained in the sound waves and on his knowledge of the words of the language in question and of the articulatory habits of the speaker.

As a matter of fact the well-known paradigmatic technique used in defining the phonemes is based on the often exaggerated precision with which isolated words can be identified. Hence, strictly speaking, the thus-defined phonemic system only pertains to the exceptional and rather abnormal situation of isolated words enunciated by one particular speaker. Application of the phonemic system to other situations, including the normal, is an as yet unproved extrapolation, notwithstanding the practical advantages of such an extrapolation for the art of writing.

In connected speech a speaker, as it were, eases the strain. He permits himself considerable overlap, especially in the vowels. No 'clustering' in his vowel-formants can be detected. Fortunately he gets away with his seemingly sloppy articulation because the listener has at his disposal other, extra-phonetic, cues derived from the context, the situation and his knowledge of the linguistic structures he may expect. In really alphabetic writing, connected speech is recorded visually with the aid of the complete set of phonemic symbols derived from the pronunciation of isolated words, a notation which suggests an articulatory precision that is decidely not present.

The seeming ease with which connected speech is recorded alphabetically by adult writers is very deceiving and lures our

attention away from a very complicated brain mechanism that places the human writer well above the level of a teletype. The alphabetic way of recording some of the aspects of connected speech is uncanningly ingenious but constitutes at the same time a great barrier to scientific progress. At the moment phonemics is completely under the spell of alphabetic spelling and pays lip-service to writing. This is frankly admitted by *K. L. Pike* who gave the following title to one of his books: "Phonemics, a technic for reducing spoken language to writing."

The current, but untenable, extrapolation in linguistics is to explain the aural identification of words as a running analysis into a familiar sequence of phonemes, thereby putting hearing into the same class as alphabetic reading. The next slippery step is to regard pronunciation as the production of a running series of phonemes. The main theme of this Congress: "The phoneme and its realization" openly shows this alphabetic bias.

Alphabetic writing is based on a conscious interpretation of the differences one thinks one hears between sound waves that are experienced as different words. These differences, however, are referred to in, often vague, articulatory terms, evidently because the articulatory data are the only data to which one has direct conscious access. Interviews with illiterates show that language users belonging to that category identify words in a subconscious manner. They cannot even answer simple questions that presuppose alphabetic training. They simply 'know' when different words are presented to them and state they 'hear' them as different.

The undeniable success of alphabetic writing and reading as practical methods based on conscious processes has, as it were, gone to the heads of those phonemicists who regard the phonemes as units that actually 'function' in a brain mechanism. They tacitly assume that, before the advent of phonemics, the phonemes did their work in the darkness of the subconscious. In their opinion the phonemes were, as it were, patiently waiting until they were brought into the light of publicity by a talent scout. Phonemes, however, have not been discovered, they have been invented, which makes all the difference.

Though the alphabetic system is the result of the combined efforts of many brain mechanisms, we may not, without further preface, infer that the brain also uses alphabetic units for its own purposes. We must reckon with the possibility that a manufacturer of lemonade drinks champagne himself. It is necessary to investigate whether the physiological properties of the brain mechanism makes such a thing as a high-speed running alphabetic analysis possible at all. Psychologists have established that the capacity of the nervous system is too low for handling the separate phonemes as they are defined by means of the paradigmatic technique. Moreover, phoneticians have shown by measurement that there is no one-to-one correlation between actual articulation and intention for the vowel phonemes in connected speech. Even in isolated words there are no such things as absolute acoustic cues that unambiguously label the speaker's intention.

Modern theory of brain mechanisms as, for instance, developed by *Frank Rosenblatt* who was greatly inspired by *D. O. Hebb*, shows how a brain mechanism can learn to recognize stimuli via a process of conditioning involving feed-back. Analysis into discrete units, so characteristic of conscious processes, is unnecessary and even undesirable in this automatic process. In the systems proposed by *Rosenblatt* in his "Neuro-dynamics" recognition of a pattern of nervous activity, be this recognition correct or not, is the result of the joint activity of, in this case, acoustic as well as situational stimuli acting on the same field of the brain.

Sound waves are, by their very nature, time functions that describe how the barometric pressure varies with time. Only a limited number of portions of a sound curve can make themselves felt in the nervous system as patterns of nervous activity. These portions follow each other in time because there is no other possibility.

Likening these portions to the alphabetic units that are also supposed to follow each other in time is tempting but nevertheless false. These portions are time intervals of possible nervous activity but the activity itself is not quantisized in discrete possible units: a practically continuous scale of patterns of nervous activity is possible in every portion. The net result is that the active length of a sound curve is shorter than the total duration of that curve.

A parrot can echo speech waves as well as barking and laughing. It cannot be expected to handle phonemes in speech and to use another system in barking or laughing. Its brain mechanism does not even recognize. It just stores the patterns of nervous activity corresponding to sound waves in a non-alphabetic manner and there is no reason to suppose that the human brain uses a fundamentally different system for storing and processing. Therefore, in my opinion, it

would have been wiser to speak of: "Phonetic reality and the phoneme" than of "The phoneme and its realisation".

Author's address: Dr. H. Mol, Prinses Beatrixlaan 28, *Oegstgeest* (Netherlands).

Discussion

Westring Christensen (Copenhague): Je suis d'accord avec M. *Mol* qu'il n'y a pas de relations très simples entre les phonèmes et les «faits» articulatoires et acoustiques, que les phonèmes n'existent pas dans le cerveau et que les phonèmes sont, dans une certaine mesure, inventés par les linguistes. Seulement je n'aime pas que M. *Mol* préfère parler de *la réalité phonétique et du phonème* au lieu de parler du *phonème et sa réalisation*, parce que la réalité phonétique en soi n'existe pas; elle est toujours interprétée par un observateur, que cela soit directement par un informateur ou par un linguiste, soit indirectement à l'aide d'instruments.

Je trouve que la «réalité» phonétique formée par une transcription phonématique est aussi «réelle» que la «réalité» phonétique observée par des instruments. Evidemment je ne nie nullement l'utilité d'examiner les aspects physiques, physiologiques ou psychologiques que propose M. *Mol*. Mais je pense que la description de la «réalité» phonétique par une transcription phonématique est beaucoup plus intéressante d'un point de vue linguistique que celle par ondes acoustiques ou méchanismes nerveux, non seulement pour le plan de l'expression dans le sens hjelmslévien, mais aussi pour le plan du contenu et pour les rapports entre les deux plans; s'il est par exemple, très compliqué de décrire les expressions de signes dans le sens hjelmslévien ou les morphèmes dans le sens américain par des traits distinctifs au lieu de phonèmes, il serait encore plus difficile de les décrire par des formants ou par des symboles de méchanismes cérébraux. Je ne crois pas qu'on puisse déduire une des descriptions de l'autre ou inversement, il faut faire les deux. Et j'aimerais qu'on arrive aussi à une description qui rend compte des deux groupes de facteurs pour le comportement des individus parlants, parce que l'interprétation humaine du discours n'est pas la même que l'interprétation animale. Il ne faut pas se faire d'illusions à ce qu'une description soit plus réelle ou plus scientifique que l'autre.

Je pense que la formule que propose M. *Mol:* «Phonetic reality and the phoneme» est aussi fausse que l'autre formule «the phoneme and its realisation», car d'une part les phonèmes n'existent que dans une notation ou transcription, et d'autre part la «réalité» phonétique en soi n'existe pas. J'aimerais qu'on trouve un autre terme que *réalité phonétique*.

Krámský (Praha): In Mr. *Mol's* lecture there were some good observations concerning the discrepancy between writing and speech as to the realization of phonemes of which words are composed. I should like to add some remarks to this interesting theme. I remember Prof. *Jakobson* saying in his lecture several years ago in Prague that some phonemes may be dropped from the word, but not any. And this is the problem: *which* phonemes may be dropped, and if they can be dropped without functional change of the meaning of the word in question, are they still phonemes? Let me give an example from Persian. There are two words in Persian meaning "four": /tʃæha:r/ and /tʃa:r/; *h* is dropped, the two *a*'s are contracted into one. Why was it possible that /tʃæha:r/ could be contracted into tʃa:r/? Perhaps because the *h* in /tʃæha:r/ has a very weak functional relevance or perhaps a non-distinctive function, so that by dropping it the word is changed but not its meaning, only its form. And the second condition is that the contracted word /tʃa:r/ does not exist in the language in a meaning different from that of the word /tʃæha:r/. If there were such words with another meaning, the reduction of sounds in the word /tʃæha:r/ could hardly take place. However, even this is not impossible in a language with a tendency to homonymy. Our observations lead us to the following conclusions:

1. We must differentiate between *systemic relevance* (that is the fact that a certain opposition exists in the system of the language) and *lexical relevance* (which depends upon the exploitation of phonemes).

2. In speech we must differentiate between *phonemes with full relevance* (that is such phonemes as are capable of changing the meaning of the word when replacing another phoneme) and *phonemes with non-full relevance* (that is such phonemes whose dropping from the word or replacing by another phoneme does not change the meaning of the word).

3. In connected speech the degree of phonemic relevance depends on the mutual relation of words in the context. If a word has a key position in a context, its distortion must not be great, its components are not redundant. On the other hand, an extreme case is the dropping of one or even more words in a sentence without changing its meaning or making it unintelligible.

The problem of the acoustic identity of the word deserves greater attention than it has been given so far; besides, it is also a problem of redundancy, and, consequently, it is of importance for the theory of communication.

Fischer-Jørgensen (Copenhague): It is evident that in current quick speech many sound features or whole sounds are slurred or omitted. As far as I have understood Mr. *Mol's* paper, he seems to conclude from this fact that we do not perceive speech in terms of phonemes. But I am not sure that this conclusion is tenable; it depends on what is meant by perception. It is probably useful to distinguish several levels of perception. We do not hear (with our ears) sounds that are not there, but we may nevertheless at a higher level interpret what we hear as a sequence of phonemes.

Hammarström (Uppsala): According to your definition "phonetic reality" can be obtained through instruments giving "articulatory" and acoustic data. I would rather suggest a definition that includes auditory data. There is no essential difference between the different ways of obtaining data. The phonetician can measure with an acoustic instrument, let us say, the length of two sounds *a* and *b*, but he can also in a very comparable way let a hearer react to the two sounds and ask him to say for instance which of them is longer.

Vachek (Praha): Mr. *Mol* seems to overlook the fact of different functions of the spoken and written norms. The latter is detatched from the extralingual reality with which the former is very closely tied; this is the reason why the former may have so much redundant elements dropped. But in case of misunderstanding we always refer to the basic, "ideal" form. Notice also *Sapirs* reference to "réalité psychologique de phonèmes": the native informants, quite ignorant of alphabetic (or any writing) give information to the linguist not in terms of sounds but in terms of phonemes.

Pohl (Bruxelles): On pourrait comparer la chaîne parlée à un accordéon. Quand la *situation* est suffisamment explicite, on peut impunément serrer l'«accordéon», quand la situation fait défaut, il est, au contraire, nécessaire de l'étendre.

J'ai annoncé des expériences sur le degré de compréhension des chiens, mais j'ai dû assez vite m'arrêter, pour de nombreuses raisons et particulièrement à cause de la quantité considérable de «matériel animal» qui aurait été nécessaire. De toute façon, en essayant le plus possible d'éviter les deux écueils du *conditionnement* et de l'*expressivité*, j'ai cherché dans quelle mesure on pouvait déformer le nom du chien sans qu'il cesse de le connaître. A première vue, et sans préjudice de ce que permettraient de conclure des recherches approfondies, il m'a semblé que l'on pouvait soumettre le «cynonyme» à des déformations et à des amputations assez importantes sans qu'il perde sa vertu communicative. Bien entendu, on ne saurait parler ici d'influence de la langue écrite.

Verh. 5. int. Kongr. Phon. Wiss., Münster 1964, pp. 431–436
(S. Karger, Basel/New York 1965).

Kansai-Universität, Osaka

Die japanischen Phoneme

Von Kobun Naito, Osaka

Phonetisch und phonemisch ist die japanische Sprache verhältnismäßig einfach. Aber da es unmöglich ist, in der kurzen Zeit auf die Einzelheiten einzugehen, werde ich in der Hauptsache nur die japanischen Vokalphoneme erörtern. Fundamentale Vokale hat das Japanische nur fünf: /i, e, a, o, u/. Doch sind alle diese fünf Vokale vervierfacht, nämlich:

kurze Vokale	/i	e	a	o	u/
lange Vokale	/i:	e:	a:	o:	u:/
Vokale mit nasaler Endung	/i˜	e˜	a˜	o˜	u˜/
Vokale mit aufhaltender Endung	/i′	e′	a′	o′	u′/

Die kurzen und die langen Vokale bedürfen wohl keiner Erklärung. Ich werde deshalb nur die Vokale mit nasaler und die mit aufhaltender Endung erläutern.

Die Vokale mit nasaler Endung sind den französischen Nasalvokalen etwas ähnlich, aber nicht gleich. Die französischen Nasalvokale sind einheitlich nasal, d. h. vom Anfang bis zum Ende gleichmäßig nasal, die japanischen Vokale mit nasaler Endung dagegen fangen eigentlich nichtnasal an und enden immer nasal. Sie sind sozusagen Diphthonge.

Die amerikanischen Phonemizisten betrachten gewöhnlich einen Diphthong als zwei aufeinanderfolgende Phoneme, die meisten europäischen Phonetiker aber betrachten einen Diphthong als ein einziges Phonem. Den japanischen Vokal mit nasaler Endung kann man ebenfalls als zwei Phoneme oder als ein Phonem betrachten. Wenn man ihn als zwei Phoneme ansieht, so ist der erste Bestandteil ein einfacher kurzer Vokal und der letzte Bestandteil etwas Nasales. Und dieses Etwas ändert sich je nach den folgenden

Lauten. Es wird bald vokalisch, bald konsonantisch, indem es doch immer nasal bleibt. Das Wort *hon* zum Beispiel, welches «Buch» bedeutet, spricht man aus wie [hoŋ] oder [hoũ], wenn es allein steht. Aber wenn ihm ein anderes Wort folgt, so ändert es sich folgendermaßen:

hon o [hoũo, hõ:o]	das Buch (Akkusativ)
hon sae [hoũsae, hõ:sae]	selbst das Buch
hon e [hoĩe, hõ:e]	zum Buch
hon ga [hoŋŋa]	das Buch (Nominativ)
hon kara [hoŋkara]	vom Buch
hon no [honno]	des Buches
hontate [hontate]	die Bücherstütze
hon mo [hommo]	auch das Buch
honbako [hombako]	der Bücherschrank

Die gewöhnlichen Konsonanten kommen im Japanischen immer vor einem Vokal vor, d. h. alle japanischen eigentlichen Konsonanten sind prävokalisch. Infolgedessen müssen alle japanischen Wörter auf einen Vokal enden. Die obenerwähnte nasale Endung wird zwar manchmal oder sogar oft konsonantisch, aber sie ist im Japanischen wesentlich verschieden von den eigentlichen prävokalischen Konsonanten. Im Deutschen, zum Beispiel in dem Wort «nun» sind der Anlaut und der Auslaut ein und dasselbe Phonem /n/. Aber im Japanischen ist das /n/-Phonem immer prävokalisch. [n]-Laut vor einem Konsonanten mit derselben Artikulationsstelle, zum Beispiel [n] in *hontate*, ist eine Variante der nasalen Vokalendung. Das japanische /n/-Phonem hat immer einen Abglitt, off-glide, zu einem Vokal. Dieser Abglitt macht den Eindruck, der diesem Phonem eigen ist und es von den anderen Nasallauten abhebt. Aber der [n]-Laut ohne Abglitt, ohne dieses Kennzeichen, gehört im Japanischen nicht zum /n/-Phonem, sondern er ist eine Variante der nasalen Vokalendung. Das gleiche gilt auch von [m] und [ŋ] (Agma). Der häufigste Fehler der Japaner in der deutschen oder englischen Aussprache ist, daß sie den Auslaut [n] wie die japanische nasale Vokalendung aussprechen.

Die nasale Vokalendung hat viele Varianten, sie ist sehr variabel. Und mir scheint es ungeeignet, so ein Variables als ein selbständiges Phonem anzusehen. Es wäre geeigneter, den Vokal mit nasaler Endung nicht als zwei aufeinanderfolgende Phoneme, sondern als ein einziges Phonem anzusehen.

Was von dem Vokal mit nasaler Endung gesagt worden ist, gilt im allgemeinen auch von dem Vokal mit aufhaltender Endung. Die aufhaltende Vokalendung ist irgendein stimmloser Verschlußlaut ohne Explosion. Kein ordentliches japanisches Wort endet auf die aufhaltende Vokalendung. Aber gelegentlich, wie beim Geschrei, kommt es vor, daß ein Wort auf die aufhaltende Vokalendung endet. In solchem Fall gebraucht man gewöhnlich den Glottisschlag, Stimmritzenverschluß, wie zum Beispiel [aʔ] (ach!). Bei den ordentlichen Wörtern kommt die aufhaltende Vokalendung immer vor stimmlosen Verschluß- und Zischlauten vor, und durch Assimilation bekommt sie dieselbe Artikulationsstelle wie die des folgenden Lautes. Dadurch entsteht die Gemination, wie zum Beispiel *atta* (war), *makka* (hochrot), *appare* (lobenswert), *assari* (einfach). In der Rechtschreibung folgen zwei gleiche Schriftzeichen aufeinander, aber nach phonemischer Auffassung wäre, im Japanischen, das erste als die Endung des vorangehenden Vokals anzusehen. Es hat eigentlich keinen Laut, sondern bedeutet eine Pause. Der Vokal mit aufhaltender Endung ist, mit anderen Worten, der Vokal mit einer Pause nach sich. Zwar ist das *s* in *assari* zum Beispiel nicht vollkommen stumm, aber das erste abglittlose *s* gilt hier für eine Pause.

Die japanische Sprache hat eine rhythmische oder Zeiteinheit, die man als Mora bezeichnen könnte. Die Silbe, deren Träger ein kurzer Vokal ist, gilt für eine Mora, und die Silbe, deren Träger ein langer Vokal oder ein Vokal mit nasaler oder aufhaltender Endung ist, gilt für zwei Moren. Also gilt bei der Silbe, deren Träger ein Vokal mit aufhaltender Endung ist, die Pause für eine Mora.

Wie vorher gesagt, sind die eigentlichen japanischen Konsonanten immer prävokalisch. Japaner sprechen von dem «geraden» und dem «krummen Laut». Aber es handelt sich hier eigentlich nicht sowohl um den Laut als vielmehr um die Verbindung der prävokalischen Konsonanten mit dem folgenden Vokal. Nehmen wir als Beispiel die Verbindungen des Konsonanten *k* mit den Vokalen: *ki*, *ke*, *ka*, *ko*, *ku*. Die Artikulationsstelle des *k* verschiebt sich je nach den folgenden Vokalen. Vor *i* ist sie beträchtlich vorwärts verschoben. Mit anderen Worten ist *k* vor *i* palatalisiert. Im Japanischen ist jeder Konsonant vor *i* immer mehr oder weniger palatalisiert. Das ist im Japanischen sozusagen natürlich. So sind die Verbindungen *ki*, *ke*, *ka*, *ko*, *ku* sogenannte gerade Laute. Aber wenn dieses palatalisierte *k* vor *a*, *o*, *u* vorkommt, so entstehen die sogenannten krummen Laute: *kya*, *kyo*, *kyu*. Orthographisch schreiben wir gewöhnlich *ky*, aber dieses *y* gilt für keinen Vokal noch Halbvokal, sondern es gilt bloß als ein Palatalisierungszeichen. *k* und *y* zusammen steht für ein palatalisiertes *k*. Wenn ein Ausländer zum Beispiel das Wort *Kyôto*, den Namen der japanischen alten Stadt, ausspricht, so klingt es gewöhnlich [kĭo:to], und dieses [kĭo:] macht auf das japanische Ohr den Eindruck von drei Moren, weil diese Silbe einen Halbvokal und einen langen Vokal enthält. Aber *kyô* gilt im Japanischen für zwei Moren, weil es nur einen langen Vokal enthält. So sind im Japanischen das nichtpalatalisierte und das palatalisierte *k* zwei verschiedene Phoneme.

Was von dem Konsonanten *k* gesagt worden ist, gilt auch von den meisten anderen Konsonanten. In der Regel hat jeder nichtpalatalisierte Konsonant ein entsprechendes palatalisiertes Gegenbild. In dieser Hinsicht ist die japanische Sprache den slawischen Sprachen ähnlich.

Japanische Konsonanten

Nichtpalatalisierte Konsonanten kommen vor *e*, *a*, *o*, *u* vor.
Palatalisierte Konsonanten kommen vor *i*, *a*, *o*, *u* vor.

nichtpalatalisiert	palatalisiert
/h/ (Variante [φ] vor *u*)	/h$_j$/ [ç]
/k/	/k$_j$/ [c]
/g/	/g$_j$/ [ɟ]
/s/	/s$_j$/ [ʃ,]*
/t/ (Variante [ts] vor *u*)	/ts$_j$/ [tʃ,]*
/d/ (nur vor *e*, *a*, *o*) } /dz/ }	/dz$_j$/ [dʒ,]*
/n/	/n$_j$/
/r/	/r$_j$/
/p/	/p$_j$/
/b/	/b$_j$/
/m/	/m$_j$/

/w/ kommt nur vor *a* vor.
/j/ kommt nur vor *a*, *o*, *u* vor.
[ŋ] und [ŋ$_j$] sind Varianten von /g/ bzw. /g$_j$/.
[z] und [ʒ,] sind Varianten von /dz/ bzw. /dz$_j$/.

* [ʃ] und [ʒ] bedeuten hier palatalisiertes [s] bzw. palatalisiertes [z].

Adresse des Autors: Prof. Kobun Naito, Kansei Universität, Fakultät d. Literatur, *Osaka Suita Senriyama* (Japan).

Discussion

Woodhead (Leeds): Whether in Japanese two vowels together – especially if they are the same – i.e. a phonetically long vowel – are better treated as a succession of two phonemes or as diphthongs. The fact that, in general, either can bear a high tone, might be evidence for treating tone as a syllable feature and each vowel element as a separate syllable.

Onishi (Tokyo): I think I have to give some comment on the "Japanische palatalisierte Konsonanten" of this speaker. He says that Japanese syllables, so-called "der krumme Laut" different from "dem geraden Laut", *kya*, *kyu*, and *kyo* are constructed, as if, with palatalized k plus vowel.

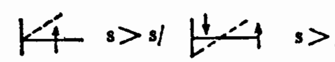
In this case, I, of course, admit that k may be palatalized in some degree; but the essential nature of this syllable is not dependent on its inclination of the consonant, but do exist on the insertion of a glide [j]; that is [kja], or may be [k$_j$ja].

In general, the definition of palatalization is the first problem. There may be two kinds of understanding palatalization. One is the strict or narrow sense of palatalization, and the other is the loose or broad sense of palatalization. And in this paper, Prof. *Naito* seems to take the latter. But, as far as he sticks to the *palatalization*, this is the matter of "*phonetisch*" and not of "*phonemisch*".

Then, I like to point out the definition, adopted and made public, by the Notation Committee of the International Phonetic Society, at Copenhagen, pre-war time, which was based on the strict sense of palatalization. That is, for example, to change [s] into [s$_j$] by palatalization, one fundamental point must be kept at the same position and the other part moves toward the palate.

While, the shift of [s] for [ʃ] should not be called palatalization, but palatals; that means another phone. The tongue-front is raised, but the tongue-tip is lowered.

To symbolize this in illustration:



So, those two can be said to belong to entirely different phones, as well as different phonemes, i.e. [s$_j$$^{1-n}$] against [ʃ$^{1-n}$].

Fujimura (Stockholm): If you consider a long vowel like [o:] in [o:ʀi:] (big) as a phoneme different from the short vowels, what is the reason to consider a diphthong like [aɪ] in [aɪsuru] (to love) as a succession of two phonemes? In consideration of the similarity in their manifestations in regard to the pitch change, for example, it seems to me more natural to treat the two cases in the same fashion.

Neustupný (Prag): Ich werde mich nur auf einige allgemeine methodische Bemerkungen beschränken:

1. Man kann leider nicht sagen, daß es heute wenige Schreibungen des japanischen phonologischen Systems gäbe. Es gibt viele, unter denen man wenigstens die Versuche von *Polivanov*, *Block* und *Hattori* erwähnen muß.

Daß viele verschiedene Beschreibungen ein und derselben Sprache vorhanden sind, ist nicht nur für das Japanische charakteristisch: es ist auch keine Spezialität der japanischen Sprache, daß die verschiedenen Autoren ganz verschiedene Systeme präsentieren, mit verschiedener Zahl und Struktur der phonologischen Einheiten.

2. Ich möchte nun die Frage stellen, ob nicht schon die Zeit da ist, einen Versuch zu machen, um die verschiedensten Systeme zusammenzubringen und zu vereinigen. Ähnlich wie in allen anderen Sprachen der Welt gibt es nämlich auch im Japanischen nur *ein* phonologisches System, das wir als solches entdecken müssen.

(1) Die erste Frage ist, ob wir immer Recht haben, wenn wir das Problem der Einphonemigkeit oder Zweiphonemigkeit, oder das Problem der Zugehörigkeit einer

Variante zu einem Phonem, ganz klar und eindeutig entscheiden wollen. Das ist z. B. der Fall bei den japanischen Sequenzen V + Nasalisation oder C + Palatalisierung + V.

Bisher hat man nämlich meistens festgestellt:

(a) es sind zwei Phoneme,

(b) es ist ein Phonem.

In Wirklichkeit gibt es jedoch verschiedene *Grade* der Einphonemigkeit und der Zweiphonemigkeit.

(2) Ich glaube, daß es wichtig ist, bei der phonologischen Beschreibung einer Sprache den *typologischen* Charakter der Sprache in Betracht zu ziehen, weil auch er auf die Struktur des phonologischen Systems Einfluß hat.

In einer flexiven Sprache ist die Gliederung einer Sequenz in Phoneme sehr klar; in einer agglutinativen Sprache ist die Grenze zwischen Monophonem und Biphonem nicht immer klar. Das ist der Fall der palatalisierten Konsonanten, der geminierten (langen) Konsonanten, der langen Vokale, der nasalisierten Vokale usw. im Japanischen.

(3) Wir wissen noch zu wenig von dem phonetischen Bau der japanischen Sprache, um die unklaren Fälle entscheiden zu können, wenn sie entscheidbar sind (und das ist, meiner Meinung nach, fast immer der Fall). Verhältnismäßig viel kennen wir von der japanischen Nasalierung (*Hattori*).

Proc. 5th int. Congr. phon. Sci., Münster 1964, pp. 437–440
(S. Karger, Basel/New York 1965).

Arabic Vowels and Vocoids
Their Characteristics and Distribution

By Raja T. Nasr, Beirut

Basic Assumptions

Two basic assumptions must be made at the outset of this paper. The first assumption is that Arabic vowels[1] and vocoids[2] are those sounds which, as a group, are produced with the least amount of friction and/or constriction. The second assumption is that velarization in Arabic phonology is phonemic with respect to the consonants, but phonetic (allophonic) with respect to the vowels.

Characteristics of Vowels and Vocoids

The Arabic vowels[3] are six in number. No glides are included in these vowels. Three of them, however, are geminates (clusters) of the same vowel. In addition to the six vowels, there are two diphthongs (combinations of different vowels). The major vocoids are seventeen in number.

Vowels	*Vocoids*	*Characteristics and examples*
/ii/	[ii]	High front close unrounded. /fiil/ [fiil] "elephant".
	[ïï]	High back close unrounded. /Siin/[4] [Sïïn] "China".
/i/	[i]	High front open unrounded. /min/ [min] "from".
	[ï]	High back open unrounded. /Sif/ [Sïf] "describe" (imp.).

[1] Phonemic term.

[2] Phonetic term.

[3] These vowels occur in most Literary Arabic dialects, although their distribution in lexical items may be different. Some colloquial dialects have eight or nine vowels.

[4] Capitals indicate velarization.

Vowels	*Vocoids*	*Characteristics and examples*
	[ɛ]	Mid front open unrounded. /bilmadrasati/ [bilmædræsæti(ɛ)] "in the school".
	[ɛ̈]	Mid back open unrounded. /biSSiini/ [bïSSïïnï(ɛ̈)] "in Chinese".
/aa/	[ææ]	Low front close unrounded. /laa/ [lææ] "no".
	[aa]	Low central open unrounded. /SaaRa/ [SaaRə] "he became".
/a/	[æ]	Low front close unrounded. /lam/ [læm] "not".
	[ə]	Mid central close unrounded. /Saffun/ [Səffun] "class".
/uu/	[uu]	High back close rounded. /suuqun/ [suuqun] "market".
/u/	[u]	High back open rounded. /huna/ [hunæ] "here".
	[o]	Mid back close rounded. /lahu/ [læhu(o)] "for him".
/ai/	[æi]	Combination of [æ] and [i]. /baitun/ [bæitun] "house".
	[əï]	Combination of [ə] and [ï]. /baiDun/ [bəïDun] "eggs".
/au/	[æu]	Combination of [æ] and [u]. /lau/ [læu] "if".
	[əu]	Combination of [ə] and [u]. Dauʔun/ [Dəuʔun] "light".

Distribution of Vowels and Vocoids

No vowel occurs in word initial position; if no other consonant is found preceding the first vowel, the glottal stop /ʔ/ will occur there. The distribution of the vowels and vocoids in medial and final positions is given below:

Vowels	*Vocoids*	*Distribution*
/ii/	[ii], [ïï]	These vocoids occur in complementary distribution, [ïï] appearing

Vowels	*Vocoids*	*Distribution*
		contiguous to velarized consonants and [ii] appearing elsewhere.
/i/	[i], [ï]	These vocoids occur in complementary distribution, [ï] appearing contiguous to velarized consonants and [i] appearing elsewhere.
	[ɛ], [ɛ̈]	These vocoids occur in free variation with [i] and [ï] respectively in final positions only.
/aa/	[ææ], [aa]	These vocoids occur in complementary distribution, [aa] appearing contiguous to velarized consonants and [ææ] appearing elsewhere.
/a/	[æ], [ə]	These vocoids occur in complementary distribution, [ə] appearing contiguous to velarized consonants and [æ] appearing elsewhere.
/uu/	[uu]	This vocoid occurs in all medial and final positions.
/u/	[u], [o]	[u] occurs in all medial and final positions; [o] occurs in free variation with [u] in final positions only.
/ai/	[æi], [əï]	These vocoids occur in complementary distribution, [əï] appearing contiguous to velarized consonants and [æi] appearing elsewhere.
/au/	[æu], [əu]	These vocoids occur in complementary distribution, [əu] appearing contiguous to velarized consonants and [æu] appearing elsewhere.

Conclusion

The characteristics and distribution of Arabic vowels and vocoids are illustrative of the realization of the phoneme. An understanding of the phoneme and the way it functions in language will aid analysts in the understanding of Arabic vowels and vocoids.

Also, conversely, the understanding of Arabic vowels and vocoids and their distribution will reveal the significance of the phoneme as the most basic functional unit in language.

Author's address: Prof. Dr. R. T. Nasr, Department of Education, Beirut College of Women, *Beirut* (Lebanon).

Verh. 5. int. Kongr. Phon. Wiss., Münster 1964, pp. 441–445
(S. Karger, Basel/New York 1965).

Das Phonem und seine Realisierung in der Kindersprache

Von Karel Ohnesorg, Brno

Zahlreiche neue Arbeiten, die in den letzten Jahrzehnten erschienen und der Kindersprache gewidmet sind, bestätigen die Richtigkeit des Gedankens, den *Leopold Stein* bereits 1925 zum Ausdruck gebracht hatte: «Lautänderungen, die phonetisch absurd erscheinen, oder die in der historischen Grammatik nicht eine Parallele finden, sind mir bisher nicht zur Beobachtung gekommen[1].» Die meisten Untersuchungen verzeichneten bisher die phonetischen Realisierungen in der Kindersprache, bzw. man verglich diese Realisierungen mit den Lautänderungen in der historischen Entwicklung der Sprachen oder mit jenen Tendenzen, die sich in der Realisierung der Umgangssprache der Gegenwart kundgeben. Verschiedene Realisierungen der Kindersprache werden in der Regel als Folge einer Ungeschicklichkeit der Sprachorgane bzw. einer artikulatorischen «Widerspenstigkeit» der Laute erklärt, mit Rücksicht auf das «principe du moindre effort». Dies ist zweifellos ein wichtiges Moment; eine rein phonetische Deutung kann jedoch nicht zur restlosen Lösung der Problematik führen. Häufig lehnen ja die Kinder eine Nachahmung ihrer eigenen (d. h. deformierten) Aussprache selbst ab und empfinden eine solche Aussprache als unrichtig, was uns beweist, daß bei ihnen ein phonologisches Sprachbewußtsein bereits entwickelt ist, und daß sie überzeugt sind, daß ihre Realisierungen der Phoneme mit den Realisierungen in der Sprache der Erwachsenen identisch sind.

Neue pedolinguistische (und pedophonetische) Arbeiten werden also die Frage lösen müssen, in welcher Entwicklungsetappe die Kinder das phonologische System und die phonologische Struktur

[1] 9, S. 106.

der Sprache wahrzunehmen beginnen, wie dies schon 1942 *Roman Jakobson* erörtert hatte[2]. Denn wie die Kinder die Sprache verstehen, bevor sie sie selbst zu gebrauchen wissen, so sind sie auch imstande, die phonologische Struktur der Sprache wahrzunehmen, bevor sie noch die einzelnen Phoneme richtig zu realisieren beginnen. Die Aneignung der Sprache bei Kindern ist nämlich keinesfalls nur eine mechanische Nachahmung der Sprachklänge der Umwelt. Wie *T. Slama-Cazacu* richtig gezeigt hatte, ist «das Kind ein Echo, aber ein selektives Echo»[3]. Das Kind ahmt nicht mechanisch nach, sondern wählt gewisse Elemente, die es (freilich im Rahmen seiner Fähigkeiten) nachahmt.

Nun können wir fest überzeugt sein, daß das erste vom Kinde wahrgenommene Sprachelement die Sprachmelodie ist. Einen zwingenden Beweis dafür brachte schon der häufig erwähnte Versuch *Tappolets:* die Reaktion des Kindes auf die Fragen «Où est la fenêtre?» und «Wo ist das Fenster?» war dieselbe, wenn die melodische Modulierung der beiden Fragen übereinstimmend war[4]. Mit Recht bezeichnet daher auch *L. Kaczmarek*[5] das erste Entwicklungsstadium der Kindersprache als das «Stadium der Melodie». Die Intonationen bilden das erste funktionelle Element in der Kindersprache und unterscheiden die Bedeutung der ersten ein- oder zweisilbigen Wörter des Kindes.

Erst später fängt das Kind an, vokalische Phoneme zu unterscheiden, deren Realisierungen zuerst nicht weniger labil als die der Konsonanten sind und eine gewisse Zeitspanne der Fixierung voraussetzen. Dabei kann angenommen werden, daß als erster der Vokal *a*, später die Vokale *e* und *o*, weiter die hohen Vokale *i*, *u* und zuletzt die für eine bestimmte Sprache charakteristischen Vokale (wie z. B. die Nasalvokale im Französischen) fixiert werden. Neue Arbeiten bestätigen hier völlig die Idee *R. Jakobsons* (*l. c.*), daß in der Kindersprache diejenigen Phoneme am spätesten auftreten, die für eine bestimmte Sprache charakteristisch sind.

Von den Konsonanten realisieren die Kinder die Verschlußlaute früher als die Engelaute, die sie durch akustisch oder artikulatorisch verwandte Okklusive ersetzen (*eten* – essen, *mato* – maso). Die Okklusive *k* und *g* werden dann bei den meisten Kindern

[2] 4.
[3] 8.
[4] 10; cf. 1, S. 317: «*gdzie oczko?*» – «*gdzie okno?*».
[5] 6.

zuerst als *t* und *d* realisiert[6]. Nachdem sich dann die Kinder eine richtige Bildung der velaren Okklusive angeeignet haben, realisieren sie diese manchmal auch dort, wo *t* und *d* stehen sollte (*gukn kak* – guten Tag, *koklet* – côtelette, *likl* – little, *kank* – tank). Das phonologische Bewußtsein entwickelt sich also bei diesem Konsonantenpaar später als bei den übrigen Verschlußlauten.

Sehr häufig werden auch stimmhafte und stimmlose Konsonanten verwechselt, und zwar selbst in jenen Sprachen, in denen dieser Gegensatz phonologisch relevant ist. Stimmlose Konsonanten werden in der Regel früher fixiert als die stimmhaften; die Kinder setzen auch häufiger stimmlose Konsonanten für stimmhafte als umgekehrt (*paum* – Baum, *kato* – gateau, *sima* – zima).

Am spätesten realisieren die Kinder den Unterschied zwischen den Liquiden, *l* und dem (artikulatorisch anspruchsvollen) *r*. Nachdem sie nämlich eine richtige Realisierung des *r* erreicht haben, gebrauchen sie manchmal dieses Phonem auch statt des richtigen *l* (*terefon* – telefon, *mýdro* – mýdlo)[7].

Wie bei den vokalischen Phonemen, realisieren die Kinder auch bei den Konsonanten diejenigen Phoneme am spätesten, die für eine bestimmte Sprache charakteristisch sind (z. B. das tschechische *ř*, das von tschechischen Kindern lange durch *ž* oder *z* substituiert wird).

Wenn das Kind die Artikulation bestimmter Laute noch nicht bewältigt hat, so folgt daraus keinesfalls, daß es diese Phoneme noch nicht wahrnimmt. Nur in dem allerersten Stadium hört das Kind die Sprache seiner Umwelt phonetisch treu und ahmt sie auch (nach seinen Fähigkeiten) mehr oder weniger treu nach. Mit der Entwicklung seines Intellektes beginnt es jedoch bald die Sprache mittels eines «phonologischen Siebes» wahrzunehmen, wie unrichtig oder labil seine eigenen Realisierungen der Phoneme auch sein mögen. Abweichende Realisierungen der Phoneme in der Kindersprache brauchen also keinesfalls nur eine Folge artikulatorischer Ungeschicklichkeit zu sein, sondern sie können auch durch eine falsche phonologische Wertung des Sprachbilds verursacht werden. Und wenn einerseits bekannt ist, daß manche mental wenig entwickelte Kinder die Sprachklänge ihrer Umwelt (ja sogar Sprachäußerungen in einer ihnen unbekannten Sprache) sehr genau (echolalisch) nachahmen können, so begegnen wir ander-

[6] 5, S. 85: «In allen Ländern haben die Kinder den Hang, *t* für *k* zu setzen.»
[7] 2, S. 361.

seits solchen Fällen, daß Kinder von hochentwickelter Intelligenz an Dyslalie leiden. (Auf diese Tatsache hat schon *H. Gutzmann sen.* hingewiesen[8].)

Alle bisherigen Untersuchungen bestätigen in vollem Umfang die einleitend angeführte Idee *L. Steins*. In konkreten Einzelfällen (z. B. in der Therapie der Dyslalie) wird man jedoch immer zuerst feststellen müssen, ob es sich um eine artikulatorische Ungeschicklichkeit bzw. schlechte Gewohnheit, oder aber um eine falsche Wahrnehmung und Wertung der oder jener Phoneme handelt.

Literatur

1. *Brenstiern Pfanhauser, St.:* Rozwój mowy dziecka. Prace filologiczne *15*, Warszawa 1930.
2. *Chvatcev, M. E.:* Logopedia. 5. Aufl. Učpedgia, Moskva 1959.
3. *Gutzmann, H.:* Die Sprachheilkunde. 3. Aufl. H. Kornfeld, Berlin 1924.
4. *Jakobson, R.:* Kindersprache, Aphasie und allgemeine Lautgesetze (Universität Uppsala 1941).
5. *Jespersen, O.:* Die Sprache. (Carl Winter, Heidelberg 1925).
6. *Kaczmarek, L.:* Kształtowanie si mowy dziecka (Towarzystwo przyjaciol nauk, Poznań 1953).
7. *Luchsinger, R.:* Über die Beziehungen der Sprache und Sprachstörungen zur sogenannten Feinmotorik. Fol. Phon. *1:* 206–224 (1948).
8. *Slama-Cazacu, T.:* Relatiile dintre gîndire si limbaj în ontogeneza (Academia, Bucuresti 1957).
9. *Stein, L.:* Das universelle Stammeln im Lichte der vergleichenden Sprachwissenschaft. Z. f. ges. Neurol. u. Psychiatrie *45:* 100–107 (1925).
10. *Tappolet, E.:* Die Sprache des Kindes (Basler Aula-Verlag, Basel 1907). – Deutsche Rundschau. Bd. 131, Berlin 1907, S. 399–411.

Adresse des Autors: Prof. Dr. Karel Ohnesorg, Universita J. E. Purkyne, Arna Nováka 1, *Brno* (ČSSR).

Discussion

Trojan (Wien): Der Vortrag von Herrn *Ohnesorg* scheint mir ein wichtiger Beitrag zu der Frage der Gültigkeit der Theorie von *R. Jakobson* vom Schichtenbau der Sprachlautsysteme zu sein. Ich erinnere hier an die Einwände von *Abrahams* und *Martinet*. Ich selbst habe während meiner langjährigen und noch fortdauernden Tätigkeit als Leiter der Sprachambulanz an der II. HNO-Klinik in Wien immer wieder die Gültigkeit der Theorie bestätigen können. Allerdings möchte ich betonen, daß es nicht auf die Reihenfolge des Lauterwerbes in der Ontogenese der Sprache ankommt, sondern auf die Tatsache, daß Dyslalien, wie ich auch in einem in der «Phonetica» erschienenen Aufsatz in einer kleinen Statistik gezeigt habe, fast ausschließlich die «späten» Laute im Sinne *Jakobson* betreffen. Man möchte auf Grund einer internationalen Enquete diese Frage zu klären suchen; das wird eine Hauptaufgabe der Biophonetischen Gesellschaft sein.

[8] 3, 7.

von Raffler Engel (Florenz): Wie Herr *Trojan* habe auch ich das Erscheinen von *i* und *u* vor *e* und *o* beobachtet. Ich glaube, es müßte klar festgelegt werden, von wann wir den ersten Vokal datieren, von den Lallauten oder dem ersten Wort?

Was das Erscheinen der Verschlußlaute vor den Engelauten betrifft, so haben meine Beobachtungen an normalen Kindern *Jakobson* nicht bestätigt: Zwei italienische Kinder sagten als erstes Wort *ča*, 'ciao'. Ein Kind sagte sogar *šiši* an Stelle von 'pipi'.

Antwort *Ohnesorg: šiši* ist onomatopäisch, und die Onomatopäen müssen gesondert betrachtet werden.

Ich bin nicht ganz sicher, daß bei jenem Kind *šiši* eine Onomatopäe war. Das Problem der Onomatopäen ist äußerst kompliziert, und wir brauchen hier die Mitarbeit der Psychologen. Mein Sohn sagte seine erste Onomatopäe, als er schon 13 Monate alt war: Ein Auto mit Sirene fuhr auf der Straße vorbei, und er machte das Geräusch nach, um mich zu fragen, was so ein Auto sei. Die Nachahmung war ein Hilfsmittel, sie war nicht als ein neues Wort gedacht.

Höffe (Dortmund): Bei all unseren Äußerungen zur Sprachmelodie sollten wir uns immer vergegenwärtigen, daß es sich hierbei nie um eine isoliert verlaufende Tonhöhenbewegung, sondern immer um eine in bestimmter Weise struktuierte Schallgestalt handelt. Damit will gesagt sein, daß ein und derselbe Melodieverlauf z. B. mit unterschiedlichem Schalldruck gesprochen eine andersartige Klangform und damit einen völlig andersartigen Ausdrucksgehalt bedingt.

Janota (Prag): Den eben erwähnten Zusammenhang zwischen Tonhöhenverlauf und Klangfarbe haben wir auch bei unseren Versuchen mit der individuellen Klangfarbe feststellen können. Wir haben unseren Hörern u. a. isolierte Tonhöhenverläufe mit einer konstanten Klangfarbe vorgelegt. Der Prozentsatz der richtigen Identifizierungen des Sprechers war sehr niedrig.

Proc. 5th int. Congr. phon. Sci., Münster 1964, pp. 446–449
(S. Karger, Basel/New York 1965).

On the Roentgenography of the Speech Organs

By Jana Ondráčková, Praha

Roentgenography of the speech organs has a relatively long tradition. We shall attempt to demonstrate its advantages for phonetics.

Roentgenography enables investigation of the articulatory formation of the speech organs not available to direct observation. As opposed to previous X-ray techniques using the usual variety of contrast materials – most frequently barium – native X-ray shots have gradually taken over the field – i.e. roentgenography without the use of any contrast material[1]. This excludes the disturbing effect of extraneous material on mechanics and the psyche when present in the superglottal cavities, interfering with the normal function of the speech organs. In connexion with the method of complex registration, i.e. synchronous sound and oscillographic recordings during exposure of the X-ray picture, results in research on the articulatory function of the speech organs are available which enable one to reproduce the sound simultaneously like on the original realisation[2].

In research on the mutual correlation between physiologic articulatory processes and their acoustic sequelae it is necessary to use methods which record the motion of speech organs and sound synchronously and continuously in the natural formation. On analysis of the roentgenograms it is then possible to compare corresponding segments from the point of view of articulation and the acoustic resultant and the auditive action from these motor acts. To this end one can use at least in part for phonetic purposes rapid roentgen

[1] A natural contrast medium is the air present in the superglottal cavities. For details see *Ondráčková, J.* and *Poch, R.* (8).

[2] The method of complex recording is described in the work of *Ondráčková, J.* and *Poch, R.* (7).

seriography with native and complex recording methods. Seriograms enable the observation not even of continuous course of motion of the speech organs but the separate phases of these motions, which, at a speed of 4 pictures/sec., give a satisfactory view of the motion of their soft parts during continuous speech.

Application of this method has enabled us to investigate which acoustic effects are associated with transitional articulatory phenomena[3] on roentgenographic (seriographic) records, and on the other hand the corresponding motor phase of the speech organs corresponds to some sort of transitional acoustic phenomenon, determined, for example, by the *Janota* segmentor[4]. In this way it can be determined which of the articulatory phenomena, acoustically manifest, are relevant for the identification of the given sound from the point of view of the hearer. The determination of these relationships is important because articulation and acoustics form an entity from the point of view of both speaker and hearer.

A further contribution to phonetic research of the action of the speech organs, quite widespread today, is sound X-ray cinematography[5]. With adequate technique native X-ray films of the natural motions of the soft parts of the speech organs and their transitional phases between sounds can be used at various filming speeds[6].

Analysis of single shots from the X-ray sound film is carried out by means of loops projected by a special adaptor for sound projection[7]. They have contributed considerably to more complete understanding of the interplay of the soft parts of the speech organs in the superglottal cavities.

From analysis of X-ray cinematographic sound records it would appear that it is not of advantage to use the term "position"

[3] Transitional phenomena from the point of view of acoustics, see for example: Speech Communication Seminar (11); *Carnochan, J.* and *Skaličková, A.* (1); *Romportl, M.* (12).

Phenomena previously labelled as transitional could be investigated only after use of adequate registration methods for acoustic and articulatory phenomena.

[4] *Janota, P.* and *Romportl, M.* (2).

[5] Films are usually presented at sittings of the Association internationale du cinéma scientifique and are listed in their catalogue.

[6] For example, *Shelton, R. L.* et al. (9); *Ondráčková, J.* (5).

[7] *Ondráčková, J.* (6).

Loops enable random uninterrupted repetition of the same phenomenon, so that characteristic signs of motion of speech organs can be identified (method of comparative analysis).

or "maximum articulation position" of the speech organs, since the soft parts are practically in continuous motion even when phonation does not occur. Not even the so-called maximum articulation phase, previously understood as the "maximum articulation position", need always be relevant for identification of the sound[8]. It is therefore of greater advantage to observe the whole syllable than the separate sounds[9]. It is important to solve the problem of the boundary between separate sounds from the point of view of articulation and acoustics, and the problem of how the various articulatory and acoustic relevant features change in time. With reference to this one must concentrate on content when using the term "transitional phenomenon".

X-ray films can be evaluated by observation on various types of trans-illuminators. For studying roentgenograms positive copies should be used[10]. Native roentgenograms can be objectively analysed, using a photoresistance adapted for reading off from the films[11].

The promising results in phonetic research enables the application of new X-ray technique, e.g. high-voltage roentgenography[12] and rotational cinefluorography[13]. Tomography has also contributed to the solution of some special phonetic problems.

References

1. *Carnochan, J.* and *Skaličková, A.:* Notes on the comparison of English-Czech phonetics (in press).
2. *Janota, P.* und *Romportl, M.:* Ein Beitrag zur phonetischen Methodik. AUC, Philologica *1:* 17–28 (1959).
3. *McDonnel, G.; Berman, H. L.* and *Lodmell, E. A.:* Supervoltage roentgenography. Amer. J. Roentgenol. *79:* 306–320 (1958).
4. *Ondráčková, J.:* Objektivní analýzy rentgenogramů mluvidel. Věstník ČSAV *71:* 558–559 (1962).
5. *Ondráčková, J.:* The movement of the tongue and the soft palate in the singing of vowels. Folia phoniat. *13:* 99–106 (1961).
6. *Ondráčková, J.:* Zur Untersuchung der physiologischen Tätigkeit der Sprechorgane in den supraglottischen Höhlen. Folia phoniat. *16:* 161–171 (1964).
7. *Ondráčková, J.* und *Poch, R.:* Zur Methodik der Untersuchung der Vokalartikulation beim Gesang. Z. Phonetik *10:* 258–268 (1957).
8. *Ondráčková, J.* and *Poch, R.:* New roentgenographic methods in the research of the activity of articulatory organs. Indian J. Radiol. *16.* 137–150 (1962).

[8] *Romportl, M.*, p. 75 (12).
[9] *Skaličková, A.*, p. 79 (10).
[10] In the complex method the original roentgenogram is a necessary part of the material.
[11] *Ondráčková, J.* (4).
[12] *McDonnel, G. et al.* (3).
[13] *Winter, F. S.* and *Lehman, J. S.* (13).

9. *Shelton, R. L. et al.:* Filming speed in cinefluorographic speech study. J. Speech Res. *6:* 19–26 (1963).
10. *Skaličková, A.:* The Korean consonants (Rozpray ČSAV, Prague 1960).
11. Speech Communication Seminar, Stockholm, Aug. 29th to Sept. 1st, 1962: I (1963).
12. *Romportl, M.:* K akustické podstatě a klasifikaci okluzív, AUC Slavica Pragensia *IV:* 75–80 (1962).
13. *Winter, F. S.* and *Lehman, J. S.:* Rotational cinefluorography. Amer. J. Roentgenol. *82:* 120–124 (1959).

Author's address: Dr. Jana Ondráčková CSc, Fonetická laborator, Ustavu pro jazyk česky ČSAV, *Praha 1* (Czechoslovakia).

Verh. 5. int. Kongr. Phon. Wiss., Münster 1964, pp. 450–452
(S. Karger, Basel/New York 1965).

Phänomene des Konsonantenwechsels in der koreanischen Sprache

Von CHAN-KI PAK, Frankfurt a. M.

1. Eines der wichtigsten Charakteristika der koreanischen Konsonantenaussprache, auf das ich hier besonders hinweisen möchte, ist, daß jeder auslautende Konsonant nur zum Verschluß des Artikulationsorgans dient; daraus ergibt sich, daß im Koreanischen nur *ein* Konsonant nach jedem Vokal stehen kann. Zum Beispiel ist es unmöglich, Lautverbindungen wie: –apt, –akt, –art, –ast, –amt auszusprechen. Der Grund dafür ist einfach: Wenn man z.B. bei «–apt» *–ap* gesprochen hat, verschließt das auslautende *p* das Artikulationsorgan schon so vollkommen, daß kein Atemhauch mehr übrigbleibt, um noch den zweiten Konsonanten *t* zu verschließen.

2. Aus diesem Aspekt versteht sich auch das zweite Charakteristikum, daß nämlich die auslautenden Konsonanten desselben artikulierenden Organs, z.B. –p und –b, –k und –g oder –t und –d als Verschlußlaute ganz den gleichen Lautwert besitzen.

Beispiel: gap = gab (i)
gak = gag (ii)
gat = gad = gas (iii)

Das bilabiale *m* von «gam» unterscheidet sich selbstverständlich von *p* oder *b* oben (i), aber das *s* von «gas» klingt ganz gleich wie *t* oder *d* oben (iii). Denn man spricht *–s* auch so vollkommen als Verschlußlaut, daß kein nachträglicher Laut übrigbleibt, obwohl man *s–* als Anlaut genau wie in anderen Sprachen reibend artikuliert.

3. Im Gegensatz zum Japanischen, in dem kein Unterschied zwischen Auslaut- und Anlautkonsonanten erkennbar ist, gibt es im Koreanischen keine Assimilation zwischen zwei zusammengefügten

Konsonanten. (Es ist hierbei zu beachten, daß im Koreanischen eine Kombination von zwei Konsonanten nur bei *einem* auslautenden und *einem* anlautenden Konsonanten vorkommt.)

Beispiel: kan–pai → kam–pai
kat–patsu → kap–patsu } japanisch
an–machen → am–machen
an–bringen → am–bringen } deutsch

Wie schnell so eine Kombination auch ausgesprochen wird, im Koreanischen wird jeder Konsonant klar und deutlich artikuliert, so daß man niemals «kanpai» und «kampai», oder an–machen und am–machen verwechseln würde.

4. *Phänomene des Konsonantenwechsels:*

	(a)	(b)	(c)
–ab → :	–am	–aB	–ab
–ap → :	–am	–aB	–ap
–ag → :	–aŋ	–aG	–ag
–ak → :	–aŋ	–aG	–ak
–ad → :	–an	–aD	–ad
–at → :	–an	–aD	–at
–as → :	–an	–aD	–as
–az → :	–an	–aD	–az

a) Wenn ein Nasallaut danach folgt.
n) Wenn ein Vokal-Substanzwort danach folgt, z. B. –d, –t, –s, –z, werden alle zu einem gleichen Laut-D, den man als Durchschnittslaut bezeichnen könnte.
c) Wenn ein Vokal-Füllwort danach folgt.

Die obigen Phänomene sind nicht als Assimilation, sondern als Wechsel anzusehen. Gründe dafür sind: Erstens z.B. das *p* oder das *t* wechselt zum *m* oder zum *n* unabhängig von dem danachstehenden Konsonanten.

Beispiel: ap–ma → am–ma
at–na → an–na } (1)

ap–na → am–na
at–ma → an–ma } (2)

Zweitens ist das Gefühl des Sprechenden nicht assimilatorisch (d.h. der Sprechende glaubt immer, das als *m* ausgesprochene *p* oder das als *n* ausgesprochene *t* buchstäblich als *p* oder *t* ausgesprochen zu haben). Drittens haben die auslautenden Konsonanten im Koreanischen besondere Eigenschaften, wie unten erklärt werden soll.

5. *Zwei divergierende Funktionen eines Phonems*

Bei jedem Konsonanten des Koreanischen kommen die zwei divergierenden Funktionen, einerseits eine vorbereitende, das Artikulationsorgan verschließende Funktion (F I), und andererseits eine resultative, das Artikulationsorgan öffnende Funktion (F II), so deutlich und typisch vor, wie man sie wohl in keiner anderen Sprache spüren kann. Ein p-Laut zum Beispiel in Funktion I (F I) unterscheidet sich nicht von einem b-Laut in F I, weil die beiden Laute als bilabiale Verschlußlaute keine nachträgliche Behauchung haben. (Hier ist es sogar nach dem Laut nicht erlaubt, die Lippen zu lockern.) Aber wenn nach p– (F I) oder b– (F I) ein Vokal-Füllwort folgt, funktioniert erst das p oder b völlig als F II, und der Unterschied wird erkennbar. Ein anlautender p-Laut hat dagegen meistens nur F II. Darum möchte ich sagen, daß ein Phonem p im Koreanischen aus einem Phonempartikel P^1– (F I) und einem Phonempartikel P^2– (F II) besteht. Und auf diese divergierenden Eigenschaften ließen sich alle oben erwähnten Phänomene zurückführen.

Adresse des Autors: Prof. Chan-Ki Pak, «Korea University», *Seoul* (Korea) z. Zt Alexander von Humboldt-Stipendiat, Beethovenstraße 36, *6 Frankfurt a. M.* (Deutschland).

C.R. 5e Congr. int. Sci. phon., Münster 1964, pp. 453–457
(S. Karger, Basel/New York 1965).

Valeur phonologique des accents serbocroates

Par Asim Peco, Belgrade

La langue serbocroate se range parmi les langues qui compensent la simplicité de leur vocalisme par la richesse et la variété de leurs accents. Cette variété d'accent date d'une période assez éloignée du développement de cette langue et se manifeste par l'existence de deux intonations comportant aussi une double quantité. Autrement dit, le serbocroate littéraire possède deux accents d'intonation montante (long ´ et bref ̀) et deux accents d'intonation descendante (long ̑ et bref ̏). Etant donné que cette accentuation, par sa nature même et par les places qu'occupent les différents accents, présente un caractère particulier, par lequel le serbocroate se distingue non seulement parmi les langues slaves, mais aussi parmi toutes les autres langues, elle est depuis longtemps l'objet des études scientifiques. Les opinions diffèrent non seulement en ce qui concerne la nature des accents s–c., surtout lorsqu'il s'agit des accents d'intonation montante, mais aussi en ce qui concerne la pertinence phonologique de ces accents, surtout lorsqu'il s'agit des accents d'intonation descendante. C'est cela justement qui nous a porté à choisir ce thème pour notre rapport. En effet, les représentants du cercle linguistique de Prague (*Trubeckoj*, *Jakobson*) considèrent que les accents s–c. d'intonation descendante ne sont pas pertinents au point de vue phonologique. A leur avis, ces accents sont des variantes phonologiques des syllabes inaccentuées (longues ou brèves), qui ne constituent que des «Grenzsignale» dans les différents mots. D'autres savants considèrent, au contraire, que les accents s–c. d'intonation descendante sont pertinents aussi au point de vue phonologique. Il est hors de doute qu'il faut faire une distinction entre l'intonation d'une voyelle accentuée, même si elle est descendante, et l'intonation d'une voyelle inaccentuée. La première est toujours pertinente au point de vue phonologique, tandis que l'autre ne l'est pas. En outre, elles diffèrent également en ce qui concerne la durée. D'après les données dont j'ai disposé, ce rapport se manifesterait de la manière suivante. La durée moyenne était chez:

a) les voyelles brèves sous l'accent descendante	10,8 cs
b) les voyelles brèves inaccentuées	8,2 cs
c) les voyelles longues sous l'accent descendante	20,5 cs
d) les voyelles longues inaccentuée	14,2 cs

Chez *W. Appel* on trouve presque le même rapport: a) = 9,6 cs; b) = 9 cs; c) = 19,9 cs; d) = 24 cs (cf. *Wiener*, Slavistisches Jahrbuch, *I:* 56 [1950]). La différence qui existe entre nous peut s'expliquer par le rythme de discours; le caractère régional de l'objet n'est pas exclu non plus.

Dans notre exposé nous nous basons sur une telle fonction phonologique des accents serbocroates. Pour nous, donc, les accents d'intonation descendante sont pertinents phonologiquement, de même que les accents d'intonation montante, bien que les premiers sont, de règle, liés à la syllabe initiale et que les autres sont relativement libre (en effet, ceux-ci ne peuvent pas se trouver seulement sur la syllabe finale des mots).

Par conséquent, dans le serbocroate littéraire, les deux intonations et les deux quantités se manifestent comme signes distinctifs et, par suite, elles sont phonologiquement pertinentes. Les exemples suivants démontrent bien cette assertion:

I. Opposition qualitatives

a) ◡̀◡ : ◡̏◡ (montant bref : descendant bref)

jàrica (blé semé au printemps)	: *jȁrica* (jeune chèvre)
klòbūk (chapeau)	: *klȍbūk* (bulle d'eau)
pàša (pacha)	: *pȁša* (paturage)

b) ◡́◡ : ◡̑◡ (montant long : descendant long)

rávan (égal, uni)	: *rȃvan* (plaine)
Lúka (nom de femme)	: *Lȗka* (nom d'homme)
dúšē (génitif de duša, âme)	: *dȗšē* (ils etouffent)

II. Oppositions quantitatives

a) ◡̑◡ : ◡̏◡ (descendant long : descendant bref)

grȃd (ville)	: *grȁd* (grêle)
lûk (arc)	: *lȕk* (oignon)
bêg (fuite)	: *bȅg* (bey)

b) ◡́◡ : ◡̀◡ (montant long : montant bref)

Góra (nom de femme)	: *gòra* (la forêt)
Zóra (nom de femme)	: *zòra* (aurore)
Kósa (nom de femme)	: *kòsa* (cheveux, pente)

III. Oppositions des mots accentués et inaccentués

ȍko (œil)	: *oko* (autour de)
krȃj (région)	: *kraj* (auprès de)
jȏj (oie!)	: *joj* (à elle, forme enclitique)

Les oppositions d'accents peuvent se manifester également dans un autre rapport: qualitative et quantitative:

a) ◡̀◡◡ : ◡◡́◡

ìskupiti (rassembler)	: *iskúpiti* (racheter)
òbaliti (humecter de bave)	: *obáliti* (renverser)

b) ◡̏◡◡ : ◡́◡◡

krȕniti (couronner)	: *krúniti* (égrainer)
kȕpiti (rassembler)	: *kúpiti* (acheter)

c) ◡̏◡◡ : ◡◡̀◡

jȁrina (laine d'agneau)	: *jarìna* (chaleur étouffante)
nȍvine (journal)	: *novìne* (innovations, nouveautés)

ou dans le lieu des accents:

a) ◡̀◡◡ : ◡◡̀◡

màlina (framboise)	: *malìna* (die kleine Anzahl, Vuk)
srèdina (mie de pain)	: *sredìna* (centre)

b) ◡́◡◡ : ◡◡́◡

Mášina (adjectif de *Mášo*)	: *mašína* (la machine)
Édina (adjectif de *Édo*)	: *Edína* (nom de femme)

Les exemples cités nous permettent de conclure que le rapport entre une intonation et l'accent d'une même quantité, sur la même syllabe, d'une autre intonation, se manifestent comme pertinence phonologique entre les mots. Cela veut dire que le rapport *á*:*à* ou *â*:*ȁ* et *à*:*ȁ* comporte toujours une différence sémantique, bien entendu, cette différence se manifeste dans les mots entiers.

Nous pourrions donner beaucoup plus d'exemples pareils où l'accent seul se manifeste comme opposition phonologique, qu'il s'agisse soit de l'endroit, de la qualité ou de la quantité, soit de ces éléments combinés, mais pour le moment ce n'est pas nécessaire. Ce qui est essentiel pour nous ici, c'est de souligner la pertinence phonologique de tous les accents de la langue littéraire serbocroate. Les exemples de type *Brâna* (nom de femme) : *brâna* (cueillie) : *brȁna* (passé simple de v. branati) : *Bràna* (nom d'homme), qui ne sont pas rares, montrent clairement que les accents sont justement les signes qui portent la distinction sémantique, et par conséquent ont la valeur phonologique.

Adresse de l'auteur: Dr Asim Peco, Doc. Filološki fakultet, St. trg 3, *Belgrade* (Yougoslavie).

Discussion

Rossi (Aix-en-Provence): Y-a-t-il en serbo-croate deux systèmes vocaliques? Un système en position atone? Un autre en position accentuée? S'il y a différence de timbre entre voyelles atones et voyelles accentuées, je ne vois aucune nécessité de faire intervenir le caractère distinctif de la place de l'accent.

De Bray (Clayton): Mr. *Peco's* description of the Serbocroatian accents at the phonemic level cannot really be disputed. But maybe for modern times, when experimental phonetics have been developed, it is not sufficient as a description of the linguistic realities of this language, which, as he states, is distinguished among European languages by the presence of word tones, and these require a description at the *phonetic* level. If word tone is really significant in the language, then the only really legitimate comparisons to prove its significance are those between words distinguished by tone *alone*, e.g. jàrica: jȁrica, rávan: râvan, and not by *length*, e.g. grâd: grȁd, nor by the position of the accent, e.g. màlina: malìna. It then requires further to be proved whether these tones are always realized identically in normal, continuous speech. The statement often made by *Belić* that the accents 'never change' (ne menjaju se) can then be questioned at the phonetic level. (See my article in the Slavonic and East European Review, Vol. XXXVIII, No. 91, June 1960.)

Mahnken (Göttingen): Zu der von Herrn *de Bray* aufgeworfenen Frage, inwieweit die skr. Akzente durch die Satzmelodie modifiziert werden: eine Modifikation der bestimmten Züge in der *phonetischen* Realisierung der 4 Akzente ist durchaus konstatierbar (z. B. schwebender oder sogar fallender Tonverlauf der Akzente ʹ und ˋ am Satzende einer Aussage, vorwiegend steigender Tonverlauf des Akzents ˆ am Anfang einer Aussage). Dagegen bleibt die phonologische Opposition zwischen ˋ und ˵ sowie ʹ und ˆ in diesen Fällen stets gewahrt. Aus den Darlegungen *Polloks* auf diesem Kongreß geht hervor, daß als das *relevante* Merkmal in der Opposition der skr. Akzente in der akustischen Ebene die Opposition zwischen einem sich in die Frequenzmodulation der gesamten Akzenteinheit *einfügenden* und einem sich gegen diese *abhebenden* Tonverlaufs*typ* angesehen werden muß (vgl. dazu *Polloks* und meine Ausführungen in Opera slawica, Bd. 3, Göttingen 1964). Dagegen kann die bei ˆ in der Mehrzahl der Fälle überwiegend fallende, bei ʹ und ˋ in der Mehrzahl der Fälle steigende *Richtung* des Tonverlaufs nur als ein *redundantes* (wenn auch recht charakteristisches) Merkmal dieser drei Akzente angesehen werden; dieses Merkmal unterliegt der Beeinflussung und Neutralisierung durch die Satzmelodie, so daß diese Akzente positionsbedingte Varianten mit anderer

Tonverlaufsrichtung aufweisen. Für den Akzent ˵ kann ohnehin die Tonverlaufs*richtung nicht* als charakteristisch angesehen werden. Zwischen ˵ und ˋ besteht die Opposition Stoßton : Gleitton. *Auditiv* stellt sich die Opposition zwischen den sogenannten fallenden und den sogenannten steigenden Akzenten des Skr. als Opposition zwischen starken (ˆ, ˵) und schwachen Akzenten (ʹ, ˋ) dar.

In Abhängigkeit von den satzmelodischen Bedingungen ändert sich die *Richtung des syntagmatischen* Tonverlaufsbogens.

Die Art des Verhältnisses zwischen dem Tonverlauf der akzentuierten Silbe und der nichtakzentuierten Silbe der Akzenteinheit bleibt dabei aber konstant (einfügend bei ʹ und ˋ, abhebend gegen diesen Bogen bei ˆ und ˵). Deshalb *ändert sich* die *Richtung* des Tonverlaufs der einzelnen Akzente in Abhängigkeit von der Gesamtrichtung des syntagmatischen Intonationsbogens, der Silbenzahl desselben und der Lage der akzentuierten Silbe innerhalb desselben (Anfang, Ende usw.). Infolge der unterschiedlichen *Distribution* der Akzente ˆ ˵ einerseits, ʹ ˋ andererseits sind die letzteren weniger häufig einer Richtungsumkehrung unterworfen als die ersteren. Die Opposition zwischen den (auditiv) *starken* Akzenten ˆ ˵ und den *schwachen* Akzenten ʹ ˋ bleibt jedoch unabhängig von der Änderung der *Richtung* des Tonverlaufs bewahrt, da sie sich in der akustischen Ebene als eine *Einfügungs*opposition darstellt (abhebend : einfügend), die von der Satzmelodie unberührt bleibt.

Verh. 5. int. Kongr. Phon. Wiss., Münster 1964, pp. 458–462
(S. Karger, Basel/New York 1965).

Phonem, Allophon und Sprachlaut in der historischen Sprachwissenschaft

Von HERBERT PENZL, Berkeley, Calif.

In der historischen Phonologie steht bei der synchronischen Analyse von Texten die Erfassung der Phoneme im Mittelpunkt[1]. Graphemische Statistik und vergleichende Lautbestimmung auf Grund von dialektisch und zeitlich verschiedenen Entsprechungen ergeben die notwendigen Grundlagen. Daß in der Ausarbeitung oder Anwendung von orthographischen Systemen anscheinend bei allen muttersprachigen Schreibern stets die Tendenz bestand, die für sie wesentlichen Lautunterschiede irgendwie zu bezeichnen, erleichtert die Erkenntnis phonemischer Oppositionen. Orthographie und Phonemsystem hängen von den ältesten Anfängen an zusammen: man denke an Wulfilas gotisches Alphabet, an den Gebrauch des lateinischen Alphabets für Altenglisch, Altsächsisch, Althochdeutsch oder für die romanischen Sprachen durch die zeitgenössischen Schreiber. Die Wiedergabe von Varianten der Phoneme, von Allophonen durch die Orthographie, z.B. in *Notkers* ahd. Orthographie[2], ist ungewöhnlich und nur in einer zweisprachigen Situation oder durch analogische Einflüsse zu erklären. Einsprachige Schreiber sind sich in der Regel der allophonischen Variation in ihrer Muttersprache absolut nicht bewußt.

Wir denken uns jeden regelmäßigen Lautwandel, der allmählich und unbewußt, nicht sprunghaft vor sich geht, an die Entwicklung von Stellungsvarianten der Phoneme geknüpft. Es ist also keine historische Lautveränderung ohne Allophone denkbar. Die Orthographie drückt in der Regel nur das phonemische Resultat aus, aber dieses bleibt ohne die Annahme der Allophonentwicklung unerklärbar. Man denke z.B. an den ahd. Umlaut: ahd. *mahti* wird zu mhd. *mähte*, ahd. *ubil* zu mhd. *übel*, ahd. *wāri* zu mhd. *wǣre* usw. Die mhd. Orthographie hat vielfach neue Lautzeichen (*ü ǣ*) für die Umlautsvokale gefunden und sie bezeichnet den Zusammenfall von *i* mit *e* in Nebensilben. Diese phonemischen Wandel sind jedoch ohne die Annahme von palatalen Allophonen von /a/ /u/ /ā/ und von mittelgaumigen Varianten von /i/ in Nebensilben im Ahd. undenkbar. Auch bei einem historischen Lautwandel wie dem ahd. Umlaut und dem ahd. Vokalzusammenfall in unbetonten Silben müssen wir also die Allophone, die zum Lautwandel führten, rekonstruieren und können für sie nur recht allgemeine phonetische Züge, z.B. palatalen Charakter, Lippenrundung bei [ü] oder relative Züge, z.B. offeneren Charakter von [ä] im Vergleich mit [e] u.dgl., nie konkretere Details der Artikulation oder absolute Züge annehmen. Kein graphemisches Material erleichtert uns bei Allophonen das Rekonstruieren. Bei keiner diachronischen Analyse ist aber ohne Allophone auszukommen. Auch das Rekonstruieren von prähistorischen Lautveränderungen erfordert unbedingt das Rekonstruieren von Allophonen, also z.B. auch bei der germanischen und bei der ahd. Konsonantenverschiebung. Synchronische Analyse und synchronisches Rekonstruieren von prähistorischen Lautsystemen soll, muß aber nicht die Annahme von Allophonen voraussetzen[5].

Haben wir also in der historischen Sprachwissenschaft nur mit Phonemen und Phonemvarianten zu tun? Wenn man zugibt, daß die Aufgabe der historischen Phonologie die realistische, objektive Erfassung aller ehemaligen, lautlichen Veränderungen ist, kann man die rein phonetischen Tatsachen nicht vernachlässigen. Auch die Sprachlaute «an sich» sind als Ziel synchronischer und diachronischer Analyse anzusehen. Freilich muß zugegeben werden, daß das verfügbare Material eine rein lautliche, ins einzelne gehende Bestimmung kaum je gestatten wird. Man wird sich mit der zögernden Feststellung einiger weniger lautlicher Tatsachen begnügen müssen. Ich versuchte am Beispiel des ahd. *r*[3] zu zeigen, daß das verfügbare Beweismaterial für eine phonetische Bestimmung kaum eine Entscheidung zuläßt, ob wir ein Zungenspitzen-r oder ein Zäpfchen-r oder beides für das Ahd. anzunehmen haben, also nicht einmal zwischen den beiden phonetischen Haupttypen der Gegenwart, abgesehen von Einzelheiten der Artikulation, mit Sicherheit zu unterscheiden ist.

Die Erkenntnis phonemischer Opposition nötigt auch zu einer phonetischen Beschreibung der Gegensätze, denn die Struktur des Systems hängt von den phonetischen Merkmalen der Unterscheidung ab. Der Versuch einer Lautbestimmung der ahd. Phoneme, die < s > und < (ʒ) z > geschrieben wurden (z.B. *ûz* «aus», *hûs* «Haus») und bis ins späte Mhd. von allen Schreibern konsequent unterschieden wurden, ist unausweichlich, aber überaus schwierig. Die Orthographie beweist die phonemische Opposition, deutet bei < z > auf ursprünglich phonemische Gleichheit mit der Affrikata /ts/. Der Lautwandel von ahd. *sk* zu mhd. *sch* [š] und der spätere Zusammenfall von *s* vor Konsonanten mit diesem /š/ deutet auf bestimmte ahd. Allophone von /s/ in diesen Stellungen. Die <s> und <z> geschriebenen Sprachlaute entziehen sich einer absoluten Identifizierung trotz allen phantasievollen Versuchen der Forschung, weil nur protokomparatives, diakomparatives (ahd. *zz* zu spätmhd. *ss*) Material, aber keine orthoepischen Beschreibungen wie bei lateinischen, griechischen, altindischen, neuhochdeutschen Lauten zur Verfügung stehen[4].

Zusammenfassend läßt sich sagen, daß in der historischen Lautlehre in synchronischer Betrachtung der Texte Phoneme und ihr System, in diachronischer Beschreibung z.B. von Phonemveränderungen die Allophone der Phoneme von größter Bedeutung und auch erfaßbar sind. Außerhalb eines phonemischen Systems sind die Sprachlaute als rein phonetische Einheiten meistens wegen des verfügbaren Beweismaterials gar nicht oder nur in ganz wenigen groben oder relativen Zügen faßbar. Die phonetische Grundlage von Phonem und Allophon und die kompromißlose Realitätsforderung der historischen Sprachwissenschaft verlangen aber ständigen Bezug auf die Sprachlaute an sich und zwingen den Sprachforscher zu Versuchen einer konkreten Lautbestimmung.

Literatur

1. *Hall, R. A., Jr.:* The oaths of Strassburg. Language *29:* 317ff. (1953); auch Language *34:* 67–69 (1958); *35:* 24f. (1959).
2. *Penzl, H.:* Zur Erklärung von Notkers Anlautgesetz. ZfdA *86:* 196ff. (1955).
3. *Penzl, H.:* Old high German <r> and its phonetic identification. Language *37:* 488–496 (1961).
4. *Penzl, H.:* Methoden der Lautbestimmung in der historischen Sprachwissenschaft. Proc. 4th int. Congr. Phonetic Sciences 1962, pp. 719–722.
5. *Twaddell, W. F.:* The prehistoric Germanic short syllabics. Language *24:* 139ff. (1948).

Adresse des Autors: Dr. H. Penzl, University of California, *Berkeley, Calif.* (USA).

Discussion

Pilch (Freiburg): Die Unterscheidung, ob wir Phoneme oder einzelne Varianten rekonstruieren, ist wichtig. Das lange diskutierte Problem der Germanistik – war der tiefe, lange Vokal des Urgermanischen [æ:] oder [a:]? – halte ich für ein Scheinproblem. Wir können nur den tiefen, langen Vokal im Fünfersystem des Urgermanischen rekonstruieren:

ī ē ā ō ū

Nach *Martinets* Prinzip der Äquidistanz müßte er ein mittlerer Vokal sein.

Martinet (Paris): Il me semble qu'on peut souvent cerner la réalisation des phonèmes d'états de langue non accessibles à l'observation directe d'un peu plus près que ne le suggère l'exposé de M. *Penzl:* dans une langue comme le vieux-haut-allemand qui venait de subir une période de rhotacisme on doit supposer une prononciation apicale de /r/ qui seule explique la confusion avec /r/ d'une sifflante apico-alvéolaire sonore (ź); ce qu'on transcrit aujourd'hui *z* dans les textes du vieux-haut-allemand provient du relâchement d'une occlusive apico-dentale, ce qui laisse supposer une prononciation dorso-alvéolaire (celle du [s] français ou allemand d'aujourd'hui), pour *s* des mêmes textes on doit supposer une prononciation apico-alvéolaire, ([ś] du castillan ou du danois contemporains) d'ou le [ś] de l'allemand *Schnee* par exemple; ce sont les emprunts qui confirment cette prononciation apico-alvéolaire pour *s* et la prononciation dorso-alvéolaire de *z*.

Heinrichs (Gießen): 1. s wurde im Mhd. wahrscheinlich nach Sch [ʃ] hin ausgesprochen. Dafür sprechen Schreibungen des 13.–15. Jahrhunderts, wie rip. 'voische' = Füße, 'geweschen' = Gewissen (falsch verbessert vom Schreiber in 'gewechen', da im Rheinischen oft fälschlich für 'sch', 'ch' gesprochen wird).

2. Es gab, zum mindesten im Mhd. schon uvulares und palatales r, je nach Stellung, ripuarisch: vorten < vorhten wird gelegentlich vochten bzw. fochten geschrieben. Mittelniederfränkischem lucht, locht = links entspricht ripuarisch lurz, lorz < lort (unverschoben), wo 'ch' offensichtlich als 'r' realisiert wurde, wie noch heute im Rheinischen üblich.

Martens (Hamburg): Mit Recht weist Herr *Penzl* darauf hin, daß *Trautmann* jene Behauptung aufgestellt hat, das uvulare [R] sei aus dem Französischen importiert worden.

Die Hinweise von Herrn *Heinrichs* untermauern die Ansicht, daß es das uvulare [R] bereits vor dem Einfluß des Französischen gegeben haben muß, wenn in Urkunden nebeneinander Formen stehen wie «vorchten – vorten – vochten» für fürchten.

Ich darf einen Fingerzeig in die Erinnerung zurückrufen, den *W. Moulton* einst gab: Bei *Jakob Böhme* steht im Kapitel 18 der «Aurora» eine ausführliche Beschreibung von der Aussprache des /r/, und *Jakob Böhme* spricht davon, daß die Zunge sich «fürchtet und zurückzieht» und daß es «im Grunde des Mundes faucht».

Wiesinger (Marburg): Zur Erhellung der phonetischen Verhältnisse der schriftlichen Überlieferung des Alt- und Mittelhochdeutschen können die gegenwärtigen deutschen Mundarten in bestimmten Fällen einen Beitrag leisten, insbesondere die konservativen Mundarten des Oberdeutschen. So bewahren z. B. die im 12. Jh. vom südbairischen Binnenland besiedelten zimbrischen Sprachinseln in Oberitalien die wahrscheinlich vorauszusetzende Aussprache der im Mhd. mit s und ʒ wiedergegebenen Laute. Während dort dem ʒ ein stimmloser alveolarer Reibelaut entspricht, der intervokalisch noch als Geminata gesprochen wird und auch in der Affrikata [ts] gilt, ist s intervokalisch, im Auslaut und im Anlaut vor Vokalen und Liquiden ein alveolopalataler Reibelaut zwischen [z] und [ʒ]. In Teilen des Hoch- und Höchstalemanischen herrschen Akzentverhältnisse, die zur Gänze Notkers Akzentzeichen entsprechen: ursprüngliche Längen werden zweigipfelig (= circumflex), ursprüngliche Kürzen (die Dehnung in offener Silbe ist – von Ausnahmen abgesehen – unterblieben) steigend (= Akut), die Diphthonge ie – uo – üe wieder zweigipfelig und die Diphthonge ei – ou – öü steigend akzentuiert. Diese Beispiele ließen sich vermehren.

Pulgram (Ann Arbor): Es ist nicht notwendig, eine programmatische Stellung einzunehmen in der Frage, ob man Allophone in alten (toten) Sprachen finden kann oder nicht. Wenn wir Glück haben, können wir es, wenn Pech, dann nicht. Es hängt dies von vielen Details ab, von den Texten, von der Geschicklichkeit des Schreibers (als Hörer), von der Distanz der Texte von der Gegenwart usw. Ein Beispiel: in den altfranzösischen Straßburger Eiden (9. Jh.) wird das Wort Bruder teils mit *d* teils mit *dh* geschrieben. Da wir annehmen müssen, daß *fratrem* zu *frère* wird, also *t* zu Null, auf dem Weg über /d/ und /ð/, stellen die Schreibungen *d* und *dh* des Textes wohl ein allophonisches Nebeneinander, eine Koexistenz, von /d/ und /ð/ dar, die eventuell zu Gunsten von /ð/ entschieden wird und schließlich zum Ausfall des Konsonanten führt.

Isačenko (Berlin): Es scheint mir wenig sinnvoll zu sein, die «Aussprache» von Lauten in älteren Sprachstadien ermitteln zu wollen. War das «m» des Indogermanischen identisch mit dem [m] des Russischen etwa in [mat,] 'Mutter'? Wir können feststellen, daß das idg. *m* Glied des folgenden Teilsystems war:

$$\mathrm{m}\begin{array}{l} \nearrow \mathrm{bh} - \mathrm{ph} \\ \searrow \mathrm{b} \;\;- \mathrm{p} \end{array}$$

Das russ. /m/ ist dagegen Glied des folgenden Teilsystems:

$$\mathrm{m} - \mathrm{m,}\begin{array}{l} \nearrow \mathrm{b} - \mathrm{b,} \\ \searrow \mathrm{p} - \mathrm{p,} \end{array}$$

Es ist unmittelbar einleuchtend, daß der phonologische Wert der beiden Laute verschieden ist, auch wenn man annehmen kann, daß die phonetische Substanz des idg. labialen Nasal unverändert blieb.

Antwort *Penzl:* Ich stimme mit Herrn *Pilch* überein, daß eine absolute phonetische Identifizierung als Teil einer Rekonstruktion eines Protosystems besonders schwierig ist. Die Diskussionsteilnehmer haben mit Recht auf die verschiedenen Methoden der Lautbestimmung hingewiesen: Herr *Martinet* auf die protokomparative Methode (ahd. *r* als Zungenspitzenlaut wegen der Ableitung von westgerm. *z*), Herr *Heinrichs* auf diachronisches Beweismaterial durch dialektische Sonderentwicklung wie ripuarisches *ch* für ahd. *r*, Herr *Wiesinger* auf die Bedeutung moderner Dialekte für die Lautbestimmung. Ich glaube, die von Herrn *Pulgram* erwähnten allophonischen Schreibungen bezeichnen schon die phonemischen Resultate. Ich möchte nicht so weit gehen wie einerseits Herr *Isačenko* und andererseits Herr *Wiesinger*, die vielleicht allzu große Skepsis bzw. allzu großen Optimismus bezüglich der Möglichkeit und der Bedeutung rein phonetischer Lautbestimmungen ausgedrückt haben. Niemand von uns ist natürlich noch der Meinung, wie es seinerzeit *E. Sievers* war, daß uns überlieferte Texte unmittelbar durch «Schallanalyse», d. h. durch gelehrtes, intuitives Nachschaffen, Lautwerte der Vergangenheit vermitteln.

C.R. 5e Congr. int. Sci. phon., Münster 1964, pp. 463–466
(S. Karger, Basel/New York 1965).

Evolution des occlusives du parler de Blaesheim, un exemple de «pression dans la chaîne» et de «pression du système»

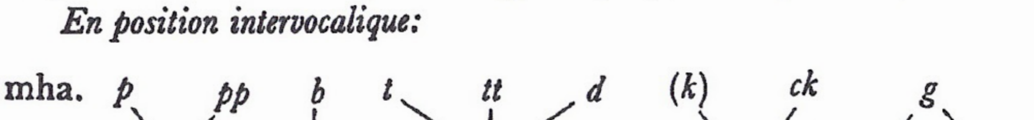
Par Marthe Philipp, Nancy

Blaesheim est un village situé à une quinzaine de kilomètres de Strasbourg dans le domaine des dialectes bas-alémaniques qui au moyen âge ont servi de base à la koiné littéraire, au moyen-haut-allemand. La langue des poètes du moyen âge (appelée par *M. Moser* «Die höfische Dichtersprache») peut, par conséquent, être considérée comme un point de départ valable pour une étude diachronique de ce parler.

L'étude de l'évolution des occlusives du parler nous permet de constater tout d'abord que les correspondances phonétiques, que l'on peut établir entre le moyen-haut-allemand et le parler de Bl., varient suivant la position de l'occlusive à l'intérieur du monème. Voici les correspondances que nous avons observées en *position initiale:*

Exemples: mha. *puppe* Bl. *bub*, mha. *bûch* Bl. *büx'*, mha. *tor* Bl. *dōr*, mha. *durst* Bl. *duršd*, mha. *kopf* Bl. *ghobf*, mha. *guot* Bl. *güəd*.

En position intervocalique:



Exemples: mha. *tepich* Bl. *debi*, mha. *suppe* Bl. *sub*, mha. *heben* Bl. *hevə*, mha. *keten* Bl. *ghed(ə)*, mha. *bette* Bl. *bed*, mha. *lëder* Bl. *ladər*, mha. *hâke* Bl. *hūgə*, mha. *snëcke* Bl. *šnag*, mha. *biegen* Bl. *biəjə*.

En position finale:

mha.	*(p)*	*b**	*t*	*d*	*k*	*g*		
Bl.	*b*		*d*		*g*	*j*	*zéro*	

Exemples: mha. *knîp* Bl. *gnibə*, mha. *leip, -bes* Bl. *lei̯b*, mha. *brët* Bl. *brad*, mha. *pfat, -des* Bl. *bfōd*, mha. *drëc* Bl. *drag*, mha. *gîge* Bl. *gī*.

En position anté- ou postconsonantique:

mha.	*p*	*b*	*t*	*d*	*k*	*g*		
Bl.	*b*		*d*		*g*	*i*	*j*	*zéro*

Exemples: mha. *plate* Bl. *blåd*, mha. *bluot* Bl. *blüəd*, mha. *trüebe* Bl. *driəb*, mha. *schenken* Bl. *šaŋgə*, mha. *singen* Bl. *siŋə*, mha. *morgen* Bl. *morjə*.

Trois changements importants se sont produits depuis le stade du moyen-haut-allemand: 1° le relâchement des représentants des anciennes douces *b g* en position intervocalique; 2° la réduction des géminées; 3° la perte de l'opposition forte ∾ douce.

En étudiant l'évolution du système phonologique et en particulier celle des occlusives, on peut observer d'une part qu'elle n'est pas la même dans toutes les positions et d'autre part que des changements différents dans chaque position aboutissent toujours au même résultat, c'est-à-dire à la création d'une série d'occlusives unique composée des trois douces sourdes *b d g*.

En position *intervocalique*, après voyelle *brève*, la série des fortes *p t k* était réduite en mha. à *t*, *p* et *k* ne se présentant pratiquement pas dans cette position. En revanche, on y trouvait les trois géminées *pp tt ck* et la série des douces *b d g*. Il semble que ces douces aient amorcé la transformation du système: le relâchement de *b g* en *v j*, en spirantes, est dû sans aucun doute à la pression des phonèmes vocaliques environnants; dans d'autres positions mha. *b g* sont restés des occlusives à Bl. La réduction des géminées *pp ck* correspond à l'élimination d'un trait pertinent devenu inutile, puisqu'elles ne s'opposaient plus, dans cette position, aux fortes simples *p k*, ni même aux douces *b g* devenues des spirantes. La confusion de *t tt* et celle de *t d*, qui ont entraîné quelques homonymies, sont des exemples de pression du système. Dans cette position, l'évolution de la série des occlusives peut être représentée de la manière suivante:

* L'évolution des correspondants de mha. *k g* nous oblige à partir du stade précédant la «Auslautsverhärtung».

mha.	+	[*t*]	+	stade inter-	**p*	[**t*]	**k*	Blaesheim:	
	pp	[*tt*]	*ck*	médiaire,				[*b* *d* *g*]	
	b	*d*	*g*	non attesté:	**β*	[**d*]	**γ*	*v*	*j*

A l'intervocalique après voyelle *longue*, il n'y avait pas de géminées en moyen-haut-allemand, mha. *p k* ne se trouvaient que dans un petit nombre d'exemples; les douces *b g* se sont relâchées en *v j* comme après voyelle brève. Les représentants de mha. *t d* ont été confondus sous la pression du système:

mha.	*(p)*	*t*	*(k)*	stade inter-	**(p)*	[**t*]	[**(k)*]	Blaesheim
	b	*d*	*g*	médiaire:	**β* ↘	[**d*]	[**γ* ↘]	[+ *d* *(g)*]

Dans cette position, une seule opposition forte ∾ douce a été éliminée: celle qui représente mha. *t–d;* les oppositions mha. *p–b*, *k–g* ont été conservées à Bl. sous la forme *b–v*, *g–j*.

En position *finale* en moyen-haut-allemand (ou intervocalique devenue finale après la chute de *e* atone), les correspondants des anciennes oppositions *p–b*, *t–d* ont été confondus, mais l'opposition *k–g* est conservée à Bl. sous la forme *g–j:*

stade du mha.	[*p*]	[*t*]	*k* ↓	Blaesheim	[*b* *d* *g*]
	[*b*]	[*d*]	*g* ↓		*j*

Après les *diphtongues* on peut observer une évolution encore différente:

stade du mha.	+	[*t*]	+	Blaesheim	[*b* *d* *g*]
	b	[*d*]	*g*		

Devant et après consonne, il y a eu coalescence des fortes et des douces au profit des douces (sauf dans quelques combinaisons):

mha.	[*p*]	[*t*]	[*k*]	Blaesheim	[*b* *d* *g*]
	[*b*]	[*d*]	[*g*]		

En position *initiale*, *p* avait un rendement très faible en mha., *t* était probablement une forte non aspirée, tandis que *k* était aspiré. La série des fortes, peu homogène, a été éliminée au profit de la série des douces: il y a eu perte de l'opposition forte ∾ douce pour les correspondants de mha. *p–b*, *t–d*, et scission de phonème pour le représentant de mha. *k* en *g* + *h*, devenu biphonématique.

mha.	*(p)*	*t*	*k* [*kh*]	stade inter-	+	+	**gh*	Blaesheim:
	b	*d*	*g*	médiaire:	**b*	**d*	**g*	$\boxed{b\quad d\quad g}$

On peut considérer que dans cette position, ces changements sont dus à la seule pression du système.

Au total, les trois occlusives orales *b d g* apparaissent dans un nombre croissant de positions, bien qu'elles soient d'origine diverse: elles représentent ou bien les anciennes fortes ou bien les anciennes douces. Cette évolution n'est d'ailleurs pas achevée à l'heure actuelle: après voyelle longue, où mha. *b g* sont représentés par des spirantes, la série des occlusives est généralement réduite à $\boxed{+\ d\ +}$. Il semble qu'à partir d'un moment donné, il y ait eu une conjoncture favorable à l'élimination de la série peu homogène représentant les anciennes fortes mha. *p t k*, à la coalescence des deux séries au profit de celle des douces, au rendement normal dans toutes les positions, à la création d'une série unique d'occlusives. Après le relâchement des représentants de mha. *b g* intervocaliques, dû à la pression dans la chaîne, il semble bien que le facteur décisif pour la suite de l'évolution ait été la pression du système, celle des unités distinctives entre elles.

Adresse de l'auteur: Marthe Philipp, Faculté des Lettres, *Nancy* (France).

Discussion

Fourquet (Paris): M^lle^ *Philipp* a présenté un cas très intéressant de regroupement des éléments phoniques, après une action phonétique qui a désorganisé le système. La nouvelle série b, d, g résulte de la fusion en un même phonème des douces en position initiale et des *fortes* en position intérieure (gan, degə < geben, decken). Des variantes combinatoires de b, g deviennent des phonèmes à elles seules (v), ou deviennent membres d'un autre phonème (f qui existait uniquement à l'initiale). Un fait analogue de regroupement caractérise la 2e mutation ainsi les allophones issus de p germanique deviennent membres d'autres phonèmes, resp. /f, ff/ et /pf, ppf/; pour nous *Pfund* et *Apfel* présentent le même phonème, mais à l'initial il correspond à *p*, à l'intervocalique à *pp* germanique.

Trost (Praha): Je me permets une seule remarque à propos de l'exposé très intéressant de M^lle^ *Philipp*. Si elle a dit qu'une certaine évolution phonologique, soit possible, mais non nécessaire, j'en dirais plus: il n'y a pas en principe, d'évolution phonologique nécessaire. L'histoire des sons est pleine de carrefours et l'évolution peut passer de l'une ou de l'autre côté.

Beyer (Strasbourg): Je crois que M^lle^ *Philipp* place la lénition de *p*, *t*, *k*, à une époque trop tardive. Des indices de cette lénition se présentent dans le Straßburger Urkundenbuch dès 1260.

Verh. 5. int. Kongr. Phon. Wiss., Münster 1964, pp. 467–473
(S. Karger, Basel/New York 1965).

Zentrale und periphere Lautsysteme

Von H. Pilch, Freiburg i. Breisgau

Innerhalb der Sprachtheorie treffen wir häufig auf Definitionen von der Form: Der Satz S gilt *in der Sprache L*, z.B. «Stehen zwei Lautelemente innerhalb der Sprache L in Opposition, so sind sie (in L) phonematisch verschieden». Der im Definiens auftretende Ausdruck *die Sprache L* ist dabei selbst ein primitiver, undefinierter Begriff. Vom umgangssprachlichen Gebrauch her können wir meistens intuitiv entscheiden, in welcher Sprache ein bestimmter Sprecher gerade spricht. Der undefinierte Begriff führt jedoch zu Verwicklungen im Falle der sogenannten *Anderthalbsprachigkeit**, d.h. bei Sprechern, die außer ihrer Muttersprache noch ein paar Brokken aus anderen Sprachen kennen und verwenden. Das ist gewiß die Mehrzahl aller Sprecher. Wir alle gebrauchen zumindest fremde Orts- und Personennamen, fremde Handels- und Markenbezeichnungen und ein paar sonstige Lehnwörter. Solches Lehngut ist oft phonematisch anders gebaut als der im engeren Sinne «einheimische» Wortschatz. Im Finnischen brechen einige moderne Entlehnungen die sonst gültigen Regeln der Vokalharmonie, z.B. *konduktööri* «Schaffner». In deutschen Entlehnungen aus dem Französischen treten phonematisch distinktive Nasalvokale auf, z.B. *grande dame*, *Pompidou*.

Es ist umstritten, ob und in welchem Umfang solche strukturell abweichenden Elemente bei der linguistischen Analyse einer Sprache berücksichtigt werden sollen. Intuitiv wissen wir zwar, daß es sich hier um fremde Elemente handelt, die irgendwie am Rande des jeweiligen sprachlichen Systems liegen. Es fehlen uns aber die Definitionen zu theoretisch sicherem Vorgehen. Zur Abgrenzung solcher «peripheren» Elemente von dem eigentlichen, «zentralen» System der Sprache L schlagen wir ein zweifaches Kriterium vor:

* Terminus nach *Ch. Hockett*, Language *32:* 468 (1956).

1. Die Klasse K von peripheren Elementen soll strukturelle Besonderheiten aufweisen, die innerhalb des untersuchten Materials *nur* ihr eignen.

2. Zumindest *ein* weiteres Kriterium muß die gleiche Klasse K als Sonderklasse ausweisen, sei es Entlehnung, sei es Gebrauch nur in bestimmten Situationen (Erregung, Trunkenheit) oder seitens bestimmter Kreise (slang, Kindersprache, Fachsprachen).

Nach diesem Doppelkriterium sind z.B. innerhalb des Deutschen nicht nur Entlehnungen aus vielen fremden Sprachen jeweils als besondere, periphere Systeme mit eigener phonematischer Struktur abgrenzbar (Nasalvokale und /ž/ in französischen, /þ/ in englischen, anlautendes /č/ in englischen, russischen und chinesischen Wörtern usw.), sondern auch Interjektionen wie /ps:t/ und /pr:/. Hier treten in Abweichung vom sonstigen deutschen Lautsystem /s/ und /r/ als Silbenkerne auf. Das zweite Kriterium besteht in der Zugehörigkeit dieser Wörter zur Klasse der Interjektionen.

Die geometrische Metapher zentrales System – peripheres System rechtfertigt sich dadurch, daß periphere Systeme stets mehr oder weniger zur Angleichung an das zentrale System tendieren (aber nicht umgekehrt). Genauer: die Sonderstrukturen peripherer Systeme werden häufig ersetzt durch intuitiv ähnlich klingende Strukturen des zentralen Systems. Die Erscheinung ist bekannt in Form der «ungebildeten» Aussprache fremder Wörter, z.B. dtsch. /oŋ/ statt frz. /õ/, dtsch. /s/ statt engl. /þ/. Sie betrifft aber auch die nicht fremden Sprachen entlehnten peripheren Systeme. In einer vom Englischen abgeleiteten Geheimsprache amerikanischer Schüler tritt an die Stelle konsonantischer Anlaute des Englischen die Gruppe /šm/. Von hierher ist das Wort *shmoo* über die 'comic strips' in die englische Gemeinsprache eingedrungen. Der Anlaut /šm/ kommt sonst im Englischen nicht vor. Die tatsächliche Aussprache des Wortes schwankt zwischen /šmu:/ und der dem zentralen englischen Lautsystem angeglichenen Form /smu:/.

Aus unseren Definitionen folgt, daß «die Sprache L» (das zentrale System) nicht identisch sein kann mit der Rede eines oder mehrerer bestimmter Sprecher. Den «reinen Sprecher», der seiner Rede gar keine fremden Brocken beimischt, dürfte es in der Wirklichkeit kaum geben. Wenn wir im Ernst die *gesamte* Rede eines Sprechers in einem einzigen System unterbringen wollten, so würde der Begriff «die Sprache L» untergehen im Begriff «Idiolekt». Größen wie «die deutsche Sprache» oder «die englische Sprache»

gäbe es für unsere Sprachtheorie nicht mehr. Dieses theoretische Ergebnis widerspräche der Erfahrung zu deutlich, als daß wir es hinnehmen könnten.

Ebensowenig läßt sich «die Sprache L» aus der Rede von Sprechern unter Ausklammern aller Lehnwörter gewinnen. Wollten wir im Ernst alle Lehnwörter aus unserem Material streichen, so blieben am Ende wohl kaum noch «Erbwörter» übrig. Jede Zeitgrenze, vor der etwa Lehnwörter gleich den Erbwörtern behandelt werden sollten, müßte – von der Sache her gesehen – ganz willkürlich festgesetzt werden. Das entscheidende Kriterium dafür, welche Lehnwörter einem Sondersystem zuzuweisen sind, muß vielmehr in ihren strukturellen Besonderheiten liegen. Die Abgrenzung von zentralen und peripheren Systemen ('co-existent phonemic systems'*) ist daher meines Erachtens sachlich nicht zu umgehen – ganz unabhängig davon, ob das hier dazu vorgeschlagene Doppelkriterium sich bewährt oder nicht.

Adresse des Autors: Prof. Dr. Herbert Pilch, Englisches Seminar der Universität, Rotteckring 4, *78 Freiburg* (Deutschland).

Discussion

Pilch ergänzt einleitend seine Darlegungen zum Thema: *Zentrale und periphäre Lautsysteme.*

Sind Wörter wie *teenager*, *dauphin*, *sotto voce* deutsch oder nicht? Wenn ja, so sind /ʒ, č, ε̃, t:/ deutsche Phoneme. Wenn nein, so werfen wir Wörter aus dem Deutschen hinaus, die laufend in deutscher Rede gebraucht werden. Wie wir uns auch entscheiden, immer geraten wir in eine Anomalie hinein. Das Problem ist grundsätzlich nicht neu, und es taucht bei jeder phonematischen Analyse auf. Erklären wir die genannten Wörter für nichtdeutsch, so bekommen wir ein einfaches deutsches Phonemsystem. Dürfen wir aber solcher Einfachheit halber über die sprachlichen Tatsachen hinweggehen? Wir können uns nicht darauf berufen, es handle sich um Lehnwörter. Viele andere Lehnwörter behalten wir nämlich drin, z. B. *Mauer* oder *achtern*. Erklären wir die Wörter umgekehrt für deutsch, so wird unser deutsches Phonemsystem bald überhaupt zerfließen. Wir werden nämlich auf Grund weiterer Lehnwörter überhaupt alle englischen, französischen und italienischen Phoneme mit in das deutsche Lautsystem aufnehmen müssen. Dazu werden weitere Phoneme kommen. Einige Deutsche sprechen z. B. Russisch, andere Schwedisch. In diesen Kreisen gelten Aussprachen wie *Berezina* /berezí:na/ statt /b,ir,iz,iná/ oder *Luleå* /lu'le:a/ statt /'luleo/ als ungebildet. Wenn wir folgerichtig das russische, schwedische und weitere fremde Lautsysteme mit in das Deutsche aufnehmen, so fragt es sich: Gibt es überhaupt so etwas wie «das deutsche Lautsystem» oder «das schwedische Lautsystem»? Tatsächlich ist von ernsthafter Seite behauptet worden, ein Phonemsystem gehöre nicht zu einer Sprache, sondern zu einem Sprecher. Das Phonemsystem jedes Sprechers umfasse alle diejenigen Laute, die er hervorbringen und voneinander unterscheiden könne. Es gibt demnach z. B. ein englisches Lautsystem für

* Terminus nach *K. L. Pike* und *Ch. C. Fries*, Language *25:* 29–50 (1949).

den Präsidenten Lyndon B. Johnson und ein anderes für Professor Kenneth L. Pike. Die Frage bleibt offen, wie denn Leute mit so verschiedenen Lautsystemen einander verstehen können.

Der Fehler liegt hier, wie ich glaube, in unvollkommener Abstraktion. Sprache wird in extrem behavioristischer Weise gleichgesetzt mit der sprachlichen Verhaltensweise einer bestimmten Gruppe von Sprechern. Die Abgrenzung dieser Gruppe bleibt dem Soziologen überlassen. Eine Schwierigkeit dieser Auffassung liegt schon darin, daß das Definiendum *Sprache* im Definiens *sprachliche Verhaltensweise* wieder auftaucht. Wir möchten dagegen den Sprachbegriff, d. h. die einzelne Sprache, als Abstraktion betrachten. Es ist eine abstrakte Struktur, zu der wir auf induktivem Wege gelangen, d. h. ausgehend von der Rede der Sprecher. Die Abstraktion *Sprache* ist also nicht dasselbe wie die Verhaltensweise der Sprecher. Sie kann es nicht sein, weil es mehrsprachige Individuen gibt. Aber auch der monoglotte Sprecher kennt in der Regel einige Wörter aus Fremdsprachen. Wir dürfen also auf keinen Fall so tun, als gehörten alle Äußerungen eines Sprechers – wenigstens im Normalfall – einem einzigen sprachlichen System an. Ob und inwieweit das der Fall ist, kann sich erst bei der Aufstellung der sprachlichen Struktur selbst erweisen. Wir brauchen also Kriterien, nach denen wir innerhalb der Rede verschiedene sprachliche Systeme unterscheiden können.

Weitere bekannte Felder für phonematisch ungewöhnliche Formen bilden die Interjektionen und die Kindersprache. Im Kymrischen steht z. B. langes /χː/ nur in der Interjektion *ach yfi* /'aχ'χəʾvi/. Dies ist innerhalb des Kymrischen der einzige Fall von phonematisch relevanter Konsonantenlänge. Im Deutschen stehen die sogenannten kurzen Vokale nie betont im absoluten Auslaut. Eine Ausnahme davon bilden die Interjektionen *aha* /a'ha/, *ha* /'ha/ und *ää* /'ʔɛ/ und auch ammensprachliche Wörter wie /'ba/, /'bɛ/.

Ein weniger bekanntes Sonderfeld bilden im Deutschen Namen von Arzneimitteln. Das sind sehr häufig drei- oder viersilbige Wörter. Sie enden auf eine betonte geschlossene Silbe. Sonst bestehen sie aus lauter offenen Silben: *Pyramidon, Irenat, Liberol, Aspirin, Favistan, Furoxon, Novadral, Aludrox.* Der Zusammenhang zwischen dieser phonematischen Struktur und dem Wortfeld *Arzneimittel* ist so eng, daß wir, wenn wir Wörter wie *Thymipin, Hämoglobin* zum erstenmal hören, sofort vermuten, es handle sich um Medikamente oder Chemikalien.

Diese Chemikaliennamen bilden ein Sondersystem nach einem doppelten Kriterium. Das erste Kriterium betrifft die phonematische Form, das zweite die semantische Eigenschaft als Wortfeld. Wesentlich ist, daß die beiden Kriterien verschiedenen Strukturbereichen angehören. Auch in unseren anderen Fällen handelt es sich um besondere phonologische Strukturen, die gleichzeitig eine außerphonologische Sonderklasse bilden, z. B. Interjektionen, Kinderstubenwörter oder Lehngut. Wir brauchen deshalb das Lehnwort *teenager* oder die Interjektion *aha* weder aus dem Deutschen hinauszuwerfen noch sie dem zentralen deutschen Lautsystem zuzuordnen. Diese Ausdrücke gehören innerhalb des Deutschen je einem besonderen, peripheren System zu. Hierher stellen wir nicht etwa alle Lehnwörter oder alle Interjektionen. Wenn sie keine phonematischen Besonderheiten aufweisen, so ist das Kriterium Nr. 1 nicht erfüllt, und die betreffenden Wörter ordnen sich dem zentralen System ein, z. B. das Lehnwort *achtern* oder die Interjektion *donnerwetter*.

Die peripheren Systeme können entweder außerhalb oder innerhalb des zentralen Systems existieren oder sich auch mit ihm überschneiden. Die Strukturen gewisser Chemikaliennamen bilden eine Untermenge der deutschen Phonemstrukturen überhaupt, d. h. des zentralen Systems. Mit russischer Aussprache gesprochenes *Berezina* enthält kein einziges deutsches Phonem. Die Ammenform /'ba/ besteht aus deutschen Phonemen, jedoch in einer im zentralen System unzulässigen Silbenstruktur.

Unser Doppelkriterium setzt nicht voraus, daß wir das zentrale System als solches schon im voraus postulieren. Wir können zunächst die gesamte Rede eines Informanten innerhalb eines Systems analysieren. Dann wird sich herausstellen, daß bestimmte, schon außerphonologisch abgrenzbare Bereiche gleichzeitig phonematische Besonderheiten aufweisen. Im allgemeinen werden diese Bereiche eng begrenzt sein. Im Grenzfall können jedoch mehrere Sondersysteme syntaktisch und semantisch etwa gleichen Umfang haben. Dann haben wir eine mehrsprachige Versuchsperson vor uns.

Martens (Hamburg): Es scheint völlig klar, daß man zwischen einem zentralen und einem peripheren Phonemsystem unterscheiden sollte. Es ist auch einleuchtend, daß die Entscheidung oft außerordentlich schwer fallen dürfte. In vielen Fällen aber kann man der Gefahr der Aufsplitterung in viele periphere Systeme entgehen, indem man einfach das zentrale System betrachtet im Gegensatz zu allen anderen «peripheren» Phonemsystemen. Es wird sich dabei ohnehin vielfach nicht um Phoneme, sondern um Phonemfolgen und Akzentverteilungen handeln, die im zentralen System nicht üblich sind. Das ist dem Durchschnittssprecher des Deutschen meistens durchaus klar. Genau das ist es ja auch, was die pharmazeutische Industrie veranlaßt, nichtdeutsch klingende Warenbezeichnungen zu wählen: es sind besondere Phonemfolgen oder Akzentverteilungen, die im zentralen System nicht üblich sind. Es soll (zum Teil) altehrwürdig «griechisch», zum Teil «modern» klingen. Das gibt ein «Verkaufsprestige».

Kiparsky (Helsinki): Fremdwörter gehen durch die «Fleischmaschine» der monoglotten Personen, die es noch gibt, und werden erst nach vollständiger Assimilation zu Lehnwörtern, in denen es nur echt einheimische Phoneme geben kann. Neue Phoneme können durch bloße Entlehnung nicht in die Sprache hineinkommen, wenn sie nicht bereits in dem einheimischen Lautsystem entstanden sind.

Daneš (Prag): Ich meine, daß wir nicht berechtigt sind, über ein zentrales und ein peripheres System einer gegebenen Sprache zu sprechen, sondern nur über Zentrum und Peripherie eines Systems; denn die Peripherie wird gerade dadurch charakterisiert, daß sie 'wenig systematisch' ist, daß sie in dem gegebenen System nicht völlig integriert ist. – Und was die Schwierigkeiten mit der Abgrenzung zwischen Zentrum und Peripherie betrifft, handelt es sich um einen Irrtum: Es gibt keine! Das Zentrum geht stufenweise in die Peripherie über ohne irgendeine scharfe Grenze; man könnte nur von einer Übergangszone zwischen beiden Komponenten sprechen. (Dasselbe gilt natürlich für das ganze Sprachsystem.)

Eva Sivertsen (Oslo): Mr. *Pilch* has made an important attempt to find some method or procedure for eliminating from our material certain elements which we would prefer not to include in our analysis, such a method must be found, since we cannot, as Mr. *Pilch* has pointed out in the previous discussion, simply decide not to consider 'loan words': what would be left of our vocabulary?

If I understand Mr. *Pilch* correctly, he suggests that we should use two criteria: one linguistic and one extra-linguistic, or perhaps one formal and one semantic. Using two independant sets of criteria seems a promising attack on the problem.

I would like to point out one possible difficulty, however. By applying these criteria we might come out with a number of sub-systems of a different nature from those that Mr. *Pilch* has in mind. In English, for example, words containing the verbal suffix *-t/-d* may have consonant clusters which occur in no other words. Tone languages may have special tonal patterns for certain grammatical categories, etc. This means that the application of these or similar criteria to our material might lead to a multiplication of systems, and reduction of our overall system into a number of sub-systems. Where would then our "central" system be?

Incidentally, I would like to point out the parallelism between this approach and that of the late *J. R. Firth* and his London school. Only, in the latter case there is no central system, if I understand this plurisystemic theory correctly.

Nickel (Kiel): 1. Hinweis auf die Schwierigkeit der Abgrenzung von zentralem und peripherem Sprachbereich allgemein (vgl. das ähnliche Problem in der Semantik, wo «Kernbedeutung» und «periphere Bedeutung» oft ebensowenig zu trennen sind). Die Abgrenzung setzt eine klare Kenntnis des zentralen Bereiches voraus. Mit welchen Methoden soll sie erlangt werden? So ist das Kriterium der Zugehörigkeit zu einer 'Sonderklasse', zu denen *H. Pilch* u. a. den Slang rechnet, auf Grund seiner unterschiedlichen Distribution in verschiedenen Gesellschaftsschichten kaum zuverlässig. Ähnliches gilt für Fremdsprachen, die oft eng mit den Sprachen ganzer Gruppen verknüpft sein können und da eine zentrale Position einnehmen, für außerhalb dieser Gruppen stehende aber peripher sind.

2. Hinweis auf die Schwierigkeiten der Trennung von 'Sprache' und 'Rede mehrerer Sprecher'. Wann wird z. B. ein 'Fremdwort' mit fremden Phonemen und Phonemfolgen Bestandteil der 'allgemeinen Sprache'? *Hockett* z. B. sieht den Phonembestand seiner Sprache schon dann erweitert, wenn eine Minorität fremde Phoneme zusammen mit Fremdwörtern übernimmt.

3. Die Richtung ist nicht immer die von der Peripherie zum Zentrum hin. Innerhalb des Wortschatzes z. B. werden auch 'zentrale' Wörter in den Bereich der Peripherie versetzt, wo sie dann oft als Archaismen ein literarisches 'Mauerblümchendasein' führen.

Fischer-Jørgensen (Kopenhagen): Es gibt zweifellos Sprachen, deren phonematische Struktur am besten dadurch beschrieben wird, daß man ein zentrales System von den peripheren Erscheinungen unterscheidet. Man müßte aber versuchen, ob man nicht dabei das extralinguistische Kriterium der Entlehnung vermeiden könnte. Jedenfalls sollte man es nur so verwenden, daß man bei der Aufstellung der strukturellen Züge, die man als peripher betrachten möchte, mit in Betracht zieht, daß die meisten nicht assimilierten Fremdwörter dadurch aus dem zentralen System ausgeschlossen werden; man sollte aber nicht das Kriterium für jedes einzelne Wort heranziehen, denn man könnte ja auch wünschen, einzelne einheimische Wörter (z. B. im Deutschen *Hermelin*) zu den Wörtern mit peripherischer Struktur zu rechnen.

Tschizewsky (Heidelberg): Es gibt objektive Grenzen für Übernahme der fremdsprachigen Phoneme in das System der Muttersprache. Selbst die slavisch sprechenden Subjekte sind nicht imstande, manche Phoneme aus anderen slavischen Sprachen zu übernehmen, so etwa Ostslaven die tschechischen langen Vokale ř, h oder silbenbildende r, l. Ähnlich steht es in den USA mit «th». – Objektive Grenzen des «Zentralsystems» sind meistens in manchen Gebieten feststellbar. Und dieses System ist ein «Filter», der nur gewisse fremde Elemente durchläßt.

Vachek (Prag): Schon im Jahre 1934 hat *Mathesius* gezeigt, daß nicht nur eine phonologische Assimilation der Fremdwörter an das einheimische Wortmaterial geschieht, sondern auch eine ganz entgegengesetzte Tendenz, die darin besteht, die Fremdwörter als solche durch phonologische Fremdsignale zu unterstreichen. – Man sollte auch diachronistische Fremdheit der lexikalischen Elemente von ihrer synchronistischen Frendheit sorgfältig scheiden – für die Entscheidung über die phonologische Fremdheit ist nur die letzte maßgebend.

Pilch (abschließende Stellungnahme): Fräulein *Sivertsens* Hinweis halte ich für sehr wesentlich. Vielleicht sollte man tatsächlich eine Vielzahl peripherer Systeme analysieren, wenn sie da sind. Sollte dann das zentrale System drohen sich aufzulösen, so müßten wir nach neuen Lösungen suchen. Für extralinguistisch halte ich die genannten Kriterien zwar nicht. Das ist aber wohl mehr eine Frage der Definition. Mir würde es – aus schon genannten Gründen – nicht einleuchten, sollte man willkürlich gewisse phonematische Eigenschaften als fremd erklären. Die hier häufig genannte Nichtanfangsbetonung im Deutschen (*Wachólder*) halte ich dazu schon deshalb für ungeeignet, weil es einfach sachlich nicht stimmt, daß das Deutsche immer Anfangsbetonung hat. Sogar Ortsnamen sind häufig nicht anfangsbetont (*Holtenáu, Labóe*). Von peripheren Systemen spreche ich deshalb, weil diese in sich ebenso strukturiert sind wie das zentrale System (Ammensprache, Arzneimittel, engl. Lehnwörter einer bestimmten «Entlehnungsschicht»). Sprecher, die *slang* als Normalsprache sprechen, kenne ich nicht. Auch diejenigen, die alle englischen Substantive in *slang* mit /šm/ anlauten lassen, sprechen so nicht den ganzen Tag.

Verh. 5. int. Kongr. Phon. Wiss., Münster 1964, pp. 474–477
(S. Karger, Basel/New York 1965).

Akzentoppositionen im Serbokroatischen

Von Karl-Heinz Pollok, Göttingen

Die Opposition zwischen den vier Akzenten des Serbokroatischen wird traditionell als eine Opposition zwischen steigenden und fallenden Akzenten (jeweils Kürze und Länge) angesehen. In der Tat zeigen experimentalphonetische Untersuchungen für die beiden «steigenden» Akzente einen steigenden Tonverlauf, für den «lang fallenden» Akzent einen überwiegend fallenden Tonverlauf. Bezüglich des «kurz fallenden» Akzentes hingegen konnten experimentalphonetische Untersuchungen den erwarteten «fallenden Tonverlauf» bisher nicht bestätigen[1]. Bei meinen eigenen, mit Hilfe des Tonhöhenschreibers unternommenen Untersuchungen konnte ich konstatieren, daß diesem Akzent ein spezifischer, für ihn als charakteristisch anzusehender Frequenzverlauf zukommt. Die trägheitsfreie Aufzeichnung des Tonhöhenschreibers zeigt bei diesem Akzent regelmäßig einen stoßtonartigen Verlauf: Nach einem anfänglichen kurzen steilen Anstieg der Frequenz bricht die Kurve um und läuft schließlich in einem kurzen Frequenzabfall aus. Dabei lassen sich zwei Varianten unterscheiden: In der weitaus größeren Zahl der Fälle (71 % des analysierten Materials[2]) ist der anfänglich steigende Ast der Kurve der längere, während nach dem Umbruch nur ein kurzer fallender Ast vorliegt (schematisch etwa). In etwa 25 % der Fälle hingegen ist umgekehrt der fallende Kurventeil der längere (schematisch etwa)[3]. Der Kürze halber bezeichne ich im

[1] Vgl. *Verfasser*, Zur Geschichte der Erforschung des serbokroatischen Akzentsystems, S. 279 f.

[2] Die Textlisten des der vorliegenden Arbeit zugrunde liegenden Materials sind abgedruckt in *I. Mahnken*, Die Struktur der Zeitgestalt des Redegebildes, S. 525–535. Untersucht wurden alle von Sprecher I und Sprecher IV gesprochenen Sätze sowie das zwischen S I und S IV geführte Gespräch.

[3] Bei den verbleibenden 4% der Wörter mit kurz fallendem Akzent (˵), in denen eine *nicht*-stoßtonartige Realisation der betonten Silbe festzustellen ist, handelt es sich durchweg um Fälle, bei denen das betreffende Wort seinen eigenen Akzent so gut wie ganz eingebüßt hat und sich enklitisch oder proklitisch an das benachbarte Wort anlehnt.

folgenden diese beiden Varianten als «steigenden Stoßton» bzw. «fallenden Stoßton». Es sei ausdrücklich vermerkt, daß in der auditiven Wahrnehmung kein Unterschied zwischen diesen beiden Varianten besteht, es sich also *nicht* um phonologische Varianten dieses Akzents handelt; vielmehr sind diese beiden Gestaltungsweisen ausschließlich durch den jeweiligen Gesamtablauf der Frequenzmodulation des betreffenden Redeabschnittes bedingt.

Die Tatsache des starken Überwiegens der Form des steigenden Stoßtons läßt sich ursächlich darauf zurückführen, daß der Akzent ˵ nur auf erster oder einziger Wortsilbe stehen kann und Wörter mit diesem Akzent nur in einer begrenzten Zahl von Fällen innerhalb einer komplexeren Akzenteinheit an nicht-erster Stelle stehen (in diesen Fällen erscheint in dem von uns analysierten Material die fallende Variante des Stoßtons[4]).

Es erweist sich somit, daß die beiden phonetischen Varianten des Akzents ˵ («steigender» bzw. «fallender» Stoßton) positionsbedingt sind (durch die prosodische Kontextbindung), das relevante Merkmal an der akustischen Gestaltung dieses Akzents aber zweifellos in dem Umbruch der Frequenzmodulation («Stoßton») gesehen werden muß. Die Opposition zwischen den beiden Akzenten ˵ und ˋ besteht also nicht in der Opposition eines fallenden und eines steigenden Tonverlaufs, sondern in der Opposition einer stoßtonartigen Frequenzmodulation (mit einem verhältnismäßig scharfen Umbruch von steigender zu fallender Bewegung) zu einer *nicht*stoßtonartig, d.h. gleittonartig (schematisch etwa) verlaufenden Frequenzveränderung. Daß dieser Gleitton beim Akzent ˋ de facto durchweg *steigend* verläuft, hat seine Ursache nicht zuletzt in der Tatsache, daß dieser Akzent nur auf nichtletzter Wortsilbe vorkommt.

Es darf hier vermerkt werden, daß sich auch bezüglich der Opposition der beiden langen Akzente des Serbokroatischen (⌒ und ´) eine Opposition zwischen den Tonverlaufs*typen* konstatieren läßt, obwohl bei diesen beiden Akzenten (infolge ihrer längeren Dauer) der insgesamt fallende bzw. steigende Charakter i.a. gleichfalls deutlich ausgeprägt ist. In dem engen mir hier zur Verfügung stehenden Rahmen ist es nicht möglich, dies eingehender darzulegen. Eine eingehendere Analyse der phonetischen Struktur der prosodischen Verhältnisse des Serbokroatischen zeigt jedoch, daß

[4] Beispiele dazu vgl. *Verfasser*, Der neuštokavische Akzent und die Struktur der Melodiegestalt der Rede, Kap. 4, § 8.

auch für die beiden langen Akzente eigentlich relevant die Opposition zweier Verlaufs*typen* ist, während die *Richtung* der Frequenzmodulation sekundären Charakter hat.

Wenn wir den kurz fallenden Akzent als stoßtonartigen Akzent bezeichnen, so bezieht sich das zunächst auf die akustische und nicht auf die auditive Ebene. Anstieg und Abfall werden in so kurzer Zeit vollzogen, daß diese, für den Akzent charakteristischen, Frequenzveränderungen in der auditiven Wahrnehmung nicht als Tonhöhen*bewegung* (innerhalb der akzentuierten Silbe) erfaßt und verfolgt werden können[5]. Der auditive Eindruck ist vielmehr einerseits maßgeblich beherrscht von dem Tonhöhen*unterschied* zwischen der akzentuierten und der ihr folgenden Silbe (woraus die Auffassung von der fallenden Natur dieses Akzents zu erklären ist), andererseits aber bestimmt durch die Schärfe bzw. Stärke dieses Akzents (die in einem Teil seiner Benennungen bzw. Definitionen [jaki, najoštriji, vrlo oštar] ihren Niederschlag gefunden hat)[6]. Diese spezifische Qualität der Schärfe bzw. Stärke der Betonung muß als das phonologisch relevante Merkmal dieses Akzents angesehen werden. Sie ist ursächlich offensichtlich auf die o.e. spezifische Art der Frequenzmodulation bei diesem Akzent (die stoßtonartige Bewegung) zurückzuführen. Bezüglich der serbokroatischen Akzentoppositionen haben wir es demnach mit der bekannten Erscheinung zu tun, daß bestimmte akustisch-physikalische Gegebenheiten in der auditiven Wahrnehmung als Qualitäten anderer Art erscheinen. Da aber ohne Zweifel als relevanter Unterschied zwischen den Akzenten ̏ und ̀ die Opposition zwischen den Tonverlaufstypen der stoßtonartigen und der gleittonhaften Frequenzmodulation angesehen werden muß, wird man auch in linguistisch-struktureller Hinsicht die Opposition zwischen den beiden Akzenten als eine Opposition zwischen Stoß- und Gleitton zu charakterisieren haben.

Literatur

Mahnken, I.: Die Struktur der Zeitgestalt des Redegebildes (Vandenhoek & Ruprecht, Göttingen 1962).

Pollok, K.-H.: Zur Geschichte der Erforschung des serbokroatischen Akzentsystems. Die Welt der Slaven *2:* 267–292 (1957).

[5] Vgl. dazu *O. Broch* im Archiv für slavische Philologie, Bd. 37, S. 215: «Die Kürze des Vokals macht jedenfalls die Veränderung der Tonhöhe innerhalb seiner (selbst) für unser Ohr kaum wahrnehmbar und damit sogar fraglich.»

[6] Vgl. *Verfasser*, Zur Geschichte der Erforschung des serbokroatischen Akzentsystems, S. 267 ff.

Pollok, K.-H.: Der neuštokavische Akzent und die Struktur der Melodiegestalt der Rede (im Druck).

Adresse des Autors: Dr. K.-H. Pollok, Seminar für Slawische Philologie, Hospitalstraße 10, *34 Göttingen* (Deutschland).

Discussion

De Bray (Clayton): Mr. *Pollok's* special contribution to the phonetic analysis of the Serbocroatian accents lies not only in confirming the contrasts in the realization of the rising as compared with the falling tones, both the long pairs and the short, but particularly in pointing out the "Umbruch", the critical transitions from rise to fall in the realization of the short falling tone. In my own independent research on the same problem I have recorded and photographed the same phenomena both of the contrasts and of the rise-fall nature of the short falling tone in certain circumstances. *Fry* and *Kostić* in their "Serbo-Croat Phonetic Reader" also pointed out before the war the contrast in the *nature* of the stress in the two short accents ('Stoß-' as opposed 'Gleitton' in Mr. *Pollok's* terminology). But is it sufficient to be content with the initial statement concerning the rising tones? Words accented on any syllable except the first are considered always to have only one of the *rising* tones, ̀ or ́. If tone is in all cases phonemically significant in Serbocroatian, it would be important to establish that these rising tones on non-first syllables are in fact regularly realized phonetically as rises. My own observations have *not* produced evidence to support such an assumption. It would be interesting to know whether Mr. *Pollok* has arrived at the same conclusions. My own feeling is that tone is not phonemically significant in non-first syllables, though even in Belgrade Serbian it is normally realized at the phonetic level as a rise, e.g. planina, novìne, upozorávati, mašína, etc. in most circumstances, though not in all.

C.R. 5e Congr. int. Sci. phon., Münster 1964, pp. 478–481
(S. Karger, Basel/New York 1965).

Le rapport du rythme et du phonème en français

Par Pera Polovina, Belgrade

Certaines définitions traitent le phonème en phénomène se déroulant sur le plan de la langue et non pas sur le plan de la parole. D'autres définitions, l'envisageant d'un autre point de vue, prennent également en considération ses réalisations dans le discours. Elles contiennent toutes une part de vérité, mais elles n'ont pas encore mis en évidence tous les éléments qui pourraient rendre la définition du phonème satisfaisante.

Les définitions du rythme sont encore moins convaincantes. Elles tiennent compte, le plus fréquemment, de l'aspect acoustique du rythme, rarement de sa source génétique. La différence entre le rythme de la langue et le rythme de la parole, existant aussi bien que la distinction entre le phonème et sa réalisation, n'est non plus précisé dans les études du rythme.

Le rapport entre le rythme et le phonème, entre la réalisation de l'un et son influence sur la réalisation de l'autre n'est pas suffisament mis en relief. Les diverses réalisations du phonème sont traités souvent dans leur dépéndance de la chaîne parlée, mais, dans ce cas, on prenait en considération leur entourage immédiat: les sons voisins. Dans les analyses des phonèmes réalisés on se bornait à les expliquer par leur place, par rapport à l'accent et dans le cadre du groupe rythmique où il se trouve, ou bien simplement suivant leur position: au début, au milieu, ou à la fin du mot. L'analyse de ses réalisations n'est pas généralement faite en fonction de leur dépendance du contexte plus vaste, de l'élan rythmique de la phrase entière, et à fortiori du mouvement du passage entier et des faits psychologiques qui les produisent.

La langue française est très intéressante pour l'étude du rapport entre le phonème et le rythme et de l'interdépendance de leur réalisation. Par exemple, l'analyse de l'*e muet* et de sa réalisation ne serait

pas complète si elle envisageait seulement sa position dans le groupe rythmique, ou même dans le cadre d'une seule phrase. Sa réalisation ou nonréalisation se fait plusieurs fois sous l'effet du rythme du passage entier. Ce son sert quelquefois par son apparition ou par sa disparition aux besoins rythmiques: à l'équilibre de la phrase et même du passage.

La réalisation des sons latents, c'est-à-dire de ceux qui apparaissent dans la liaison, est aussi expliquée de différents points de vue: historique, social, phonique, etc., mais leur rapport avec le rythme est négligé. Or, si l'on dit que la liaison apparaît plus souvent dans le discours soutenu, le style élevé, ne faudrait-il pas étudier aussi quel est l'effet de son apparition, à quelle fin elle sert. Ne serait-ce pas aussi, entre autres, pour soutenir la sonorité de la phrase, ou une certaine forme du rythme continu?

En étudiant le rythme de 400 phrases liées dans trois textes de Camus par la méthode auditive et instrumentale (lues par 7 Français cultivés), et en prenant en considération l'effet du contexte entier du passage, nous avons remarqué certains indices de l'influence du rythme du contexte sur la réalisation des phonèmes en français. Le retour de l'accent rythmique (qui influence, bien sûr, la réalisation du phonème) se fait de la façon suivante: Dans le texte indifférent l'accent rythmique tombe sur les mêmes voyelles, dans la lecture de tous nos lecteurs, dans 89% de cas. Dans le texte affectif cet accent frappe les mêmes sons dans 76% de cas. D'autre part les groupes rythmiques, en raison du rythme du contexte entier, sont plus courts dans le texte contenant moins de sensibilité, tandis que les groupes rythmiques sont plus longs si le texte exprime l'enthousiasme, l'admiration ou l'amour. Le pourcentage de décalage n'est donc pas trop grand dans le texte dénué d'éléments affectifs, et il est un peu plus grand dans le texte imprégné de sentiments. Pourtant ce pourcentage n'est non plus si petit pour ne pas attirer l'attention des phonéticiens.

La place de l'accent rythmique dans la diction de nos sujets n'est pas la chose essentielle que nous voulions mentionner. L'analyse d'un contexte plus vaste que la phrase a indiqué que l'accent d'insistance dans la lecture de tous nos Français ne tombaient pas sur les sons d'une façon tout à fait individuelle, et qu'il ne se promenait pas de manière arbitraire. Sa place est déterminée dans plusieurs cas par le contexte. Dans la lecture de nos sujets, cet accent tombe sur les mêmes mots dans 13,6% de cas, c'est-à-dire sur les mots que le contexte du passage met en évidence et impose à tous les lecteurs. D'autre part il tombe sur le même son de ces mots (et ce n'est pas toujours la consonne) dans 71 % de cas, tandis que si l'on observe la place de cet accent dans tous les mots qui sont frappés par lui (y compris les mots que le contexte ne met pas en evidence), on se rend compte qu'il frappe les mêmes sons dans 9,6 % de cas. Il serait, par conséquent intéressant d'étudier davantage l'apparition de cet accent et son influence sur la réalisation des phonèmes suivant le contexte et son rythme; d'analyser son rôle rythmique.

Si les accents et leur place déterminent, dans une certaine mesure, la réalisation des phonèmes, les phonèmes de leur côté, par leur structure et leur distribution, propres à la langue française, influencent le relief et la durée du rythme des syllabes, des phrases et du contexte. Le système combinatoire des sons français a été étudié sur le plan des mots, mais l'étude de la combinaison des sons dans les transitions entre les mots qui forment les groupes rythmiques n'a pas été faite, si nous ne nous abusons. Nous avons observé la combinaison des sons qui se suivent quand les différents mots forment les groupes rythmiques dans les textes mentionnés et dans certains dialogues de la langue courante. Nous avons remarqué que les syllabes formées par suite du contact entre les mots constituant les groupes rythmiques, sont composées comme suit: consonne + voyelle, ou bien voyelle + consonne. Ces dernières sont plus fréquentes. Si, au milieu du groupe rythmique il y a rencontre de certaines consonnes il apparaît souvent entre elles un *e muet*, par exemple entre *d* et la consonne suivante, *ʒ* et la consonne suivante, *k* et la consonne suivante, mais moins cependant après *m*, *n* et les autres consonnes. Les consonnes *l* et *r* se combinent très facilement avec les autres consonnes et ce sont elles qui se trouvent justement le plus souvent dans les combinaisons avec les autres consonnes aux transitions entre les mots formant les groupes rythmiques. Une autre distribution des sons caractéristique au rencontre des mots qui créent le groupe rythmique est faite par la suite de *e* (ouvert ou fermé) + consonne. Cela provient, nous semble-t-il, de l'emploi fréquent des formes verbales terminant en *e*, *ε*, des mots *et*, *mais*, des articles au pluriel, de l'emploi fréquent du verbe *être*, etc. Cette distribution des sons, influence le cours du rythme et ses formes surtout là, où, pour d'autres raisons, on lie ou on ne lie pas les mots dans un même groupe rythmique (déterminé + déterminant, conjonctif + un autre mot). Le système de distribution des sons dans la langue française soumet à

ses tendances certaines réalisations de groupes rythmiques, et par conséquent du rythme de la parole en général – de même que le tempo et les autres facteurs du rythme influencent les réalisations des sons.

Notre communication ne prétend pas donner les solutions du problème du rapport entre le rythme et le phonème en français, elle ne fait qu'indiquer la valeur du contexte dans l'étude du phonème. Celui-ci se réalise sous une forme détérminée, dans une certaine mesure, en fonction du contexte du passage entier. Pour la découverte des tendances provenant de l'interdépendance du phonème et du rythme, étudiée dans le contexte plus vaste que la phrase il faudrait ajouter d'autres recherches. Nous n'apportons que certaines observations dont le but serait de suggérer l'aspect du problème et de la méthode.

Adresse de l'auteur: Dr Polovina Pera, Doc., Bulevar Revolucije 76, *Belgrade* (Jougoslavie).

Verh. 5. int. Kongr. Phon. Wiss., Münster 1964, pp. 482–485
(S. Karger, Basel/New York 1965).

Die Entwicklung vom Laut zum Phonem in der Kindersprache

Von Walburga von Raffler Engel, Florenz

Beobachtungen an verschiedenen Kleinkindern und dann eine spezifische Studie meines jetzt drei Jahre und zwei Monate alten Sohnes mit fast täglichen schriftlichen Annotationen seit seinem ersten Schrei ergeben Daten, die bekannte Tatsachen bestätigen und Neues ans Licht bringen, insbesondere die Wichtigkeit der Summlaute.

Schreien und Weinen sind wohl Ausdruck, aber nicht Sprache. Sie dauern weiter an, wenn auch die Sprache vollkommen perfekt ist. Gurgellaute sind auch keine Vorläufer der Sprache. Sie werden fortgesetzt durch Laute des Wohlgefallens.

Die Lallaute dagegen erscheinen als regelrechte Sprachversuche. Sie verschwinden, wenn die Sprache gemeistert ist, und erscheinen nur bei Sprachschwierigkeiten, wie Stottern oder dem Erlernen einer Fremdsprache wieder. Außerdem treten die Lallaute zu einer Zeit auf, wo das Kind schon einiges Gesprochenes versteht. Die aktive Sprache entsteht sehr schnell nach der passiven, nur braucht letztere bedeutend länger, um sich zu entwickeln.

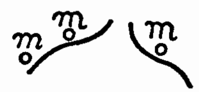
Wenn nach der Periode des Summens, der erste *m*- oder *b*-Laut auch wahrscheinlich noch kein präzises Wort sagen will, so erscheint er mir doch als ein bewußter Versuch, sich sprachlich auszudrücken oder wenigstens die Sprache nachzuahmen. Das erste wortartige Lautgebilde meines Söhnchens war *am*, wenn er hungrig war. Gleich eine Woche darauf, mit neun Monaten, sagte er sein erstes richtiggehendes Wort *pappa*, das italienische Wort für Essen (italienisch ist die erste Sprache meines Kindes). Er setzte sich in eine Ecke des Laufgitters und sagte das Wort immer klarer. Dann stellte er sich strahlend auf und sagte *la páppa*, dasselbe Wort mit Artikel, so wie er es gewöhnlich von uns hörte. Denselben konzen-

trierten Gesichtsausdruck des Wortprobierens hatte er bei seinen Lallauten gehabt und dasselbe verschmitzte Lächeln, wenn ich mich ihm in diesen Momenten näherte. Wenn man ein vier Monate altes Baby anspricht, bekommt man auch Lallaute wie eine Antwort zu hören.

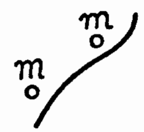
Die ersten Worte eines Kindes halte ich nur für Nachahmung des Gehörten; aber schon zwei Monate nachdem mein Sohn sein erstes Wort hervorgebracht hatte, konnte ich feststellen, wie er meine Lippenbewegungen beobachtete. Kurz darauf vereinigte er Ohr- und Augenaufmerksamkeit und versuchte mit bewußter Anstrengung seine «Worte» richtiger auszusprechen. Wir haben also zuerst den Laut ohne Bedeutung und dann das volle Wort. Das Wort ist zuerst rein Ohr bedingt und dann auch Auge bedingt (Lippennachahmung), um dann gleichzeitig von Ohr und Auge abzuhängen.

Mit dem ersten Wort bzw. dem zweiten, je nachdem, ob semantische Erwägungen in der Phonemdefinition ein- oder ausgeschlossen werden, sind wir schon auf dem Gebiet der Phonemik, und zwar sind wir auf dem globalen Gebiet der Phonemik, der segmentalen sowie der suprasegmentalen, die von meinem Sohn gewiß nicht als verschieden empfunden worden sind. Nachdem der Kleine schon einige Zeit Mamma (mámă) sagen konnte, kam er um seinen Vater (Papà) zu rufen mit mamá heraus. Erst danach korrigierte er allmählich das *m* bis es *p* wurde. Hier war der erste phoneto-phonemische Unterschied nicht in der Artikulation, sondern im Tonfall. *Die melodischen Faktoren erschienen also vor den artikulatorischen.* Als ich diese bemerkt hatte, kam mir ins Gedächtnis, daß jedes bis dahin gesagte Wort seine melodische Eigenheit gehabt hatte. Der Name unseres Hundes, Achille, war als ein trisyllabisches *a* herausgekommen.

Dann dachte ich zurück an die Zeit vor dem Eintreten der Lalllaute. In der Periode zwischen den Gurgellauten und den Lalllauten kann man cooing, eine Art labiales Summen beobachten. Dieses Summen, was nichts mit Singen zu tun hat, wurde sehr schnell satzmelodisch. Meiner Ansicht nach ist bis jetzt diesen Summlauten nicht genug Wichtigkeit zugeschrieben worden. (Hier möchte ich kurz bemerken, daß mein Sohn, der offensichtlich ein ausgezeichnetes Sprachgefühl besitzt, leider aber reichlich unmusikalisch ist.) Der Kleine – wie alle Kinder seines Alters – hielt summend ganze Monologe mit Fragen und Antworten usw.:



Nicht lange nach dieser Summperiode gebrauchte mein Sohn die nasale Bilabiale als ein klares und bewußtes Ausdrucksmittel: Auf die Dinge, die er mir zeigen wollte, deutete er mit dem Finger und sagte m; wenn er fragte, ob er sie haben dürfte, ging der Ton aufwärts



Nach dem Saugen, wenn die Lippennerven phonetisch artikulieren können und das bilabiale *m* oder *b* (als vokalisierten Konsonant *m*, *b*, oder zusammen mit šwa, *əm* usw., oder frikativisiert wie *bw* usw.) formen, tritt als erste Spracherscheinung die *Satzmelodie* auf. Danach die *Wortmelodie*. Dann kommt dazu der variierte *Laut*, der zum Wort ausgearbeitet wird. Das Wort hat gleich den ihm eigenen Ton. Tonstufe und Tonstärke, die ja sehr verflochten sind, erscheinen gleichzeitig. Das schwierige Wort Papà hatte in allen seinen Entwicklungsstufen stets den Akzent auf der letzten Silbe, die auch stets den richtigen Ton hatte. Mit elf Monaten war es z.B. *da: dá*. Niemals wurden die Worte Papà und pappa verwechselt. Das Wort pappa wurde mit neun Monaten fehlerlos ausgesprochen, während der Kleine dreizehn Monate alt war, bevor er korrekt Papà sagen konnte. Meiner Ansicht nach tritt also das segmentale Phonem erst nach dem suprasegmentalen auf.

In einer mehr fortgeschrittenen Sprachperiode haben wir dann allmählich ein immer größeres Inventar der segmentalen Phoneme, eventuell zuerst ohne Allophone. Es handelt sich noch um eine Phonemik sui generis. Das Kind beherrscht nur wenige Laute, und diese mögen phonetisch falsch sein; aber seine Laute sind so organisiert, daß sie sich semantisch unterscheiden. Diese Feststellung zeigt übrigens wieder, daß man eigentlich doch nicht von der Bedeutung absehen kann, wenn man sprachliche Probleme studiert.

Kurzgefaßt sind die Entwicklungsstufen folgende: Der erste, noch ausdrucklose, rein phonetische Sprachversuch ist die Bilabiale. Danach kommt die Satzmelodie, deren Sinnbedeutung schwer zu klären ist. Dieser folgt die melodie-phonemische Bilabiale (*m* sagend

und auf Dinge deutend) und dann erscheint das melodisch-artikulatorische Wort bzw. «Satzwort».

Wie bei dem Beginn des Verstehvermögens scheint also auch bei der Sprache der Weg vom großen Allgemeinen zur Einzelheit zu führen: die Nachahmung des Wortes erscheint vor der des Lautes.

Adresse der Autorin: Frau Prof. Dr. Walburga von Raffler Engel, Via Bolognese 269, *Firenze* (Italien).

Verh. 5. int. Kongr. Phon. Wiss., Münster 1964, pp. 486–489
(S. Karger, Basel/New York 1965).

Department of Germanic Languages, University of Washington, Seattle

Informationstheoretische Berechnungen der phonemischen Bedeutungsfunktion

Von C. Reed, Seattle, Wash.

Der klassische Begriff des Phonems entwickelt sich aus der Zusammenstellung von verschiedenen Inventaren der sprachlichen Gegensätze, die vom Standpunkt der phonetischen Merkmale aufgefaßt werden. Diese Inventare muß man zunächst aus streng definierten sprachlichen Umgebungen aufstellen. Die Identifizierung der Allophone und deren Klassifizierung in einer Minimalliste der Phoneme lassen sich alsdann entweder durch physiologische oder durch akustische Kriterien verwirklichen. Dabei bemerkt man sofort, daß gewisse Phoneme in diesem oder jenem Inventar keine Vertreter haben, daß auch gewisse phonetische Merkmale diesem oder jenem Inventar gänzlich fehlen.

Wie *Friedrich Kolmar-Kulleschitz*[2] schon erklärt hat, ist das Phonem die kleinste Einheit der semantischen Struktur, die tatsächlich alle Sprachebenen durchzieht. Er spricht vom «semantischen Gehalt der Phoneme», der «sich erst durch Vertauschung eines Phonems in einem Morphem als ‚Andersheit' erweist» (S. 73). Das Phonem hat also keine Bedeutung an sich, trägt jedoch zur spezifischen Bedeutung eines größeren Kontextes bei. Seine Leistung kann in diesem Sinne als eine bedeutungsunterscheidende bezeichnet werden.

Bei der phonemischen Analyse befaßt man sich vor allem mit Problemen der phonemischen Verteilung – mit dem statistischen Aspekt der phonologischen Erscheinungen. Die Klassifizierung der sogenannten Varianten in einem allesumschließenden Inventar läßt schon erkennen, daß nicht alle belegbaren Gegensätze in allen phonemischen Umgebungen auftreten. Am Ende der Silbe oder des

Wortes treten im Deutschen zum Beispiel keine Gegensätze zwischen stimmlosen und stimmhaften Verschluß- oder Reibelauten auf: der Gegensatz zwischen /s/ und /z/ besteht nur zwischenvokalisch (/raisen/:/raizen/). Das Phonem /h/ kommt nur am Anfang einer Silbe vor betontem Vokale vor; /ŋ/ nur nach gewissen Vokalen. Einige Konsonanten nehmen an Konsonantenverbindungen keinen Teil. Das alles gehört zur Beschreibung der phonologischen Struktur – zur *langue* – und steht im Zusammenhang mit den statistischen Beschränkungen der Sprache überhaupt.

Obgleich *Trubetzkoy* in seinen Bemerkungen über die phonologische Statistik auf die Stabilität der Häufigkeitsdaten verschiedener Sprachen (vornehmlich des Deutschen) hingewiesen hat, blieb er bei der de Saussureschen Dichotomie und behandelte die Sache als Gegenstand der *parole*[5]. Es ist das Verdienst *G. Herdans*, dieses Mißverständnis endlich aufgeklärt zu haben. Denn die Struktur einer Sprache besteht, wie er meint, nicht nur aus phonetischer Stabilität, sondern auch aus einer Stabilität des Sprachgebrauchs ([1], S. 79).

Informationstheoretische Untersuchungen der phonemischen Bedeutungsfunktion behandeln nicht nur die Wahrscheinlichkeiten der verschiedenen Möglichkeiten – je nach den stellungsbedingten Varianten –, sondern auch die Beschränkungen der grammatischen Struktur, wie sie sich auf längere Phonemkombinationen beziehen. Nach einem beliebigen Anfang der sprachlichen Äußerung hängt alles weitere von den kumulativen Möglichkeiten des Vorhergesagten ab. Die Struktur des Wortes, des Satzteils und des Satzes werden durch progressiv wechselnde Möglichkeiten der Phoneminventarvarianten charakterisiert. Dabei spielt das Phonem eine zweifache Rolle: erstens als Signal einer «Andersheit», zweitens als eine Art Rückkopplung gegen sprachliche Entropie.

Zwischen einer maximalen Entropie (d.h. «Ungewißheit» der verfügbaren Möglichkeiten) und völliger Redundanz (d.h. Aufhebung der anderswo auftretenden Unterscheidungen) erkennt man stufenweise eine ganze Reihe von Zwischenstadien, deren Informationsinhalt jedesmal stochastisch aus dem schon Gegebenen hervorgeht. Die Bedeutungsfunktion des einzelnen Phonems hängt also nicht nur von seiner phonologischen Umgebung ab, sondern auch von der ganzen Sprachform, ja sogar dem ganzen Sprachinhalt.

Kommunikationsforscher haben wiederholt betont, daß der Begriff «Information» vom mathematischen Standpunkt aus gesehen nichts mit dem der herkömmlichen Semantik zu tun habe ([3], S. 155 bis 168; [4]). Wollen wir aber diesen Begriff als Modellvorstellung gelten lassen, so dürfen wir wenigstens auf die Möglichkeit hoffen, daß sich eine Berechnung der semantischen Belastung jedes Phonems durchführen ließe. Mit einem «Mittelwert» der Information in diesem Sinne ist doch nicht geholfen. Denn Phoneme haben bekanntlich verschiedene Funktionen. Man denke z.B. an das deutsche /ə/, das nur in unbetonten Silben (einfach als Silbenträger) vorkommt. Es ist das häufigste Phonem (10 bis 12%, sogar 14% bei sehr umständlicher Rede), funktioniert hauptsächlich nur auf grammatischer Ebene und hat im allgemeinen keine vokalische Opposition. Seine Häufigkeit in den betonten Silben ist Null; von einem Mittelwert zu reden wäre also irreführend.

Ähnlich verhalten sich die häufigsten Konsonanten /n/, /r/, /t/ und /s/, die öfters von den Flexionsendungen in Anspruch genommen werden. Dort sind sie manchmal vollständig überflüssig, d.h. «redundant», und haben demnach keinen «Informationswert». Das gilt besonders für /n/ und /t/ im Partizip Perfekt sowie /n/ in der Deklination des Adjektivs usw. In der deutschen Umgangssprache kann es sogar geschehen, daß die Opposition zwischen Endungs-/n/ und -/r/ aufgehoben wird. Jedenfalls funktionieren die häufigsten deutschen Phoneme bezüglich der Bedeutungsunterscheidung zum größten Teil auf einer grammatischen Ebene.

Um eine gründliche Untersuchung dieses Problems zu unternehmen, müßte man zunächst eine genaue Analyse der sprachlichen Elemente (der Phonologie, Morphologie und Syntax) durchführen. Zweitens bedarf es einer vollständigen Wahrscheinlichkeitsrechnung dieser Elemente nach ihren grammatisch geregelten Stellungen. Dazu braucht man im Deutschen z.B. neue Häufigkeitsperspektiven, die nicht mehr die alte Wort- und Buchstabenzählung zum Maßstab haben, sondern die Lautung selbst und deren wissenschaftlich erschlossene Einheiten. Solch eine mühsame Arbeit würde sich reichlich lohnen, denn aus derartigen statistischen Studien dürfte man sehr einleuchtende Aufschlüsse über die Struktur, Leistung und geschichtliche Entwicklung der Sprache erhalten.

Literatur

1. *Herdan, G.:* Language as choice and chance (P. Noordhoff N.V., Groningen 1956).
2. *Kolmar-Kulleschitz, F.:* Ist das Phonem ein Zeichen? Phonetica *5:* 65–75 (1960).

3. *Schmetterer, L.:* Sprache und Informationstheorie. Sprache und Wissenschaft (Vandenhoek u. Ruprecht, Göttingen 1960).
4. *Shannon, C. E.* and *Weaver, W.:* The mathematical theory of communication (University of Illinois Press, Urbana 1949).
5. *Trubetzkoy, N. S.:* Grundzüge der Phonologie. TCLP VII, Prag 1939.

Adresse des Autors: Prof. Carroll E. Reed, Department of Germanic Languages, University of Washington, *Seattle 5, Washington* (USA).

Verh. 5. int. Kongr. Phon. Wiss., Münster 1964, pp. 490–494
(S. Karger, Basel/New York 1965).

Aus dem Institut für Phonetik der Universität Münster

Realisationsgrad und Verbindlichkeit lautlicher Distinktionen

Von Helmut Richter, Münster

Das Ziel der Phonometrie, Sprachen hinsichtlich lautlicher Distinktionen zu differenzieren[1, 4], läuft darauf hinaus, die funktionelle Bedeutsamkeit der untersuchten Distinktion in einer Reihe kommunikativer Systeme zu vergleichen. Die Funktion einer lautlichen Distinktion manifestiert sich in konkreten sprachlichen und situativen Kontexten. Wir definieren daher als *Realisationsgrad* einer (binären) Distinktion die absolute oder relative Differenz zwischen M_i, dem i-ten Meßwert zu einer phonematischen Distinktionsklasse, und m_k, einem kontextuell benachbarten[2] Meßwert zur anderen phonematischen Distinktionsklasse: $r = M_i - m_k$ bzw. $r_{rel} = (M_i - m_k)/m_k$*.

Mit dieser Begriffsbildung ist ein Anschluß an die geläufige phonematische Fragestellung gewonnen, ob eine Distinktion in bestimmten Umgebungen aufgehoben ist. Die formal-systematische Aufhebbarkeit der Distinktion erscheint allerdings lediglich als Sonderfall einer generellen Variabilität der *Verbindlichkeit* lautlicher Distinktionen in Abhängigkeit vom Kontext. Wir zögern nicht, auch die verschiedenen *Grade* der Distinktionsrealisierung als Ausdruck wiederum graduell verschiedener Verbindlichkeiten der Distinktion zu betrachten.

Freilich ist es nicht möglich, aus einem Realisationsgrad die Verbindlichkeit der Distinktion zu erschließen. Die aussagenlogische Verbindung 'r' → 'v', also die Implikation der Aussage, daß eine Verbindlichkeit des Grades v vorliegt ('v'), durch die Aussage, daß ein Realisationsgrad r vorliegt ('r'), ist falsch, da Zeile 2 (Spalte 3)

* Beim Zwirnerschen Quantitätsquotienten wird zu dieser Größe 1 addiert.

der Wahrheitsmatrix zutrifft. Die Realisierung einer Distinktion kann an einer bestimmten Stelle einer Äußerung den Eindruck der Überprägnanz genau dadurch hervorrufen, daß der Sprecher – etwa zwischen [stimmhaft] und [stimmlos] – einen ebenso scharfen Unterschied macht wie an einer anderen Stelle, «wo es viel mehr darauf ankam». Die Verbindlichkeit ist keine *notwendige* Bedingung des Realisationsgrades der Distinktion.

Da sich auch die umgekehrte Implikation 'v' → 'r' als falsch erweisen läßt, ist der jeweilige Grad der Verbindlichkeit auch keine *hinreichende* Bedingung des Realisationsgrades. Es genügt, sich einen Sprecher vorzustellen, der die gleiche Äußerung in gleichartigen Situationen mit dem einzigen Unterschied verschieden scharfer Realisierung einer Distinktion erzeugt, um einzusehen, daß Zeile 3 (Spalte 4) der Wahrheitsmatrix erfüllbar ist.

Tabelle I

Wahrheitsmatrix

'r'	'v'	'r' → 'v'	'v' → 'r'	'ř'	'v'	'ř' ⇄ 'v'
wahr	wahr	wahr	wahr	wahr	wahr	wahr
wahr	falsch	falsch	wahr	wahr	falsch	falsch
falsch	wahr	wahr	falsch	falsch	wahr	falsch
falsch	falsch	wahr	wahr	falsch	falsch	wahr

Diese logischen Schwierigkeiten fallen fort, wenn man den Begriff des *minimalen Realisationsgrades* einführt. Aussagen über bestimmte Größen ř des minimalen Realisationsgrades ('ř') sind Aussagen über die jeweilige Verbindlichkeit der Distinktion äquivalent: 'ř' ⇄ 'v'. Die graduelle Verbindlichkeit ist notwendige und hinreichende Bedingung des minimalen Realisationsgrades, dieser ein brauchbares Maß der Verbindlichkeit. Die Zeilen 2 und 3 (Spalte 7) der Wahrheitsmatrix treffen nicht zu, denn weder lassen sich zu verschiedenen Verbindlichkeiten gleiche Minima der Distinktionsrealisierung denken, noch zu gleichen Verbindlichkeiten verschiedene Minima.

Natürlich eröffnet nur die Statistik einen Zugang zu dieser Größe. Die theoretischen Verteilungen, die man zur Interpretation der beobachteten Variabilität heranziehen kann, bestimmen für jeden noch so kleinen Wert von r eine Wahrscheinlichkeit der

Unterschreitung. Offenbar stimmt diese Eigenschaft mit der linguistischen Realität gut überein. In Analogie zu der bei Signifikanzprüfungen üblichen Regelung, einen «Fehler erster Art» (*Neyman*[3]) in Höhe von 5% zuzulassen, beziehen wir den *minimalen statistischen Realisationsgrad* $\check{r}_{st}$ auf eine Unterschreitungswahrscheinlichkeit $P = 0{,}05$ als Kriterium. Dieses muß allerdings relativiert werden.

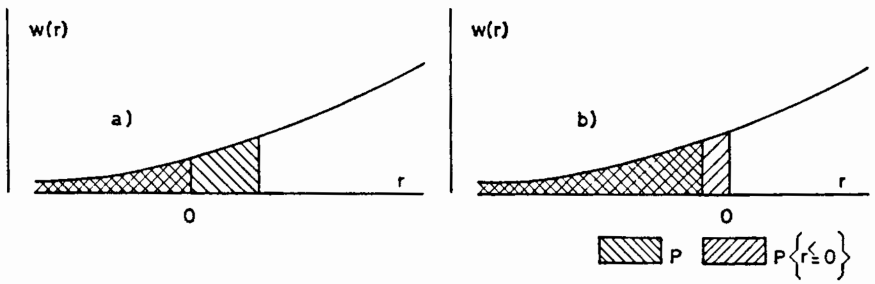


Abb. 1. Unterschreitungswahrscheinlichkeit P.

a) $P\{r \leqq 0\} < P$
b) $P\{r \leqq 0\} > P$

In der Praxis begegnen Verteilungen, wo $P\{r \leqq 0\}$, die Unterschreitungswahrscheinlichkeit von $r = 0$, größer ist als 5% (Abb. 1). Geht man aber davon aus, daß sich die Realisation selbst nur in $100 \cdot P$% aller Fälle unterhalb des Minimums bewegt, so kann der Fall $P\{r \leqq 0\} > P$ bei Lebendigkeit der Distinktion in der Sprachgemeinschaft überhaupt nur eintreten, weil durch den Erfassungsmodus der Realisationsgrade eine zusätzliche Ursache ihrer Variation entstanden ist. In praxi entsprechen dem Schwierigkeiten, die höchstzulässige Klammerbreite für kontextuelle Nachbarschaft festzulegen. Neben der P zugeordneten realisatorischen Unterschreitung muß eine *methodische Variabilität* der Realisationsgrade berücksichtigt werden.

Statt der unsinnigen Angabe eines unter 0 liegenden Realisationsminimums setzt man bei $P\{r \leqq 0\} > P$ den minimalen statistischen Realisationsgrad spontan mit 0 an. Damit hat sich aber das Wahrscheinlichkeitskriterium derart verändert, daß eine zusammengesetzte Unterschreitungswahrscheinlichkeit $P' = P + (P\{r \leqq 0\} - P)$ zugelassen wird. Nun könnte eine solche Korrektur «nach oben» auch in jedem Fall $P\{r \leqq 0\} < P$ angezeigt sein, ebensogut aber eine gegensinnige oder gar keine Korrektur. Die methodische Variabilität der Realisationsgrade kann sich nicht nur, wie bei

$P\{r \leqq 0\} > P$ offenkundig, im gleichen Sinn wie die realisatorische Unterschreitung auswirken, sondern auch im entgegengesetzten Sinn oder überhaupt nicht. Wir definieren daher einen *methodischen Sicherheitsbereich* zu beiden Seiten des minimalen statistischen Realisationsgrades $\check{r}_{st}$, dessen Grenzen durch Erhöhung und Verringerung von P bestimmt werden.

Es muß wiederum Sache der Übereinkunft bleiben, in welchem Ausmaß P verändert werden soll. Wir schlagen ein Kriterium $P \pm 0{,}025$ vor und geben die zu 0,025 bzw. 0,075 gehörigen Merkmalswerte als untere Bereichsgrenze r_1 bzw. obere Bereichsgrenze r_2 an, sofern sie größer oder gleich 0 sind. Negative Werte von r_1 und $\check{r}_{st}$ werden gleich 0 gesetzt.

Tabelle II. Methodische Sicherheitsbereiche zu $\check{r}_{st}$ in Abhängigkeit von $P\{r \leqq 0\}$.

a)	$P\{r \leqq 0\} \leqq 0{,}025$	$0 \leqq r_1 < \check{r}_{st} < r_2$
b)	$0{,}025 < P\{r \leqq 0\} \leqq P$	$0 \leqq \check{r}_{st} < r_2$
c)	$P < P\{r \leqq 0\} < 0{,}075$	$0 < r_2$
d)	$P\{r \leqq 0\} \geqq 0{,}075$	$r_2 \leqq 0$*

* Distinktion nicht verbindlich.

In diesen Vorschriften ist die spontane Korrektur $P + (P\{r \leqq 0\} - P)$ für die Fälle $0{,}05 < P\{r \leqq 0\} < 0{,}075$ enthalten (Tab. II; Abb. 2).

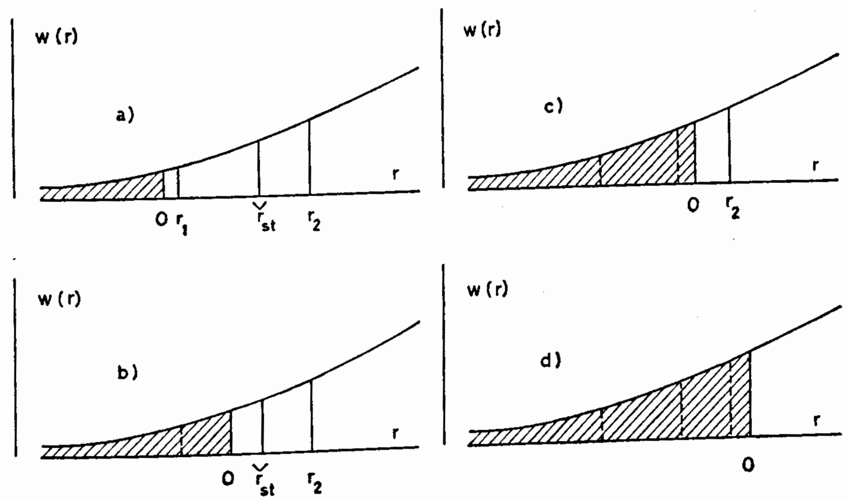


Abb. 2. Methodische Sicherheitsbereiche zu $\check{r}_{st}$ in Abhängigkeit von $P\,(r \leqq 0)$ (vgl. Tab. II).

Bei größeren Wahrscheinlichkeiten von $r \leqq 0$ verliert die Angabe eines Realisationsminimums zunehmend an Adäquatheit; aus der schematischen Anwendung der Korrektur $P + (P\{r \leqq 0\} - P)$ ist nicht auf massivere Effekte der methodischen Variabilität zu schließen, als sie mit $\pm 2{,}5\,\%$ angenommen werden. Mit anderen Worten: hat $r = 0$ insgesamt eine Unterschreitungswahrscheinlichkeit von $7{,}5\,\%$ und mehr, dann verwerfen wir die Hypothese, daß die Distinktion in der untersuchten Klasse von Umgebungen oder Situationen verbindlich ist.

Bei Gaußverteilung der Realisationsgrade z.B. ist

$$\check{r}_{st} = \mu_r - 1{,}64\,\sigma_r \quad (\mu_r \geqq 1{,}64\,\sigma_r)$$

zwischen $$r_1 = \mu_r - 1{,}96\,\sigma_r \quad (\mu_r \geqq 1{,}96\,\sigma_r)$$

und $$r_2 = \mu_r - 1{,}44\,\sigma_r \quad (\mu_r \geqq 1{,}44\,\sigma_r).$$

Es sei betont, daß der minimale statistische Realisationsgrad von der Verteilungsform unabhängig ist, seine Berechnung und ein Vergleich verschiedener Distinktionen also nicht die Gaußverteilung voraussetzen. $\check{r}_{st}$, r_1 und r_2 könnten notfalls auch an Hand der beobachteten relativen Häufigkeiten der Realisationsgrade bestimmt werden. Schon um die Homogenität des jeweiligen Kollektivs beurteilen zu können, ist es jedoch ratsam, die Anwendbarkeit einer bestimmten theoretischen Verteilung zu überprüfen. Damit reduziert sich dann auch die durch den Stichprobencharakter der empirischen Kollektive bedingte Unsicherheit über den prozentualen Anteil der kritischen Fälle im statistischen «Universum».

Literatur

1. *Bethge, W.:* Beziehungen der Generation zur Quantität in den deutschen Mundarten. Phonetica *9:* 200–208 (1963).
2. *Richter, H.:* Der Realisationsgrad der Quantitätsdistinktion und das Problem der Isophonie. Phonetica (in Vorbereitung).
3. *Weber, E.:* Das Ergebnis-Folge-Verfahren (Sequenzanalyse). Grundlagen und Anwendungen. Wiss. Z. Humboldt-Univ. Berlin, Math.-Nat. Reihe *7:* 559–581 (1957/58).
4. *Zwirner, E.:* Phonometrische Isophonen der Quantität der deutschen Mundarten. Phonetica, Suppl. ad Vol. *4:* 93–125 (1959).

Adresse des Autors: Diplom-Psychologe Helmut Richter, Institut für Phonetik an der Universität Münster, Steinfurter Straße 107, *Münster i. Westfalen* (Deutschland).

Proc. 5th int. Congr. phon. Sci., Münster 1964, pp. 495–501
(S. Karger, Basel/New York 1965).

The Phonemes of a Dialectal Area, Perceived by Phoneticians and by the Speakers Themselves

By K. Ringgaard, Aarhus

1. At the Institut for jysk sprog- og kulturforskning, University of Aarhus, Denmark, material is being collected both for a Dictionary and for a Linguistic Atlas of the dialects of Jutland. One of the methods used is questionnaires sent to between 500 and 600 dialect speaking informants who do not know any phonetic alphabet but answer the questions as well as they can in Danish orthography. 1959 they were asked their pronunciation of the word *høj* (high). The answers are plotted in the map of figure 1. Merely orthographic variants have been given the same sign. The signs have been placed parish by parish, so that identical information from the same parish have been given only one sign.

As you will see the resulting map is a very clear one, with sharp boundaries and general agreement in the areas. This agreement must have some foundation in the pronunciation, and there can be no doubt that the answers give us the speakers' perception of the phonemes of their own speech. In other words it is a phonemic transcription.

2. 1880 and 1887 *Wenker* sent out his famous Fragebogen in Germany, to which then belonged the Danish speaking North-Slesvig. The questionnaires were sent to each school in a parish and were filled in by the teacher. The word 'high' occurs in question 29. The answers are plotted on figure 2. In this case *all* answers are shown as it was easy to localize them to a school district inside a parish. The answers are of course written in common orthography and as before merely orthographic variants are given by the same sign. The resulting map corresponds well to that of figure 1. But in details there are some discrepancies. A closer scrutiny reveals, however, that they can almost all be ascribed to the teacher. It seems probable

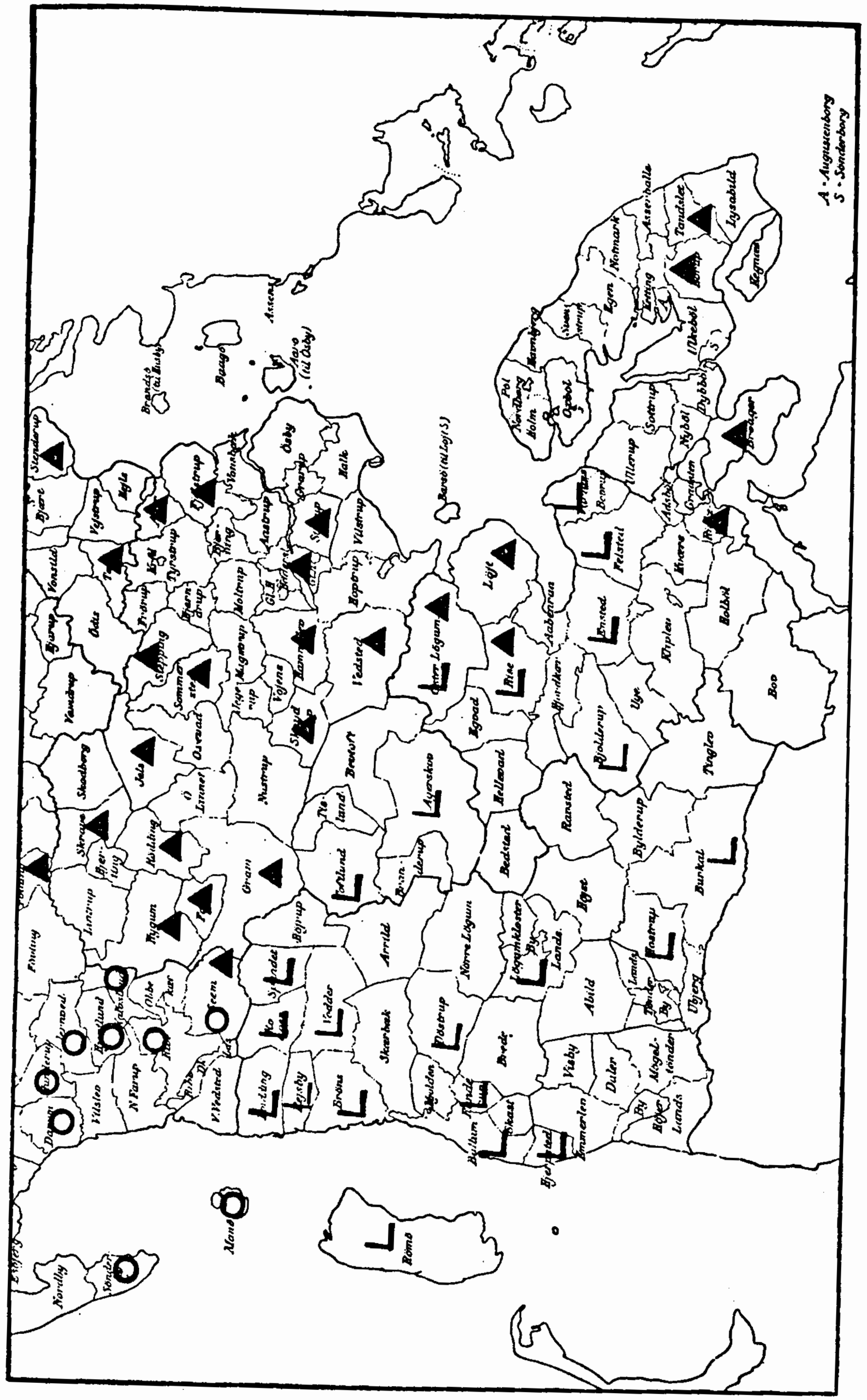


Fig. 1. The North-Slesvig section of the dialectal pronunciation in Jutland of the Danish word 'høj' according to informants 1959.

that if the answers had been given us directly from the speakers without any intermediate link the two maps would have corresponded closely to each other.

3. In the 1920's and 1930's the Danish *Stednavneudvalg* (Place Name Society) sent out fieldworkers to get a phonetical notation of the place names of North-Slesvig. They are published in "Sønderjyske Stednavne". Kbh. 1931ff. Here the word '*høj*' is found a lot of times. As was the case in figure 1 the signs have been placed parish by parish, and identical examples inside a parish are represented by only one sign. In order to facilitate the comparison with figures 1 and 2 only the *vowel quality* is shown on figure 3 and no regard has been payed to suprasegmental fenomena. Nevertheless it has been necessary to use a lot of different signs. This is what was to be expected and we begin our comparison of the phonemic and phonetic maps in good spirit, hoping to get information of the phoneme and its realization. This is, however, not the case. We do not find any meaningful clear-cut distribution of the pronunciation, only a great confusion. And our scrutiny reveals several surprising things. In the western region we find a clear variation between the vowel *y* and a more open quality conditioned by a following palatalized t, but it is a surprise that in five not neighbouring parishes we have an opening to *ø* which we would have expected to be /hø/ and not /hy/. An additional surprise is it that one of the instances is in the /høj/-area. Still more surprising is it that in the eastern region we find the same variation in six parishes in which it has hitherto not been known to exist. It seems to have wedged itself in through a hy-area.

In the /høj/-region of the east we find that on the island of Als the pronunciation is exceptionally unvarying, a stability which one would not have thought possible, while on the other hand the northeastern part shows us an equally exceptional variability. If we had only had the Wenker-questionnaire we would here have found an interesting example of a geographical penetration of the /hy/-pronunciation, occurring in recent years. The 1959 questionnaire makes, however, this interpretation impossible.

But quite a new light is thrown upon these futile speculations when we turn to figure 4. This is a map of the fieldworkers. As it will be seen there are mainly three, all of them trained phoneticians. Fieldworker number 1, represented by a triangel is a native of the western area. But he seems to have been unable to liberate himself

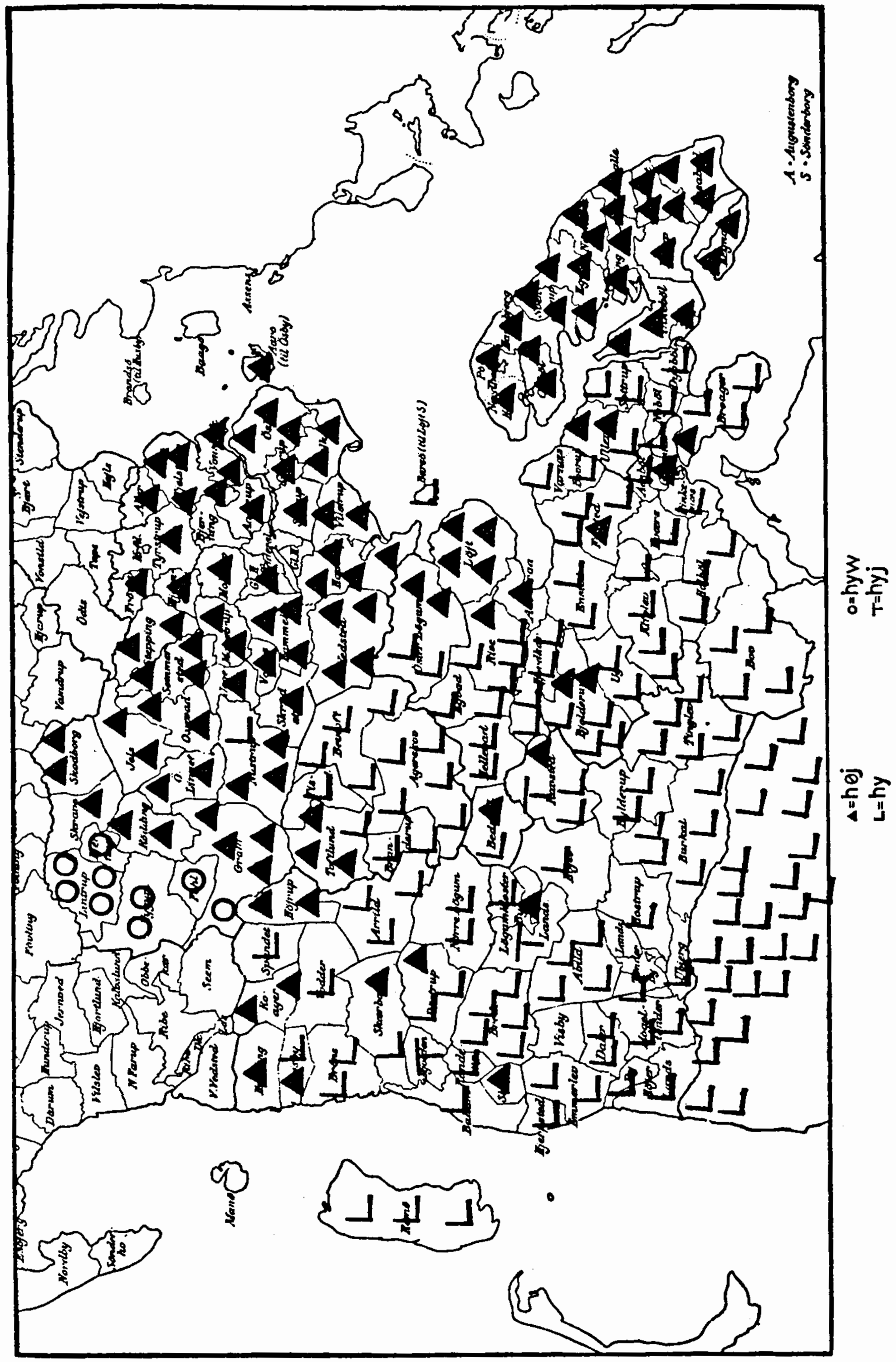


Fig. 2. The pronunciation of the word 'høj' in North-Slesvig according to the Wenker-questionnaire 1880 and 1887.

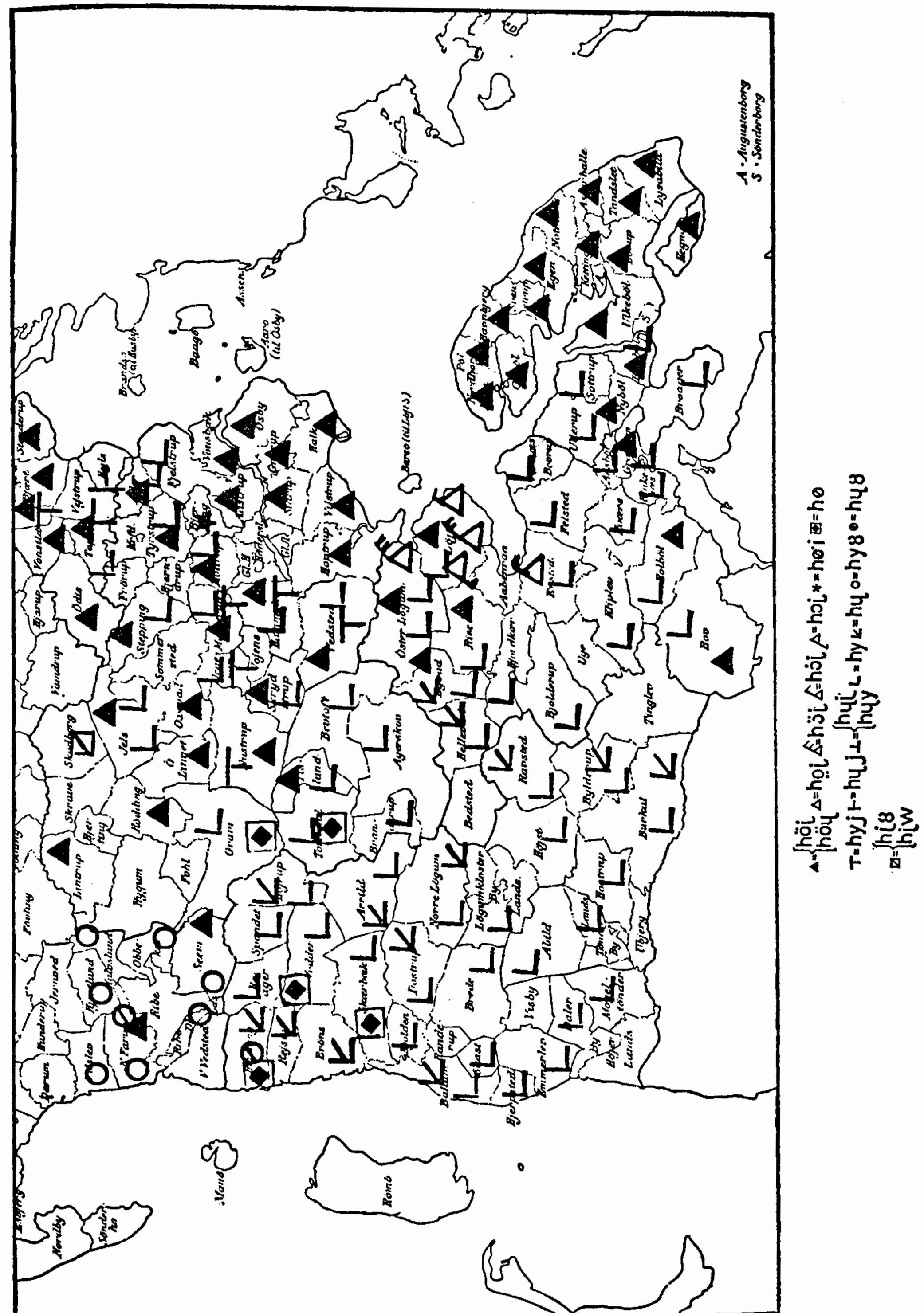


Fig. 3. The pronunciation of the word 'høj' in North-Slesvig according to "Sønderjyske Santedvne" 1931 ff.

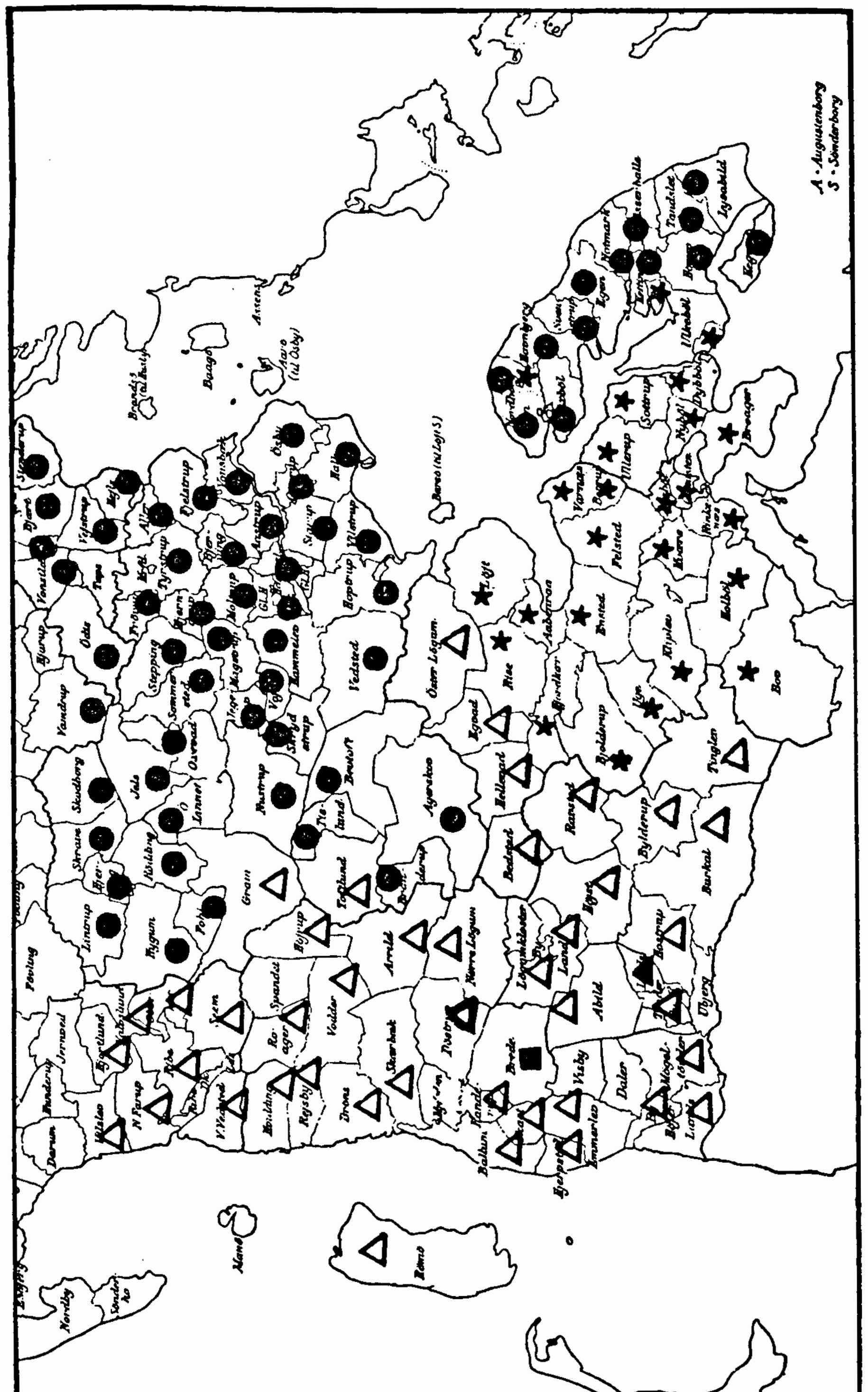


Fig. 4. The fieldworkers of "Sønderjyske Stednavne".

of his native speech with its variations which he has heard everywhere he has been, hence the curious above mentioned wedge into the eastern part.

It is interesting that he never has noted the slightly diphthongized pronunciation *hyj*, given so often by fieldworker number 3 and by three of the teachers of the Wenker-questionnaire. His failing to note it may perhaps be taken to show, that although he intended to use narrow transcription he has nevertheless transcribed his own speech phonemically. Fieldworker number 2 represented by a star seems to be rather reliable. Fieldworker number 3 seems on the other hand to be the most unreliable of them all. While he was on the island of Als he seems to have made the resolution to disregard lesser variations although he was using narrow transcription. But when he came to the north-eastern part he seems to have made the opposite resolution and tried to note faithfully what he heard, or believed he heard.

The very sad conclusion is then that *the narrow transcriptions of the phoneticians do not tell us so very much about the actual dialectal realizations of the phonemes* but tell us more about the fieldworkers themselves, about their native pronunciations and about their confusion when coming to new regions.

By this I do not want to say that we should base our work solely upon direct information from lay speakers. There are a lot of things they cannot inform us about in traditional orthography.

Neither do I want to say that we should always distrust phoneticians. If they are well acquainted with the dialect or language in question, especially if they are native speakers, they can give us the same as the layman and a lot more.

But I think that a study should always be the result of a close cooperation between the native speakers and a trained scholar who knows the language and knows what he is doing.

And I do want to say that I entirely distrust the information from fieldworkers whose material is derived from some hours' tape recording or some week's stay at a village. *They can give us neither a phonemic nor a reliable phonetic transcription.*

It is sad that most of the material for linguistic studies and atlases throughout the world is obtained in just this way at present.

Author's address: Dr. K. Ringgaard, Institut for Jysk Sprog- og Kulturforskning, Universitetet *Aarhus* (Denmark).

Proc. 5th int. Congr. phon. Sci., Münster 1964, pp. 502–505
(S. Karger, Basel/New York 1965).

Recoding Speech for the Deaf and Hard of Hearing

By Arne Risberg, Stockholm

For the profoundly deaf and the severe hard of hearing lip-reading is the principal means of comprehending verbal communication. Some individuals acquire very good results in this art but for others, and especially for the congenitally deaf, lip-reading is a very poor supplement for hearing.

During the last years many investigators have tried to find out what specific ability characterizes a good lip-reader. The results are sometimes contradictory. Other investigators have applied the theory and the methods of modern structural linguistics to the lip-reading situation. The aim has been to find the visual cues to the phonemes. *Woodward* and *Barber*[10] studied the visual discriminability of the English consonant phonemes in $C_1V \cdot C_2V$ combinations, and their results show that the consonant phonemes can be grouped in 4 units comprising the following phonemes. Unit 1: p b m; Unit 2: W, w, r; Unit 3: f, v; Unit 4: t d n l θ ð s z č j š ž y k g h. Discrimination is possible between the units but not within them. The 24 different phonemes in English have thus been reduced to only 4 for the lip-reader. The fact that lip-reading is possible indicates that the lip-reader must read on a higher level than the phoneme and also that he must make use of the redundancy in the language in a very efficient way.

Numbers[5] and others have shown that a combined use of residual hearing and lip-reading gives better results than only lip-reading even if hearing alone cannot be used for comprehending speech. The explanation seems to be that the information transmitted by hearing gives data to reduce some of the uncertainty in the visual message.

The information that can be transmitted by a very small residual hearing is very limited and it is probable that only amplifying the

speech sounds is not the most efficient way of transmission. The development of techniques for automatic analysis of speech parameters such as voicing, fricative, pitch, and so on, gives us a new method for transmitting these parameters to the deaf and hard of hearing. The parameters can be recoded to signals that are better perceived than the mere amplified sounds. The parameters can either be recoded into new acoustic signals, or into tactual and visual forms. How this recoding shall be done to get the best possible result for a specific type of loss of hearing must be studied first by investigating the coding dimensions that can be used for the different senses and then by building practical recoding systems. In some cases when the residual hearing is relatively good it is possible to recode speech so that simultaneous lip-reading is not necessary. It is less likely that a recoding of speech to tactual or visual signals can give sufficient information but it is possible that some individuals after long training can perceive everyday speech at a reduced speaking rate and tactual recoding can then be used for communication with the deaf and blind.

The recoding of speech after these principles can either be used as an aid in speech perception or as a means for speech correction. In both instances the instrumentation and the technique is basically the same. In speech correction work the subject's own voice is processed in the same way as that of the therapist's and the perceived visual, tactile, or auditive patterns serve as a feedback for correcting the pronunciation of sounds and words. As a device for speech reception the auditive and tactual channels are of primary interest since the visual channel would interfere with lip-reading. The personal aid should also be a light-weight portable system. No such limitations need to be considered in speech training.

Experiments with recoding of speech for the deaf and hard of hearing of different degrees of loss of hearing have been made at several laboratories[2, 4, 7, 9]. Some of these experiments have been directed towards speech correction only[1]. Very few results have been published.

At the Speech Transmission Laboratory these problems have been taken up for a general study. The work on tactual speech transmission published by *Pickett* at the Fourth International Congress of Phonetic Sciences in Helsinki, 1961, and in later publications[6] is continued and extended. Visual speech transmission using simple instantaneous spectrum analyses for speech correction purposes have been tried with hard of hearing subjects and normal hearing children with speech defects. The results are very promising.

The technical problem of extracting signals corresponding to the distinctive features[3] is studied and possible coding dimensions for these signals both as tactual, visual, and auditory signals are investigated.

Results obtained were reviewed at the congress.

References

1. *Anderson, F.:* An experimental pitch indicator for training deaf scholars. J. acoust. Soc. Amer. *32:* 1065–1074 (1960).
2. *Gault, R. H.:* The interpretation of speech by tactual and visual impression. Arch. Otolaryng. *4:* 228–239 (1926).
3. *Jakobson, R.; Fant, G.* and *Halle, M.:* Preliminaries to speech analysis. The distinctive features and their correlates. Acoustics Laboratory, M.I.T., Technical Report No. 13 (Jan. 1952).
4. *Johansson, B.:* A new coding amplifier system for the severely hard of hearing; Proc. 3rd int. Congr. Acoustics, Stuttgart 1959; vol. 2, pp. 655–657 (Elsevier, Amsterdam 1961).
5. *Numbers, Mary E.* and *Hudgins, C. V.:* Speech perception in present day education for deaf children. Volta Rev. Reprint No. 599 (Sept. 1948).
6. *Pickett, J. M.:* Transmitting speech sounds by a tactual vocoder and by lip-reading. Royal Institute of Technology, Speech Transmission Laboratory Report No. 27 (March 1963).
7. *Pimonow, L.:* L'application de la parole synthetique dans la correction auditive. Acustica *12:* 285–290 (1962), English translation J. Aud. Res. *3:* 73–82 (1963).
8. *Potter, R. K.; Kopp, G. A.* and *Green, H. C.:* Visible Speech (Van Nostrand, New York 1947).
9. *Wedenberg, E.:* Auditory training of severely hard of hearing using a coding amplifier. Proc. 3rd int. Congr. Acoustics, Stuttgart 1959, vol. 2, pp. 658–660 (Elsevier, Amsterdam 1961).
10. *Woodward, Mary F.* and *Barber, Caroll G.:* Phoneme perception in lipreading. J. Speech and Hearing Res. *3:* 212–222 (1960).

Author's address: Mr. A. Risberg, Speech Transmission Laboratory, Royal Institute of Technology, *Stockholm 70* (Sweden).

Discussion

Delattre (S. Barbara): If a 1000 cps low pass filter still allows a distinction such as [t/s] to be made, it must mean that there exist cues in the first and second formant transitions that really count. I can think of two, offhand. The first transition has a higher locus and a slower tempo, and the second formant has a lower locus, for [s] than for [t].

Denes (Murray Hill): Your last slide, showing the high articulation scores for distinguishing between /s/, /st/ and /t/, for speech low-pass filtered at 700 c/s, surprised me greatly. The voice energy for these sounds is largely concentrated at frequencies well above this cut-off and I wonder if you could say how sharp your low-pass filter was and also what acoustic feature you think the subjects utilize in making these distinctions.

With regards to the question of the value of a non-portable re-coding device for aiding in speech recognition of the deaf, such devices would be useful with telephones. During normal face-to-face conversation the deaf person can supplement his deficient hearing by lip-reading; this is not possible when he uses the telephone. Under the conditions a re-coding device, with a suitable display, could take the place of lip-reading and make speech communication possible even if the display of the re-coding device on its own were not sufficient.

Lebrun (Bruxelles): In answer to the question "How can deaf or hard of hearing people who lip-read distinguish between [p], [b], and [m]?" I should like to suggest that perhaps they use the following cue:

When [p] is pronounced, the lips of the locutor – at least, if he speaks French – are pressed harder against one another than when [b] or [m] is pronounced. On the other hand, when [b] is articulated, the lips are slightly protruded, whereas they are slightly retracted when [m] is uttered. The slight protrusion of the lips during the "tenue" of the [b] (or [p]) may to some extend anticipate the explosion which is to follow, or else result from the relatively high pressure within the mouth. In the case of [m], the slight retraction of the lips may perhaps be due to the relatively low pressure in the mouth cavity.

Verh. 5. int. Kongr. Phon. Wiss., Münster 1964, pp. 506–510
(S. Karger, Basel/New York 1965).

Zu akustischen Korrelaten der distinktiven Merkmale

Von Milan Romportl, Praha/Göttingen

0. In diesem Referat soll an die bekannten Untersuchungen der von Prof. *Jakobson* geführten und durch dessen ältere Arbeiten inspirierten M.I.T.-Gruppe angeknüpft[1, 2, 3] und nachgeprüft werden, 1. ob akustische Korrelate der distinktiven Merkmale überhaupt zu finden sind und – im positiven Fall – 2. was für eine akustische Struktur sie aufweisen.

1. Es sei hier vorausgesetzt, daß auf Grund einer richtig durchgeführten distributiven Analyse und unter Anwendung richtiger Regeln[8, 7] die Phoneme einer bestimmten Sprache richtig festgestellt worden sind.

1.1 Wenn man die Realisierungen eines auf diese Weise bestimmten Phonems akustisch analysiert, entdeckt man in einer jeden von diesen Realisierungen eine Menge von verschiedenen akustischen Elementen, Zügen (weiter: Zug).

Diese Züge können verschiedenen Charakter aufweisen. Als akustischer Zug kann z.B. die An- oder Abwesenheit der akustischen Realisierung überhaupt, ihre Gesamtdauer und Gesamtintensität, der Charakter des akustischen Spektrums (kontinuierlich oder diskret), Intensität und Frequenz von dessen Bestandteilen, ferner deren Verhältnis, Dauer und Reihenfolge usw. auftreten.

1.11 Man könnte vielleicht noch eine weitere Tatsache erwähnen, obwohl bei unserer Stellungnahme diese nicht notwendigerweise in Betracht gezogen werden muß, da sie für eine synchronische Betrachtungsweise unberücksichtigt bleiben könnte. Es handelt sich nämlich darum, daß einerseits nur einige der akustischen Züge als «primär» anzusehen sind, indem sie gewissermaßen die artikulatorische «Absicht» akustisch «verkörpern» und als akustisches Symbol dieses Zuges angesehen werden könnten (z.B. Konzentration der Gesamtenergie im tieferen Bereich des akustischen Lautspektrums), während, andererseits, die übrigen Züge als «sekundär» betrachtet werden können, da sie nur Begleitzüge sind und als bloße Symptome eines anderswohin zielenden artikulatorisch-akustischen Geschehens zu bewerten sind (z.B. Frequenzänderungen der Vokalformanten in der Nähe von Konsonanten).

1.2 Durch physiologisch-akustische Untersuchungen (die von uns als Gehöranalyse bezeichnet werden)[4] wird von den durch eine akustische Analyse entdeckten Zügen jene Anzahl (Menge) dieser Züge isoliert, die vom Hörer erfaßt werden (bzw. erfaßt werden können).

Dabei wird außer der Einschränkung des in Betracht gezogenen akustischen Feldes auf die dem Gehörfeld entsprechenden Frequenz- und Intensitätsbereiche auch z.B. die bekannte Erscheinung einer Poststimulusadaptierung des Gehörs[6] zur Geltung kommen (wenn z.B. ein Konsonant von einer um 30 dB geringeren spezifischen Lautstärke einem Vokal folgt).

1.21 Man könnte voraussetzen, daß ein Durchschnitt der Mengen solcher Elemente (A_k) in allen Realisierungen eines Phonems (bzw. in allen in einer und derselben Position vorkommenden Realisierungen eines Phonems – vgl. weiter) eine neue Menge bilden wird, die alle für das Phonem (bzw. für die in jener Position vorkommende Variante, Allophon) «charakteristischen» Züge enthalten wird, d.h. $A = \bigcap_{k=1}^{n} A_k$

Wenn wir unsere Analysen weiter fortsetzen würden, so würden wir in neuen Durchschnitten von Mengen, die immer zwei verschiedenen Phonemen (bzw. zwei entsprechenden Allophonen zweier verschiedener Phoneme) entsprechen (A, B, C, ...) diejenigen Zügebündel ($\mathit{A} = A \cap B$, $\mathit{B} = A \cap C$), ... $\mathit{L} = C \cap D$, ...) bestimmen können, die diesen Phonemen gemeinsam sind, bzw. – was wichtiger sein wird – diejenigen der beiden Phoneme differenzierenden Züge (Zügebündel), die dem Komplement dieser letzteren Mengen entsprechen, also $\mathit{A}_B = A - (A \cap B)$, $\mathit{B}_A = B - (A \cap B)$ usw.

Die Fortsetzung der Analyse führt zu den kleinsten Zügebündeln, von denen man wohl annehmen darf, daß sie den distinktiven Merkmalen entsprechen. Es kann unter Umständen der Fall eintreten, daß für verschiedene Positionen (α, β, γ, ... – z.B. An-, In- und Auslaut) einem und demselben unbestreitbaren Merkmal verschiedene akustische Züge (Zügebündel) entsprechen, so daß man für Korrelate des distinktiven Merkmals erst das ganze Mengensystem halten dürfte, d.h. $\mathfrak{A} = \{A_\alpha, A_\beta, A_\gamma\}$.

Dabei sollte gelten, daß alle Phonempaare, für deren Paarglieder (für jede einzelne Position) die Bedingung $(A-[A \cap B]) : (B-[A \cap B]) = a : \bar{a}$ gilt, wobei selbstverständlich die Bedingung $a \in A$, $\bar{a} \in B$ bewahrt werden muß, derselben Korrelation angehören.

1.22 Man muß betonen, daß unsere Erwägungen nur für solche Fälle gelten können, wo ein Phonem wirklich realisiert wird und seine Funktion als Mitglied einer Opposition in der Merkmalebene völlig ausübt.

2. Jedes distinktive Merkmal kann gewöhnlich durch mehrere akustische Züge (durch eine Gruppe dieser Züge) realisiert werden. Das Merkmal weist dabei unbestreitbar eine Redundanz auf, die man mathematisch ausdrücken könnte.

Das Verhältnis einzelner Züge im Rahmen eines Merkmals kann unterschiedlich sein.

2.1 Außer den Zügen, die das Zentrum der Struktur bilden und die wir als intern (bzw. zentral) bezeichnen werden, gibt es öfters Züge, die das Merkmal nur unverbindlich begleiten; diese werden hier als extern bezeichnet. Zu diesen gehören manchmal einige der sekundären, symptomatischen Züge, die wir auch als irrelevant bezeichnen.

Das Verhältnis der internen Züge (A, B) zu den externen (irrelevanten) Zügen (c) wird man mit dem Schema $\left\{{A \atop (c)}\right\}$ darstellen.

2.2 Aber auch die Wichtigkeit der internen Züge ist nicht immer dieselbe.

2.21 Außer den hierarchisch höchsten Zügen, die hier mit Anwendung des traditionellen Terminus als relevant (A) bezeichnet werden, sind öfters auch solche Züge vorhanden, die unbestreitbar als Bestandteile des Merkmals angesehen werden müssen, doch in der Hierarchie der Züge niedriger stehen. Wir bezeichnen sie hier als redundant (B) und ihr Verhältnis zu den relevanten Zügen wird mit dem Schema $\left\{{A \atop B}\right\}$ ausgedrückt.

2.22 Wenn zwei oder mehrere Züge relevant sind und über ihre Unter- oder Überordnung nicht entschieden werden kann, so wird zum Ausdruck dieses Verhältnisses das Schema $\{A_1{-}A_2\}$ angewendet. Dabei muß betont werden, daß das Ganze des Merkmals auch in diesem Fall eine Redundanz aufweist.

2.3 In verschiedenen Sprachen können gleichbezeichnete Merkmale eine völlig unterschiedliche Struktur aufweisen.

Außer dem allgemein bekannten Beispiel der verschiedenen Struktur eines distinktiven prosodischen Merkmals «akzentuiert-unakzentuiert» in verschiedenen Sprachen mit der sogenannten dynamischen Betonung kann man beispielsweise die akustische Struktur des bekannten distinktiven Merkmals «Stimmbeteiligung» im Tschechischen oder Russischen durch das Schema

$\left\{\begin{matrix}\text{Stimmbeteiligung}\\ \text{Intensität}\\ \text{(Dauer)}\end{matrix}\right\}$ ausdrücken, dagegen das Merkmal, welches im

Deutschen öfters gleich bezeichnet wird, durch das Schema

$\left\{\begin{matrix}\text{Intensität-Aspirierung}\\ \text{Stimmbeteiligung}\end{matrix}\right\}$, wenn nicht ein Schema

$\left\{\begin{matrix}\text{Intensität-Aspirierung}\\ \text{(Stimmbeteiligung)}\end{matrix}\right\}$ noch passender sein sollte[5].

Literatur

1. *Jakobson, R.:* Observations sur le classement phonologique des consonnes. Proceedings of the Third Int. Congress of Phon. Sciences; pp. 34–41 (University, Ghent 1939).
2. *Jakobson, R.; Fant, C. G. M.* and *Halle, M.:* Preliminaries to speech analysis. M.I.T., Acoust. Lab. *13:* 3 (1955).
3. *Jakobson, R.* and *Halle, M.:* Fundamentals of language (Mouton, 's-Gravenhage 1956).
4. *Janota, P.* und *Romportl, M.:* Ein Beitrag zur phonetischen Methodik. Acta Univ. Carolinae, Philol. *1:* 17–28 (1959).
5. *Romportl, M.:* K akustické podstatě a klasifikaci okluzív. Slavica Pragensia *4:* 75–80 (1952). – Zur akustischen Analyse und Klassifizierung der Nasale. Wiss. Z. Halle *9:* 1653–1660 (1962). – Zvukový rozbor ruštiny (AUC, Praha 1962). – Zur akustischen Struktur der distinktiven Merkmale. Z. Phonetik *16:* 191–198 (1963).
6. *Sedláček, K.:* Základy audiologie (Státní zdravotnické nakladatelství, Praha 1956).
7. *Trnka, B.:* Určování fonému. Acta Univ. Carolinae, Phil. Hist. *7:* 16–22 (1954).
8. *Trubetzkoy, N. S.:* Grundzüge der Phonologie. TCLP 7 (Praha 1939).

Adresse des Autors: Prof. Dr. M. Romportl, Havlovská 40, *Praha 6 – Dejvice* (ČSSR).

Discussion

Horálek (Prag): Die distinktiven Züge sind akustische Realität; deswegen ist man eigentlich nicht berechtigt, von den akustischen Korrelaten der distinktiven Züge zu reden. Fraglich ist auch, ob es sich bei solchen komplexen Qualitäten, wie z.B. Stimmbeteiligung, um eine Summe (Menge) von Elementen handelt oder um eine Gestaltqualität (die dieselbe bleiben kann, obschon alle Bestandteile durch andere ersetzt werden können).

Eli Fischer-Jørgensen (Virum): Ich möchte meiner allgemeinen Zustimmung zu den Ausführungen von Herrn *Romportl* Ausdruck geben. Der Terminus «distinctive feature» ist mehrdeutig und müßte durch verschiedene genauer abgegrenzte Termini abgelöst werden. Ob dabei die Unterscheidung zwischen «Merkmal» und «Zug» terminologisch ganz glücklich ist, scheint mir etwas zweifelhaft; aber jedenfalls müßte man folgende Phänomene unterscheiden: 1. die minimale Opposition in einer konkreten Sprache, 2. ein Glied einer solchen Opposition, 3. die phonetischen Eigenschaften, wodurch sich eine solche minimale Opposition manifestiert (und dabei können mehrere solche Eigenschaften entweder zugleich oder in verschiedenen Stellungen relevant sein), 4. die allgemeinen phonetischen Dimensionen, die man zur Beschreibung einer konkreten Sprache heranziehen kann.

Antwort *Romportl:* Im ersten Teil seiner Bemerkung hat Herr *Horálek* nicht völlig unrecht. Man könnte wirklich über die Richtigkeit der angewandten Termini streiten, doch die Diskussion wäre zu lang und würde u.a. schon die Terminologie unserer Vorgänger betreffen. – Was ich aber ablehnen muß, ist die Vermutung, daß hier als Korrelate der distinktiven Merkmale nur Summen von akustischen Zügen angesehen werden. Es handelt sich immer um Mengen (bzw. Durchschnitte und Komplemente der Durchschnitte der Mengen), die als eine Art von geordneten Mengen anzusehen sind und eine hierarchisierte Struktur aufweisen.

C.R. 5e Congr. int. Sci. phon., Münster 1964, p. 511
(S. Karger, Basel/New York 1965).

Remarques sur la phonologie historique du roumain

(Résumé)

Par A. Rosetti

L'auteur montre, tout d'abord, ce qui, dans la masse des faits, sera retenu pour son exposé, à savoir la manière dont les changements phonétiques ont été réalisés et la fonction des changements dans le système de la langue.

La confusion des timbres vocaliques a provoqué la disparition de quelques voyelles. Rôle des voyelles prépalatales dans l'évolution des consonnes, et des occlusives nasales, dans l'évolution des voyelles. Métaphonie des voyelles *e* et *o*, qui a provoqué la diphtongaison de ces voyelles en *ea'* et *oa'*. La syncope des voyelles, à l'intérieur du mot phonétique, a eu pour suite la création de groupes de consonnes nouveaux. Adjonction des voyelles *ă* et *î*, spécifiques du roumain, à l'inventaire existant. Amuïssement de l'-*i*, marque du pluriel des substantifs masculins. Les semi-voyelles *e* et *o*. Consonnes à sonorité pertinente. Devant *e*, *i*, les consonnes vélaires et les consonnes dentales deviennent des fricatives et des affriquées. Emprunt de *h* au slave. Disparition des consonnes finales et de *u*: lat. *lupus* > dr. *lup*.

Adresse de l'auteur: Prof. A. Rosetti, 56 Str. Dionisie Lupu, *Bucarest 22* (Roumanie).

C.R. 5e Congr. int. Sci. phon., Münster 1964, pp. 512–516
(S. Karger, Basel/New York 1965).

Analyse spectrographique et interprétation fonctionnelle de la nasalité dans un parler de l'Italie du nord (Rossano, prov. Massa Carrara)

Par MARIO ROSSI, Marseille

M. *P. Delattre*[1] note comme indices acoustiques de la nasalité vocalique: 1° La perte d'intensité de F1; la création d'un nouveau formant (Fn1) vers 250 Herz; 3° la montée de F3 et la descente de F4. La réduction d'intensité de F1 est le plus important. M. *Shirô Hattori*[2], de son côté, a cru voir, dans les dépressions ou «antiformants» observés sur les spectrogrammes de voyelles nasales, des indices révélateurs de la nasalité.

Comment, sur le plan acoustique, se réalise la nasalité vocalique à Rossano (aire linguistique: Emilie occidentale)? Les résultats de l'analyse acoustique confirment-ils ceux de l'interprétation auditive? Y a-t-il parallélisme entre la production et la réception de l'onde sonore, entre la réalité physique et la structure linguistique?

I° – *L'analyse spectrographique**

Si l'on compare le spectre des voyelles nasales à celui des voyelles orales correspondantes, dans le même contexte phonique en position accentuée (Fondamental: 120 Hz), on arrive aux conclusions suivantes:

A. – Influence de la nasalité sur F1

a) Ecrêtage du fondamental pour toutes les nasales.

b) Ecrêtage des harmoniques:

* Sur le Sona-Graph de l'Institut de Phonétique de Grenoble. Sujet interrogé: un homme de 40 ans, habitant de Rossano. Documents sonores obtenus sur Ampex 601, à partir de la Chambre sourde. Les signes [õ], [ẽ] et [ũ], représentent respectivement et à peu de choses près les voyelles nasales du français dans les mots «bon», «pain» et «brun».

- pour [ã], écrêtage des harmoniques 1, 2, 5, 6, c'est-à-dire autour de 240, 360, 720 et 840 Hz (voir fig. 1 et 2, 3 et 4);
- pour [õ], 1, 4, 5, c'est-à-dire vers 240, 600, 700 Hz (voir fig. 5 et 6) (toutes les figures sont supprimées);
- pour [ẽ], 1, 5, c'est-à-dire vers 240 et 720 Hz (voir fig. 7 et 8).

c) La perte d'intensité du 5e harmonique annonce une dépression pour toutes les nasales entre 700 et 900 Hz.

d) Appauvrissement du sommet de F1 qui généralement comprend:

- deux harmoniques pour [a], le 3e et le 4e; et un harmonique pour [ã], le 3e *ou* le 4e (voir fig. 1 et 2, 3 et 4);
- trois harmoniques pour [ɔ] et [ɛ], les 2e, 3e et 4e; et deux harmoniques pour [õ] et [ẽ], les 2e et 3e (fig. 5 et 6, 7 et 8).

Le premier formant de la nasale est généralement moins riche en fréquences que celui de l'orale correspondante: son spectre est plus étroit (200 à 400 Hz en moins). La comparaison entre orales et nasales est établie dans le même contexte consonantique, car la largeur du spectre de F1 semble dépendre de l'entourage phonique: le spectre est plus étroit dans un contexte labial que dans un contexte palatodental.

B. – Influence de la nasalité sur F2

Généralement on observe une perte d'intensité de F2 et un appauvrissement de son spectre en harmoniques.

Les indices acoustiques de la nasalité mis en évidence par M. *Delattre* (perte d'intensité de F1 de 12 à 15 db, Fn1 à 250 Hz) ne se trouvent jamais réunis dans la production des voyelles nasales de notre sujet; Fn1 est très rare; l'intensité de F1 (sommet de l'enveloppe) ne diminue guère (4 db au maximum). La nasalité se traduit ici surtout par un appauvrissement fréquentiel du spectre, par la perte d'intensité de certains harmoniques dans les zones des 120, 240 et 720 Hz, par des dépressions entre 700 et 900 Hz.

Les résultats de l'analyse acoustique semblent confirmer les données de l'interprétation auditive: la plupart des voyelles nasales sont incomplètes, elles sont généralement suivies d'un appendice consonantique. La nasalité vocalique a cependant une valeur acoustique suffisante pour pouvoir être utilisée dans la langue, car elle est très distinctement perçue dans tous les cas. Il ne s'agit pas de simples voyelles «nasalisées» (*Straka*[6]).

IIº – *L'interprétation fonctionnelle*

En position implosive, à l'intérieur du mot, les oppositions entre /*m*/, /*n*/, /*ɲ*/ se neutralisent, le lieu d'articulation de la consonne nasale est déterminé par celle qui suit. A la finale absolue, on a toujours [*ŋ*]. Il est difficile d'accorder le statut de phonème à la nasale vélaire, qui apparaît dans des contextes où les autres consonnes nasales ne peuvent se rencontrer.

Le parler de Rossano connaît sous l'accent 9 phonèmes vocaliques oraux (*i*, *e*, *ɛ*, *a*, *ɔ*, *o*, *u*, *y*, *ø*) et 4 voyelles nasales (*ã*, *ɛ̃*, *ɔ̃*, *œ̃*) suivies généralement d'un appendice consonantique nasal. Les voyelles nasales se rencontrent

a) devant consonne: à l'initiale du mot, à l'intérieur du mot et dans la chaîne parlée;
b) devant voyelle: dans la chaîne parlée;
c) à la finale absolue.

A la finale absolue, très souvent, la nasale vélaire disparaît ou se réduit à une simple résonnance nasale, non suivie d'occlusion, qui s'étend au-delà du segment vocalique et qui contribue à l'allonger (fig. 1, 4, 5). Dans la chaîne parlée, devant voyelle ou consonne, l'appendice consonantique nasal disparaît la plupart du temps (fig. 9, 10, 11).

Il serait tentant d'expliquer fonctionnellement la nasalité à Rossano en posant, comme le fait M. *Jorge Morais-Barbosa*[5] pour le portugais, un archiphonème /*N*/ qui se réaliserait tantôt vocaliquement, tantôt consonantiquement. Mais cette interprétation ne rend pas compte de tous les faits de nasalité dans notre parler. M. *Morais-Barbosa*([5], p. 707) écrit que la réalisation de /*N*/ «se fait comme il a été indiqué ou bien par une consonne nasale, ou bien par une simple résonnance nasale sur la voyelle précédente qui ne cesse pas pour autant d'être phonologiquement neutre, c'est-à-dire, ni orale ni nasale, puisque l'oralité ne serait pertinente qu'au cas où la nasalité le serait aussi».

Or a) dans l'exemple (fig. 12) où l'appendice consonantique a disparu, la voyelle orale accompagnée de résonance nasale constitue bien une voyelle nasale qui peut s'opposer à une voyelle orale dans le même contexte.

b) Il n'est pas indifférent, dans la chaîne parlée que /*N*/ se réalise consonantiquement ou vocaliquement: les segments [*fã*, *fan*, *fam*] ont un sens tout à fait différent. La nasalité vocalique peut

donc être pertinente. Mais [*fã*] peut aussi se réaliser [*fãŋ*] et [*ŋ*] n'est pas une variante de /*n*/. Devons-nous supposer un phonème /*ŋ*/? Il semble que non: nous pouvons en effet supprimer /*ŋ*/ sans que le sens ne change et la phrase reste parfaitement compréhensible. Par contre, si nous supprimons /*m*/ ou /*n*/ ou si nous reportons la nasalité sur la voyelle précédente, dans les deux derniers exemples, le sens change aussitôt. Il serait d'ailleurs paradoxal de pouvoir prouver l'identité phonologique de [*ŋ*] par les seuls exemples où précisément il se réalise rarement. Il est préférable de considérer /*ŋ*/ comme une réalisation du phonème vocalique nasal. D'autres exemples confirment cette hypothèse: [*lãŋna*] n'est pas une variante de [*lanna*], car la langue de Rossano exclut les géminées à l'intérieur du mot. Le plus souvent d'ailleurs [*lãŋna*] se réalise sans la nasale vélaire; /*ã*/ peut s'opposer à /*a*/ dans le même contexte.

Nous supposerons donc dans notre parler l'existence, à côté des 9 phonèmes vocaliques oraux, de 4 phonèmes vocaliques nasaux /*ã*, *ẽ*, *õ*, *ũ*/.

Nous considérerons [*m*, *n*, *ɲ*, *ŋ*] implosifs, à l'intérieur du mot, comme des sons de passage entre la voyelle nasale et la consonne qui suit. Dans tous les autres cas nous dirons que [*ãŋ*, *ɛ̃ŋ*, *ɔ̃ŋ*, *œ̃ŋ*] sont des variantes de /*ã*, *ɛ̃*, *ɔ̃*, *œ̃*/.

Ainsi les résultats de l'analyse structurale ne coïncident pas avec ceux de l'analyse acoustique. Les deux plans acoustique et fonctionnel ne sont pas forcément parallèles. Il peut y avoir décalage entre structure linguistique et réalité acoustique, celle-ci étant sujette à des variations dans l'espace par villages, dans le temps par générations sans que le système – plus stable – ne change aussitôt. L'essentiel est d'abord de savoir comment la langue structure et interprète les données du continuum physique.

Nous avons ici volontairement renversé les niveaux de l'analyse phonétique, en passant du plan acoustique au plan structural. Mais nous nous sommes bien gardés de chercher dans l'onde sonore analysée «la contre-partie exacte des faits de langue» (*Malmberg*[3]). Ce faisant, nous avons montré, je crois, le bien-fondé de la théorie de M. *Malmberg* sur les niveaux d'abstraction dans l'analyse phonétique et phonémique (*Malmberg*[3], pp. 456–457; *Malmberg*[4], pp. 220–243).

Bibliographie

1. *Delattre, P.:* Les attributs acoustiques de la nasalité vocalique et consonantique. Stud. linguist. *8:* 103–109 (1954).

2. *Hattori, Shirô:* Nasalization of vowels in relation to nasals. JASA 30 (1958).
3. *Malmberg, B.:* Analyse instrumentale et structurale des faits d'accents. Proceedings of the IVth int. Congr. Phon. Sci., Helsinki 1961, p. 456.
4. *Malmberg, B.:* Levels of abstraction in phonetic and phonemic analysis. Phonetica *8:* 220–243 (1961).
5. *Morais-Barbosa, J.:* Les voyelles nasales portugaises: interpretation phonologique. Proc. IVth int. Congr. Phon. Sci., Helsinki 1961, pp. 691–709.
6. *Straka, G.:* Les voyelles nasales. Rev. linguist. Rom. *19:* 270 (1955).

Adresse de l'auteur: Dr M. Rossi, 19, bd Joseph Fabre, *Marseille XII* (France).

Proc. 5th int. Congr. phon. Sci., Münster 1964, pp. 517–520
(S. Karger, Basel/New York 1965).

A Note on the Rise-Fall Nuclear Glide in English Intonation

By Maria Schubiger, Zürich

We know a good deal about the function of the fall-rise nuclear glide (FR), while the rise-fall (RF) has so far been given much less attention. It is the purpose of this paper to consider one aspect of the latter. Only statements will be taken into account. For reasons of comparison a few well-known facts concerning the FR will have to be stated, too.

It has repeatedly been pointed out that the FR can have a concessive or, as some would prefer to call it, a limitative function. When used in a retort, it restricts the speaker's agreement; e.g. A. *It 'wasn't a 'good 'winter for ˋskiing in ˌSwitzerland, ˋwas it?* B. *We had 'more 'snow than in ˇEngland.* B. broadly agrees with A. but makes a statement which qualifies his agreement. Here the function of the FR corresponds to that of *at least, anyway, at any rate.* When used in a qualifying answer to a general question the FR has a similar function. If the answer is a qualified affirmative, it points to the negative, and vice versa; e.g. *I can ˇread it* as an answer to *'Do you 'know ˏDutch?* is a qualified affirmative. *It 'isn't esˇsential* in reply to *'Should I 'learn the 'language of the ˏnatives?* is a qualified negative. In both cases the FR suggests a sequence beginning with *but* or *though.*

Another relation which for want of a more satisfactory term we call concessive is the converse of the one just mentioned. It can be expressed by *even. Even non-specialists could follow* means... *of whom it was to be expected least. At least the specialists could follow* means... *of whom it was to be expected first and foremost.* The particle *even*, too, is sparingly used in colloquial speech, intonation being in many cases the bearer of the concessive relation. Intonation can achieve this in several ways. One – and this has been pointed out before – is the tonetic stress pattern: nuclear stress on the word that could be modified by *even*, subdued non-nuclear stresses; e.g. *`Homer ,sometimes ,nods. I ˏdidn't ·tell my `husband. We ˏdidn't be·lieve him for a `minute.* Compare with this the non-concessive *We 'didn't be`lieve him for a ,minute.* I am quoting this last example from an article by Professor *Bolinger*.

As these examples show, the effect depends on the presence in the sentence of at least one or two normally stressed words, which are wholly or partly de-stressed, in order to place in relief the word which the *even* connotation refers to. It is true that the destressing is not always essential. If the word bearing the nuclear glide is normally post-nuclear, as in the last example, the place of the nucleus in itself suggests *even*; e.g. *We 'didn't be'lieve him for a `minute.* But *I 'didn't 'tell my `husband* with two full prenuclear stresses and a non-emphatic nuclear fall does not suggest *even*. Nor can concession be expressed by the tonetic stress pattern alone in short sentences with only one normally stressed word. *There'll be `more* as a retort to *There'll be about `ten, I sup,pose*, or *I should be `glad to* in answer to *'Would you 'mind ˏhelping?* do not suggest *even*. But they can be made to suggest it by bearing a RF instead of a simple F nucleus. The RF can suggest that the speaker is impressed and wants to impress his interlocutor. It can also convey complacency or censoriousness. Now one or several of these connotations are very often attached to the *even* connotation. In a retort *even* often suggests that the speaker's utterance adds something unexpected to his interlocutor's statement. In an answer to a general question it suggests that the answer corresponds to an expanded affirmative or negative. So the RF, whose basic function is purely expressive, can, in favourable circumstances, assume a grammatical function as well: the mood it expresses is interpreted also as the syntactical relation which easily engenders this mood. *There'll be ^more* as a retort to *There'll be about `ten, I sup,pose* comes to mean *even more* (possibly with the implication: so you'd better provide enough seating accomodation). *I should be ^glad to* in answer to *'Would you 'mind ˏhelping?* suggests: not only do I not refuse to help; I am even glad to do it. I am quoting these two RF sentences – and some of the following ones – from *O'Connor* and *Arnold's* Intonation of Colloquial English, where the RF is allotted ample space. The RF is the main bearer of the *even* relation also in those cases where the pre-nuclear stresses are not subdued; e.g. A. *'Have you ˏfinished it?* B. *I 'haven't be^gun it.* Or: *It was `lovely in ,Scotland ,last ,winter; we had 'more 'snow than in ^Switzerland.* Here the second sentence is not a retort but an expansion of the speaker's own previous statement. Sometimes there is a RF in utterances where the tonetic stress pattern would be sufficient to suggest the *even* relation: e.g. *It's `useless ,writing a ,letter; a ^telegram wouldn't ,reach him in ,time.* Needless to say that in many utterances where the particle *even* is enunciated the nucleus is a RF. In *O'Connor* and *Arnold* there are several instances of it.

We have seen that there is a certain parallelism between the intonation of utterances expressing limitation and that of utterances expressing expansion. Both suggest these grammatical relations by means of a two-directional instead of a one-directional nuclear glide, limitation by adding a rise to the fall, expansion by placing a rise before the fall. But there the parallelism ends. The R part of the FR is basically an intellectual device, it expresses incompleteness, need of supplementation. Consequently the FR can suggest *at least* both in emotional and in relatively unemotional utterances. The sentences I quoted at the beginning can be quite matter-of-fact; e.g. A. *'Do you 'know ˏDutch?* B. *I can ˇread it.* Or: *'Should I 'learn the 'language of the ˏnatives?* B. *It 'isn't esˇsential.* The RF, on the other hand, is a purely emotional variant of the F. Its counterpart in the domain of incompleteness is the RFR. Therefore the *even* relation can be expressed by means of the RF only in emotional speech. The unemotional counterpart of *There'll be ^more* is *There'll 'even be `more.* The tonetic stress pattern with an F nucleus suggesting *even*, on the other hand, can sometimes be made to sound quite matter-of-fact, especially in those cases where the nucleus falls on an element of the sentence that does not currently bear it. *`Homer ,sometimes ,nods* or *They 'wouldn't be 'happy if they had `money* can be said quite unemotionally.

These remarks do not by any means exhaust the subject. There are other similar relations that favour the RF. One of them I briefly pointed out in my contribution to a discussion on the growing tendency to stress prepositions and give them a RF nuclear stress (in English Studies 1963; p. 275). I will quote only one instance: *You 'say this 'isn't an 'honest `business. Then 'why did you 'stay ^in it?* Here the RF appears because the contrary of what would be the right reaction to a given situation has happened. There is a good deal more scope for investigation even in this highly restricted field within the vast domain of English intonation.

Author's address: Dr. M. Schubiger, Hochstraße 81, *Zürich 44* (Switzerland).

Discussion

Danes (Praha): It is a matter of fact that languages differ in their sentence intonation; but it is also true that they are very often similar in respect to various functions of the intonation. Now, the system of terms and the metalanguage by means of which different investigators try to describe the great variety of many and many subtle semantic nuances is very inconsistent and impressionistic. Such a situation is very unpleasant, esp. if we try to compare different languages. I suggest, therefore, that we should try to elaborate an exact system of terms by means of which the various subtle semantic values of intonation could be described more exactly.

Lebrun (Bruxelles): The use of the word *concessive* to refer to such clauses as "Whatever you may say (, I won't go)" is rather unfortunate because these sentences obviously express no concession.

The use of the word *concessive* to refer to the second intonational pattern (rise-fall) is similarly regrettable, because this pattern does not denote any concession: it enhances, or expands, a preceding statement.

On the other hand, some of the sentences quoted by Miss *Schubiger* to illustrate the first pattern (fall-rise) do express a concession (e.g. "It isn't essential"). If one insists on using the word *concessive*, it would seem that one ought to use it to refer to the first rather than to the second intonational pattern.

Jürgensen (Copenhagen): Questioned the usefulness of the term 'concessive'.

He further found that it would be a little surprising if 'rise-fall' had the connotation suggested seeing that rise-fall does not carry this suggestion in, say, excited recitation of one^, two^, three^, and similar situations.

Partridge (Johannesburg): I suggest that in place of the term 'concessive', the speaker should use the term 'contradistinctive' for the phenomenon she has been discussing.

Proc. 5th int. Congr. phon. Sci., Münster 1964, pp. 521–524
(S. Karger, Basel/New York 1965).

Syllable-division, Duremes and Juncture in English

By A. E. Sharp, London

The descriptive technique leading to the formulation of so-called 'junctural' contrasts has been applied since its conception to a variety of languages, and might seem by this time an indispensable item in any phonological tool-kit. Yet in the description of English, where junctural terminology has been perhaps most insistently utilized, what might be called the 'basic' phenomenon of 'plus juncture' remains highly controversial. Its incidence has been stigmatized as 'optional'[1], 'dispensable'[1], 'vestigial'[2], 'sporadic'[2] and – by implication – inefficient[3], impossible[4] and useless[4]. Its use as a segmentator has been assailed as productive of most undesirable consequences. In some quarters, therefore, the whole technique is in danger of falling into disrepute. I hope that the necessarily brief observations that follow will help to clarify, rather than further confuse, the position.

The feature which, according to several American accounts, unites plus juncture with the other, 'terminal', junctures is one of timing, described in a recent survey as 'a stretching of preceding segmental phonemes'[5]. In some places, we are told, where no such stretching is audible, a 'plus' may nevertheless be instrumentally demonstrated; its presence is then in practice to be inferred from the occurrence of particular marginal allophones at the stated point. Now if juncture is to be given phonemic status, it is easy to understand the motivation behind the insistence on a constant feature of some kind: but a strong case can be made for regarding the allophonic variation in question on the one hand, and prolongation

---

[1] See Lit. 6, pp. 59–60.
[2] See Lit. 1, p. 28.
[3] See Lit. 4, pp. 171–176.
[4] See Lit. 2, p. 4 for juncture as viewed by "production-phonemics".
[5] See Lit. 8, p. 17 and the reference thereat.

phenomena on the other, as two primarily different things. Such a view requires, of course, that the allophonic variation be conditioned by something else; and the only possible candidate in this context is our old friend, syllable-division.

Here it must suffice to suggest that syllable-division is, as it were, the prerequisite for juncture, and establishes a point at which junctural contrasts may operate. These contrasts may then be viewed as essentially durational and be taken to comprise a system of 'duremes' whose several exponents would be different degrees of prolongation operative over a stated domain. The commutational possibilities would vary from place to place in structure, and at some points syllable-division alone, without additional prolongation, would be in question. I say *additional* because the allophonic variation correlated with the location of a syllable boundary may itself involve duration.

Thus duremes would occur only at points of syllable-division manifesting durational features not accounted for either by syllable-division itself or by any other factor such as stress. The precise terminology is less important than the distinction involved: if syllable-division is taken as the limiting case of juncture, it may be known as 'syllable juncture' and the duremes (or the *other* duremes if syllable-division is a kind of zero dureme) may have other names such as 'plus juncture', 'double-plus juncture', etc., where the 'plus' would have greater appropriateness than hitherto and be made to earn its name.

If in this way syllable-boundary and dureme were kept clearly distinct, difficulty would no longer arise over those cases where the boundary is relatively obvious but no prolongation seems either auditorily or instrumentally detectable. For example, it need occasion neither surprise nor dismay that 'an' in 'an aim' appears to have a shorter, rather than a longer, 'n' than 'name' in 'a name'[6]. If, of course, we re-define plus juncture, as is perhaps the trend, so that its domain extends both ways, we may escape from this particular difficulty, but we may still need to show that either phrase is longer overall than a single word of similar segmental and syllabic structure.

So much for the relationship between juncture and syllable-division[7]. In American doctrine, however, there has been a close

[6] See Lit. 3.

[7] See further, however, my earlier remarks in Lit. 5, pp. 104–135, especially 126ff.

connexion between juncture and intonation in that some features of the pitch patterns of English have been attributed to terminal junctures, themselves sometimes defined primarily in terms of contrastive degrees of prolongation greater than that credited to plus juncture. Now the British configurational tradition in intonational analysis is quite different from the system of pitch levels espoused by most American linguists, and this is not the place to argue their relative merits. It would seem, however, infinitely more satisfactory to let the intonation patterns stand on their own feet than to posit that they arise in part from supposedly contrastive timings.

In conclusion, a few words on the communicative relevance of junctural phenomena in English, which bears closely on the general viability of junctural formulations. On this score controversy centres around syllable and/or plus juncture, and the terminals may be disregarded. It seems imperative, in this connexion as in many others, to insist on the relativity of the phonetic distinctiveness of the exponents of phonological contrasts. Even two phonologically parallel minimal pairs may differ greatly in discriminability: the fact that in English the location of a syllable boundary with respect to, say, a nasal is less readily perceptible than is the case with a voiceless plosive[8] does not mean that we must abandon all thoughts of establishing contrasts at such points, still less that we must renounce such contrasts altogether. We should do well inter alia to re-examine all points where analogy predicts a syllable boundary which in practice is not easily perceptible. In any case, those who would discredit juncture offer us no acceptable alternative: whatever may be necessary at the moment, we cannot in the long run honourably remain satisfied with orthography-based delimitations unless these can be genuinely validated for speech. In the absence of such validation we must continue to search for some kind of phonological truth, though where alternative interpretations are available we shall naturally prefer that which is most congruent with our grammatical analysis. The search will be materially shortened if all participants avoid over-generalization and make clearer than has hitherto generally been the case what precisely in their view is the relationship of juncture to the other variables of phonology and for precisely what phonetic features it may be held accountable.

[8] See Lit. 4, p. 175.

References

1. *Bolinger, D. L.:* Length, vowel, juncture. Linguistics *1* (1963).
2. *Hill, A.:* Various kinds of phonemes. Studies in Linguistics *16:* 1 (1962).
3. *Lehiste, I.:* An acoustic-phonetic study of internal open juncture. Supplement to Phonetica *5* (1960).
4. *O'Connor, J. D.* and *Tooley, O. M.:* The perceptibility of certain word-boundaries. In Honour of Daniel Jones (Longmans, London 1964).
5. *Sharp, A. E.:* The analysis of stress and juncture in English. Transactions of the Philological Society 1960 (Blackwell, Oxford).
6. *Strang, B. M. H.:* Modern English structure (Edward Arnold, London 1962).
7. Second Texas Conference on Problems of Linguistic Analysis in English (Austin, Texas 1962).
8. *Trager, L.:* Some thoughts on "juncture". Studies in Linguistics *16:* 1 (1962).

Author's address: Prof. A. E. Sharp, School of Oriental and African Studies, University of *London, W.C. 1* (England).

Proc. 5th int. Congr. phon. Sci., Münster 1964, pp. 525–527
(S. Karger, Basel/New York 1965).

University of Hawaii

Some Experiments on Chinese (Mandarin) Tone Sandhi*

By YAO SHEN, Honululu

The present paper concerns the phonetic nature of morphophonemic Tone-2 of Mandarin Chinese. It deals with the morphophonemic Tone-2 derived from lexical Tone-3 in sequences of 33 when the second lexical 3 occurs as a neutral tone or Tone-0. Spectrograms of reduplicated forms are used to determine the phonetic nature. The speakers are two women whose first language is Mandarin Chinese.

It is generally said that the phonetic nature of lexical Tone-2 is mid-high and that of lexical Tone-3 is mid-low-high. The phonemic distinction between these two tones is either time or frequency or the "low", which will be called the "dip" in this paper, in Tone-3. It is fairly unknown, however, that spectrograms show that either tone can be longer or shorter than the other in time, either tone can be higher or lower than the other in frequency, and both tones have a dip. The distinctive features can now be described as follows: 1. Tone-3 dips late; Tone-2 dips early. 2. Tone-3 dips more; Tone-2 dips less. 3. Tone-3 rises less from vocalic-start to the end; Tone-2 rises more. (See A for Tone-3 and B for Tone-2.)

In a sequence of two lexical 3's: 33, the first 3 can occur as morphophonemic 2. The sequence is then morphophonemic 23: M23. It has been found that this morphophonemic 2 of M23 also has the same three distinctive features as lexical 2. (See C.) The results of the above study which concerns the phonetic, phonemic,

* Indebtedness goes to Mrs. *Jessica Chao*, one of the two Mandarin speakers; Miss *June Shoup* of Communication Sciences Laboratory, University of Michigan, Ann Arbor, Michigan; Mr. *Leslie Hanzely*, electronic technician of Institute of Languages and Linguistics, Georgetown University, Washington, D.C.

and morphophonemic 2 were presented in a paper by the present writer at the First World Congress of Phoneticians in Tokyo, Japan, August, 1960, and the paper was published in *Study of Sounds*, Volume IX (1961), Tokyo, Japan.

In a different context a lexical 3 followed by another 3 can occur as morphophonemic 3 followed by Tone-0: M30. For example: jyejye "elder sister", nainai "paternal grandmother", yangyang "to itch". This M3 of M30 has been described by *Y. R. Chao* and *Samual Martin* to have a dip but with no rise. *Mandarin Reader* describes it as the same as lexical 3, that is, it rises after the dip.

A lexical 3 followed by another 3 can also occur as morphophonemic 2 followed by Tone-0: M20. For example: swoyi "consequently", syaujye "Miss", keyi "can or be able to". This M2 of M20 has been described by *Chao* and *Martin* and in *Mandarin Reader* to have no dip.

This present work is to see if the morphophonemic 2 of M20 has a dip. If it does not, then the M2 of M20 is different from the M2 of M23. If it does, then is it like M3 of M30 or is it like M2 of M23?

Procedure: Minimal pairs of lexical Tone-3 and Tone-2 of the same finals were used, since the above mentioned study has proved that the tone occurs in the "final", that is, somewhere between the vocalic-start and the end.

Spectrograms similar to the following were run, and the 10th harmonics were traced.

1. Lexical 3 followed by "de": M30. Example: syede "written". D shows M3.
2. Reduplicated forms of 3: M30. Example: jyejye "elder sister". E shows M3.
3. Lexical 2 followed by "de": 20. Example: syede "oblique". F shows 2.
4. Reduplicated forms of 2: 20. Example: syesye "babytalk for 'shoes'". G shows 2.
5. Reduplicated forms of 3: M20. Example: syesye "write a bit". H shows M2.
6. Reduplicated forms of 2 in a stream of speech: 20. Example: changchang "taste a bit" in lai 'changchang' hauchr bu hauchr "Come to 'taste a bit' to see if it tastes good or not". I shows 2.
7. Reduplicated forms of 3 in a stream of speech: M20. Example: tangtang "lie down a bit" in shangfeng 'tangtang' jyou hau le "Cold will get well if 'lie down a bit'". J shows M2.

Description

1. Lexical 2 of 20 dips and rises. (See F, G, and I.)
2. M2 of M20 also dips and rises. (See H and J.) This disagrees with the descriptions given by *Chao* and *Martin* and in *Mandarin Reader*.
3. M3 of M30 dips but does not rise. (See D and E.) This agrees with the descriptions given by *Chao* and *Martin* but does not agree with that given in *Mandarin Reader*.

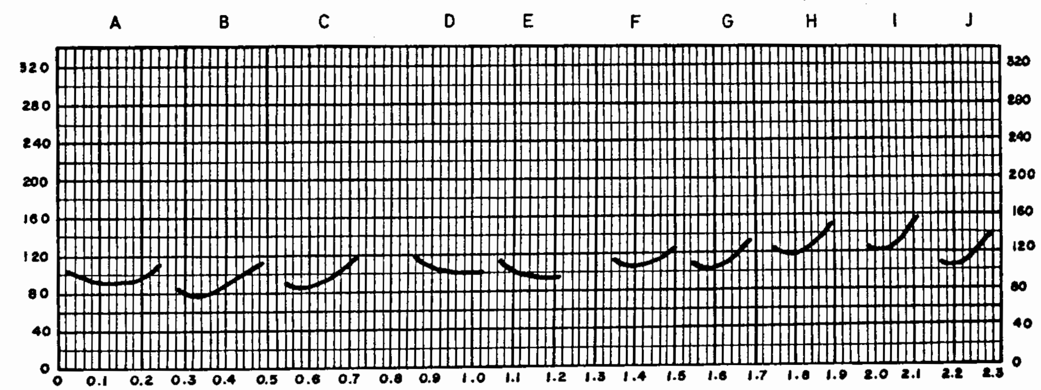


Fig. 1. Spectrographic illustrations.

Analysis

1. M2 of M20 (H and J) is not like Tone-3 in isolation (A).
2. M2 of M20 (H and J) is not like M3 of M30 (D and E).
3. M2 of M20 (H and J) is like Tone-2 in isolation (B).
4. M2 of M20 (H and J) is like 2 in 20 (F, G, and I).
5. M2 of M20 (H and J) is like M2 in M23 (C).

Conclusion: Morphophonemic 2 of M20 derived from 33 dips and rises.

Author's address: Dr. Y. Shen, University of Hawaii, Classroom Bldg. A, *Honolulu 14* (Hawaii).

Discussion

Weingartner (Hamburg): Research on the difference between Chinese tones 2 and 3 showed me no dips in tone 3 and dips in tone 2, especially at morphemes that begin with a vowel (which I write øv = Zero + vowel).

At least for these cases the distinction from a dip might not hold.

Verh. 5. int. Kongr. Phon. Wiss., Münster 1964, pp. 528–531
(S. Karger, Basel/New York 1965).

Die Möglichkeiten der Beeinflussung des alttschechischen phonologischen Systems durch das Mittelhochdeutsche

Von E. SKÁLA, Prag

Über das Problem der Realisierung von gemeinsamen lautlichen Veränderungen im Mittelhochdeutschen und Alttschechischen sind die Meinungen geteilt. Bewirkte das enge Zusammenleben der Tschechen und Deutschen in Böhmen, vor allem in Prag, in der Epoche der Entstehung der beiden Schriftsprachen tatsächlich die Durchführung von gemeinsamen lautlichen Veränderungen? Während *J. Gebauer* den deutschen Einfluß auf die Durchführung des Umlautes *a* > *ě*, der Diphthongierung *ú* > *au*, *ý* > *ej* und *ó* > *uo* > *ů* im Alttschechischen annahm[1], wurde diese Möglichkeit von *A. Beer*, *A. Kraus* und *F. Trávníček* bestritten[2]. Der ganze Fragenkomplex umfaßt noch die Monophthongierungen *ie* > *í*, *uo* > *ů* und die Beseitigung der Korrelation *y* – *i* und *ł* – *l*. Dieses Paar war das letzte, das die Mouillierungskorrelation aufgab, ähnlich wie *b*, *p*, *v*, *f*, *m*, *d*, *t*, *n*, *z*, *s*. *M. Komárek* nimmt in Anlehnung an die bereits im Jahre 1412 vom tschechischen Reformator Jan Hus ausgesprochene Meinung den deutschen Einfluß auf die Aufhebung der Mouillierungskorrelation als sehr wahrscheinlich an[3]. *Komárek* lehnt jedoch mit *Trávníček* die Annahme *Gebauers* ab, daß die alttschechische Diphthongierung *ó* > *uo*, *o* > *uo* von der ahd. Diphthongierung *ô* > *uo* beeinflußt wurde. Gegen *A. Kraus*, dem der Einfluß der mhd. Diphthongierung *û* > *ou* > *au* auf die alttschechische Diphthongierung *ú* > *ou* wahrscheinlich erschien, nimmt *Komárek* an, daß der Impuls zur Diphthongierung *ú* > *ou* von der ähnlichen Veränderung *ý* > *ei* ausging, die tatsächlich schon im 12. Jahrhundert belegt ist: *Teinéz* «ON Týnec», während *ú* > *ou* erst im 13. Jahrhundert belegt ist[4]. *N. S. Trubetzkoy* erkannte, daß sich der Zug zur Diphthongierung nicht als Wirkung des stark zentralisierenden Akzents, sondern der Silbenschnittkorrelation erweist[5]. Den Grund für die alttschechische Diphthongierung sieht *Komárek* in inneren Voraussetzungen des alttschechischen Lautsystems und in analogen Diphthongierungen in ost- und südslawischen Sprachen, wo kein deutscher Einfluß anzunehmen ist[6].

[1] *Gebauer*, S. 121, 263, 283, 246–247.

[2] *Beer*, S. 1–25. *Kraus*, S. 47; ders. in: Listy filologické XXXII, Praha 1905, S. 475 bis 476 und ebenda, Jg. XXXIII, S. 62–63. *Trávníček*, S. 72.

[3] *Komárek*, S. 108.

Es ist die Frage berechtigt, warum innere Gründe einmal gelten sollen und ein anderes Mal nicht. *A. Lamprecht* faßt das Problem des deutschen Einflusses auf die alttschechische Diphthongierung so auf, daß er zwar gewissermaßen deren Systemgebundenheit anerkennt, daß er jedoch zugleich die Ausbreitung des deutschen Einflusses vom zweisprachigen Prager Zentrum für notwendig hält[7]. Die Konfrontierung der Ergebnisse von *Komárek* und *K. B. Lindgren* läßt den Schluß zu, daß die Diphthongierung in beiden Sprachen parallel verlief. Beide Autoren nehmen an, daß die Diphthongierung in der gesprochenen Sprache um 1500 beendet war. Während sich *ý* in der tschechischen Schriftsprache bis heute halten konnte, lassen vereinzelte Belege, noch nach 1500, selbst in obd. Texten, die Schrifttradition erkennen[8]. *Lindgren* wies nach, daß im Deutschen kein Unterschied in der Diphthongierung zwischen Wörtern besteht, die der Apokope ausgesetzt sind und den anderen Fällen. Eine Differenz ist jedoch immer bei einsilbigen Wörtern festzustellen, die einen höheren Prozentsatz von Diphthongierungen aufweisen, so daß angenommen werden darf, daß die Diphthongierung zuerst und am stärksten in einsilbigen Wörtern einsetzte[9]. Für unsere Frage erscheint wichtig, daß sich in deutschsprachigen Quellen aus Böhmen die Diphthongierung langsamer durchsetzt als im Bairischen. In der deutschen Fassung der Chronik Dalimils vom Jahre 1389 erreicht die Diphthongierung *î* > *ai* 7%, *û* > *au* 10%. Um 1400 weist das Schriftbild in Lutwins Adam und Eva noch keine

[4] *Bergmann*, S. 236.

[5] *Trubetzkoy*, S. 176.

[6] *Komárek*, S. 104.

[7] *Lamprecht*, S. 91 und 93.

[8] So z.B. *zitt* «Zeit» im Jahre 1577 oder *stritt* «Streit» noch im Jahre 1653 in Egerer Texten. Vgl. *E. Skála*, S. 146.

[9] *Lindgren*, S. 54.

Diphthonge auf[10]. In der Egerer und Prager Kanzlei setzt sich die Diphthongierung *û* > *au* schneller durch als diejenige von *î* > *ai*[11]. Auch dieser Umstand spricht gegen die Möglichkeit eines Einflusses der deutschen Diphthongierung auf die tschechische, da im Alttschechischen zuerst *î* diphthongiert wurde.

Die Realisierung von gemeinsamen lautlichen Veränderungen im Mittelhochdeutschen und im Alttschechischen ist auf innere, systemgebundene Entwicklungstendenzen zurückzuführen. So wird auch die alttschechische Monophthongierung *ie* > *í* und *uo* > *ů*, oder die Veränderung des bilabialen *w* in labiodentales *v* allgemein aufgefaßt, wo ein deutscher Einfluß chronologisch genau so möglich sein sollte. Ebenso wird durch innere Entwicklungstendenzen die Aufhebung der Mouillierungskorrelation der Konsonanten im Alttschechischen zu erklären sein. Die Mouillierungskorrelation fehlt nämlich nicht nur im Tschechischen, im Slowakischen, im Kaschubischen, im Ober- und Niedersorbischen und im Slowenischen, wo man den Einfluß mhd. unmouillierter Konsonanten annehmen könnte, sondern auch im Serbokroatischen und Westbulgarischen, wo an einen solchen Einfluß nicht zu denken ist[12].

Literatur

Beer, A.: O stopách vlivu německého v češtině staré. S. Ber. königl. böhm. Ges. Wiss. *7:* 1–25 (1905).

Bergmann, F.: K chronologii některých staročeských zjevů mluvnických. Listy filologické *48* (1921).

Gebauer, J.: Historická mluvnice jazyka českého; I (Praha 1894).

Jakobson, R.: Über die phonologischen Sprachbünde. TCLP *IV* (1931).

Komárek, M.: Historická mluvnice česká. I. Hláskosloví (Praha 1958).

Kraus, A.: Jan z Michalovic. Německá báseň třináctého věku (Praha 1888).

Lamprecht, A.: K otázce vlivu jazyka na jazyk. Sborník prací filosofické fakulty brněnské university. Řada jazykovědná (Brno 1958).

Lindgren, K. B.: Die Ausbreitung der nhd. Diphthongierung bis 1500. Ann. Acad. Sci. Fenn. Ser. B, vol. 123/2 (Helsinki 1961).

Mourek, V. E.: Zum Prager Deutsch des XIV. Jahrhunderts. S. Ber. königl. böhm. Ges. Wiss. *3:* 41–44 (1902).

Skála, E.: Die Entwicklung der Kanzleisprache in Eger (1310–1660). Habil.-Schr. (Leipzig 1961).

Trávníček, F.: Historická mluvnice československá (Praha 1935).

Trubetzkoy, N. S.: Grundzüge der Phonologie. TCLP *VII* (1939).

Adresse des Autors: Dr. phil. habil. Emil Skála, Lumírova 1, *Prag 2* (ČSSR).

[10] *Ebenda,* S. 17.
[11] *Mourek,* S. 41–44; *Skála,* S. 37, 38.
[12] *Jakobson,* S. 236–237.

Discussion

Vachek (Praha): believes that the speaker's remarks were convincing. Still, it is possible that some foreign influence may be admitted if it coincided with the needs and wants of the system of language exposed to that influence.

Heinrichs (Gießen): Auch wenn man Herrn *Skálas* Beweisführung durchaus anerkennt, erregt es doch Verwunderung, daß in zwei verschiedenen, aber benachbarten und ineinander verwobenen Sprachen ähnliche oder gleiche Lautentwicklungen stattgefunden haben. Ähnliches kann man auch im germanisch-romanischen Grenzgebiet feststellen. (Z.B. lat. altus ∾ frz. haut, dt. kalt ∾ ndl. koud, rhein. ka:t; directum > drextu > droit; recht > rhein. reit, rɛ:t; frz. vin, – rhein. wiŋ usw.) Man fragt sich, ob nicht noch etwas anderes hinzugekommen sei, damit die Entwicklung so gelaufen ist. Man könnte z.B. daran denken, daß die ei- und au-Aussprache als die vornehmere galt, entweder zuerst bei den Deutschen oder Tschechen oder auch bei zweisprachigen Menschen der sprachlichen Hochschicht, seien es Tschechen oder Deutsche. Wie dem auch sei, die gleiche oder ähnliche Entwicklung in zwei fremden Sprachen bleibt verwunderlich und verlangt nach einer Erklärung, die natürlich keine Mystifikation sein darf.

Ising (Berlin): weist auf Ergebnisse der Sprachinselforschung hin, die in manchen Fällen auffällige Übereinstimmungen zwischen fremdsprachlichen Lautungen und deutschen Siedlungsmundarten erkennen lassen. Er stellt außerdem die Frage nach der sprachsoziologischen Rolle der Zweisprachigkeit und nach ihrer Verbreitung im mittelalterlichen Böhmen.

Antwort *Skála:* 1. Dazu, was Herr *Heinrichs* über die Verhältnisse am Niederrhein sagt, wurde von *E. Skála* bemerkt, daß diese gemeinsamen Entwicklungen auch anderswo (im Obersorbischen, im Slowakischen, im Slowenischen) zu beobachten sind, daß sie jedoch durch das eigene System in diesen Sprachen Eingang fanden. Andere Argumente lassen eine rationelle Deutung kaum zu.

2. Zum Diskussionsbeitrag von Herrn *Ising* wurde auf die Existenz des Systems der deutschen Schriftsprache in Regensburg und Nürnberg bereits um 1280, also vor dem Prager Deutsch, hingewiesen. An dieses System, nicht an Siedlermischung oder sprachsoziologische Überlegungen, ist bei der Beurteilung des Prager Deutsch zu denken.

Verh. 5. int. Kongr. Phon. Wiss., Münster 1964, pp. 532–535
(S. Karger, Basel/New York 1965).

Die Bestimmung der Stimmkategorien mittels Resonanzröhren

Von ANTTI SOVIJÄRVI, Helsinki

Vor einigen Jahren habe ich während der Stimmbehandlung mehrerer Kinder, die an *Rhinolalia aperta functionalis* litten, zufälligerweise für die gewöhnlichen Blasübungen auch solche Glasröhren anwenden lassen, die bei stimmhaftem Blasen ins Wasser eine kräftige Resonanz erweckten. Dabei konnte ich feststellen, daß diese Resonanz nicht immer ohne Störungen entstand. Als ich die Ursachen dieser Resonanzstörungen näher untersuchte, leuchtete es mir ein, daß mindestens die Länge der Röhre eine wichtige Rolle spielte. Meine Beobachtungen führten weiter auf folgenden Weg: die Länge der Röhre, mit der eine unstörbare Resonanz hervorgebracht werden konnte, war deutlich von dem Alter der behandelten Kinder (zwischen 8 und 12 Jahre) abhängig. (Der innere Durchmesser der Röhren war 8 oder 9 mm.)

Das Alter der Kinder	Die passenden Röhrenlängen (die Dicke der Glaswand 1 mm)				
8–10 Jahre	24 cm	–	24,5 cm	–	25 cm
11–12 Jahre	25 cm	–	25,5 cm	–	26 cm

Dadurch stellte ich folgende neue Fragen: Weshalb konnte ein Kind *A*, das z.B. 9 Jahre alt war, nur mit einer 24 cm langen Röhre eine einwandfreie Resonanz in beliebiger Tonhöhe hervorbringen, wogegen die Kinder *B* und *C*, die ebenso alt waren wie das Kind *A*, etwas längere Röhren von 24,5 bzw. 25 cm zur Erzeugung einer unstörbaren Resonanz anwenden mußten? Könnten solche individuellen Verschiedenheiten vielleicht auf Eigenschaften der Kinderstimme hinweisen, die mit den *verschiedenen Stimmkategorien der Erwachsenen* vergleichbar waren?

Unmittelbar nach dem Auftauchen des Problems habe ich be-

gonnen, die Korrelation zwischen den Stimmkategorien von Männern und Frauen und den verschiedenen Röhrenlängen (mit Durchmessern von 8 bis 15 mm) zu untersuchen. Als Versuchspersonen fungierten zuerst diejenigen Erwachsenen, die schon vorher auf Grund ihrer Praxis als Kunst- oder Chorsänger wußten, welche Stimmkategorie sie vertreten. Als ich solche Sänger dutzendweise im Hinblick auf die erwähnte Korrelation untersucht hatte, stellte ich fest,

1. daß der Röhrendurchmesser (zwischen 8–15 mm) keine wesentliche Rolle bei der Entstehung der einwandfreien und von der Tonhöhe unabhängigen Resonanz spielt und
2. daß aber die Länge der Röhren in folgender beinahe ausnahmslos eindeutigen Korrelation mit den Stimmkategorien steht:

		Der höchste Ton des Brustregisters
a) Tenor und Sopran	26 cm	*a*/ a^1
b) Tenorbariton und dramatischer Sopran	26,5 cm	*g*/ g^1
c) Bariton und Mezzosopran	27 cm	*f*/ f^1
d) Baßbariton und Mezzoalt	27,5 cm	*e*/ e^1
e) Baß und Alt	28 cm	*d*/ d^1

Ich konnte keine grundsätzlichen Ausnahmen in dieser Beziehung feststellen; denn Resonanzstörungen traten nur ein, wenn die betreffende Versuchsperson stets oder teilweise die Gesangstöne mit einer *erhöhten* Kehlkopfstellung phonierte.

Unter der Voraussetzung, daß die so geprüfte Korrelation wirklich stichhaltig ist, wären wir zu der praktischen Folgerung berechtigt, daß umgekehrt die individuelle Stimmkategorie einer erwachsenen Versuchsperson mit Hilfe der 26 bis 28 cm langen Glasröhren bestimmt werden könne. Dabei muß man jedoch darauf achten, daß das Anheben des Kehlkopfes der Versuchsperson vermieden wird.

Während der zwei letzten Jahre habe ich in meiner stimmphysiologischen Übungsbehandlung zuerst die richtige Stimmkategorie des Schülers oder des Patienten mittels des hier erklärten Verfahrens bestimmt und dann mit Hilfe der entsprechenden Glasröhre starke und sehr günstig wirkende Resonanzübungen durchführen lassen; bei diesen Übungen habe ich meistens in folgenden «Wörtern» ein langes *bb* phonieren lassen: [*jubbu: jübbü: jibbi:*]. (Diese Übungswörter beginnen mit dem Halbvokal *j*, um den harten Einsatz zu vermeiden.) Um die Sprech- und Singstimmengebung zu entwickeln, kann man nach meiner Erfahrung am besten gleich mit diesen Röhrenresonanzübungen beginnen. Einige Schüler und Schülerinnen, die vorher das Problem ihrer individuellen Stimmkategorie nicht haben lösen können, wurden mittels der Röhrenresonanz-Übung davon überzeugt, welche Stimmkategorie sie eigentlich vertreten. Die ausgeglichen klingende Stimme solcher Schüler oder Schülerinnen beginnt bald dem bekannten Umfang der bestimmten Stimmkategorie zu entsprechen, vorausgesetzt, daß man täglich und während *einiger Wochen* diese Übungen durchgeführt hat und daß keine eigentlichen Stimmstörungen vorhanden waren, als die Übungen begonnen wurden. Ich habe immer positive Erfahrungen bei der Anwendung dieses Resonanzverfahrens gemacht. Bis heute habe ich in Finnland über 100 Schüler oder Patienten damit behandelt, und einige Kollegen in Schweden, Norwegen und Dänemark, denen ich bisher von diesem Verfahren erzählt habe, haben meines Wissens ebenso erfreuliche Resultate erzielt.

Die theoretischen Grundlagen des Verfahrens habe ich bis heute noch nicht ausführlich bearbeitet. Aber ich habe folgende und offenbar sehr wichtige Korrelationen zwischen den Längen der Glasröhren und den Dimensionen der Sprechwerkzeuge festgestellt. Der bucco-tracheale Abstand zwischen den Spitzen der Vorderzähne und der Bifurkation der Luftröhre variiert nach den Meßresultaten mehrerer Röntgenaufnahmen von singenden erwachsenen Versuchspersonen zwischen 26 und 28 cm, und zwar dergestalt, daß dieser Abstand mit der jeweils verwendeten Röhrenlänge und der entsprechenden Stimmkategorie der Versuchsperson übereinstimmt. Nur in solchen Fällen habe ich kleine Ausnahmen festgestellt, in denen das *Spatium hyothyreoideum* der Versuchsperson sich wegen hyperkinetischer Spannungen einige Millimeter verengert hat oder in denen die Versuchsperson mit einer erhöhten Kehlkopfstellung gesungen hat. Um diese Nachteile soweit wie möglich zu vermeiden, habe ich in der letzten Zeit nur solche Versuchspersonen für die Röntgenaufnahmen ausgewählt, die beim Singen keine solche Mängel aufwiesen. Damit sind auch die wenigen Ausnahmen bei den Meßergebnissen verschwunden.

Zum Schluß erwähne ich nur ein Beispiel von meinen Meßresultaten. Die Versuchsperson (Sopran L.K.) hat beim Singen

(des Vokals *u* in der Mittellage) einen Abstand zwischen den Vorderzähnen und der Spitze des Zäpfchens von 7,3 cm, einen Abstand zwischen dem Zäpfchen und der Oberfläche der Stimmlippen von 8,2 cm und einen Abstand zwischen den Stimmlippen und der Bifurkation von 10,5 cm gehabt. Der ganze bucco-tracheale Abstand ist also 26 cm gewesen.

Adresse des Autors: Prof. A. Sovijärvi, Phonetisches Institut der Universität, Hallituskatu 11–13, *Helsinki* (Finnland).

Discussion

Trojan (Wien): Um ein Bild von den Wirkungen der von Herrn Sovijärvi angegebenen Resonanzröhrenmethode zu gewinnen, habe ich meine Mitarbeiterin *Hertha Weihs*, die sich durch ihre Arbeiten zur Pathologie und Charakterologie der Stimme bekannt gemacht hat, gebeten, diese Methode einer Überprüfung zu unterziehen. Ihr Ergebnis beschreiben die folgenden Ausführungen:

«Bei der von Herrn Sovijärvi angegebenen Resonanzröhrenmethode empfinden nach meinen vorläufigen Untersuchungen die Patienten eine merkliche Erleichterung der Stimmgebung, Verstärkung der Resonanz und eine deutliche Hartwandigkeit der supraglottischen Räume, und zwar im Sinne eines angenehmen Spannungsgefühles, das noch bis zu acht Stunden nachher anhalten kann, insbesondere im Bereich der Gaumensegel-, Zungen- und Wangenmuskulatur. Besonders nachhaltig werden im günstigen Sinne die Nackenmuskulatur und die Mm. sternocleidomasteodei, die propriozeptive Steuerung der Phonation und die Nasenatmung beeinflußt. Allerdings sind bestimmte Voraussetzungen für die Übungen notwendig: der subglottische Atemdruck muß bei inspiratorischer Spannung optimal beibehalten und der Ton somit ohne Luftverschwendung gebildet werden; auch muß der Wasserspiegel im Röhrchen konstant bleiben, so daß keine Luftblasen aufgeworfen werden. Die entsprechende Röhrenlänge ist im Hinblick auf die Stimmgattung nach meinen Erfahrungen tatsächlich Voraussetzung für die beschriebenen günstigen Wirkungen, die von den Patienten spontan angegeben werden. Unter den 20 von mir untersuchten Fällen befand sich auch eine zentral-bedingte Rhinolalia aperta; auch dieser Fall hat günstig angesprochen.»

Verh. 5. int. Kongr. Phon. Wiss., Münster 1964, pp. 536–541
(S. Karger, Basel/New York 1965).

Aus dem Deutschen Seminar der Universität Erlangen-Nürnberg

Dialektentstehung als Ergebnis von Systemmischungen

Von H. Steger, Münster

Synchronie und Diachronie in den mundartlichen Entsprechungen der mittelhochdeutschen Reihen /ē/ – /œ/ – /ō/ und /ei/ – /öu/ – /ou/ im Oberostfränkischen

In einem Teil der Mundarten des östlichen Franken, hier oberostfränkisch genannt, zeigen die in der mhd. Literaturperiode ē–œ–ō geschriebenen Laute in der älteren Mundart die Entsprechungen ī–ȳ–ū. Heutiges Fehlen dieser Hebung des Vokals in bestimmten Positionen (vor r, n) und bei Verkehrswörtern bleibt unberücksichtigt. In den mdal. Phonemen /ī/ – /ȳ/ – /ū/ haben sich außer den genannten mhd. Lauten auch die mdal. Entsprechungen zu mhd. e, ö, o in Dehnung (e^{d}, $ö^{d}$, o^{d}) sowie i^{d}, $ü^{d}$, u^{d} und die mdal. Reflexe von mhd. ie, üe, uo (außer vor r und h) vereinigt. Auf der anderen Seite treffen wir die mdal. Realisation ā als Entsprechung für die mhd. ei, öu, ou geschriebenen Laute. Wir gewinnen ein mdal. Phonem /ā/.

Beide Erscheinungen zusammengenommen ergeben ein mdal. Langvokalteilsystem folgender Gestalt:

ī ȳ ū

ā

Die noch fehlende Reihe der Langvokale mittlerer Zungenstellung /ē/ – /œ/ – /ō/ werden im Oberostfr. von den Entsprechungen für mhd. $ë^{d}$, $ä^{d}$, æ sowie a^{d}, â gebildet; sie bleibt für unsere Betrachtung beiseite[1]. Ebenso soll in allen Reihen der Umlaut nicht

[1] Vgl. dazu *H. Steger*, Sprachraumbildung....

beachtet werden, da er als nachträgliche Abspaltung von dem velaren Ast hier nicht wichtig ist.

Wir erinnern uns, daß die mhd. Reihe /ē/ – /œ/ – /ō/ aus der germ. Reihe /ai/ – /au/ unter bestimmten Bedingungen hervorgegangen ist, ferner daran, daß die Schreibungen ei und ou für germ. ai–au in ahd. Zeit vom rheinischen Westen ausgegangen sind[2].

Diese Vorgänge lassen sich so darstellen:

germ. ai > mhd. /ē/ vor h, r, w
> mhd. /ei/ in den übrigen Positionen

germ. au > mhd. /ō/ vor h und allen Dentalen
> mhd. /ou/ in den übrigen Positionen

Da die Stellung von ai und au vor bestimmten Lauten darüber entscheidet, ob die monophthongische oder die diphthongische Variante geschrieben wird, handelt es sich zunächst zweifelsohne um eine allophonematische Differenzierung der Reihe, die keiner eigenen Zeichen bedürfte. Von selbständigen Phonemen kann man erst sprechen, seitdem etwa durch den Schwund alter Halbvokale und durch die Vorgänge der 2. Lautverschiebung die Stellungsabhängigkeit der Realisation nicht mehr vorhanden war.

Nicht übersehen werden darf auch, daß die Ausgliederung einer selbständigen Phonemreihe durch die Zeichen ⟨ê⟩ und ⟨ô⟩ im Schriftbild sichtbar gemacht wird. Ihre tatsächliche Realisation wird damit nicht unter allen Umständen beschrieben[3].

Werfen wir nun einen Blick auf die Mundartkarten für die Lautreihen mhd. ē–œ–ō einerseits und mhd. ei–öu–ou andererseits, so fällt ins Auge, daß die Differenzierung des germ. ai und au im gesamten md. Bereich (Rheinland, Oberhessen, Thüringen, Obersachsen) zur Realisation

germ. ai > mhd. */ē/ > mdal. md. /ī/ [weh]
> mhd. */ei/ > mdal. md. /ē/ [heiß]

germ. au > mhd. */ō/ > mdal. md. /ū/ [hoch]
> mhd. */ou/ > mdal. md. /ō/ [auch]

neigt[4].

[2] *H. Brinkmann*, bes. S. 158ff. (Lit.).

[3] Vgl. jetzt die umsichtige Studie von *A. van der Lee*, bes. S. 145ff. – *H. Penzl*, S. 174ff. – *W. G. Moulton*, bes. S. 15ff., 19 .

[4] DSA Kten. 16, 33, 87–90. Ferner handschriftliches Material.

Auch dort, wo germ. ai durch diphthongische Entsprechungen wiedergegeben wird (ei, e^1), bleibt zusätzlich der Unterschied in der Hebungsstufe gegeben.

Im allgemeinen ist jedoch die Opposition Monophthong-Diphthong in eine Opposition hochzungiger Monophthong – mittelhochzungiger Monophthong umgesetzt.

Der Blick auf Oberdeutschland läßt erkennen, daß hier die Entsprechung zu germ. ai (> mhd. *ē) und germ. au (> mhd. *ō) jeweils durch eine vielfältige Lautskala reflektiert werden (ę̄, *ǟv*, *ǟ*i bzw. ǭ, å*v*, åu); aber es fehlt ihnen durchwegs das Charakteristikum der Hebung.

Unter den mdal. obdt. Entsprechungen von germ. ai und au ragt ā für germ. au (> mhd. *ou) hervor. ā, ai, å*a*, åi als Entsprechungen von germ. ai (> mhd. *ei) sind zwar differenziert, zeigen aber durchwegs Tiefzungenstellung des ersten Bestandteils. Verglichen mit dem Md. fällt damit auf, daß ebenfalls ein Hebungsunterschied die Opposition bildet, daß aber, verglichen mit der md. Differenzierung, eine jeweils tiefere Stufe der Zungenstellung eingehalten wird. Man wird zur Verdeutlichung die obdt.Form der Differenzierung von germ. ai und au in folgendem Bild idealisieren dürfen:

germ. ai > mhd. */ē/ > mdal. obdt. /ę̄()/ [weh]
> mhd. */ei/ > mdal. obdt. /ā()/ [heiß]

germ. au > mhd. */ō/ > mdal. obdt. /ǭ()/ [hoch]
> mhd. */ou/ > mdal. obdt. /ā()/ [auch]

Es scheint, daß uns hier ein alter obdt.-md. Systemgegensatz vor Augen tritt; denn man wird es für wenig wahrscheinlich halten, daß eine nachträgliche obdt. Senkung von dem schriftmhd. ei und ou zu den tiefzungigen Entsprechungen des Obdt. geführt hätte, vielmehr wird man die schon von *Schwarz* und anderen vertretene These gestärkt sehen, daß das Obdt. in den Grundmundarten immer beim germ. Stande von ai und au geblieben war, während das Md. offenbar in merowingischer und karolingischer Zeit eine Hebung durchführte, die im Schriftbild ei und ou ihren Ausdruck fand und als Schriftform auch nach Oberdeutschland ausstrahlte [5].

Die Fülle der entstehenden Fragen kann hier nicht behandelt werden. Wir notieren als offen besonders :

[5] *E. Schwarz*, (1955), S. 62. Ders., (1962), S. 155ff., Abb. 32.

1. In welchem Zusammenhang steht die beobachtete Erscheinung der Differenzierung von germ. ai und au mit dem scheinbar nord-südlichen Fortschreiten der Monophthongierung dieser beiden Laute, die im niederdeutschen Bereich in allen Positionen zu Monophthongen geführt hat?

2. Ist die md. Differenzierung von ai in ē(i) und ī sowie von au in ō(u) und ū schon auf der germ. Stufe, zur Zeit der allophonematischen Regelung, als unterschiedliche Zungenhöhendifferenzierung vorhanden gewesen, so daß im Md. eine nachträgliche Hebung des Gesamtsystems um eine Stufe stattfand? Bei dieser Frage ist auch in Betracht zu ziehen, daß auch die Weiterentwicklung von germ. $ē_2$ [Brief] und ō [gut] vom md. Bereich ausgehend zu gehobenen Diphthongen geführt hat.

Im Augenblick kommt es uns nur darauf an zu fragen, wie vor dem nun entworfenen Bild die eingangs dargestellten oberostfr. Verhältnisse erklärt werden können. Denn im Oberostfr. scheint mdal. ā aus mhd. *ei, *ou dem oberdt. Muster zu entsprechen, während mdal. oberostfr. ī, ȳ, ū aus mhd. *ē, *œ, *ō das md. Muster wiederholen.

germ. ai > mhd. */ē/ > mdal. oberostfr. /ī/ [weh]
> mhd. */ei/ > mdal. oberostfr. /ā/ [heiß]

germ. au > mhd. */ō/ > mdal. oberostfr. /ū/ [hoch]
> mhd. */ou/ > mdal. oberostfr. /ā/ [auch]

Gegenüber den mhd. Graphemen ē und ō liegt also eine Hebung vor, gegenüber den Graphemen ei und ou eine Senkung.

Angesichts des beschriebenen gleichartigen Grundverhaltens der md. und oberdt. Mundarten bei der Ausgliederung einer neuen Phonemreihe aus der /ai/ – /au/-Reihe kann man dies kaum für ursprünglich halten, man muß vielmehr daran denken, daß die Eigenart des oberostfr. Systems als die Folge einer Mischung fertiger Systeme anzusehen ist, nämlich des md.-thür. und des oberdt.-alem.-bair. Dabei hätte beim Sprachausgleich das Md. die Hebung der schriftmhd. Langvokale mittelhoher Zungenstellung und das Oberdt. die Beibehaltung der Tiefzungenstellung bei den germ. Diphthongen mit tiefzungigem ersten Bestandteil beigesteuert.

Mit sprachlicher Argumentation wäre damit ein Hinweis auf die Entstehung des Oberostfr. aus der Mischung thür. und oberdt. Mundarten gegeben.

Es zeigt sich, daß es sich beim Oberostfr. nicht um die Entfaltung germ. Sprache handeln wird, die durch die völkerwanderungszeitliche Landnahme elbgermanischer Stämme in das Land um Regnitz und Obermain kam. Vielmehr wird erst bei der endgültigen Aufsiedlung dieses Landstriches nach dem (8.)–9. Jahrhundert durch die Mischung nordw. und südl. benachbarter Dialekte das Bild Oberostfr.'s als Sprachlandschaft geprägt.

Der sprachliche Gesamtbefund zeigt dabei in der Konfrontierung mit den historischen Quellen tatsächlich, daß bei diesen Vorgängen der Sprachraumbildung in Oberostfr. wesentlich thür.-grabfeldische und südfr.-bair. Siedlungszusammenhänge heraustreten[6].

Es scheint, daß durch die Beobachtung von Systemmischungsvorgängen die Mundartforschung vorankommen kann.

Literatur

Brinkmann, H.: Sprachwandel und Sprachbewegungen in althochdeutscher Zeit (Frommann, Jena 1931).

van der Lee, A.: Zur Aussprache der gotischen Digraphen ai und au. Festgabe für *L. L. Hammerich* (Naturmetodenssproginstitut, Kopenhagen 1962).

Moulton, W. G.: Zur Geschichte des deutschen Vokalsystems. PBB *83:* 1–35 (1961).

Penzl, H.: The development of Germanic ai and au in old high German. Germanic Review *22:* 174 (1947).

Schwarz, E.: Die elbgermanische Grundlage des Ostfränkischen. JffL *15:* 62 (1955). – Sudetendeutsche Sprachräume 2 (Handbuch der Sudetendeutschen Kulturgeschichte 2) (Lerche, München 1962).

Steger, H.: Sprachraumbildung und Landesgeschichte im östlichen Franken. Schr. d. Inst. f. fränk. Landesforschung 12 (Degener, Neustadt/Aisch 1964).

Adresse des Autors: Doz. Dr. H. Steger, Germanistisches Institut der Universität, Domplatz 20–22, *44 Münster* (Deutschland).

Discussion

Wiesinger (Marburg): Die Voraussetzung, daß bei Weiterentwicklungen der einzelnen mhd. Reihen eine das gesamte System beherrschende Tendenz als Senkung (Teiloberdeutsch) oder Hebung (Teilmitteldeutsch) vorliegen müsse, scheint keine allgemeine Gültigkeit zu besitzen. Die Monophthong- und Diphthongsysteme der oberdeutschen Mundarten zeigen gerade bezüglich der Reihen mhd. ei – ou – öü und mhd. ê–ô–ö verschiedenes Verhalten. Überhaupt kann im Osten des Alemannischen seit etwa 1000 und im Bairischen seit etwa 1150 nicht von einer Reihe mhd. ei – ou – öü gesprochen werden, da eine Gliederverschiebung stattgefunden hat. Mhd. ei (= ei[1]) wurde dort vom neuen Kontraktionsdiphthong ei[2] aus der ahd. Lautfolge *egi* aus dem ursprünglichen Reihenzusammenhang in die Isolierung gedrängt, so daß die Reihe ei[2] – ou – öü heißt,

[6] Vgl. *Steger.*

der ei[1] isoliert gegenübersteht. Im alem. Osten und im Bairischen bestehen in mhd. Zeit offene /ē – ō – ö/ und teils offene, teils geschlossene Diphthonge /ei – ou – öü/ und isoliertes /ai/. Diese Konstellation erlaubt verschiedene Entwicklungen. Es kommen in Frage: 1. der Zusammenfall durch Diphthongierung und Senkung in /ai/ – /au/ im zentralen Schwäbischen. 2. die Hebung der einen Reihe bei unverändertem Öffnungsgrad der anderen, im südlichen Ostschwäbischen. 3. die Senkung der einen Reihe bei verschiedenem Öffnungsgrad der anderen, im Nordbairischen.

Trost (Prag): Ich stimme mit Herrn *Wiesinger* darin überein, daß die «Reihenschritte» nicht selten gestört sind. Ich möchte nicht in den Reihenschritten das Um und Auf der Phonologie erblicken und auch nicht in Lautähnlichkeit das Um und Auf des Lautwandels. Entscheidend ist die Tendenz zur Aufrechterhaltung von Differenzen. Systemmischung ist zweifellos von Bedeutung, doch liegt sie kaum jemals darin, daß ein System A zu soundso viel Prozent aus dem System a und im übrigen aus dem System b besteht. Wenn ein System irgend etwas aus einem anderen System übernimmt, so kommt es natürlich zur Anpassung des Alten an das Neue, und es entsteht in der Regel ein neues Ganzes.

Beyer (Strasbourg): Zu dem Vortrag von Herrn *Steger* habe ich Folgendes zu bemerken:

1. Die Hebung vom mhd. *ê*, *ô* zu *i:* und *u:* ist sicherlich ein mitteldeutsches Merkmal. Diese Wandlung ist belegt von Luxemburg und Westlothringen bis Schlesien. In gewissen mitteldeutschen Gebieten hat sich *i:* (<*ê*) zu *ei* weiterentwickelt; so auch *u:* (< *ô*) zu *ou:*. Die Diphthongierung ist gegebenenfalls an die Stellung oder an die Umgebung gebunden.

Es stellt sich nun die Frage, ob das nord- und mittelbairische *ou* (*öu*) nicht aus *u:* hervorgegangen ist; dieses *u:* könnte mitteldeutscher Import gewesen sein.

Sicher scheint zu sein, daß die mitteldeutsche Tendenz zur Hebung von *ê*, *ô* in das Elsaß eingedrungen ist, wo geschlossene *e:*, *o:* bis südlich Colmar belegt sind.

2. Was das oberostfränkische *a* anbelangt, so ist es nicht möglich anzunehmen, daß es direkt aus germ. *au* herzuleiten ist. Für das gesamte Oberdeutsche müssen wir mhd. *ou* ansetzen, denn geschlossene Lautungen gibt es noch in vielen oberdeutschen Landstrichen.

Im Oberostfränkischen scheint *au* (< *ou*) unter dem Druck des bairischen Diphthongs zu *a* ausgewichen zu sein (baux «Bauch»/ rax «Rauch»).

C.R. 5e Congr. int. Sci. phon., Münster 1964, pp. 542–543
(S. Karger, Basel/New York 1965).

Le phonème, sa projection psycho-sensorielle, sa réponse psycho-motrice

Par A. Tomatis, Paris

Le phonème est un produit de l'acte parlé.

Or, cet acte parlé est un acte moteur. Il répond à l'exécution d'un déterminisme qui a fait éclore dans les mécanismes complexes de l'élaboration du geste articulatoire tout un ensemble d'impulsions volontaires qui vont jusqu'à réaliser à la perfection ce qu'une représentation motrice a évoqué.

Un souci d'analyse prétend isoler ce phonème en une entité qui se voit dès lors accorder des critères physico-acoustiques qualitativement et quantitativement définis.

Pourtant, ce produit, aussi mouvant que la vie qui le crée, s'accommode fort peu du cadre formel et temporel dans lequel on le veut insérer et qui lui retire, par sa rigidité, toute raison d'exister.

Aussi, croyons-nous qu'il serait bon d'observer, vis-à-vis du langage, une attitude clinique qui veut, dans une sémiologie précise, mettre en évidence cette organisation interne corporelle d'où émane l'acte parlé. Avec ce recul, la distance d'observation devient suffisante pour que le corps en sa totalité apparaisse comme l'instrument du langage.

Mais quel instrument exceptionnel, qui sait traduire selon notre désir, en fonction de notre volonté, tel ou tel trait saillant d'une pensée qui nous habite, en des termes informationnels transmissibles à l'autre. Au surplus, cet instrument exceptionnel va non seulement nous permettre la communication avec autrui, mais aussi nous assurer la possibilité de contrôler cette coulée verbale, grâce à des affecteurs neurologiques qui éveillent notre conscience sur les qualités de l'acte ainsi élaboré.

En la circonstance, le corps parlant devient un moi objet du langage, obéissant et asservi en quelque sorte au moi pensant qui supervise notre ego.

Le langage, dès lors, apparaît comme émanant de l'acte moteur le plus élaboré et le mieux contrôlé que nous sachions réaliser. Mais déjà ce dernier fait appel à un acte cristallisé suivant une représentation antérieure, autrement dit elle se réfère à un acte mémorisé ou tout au moins retrouvé grâce à la mémorisation d'une «image motrice», afin qu'il puisse adhérer à tout moment au schéma verbal souhaité, nos affecteurs sensoriels en assurant le contrôle.

Le premier d'entre eux est l'oreille; ensuite le corps y participe en totalité, par la sensibilité cutanée.

Ces affecteurs ont deux rôles:

– en premier lieu, ils permettent de capter l'information verbale qui leur parvient de l'extérieur; à eux d'en apprécier et d'en dépouiller le codage;

– en deuxième lieu, ils rendent possible la régulation au cours de l'information émise pour l'extérieur, en direction de l'autre, vers autrui, en jouant le rôle de capteur auto-informateur.

En définitive, dans le jeu de notre mémorisation, notre système sensoriel seul a été sollicité à l'appel de l'image motrice venant de l'extérieur; on a joué de lui comme on joue d'un instrument, et, lors de la reproduction de la structure vocale désirée, les affecteurs se sont apprêtés à contrôler la même identité informative que celle qu'avait fait surgir chez eux la même excitation sensorielle.

Ainsi, tout élément quel qu'il soit, prélevé dans la chaîne parlée, et notamment le phonème que l'on veut isoler, se comporte comme un stimulus complexe qui détermine une image sensorielle et c'est à la recherche de cette image enclenchée par nos affecteurs que nous parvenons à réaliser notre acte moteur dont le but informationnel n'est, en tout premier lieu, que celui d'informer les sens de celui qui parle.

Adresse de l'auteur: Dr A. Tomatis, 78, av. Raymond-Poincaré, *Paris 16e* (France).

Discussion

Francescato (Amsterdam) souligne la necessité de contractes plus intimes entre les linguistes et les psychologues, physiologues, etc. Cette necessité est démontrée par l'usage du mot «phonème», du reste défini dès le commencement, dont a fait usage M. *Tomatis*, et qui ne correspond pas à la terminologie des linguistes.

Pohl (Bruxelles): Comme pour la plupart d'entre vous, la communication de M. *Tomatis* a été une révélation, mais je ne peux cacher que j'ai été choqué par le mot *phonème*.

Les mots *sons* et *allophones* ne conviennent peut-être pas non plus parfaitement pour désigner le tronçon de la chaîne parlée auquel songe M. *Tomatis*. Le son, par exemple, peut désigner autre chose qu'un élément du language. Je me demande si on ne peut pas se mettre d'accord en se servant d'un mot qui existe, mais qui est peut-être pas encore très répandu, celui de *phonète*.

Verh. 5. int. Kongr. Phon. Wiss., Münster 1964, pp. 544–547
(S. Karger, Basel/New York 1965).

Grundfragen der Biophonetik

Von Felix Trojan, Wien

Das Anliegen der klassischen, vorexperimentellen Phonetik war die Erforschung der Lautbildung in der verstandesmäßigen Rede. Diese Linie setzt sich in der älteren experimentellen Phonetik wie in der Phonologie, Phonometrie und Informationstheorie fort. Nach *Trubetzkoy* erfordert die phonologische Beschreibung einer Sprache in erster Linie eine Untersuchung ihrer Darstellungsebene; daneben soll freilich auch die «phonologische Lautstilistik», und zwar sowohl der Kundgabe- wie der Appellfunktion, berücksichtigt werden. Beschränkt sich eine solche phonologische Lautstilistik auf die konventionell gewordenen Einwirkungen des Affektlebens auf die Phoneme, so hatte *Meyer-Eppler* kurz vor seinem Tod weit Umfassenderes im Sinn: er plante in Zusammenarbeit mit dem Vortragenden eine Einbeziehung des Gesamtgebietes des stimmlichen Ausdrucks in die Informationstheorie.

Tatsächlich hat die biologische Betrachtung der in den Bereich des Gehörs eingreifenden funktionellen Lautgebung (die von den funktionslosen Wirkungen lebender Organismen auf das Gehör streng zu sondern ist) den Horizont nach zwei Seiten hin erweitert. Einerseits wurden die Leistungen der physischen Generatoren des stimmlichen Ausdrucks, vor allem die Funktionsrichtungen der vegetativen Steuerung, die subkortikale Lust-Unlust-Diakrise und die Registerbildung mit ihrer expressiven Symbolik ins Licht gerückt (*Trojan*, *Kaiser*, *H. Weihs*, *Fónagy* u.a.), anderseits hat die Erforschung der schier unübersehbaren Mannigfaltigkeit im anatomischen Bau der Stimmwerkzeuge bei den verschiedenen Tierarten (*Göppert*, *Némay*, *Negus*, *Kelemen*, *DuBrul*) und die Beobachtung und Registrierung der tierlichen Stimmfunktionen (*Regen*, *Schwartzkopf*, *Frings*, *Busnel*, *Tembrock*) wissenschaftliches Neuland erschlossen, von dem bisher so gut wie keine Wege zur Humanphonetik führen. Dennoch müssen enge Beziehungen zwischen der «Bioakustik», ins-

besondere zwischen der Lautgebung der Primaten und dem stimmlichen Ausdruck beim Menschen bestehen. Die Theorie dieses stimmlichen Ausdrucks umfaßt nicht nur den Ausdruck der Sprechstimme, sondern auch den vorsprachlichen Ausdruck beim Kinde. Obwohl die Generatoren hier wie dort dieselben sind, wird man den Brückenschlag doch zweckmäßig (wie auch *Tembrock* meint) vom vorsprachlichen Ausdruck her in Angriff nehmen. Von Bedeutung ist in diesem Zusammenhang auch die Frage, wieweit der stimmliche Ausdruck beim Menschen international gleichartig ist und wie in den einzelnen nationalen Kulturen die Konvention das physiologisch Bedingte steigert oder hemmt.

Unabtrennbar von dem hier so weit gefaßten Fragenkreis des stimmlichen Ausdrucks ist ferner die in letzter Zeit wieder mehrfach aufgeworfene Frage des Sprachursprungs. Vom Standpunkt der Theorie des stimmlichen Ausdrucks und der Lehre von den Primär- und Sekundärfunktionen aus gesehen, erscheint es als unmöglich, die Sprache vom stimmlichen Ausdruck herzuleiten. Licht erhält dieses Problem auch von seiten der vergleichenden Anatomie des Kehlkopfes. Nach *Kelemen* wie nach *DuBrul* ist die Entstehung der Lautsprache an bestimmte periphere Voraussetzungen gebunden, so vor allem an die Erschließung des Mundraumes für die Lautbildung durch Aufgabe der retrovelaren Stellung der Epiglottis. *Kelemen* mißt diesen peripheren Voraussetzungen größere Bedeutung zu als der Entwicklung des Gehirns. *Bryan* (Current Anthropology 4, 1963) schließt sich ihm an und schlägt anstelle der bisherigen Bemühungen, Primaten eine menschliche Sprache zu lehren, Versuche vor, durch die sie ihre arteigenen Laute mit Bedeutungen zu verbinden lernen.

Ist nun auch die Sprache nicht aus dem stimmlichen Ausdruck hervorgegangen, so hat er sie doch mitgeformt. Das zeigt sich vor allem darin, daß in den phonologischen Oppositionen der Sprache die Thematik der schon genannten Generatoren des stimmlichen Ausdrucks als bloßes Formelement aufgenommen und variiert wird. Für den am stärksten wohl in der deutschen Hochsprache ausgeprägten Gegensatz von losem und festem Anschluß habe ich dies vor kurzem durch einen psychologischen Versuch aufgewiesen. Ähnliches gilt auch für den Gegensatz von Vokalismus und Konsonantismus, insbesondere den stimmlosen Verschluß- und Engenlauten, und für die Entgegensetzung dieser und anderer stimmloser Laute zu ihrem stimmhaften Korrelaten. Alle diese sprechmotorisch bedingten Oppositionen stehen in Zusammenhang mit der tonischen Polarität des Vegetativums, zu dem Gegensatz ergo- und trophotroper Phasen und ihrem stimmlichen Ausdruck.

Wäre eine beliebige Sprache nichts anderes als ein Informationssystem, dann hätte sie keinen in einer bestimmten Weise «menschlich anmutenden» Charakter. Wie sich im Ausdruck jeder Persönlichkeit subkortikale und kortikale Funktionen manifestieren, so auch im kollektiven Ausdruck der Sprache die gesamten in einer Sprachgemeinschaft vorherrschenden Wesenszüge. Damit aber erschließt sich der Aufgabenkreis einer Diagnostik der Einzelsprachen und Dialekte als Ausdrucksgestaltungen von Sprachkollektiven. Um ein konkretes Beispiel zu bringen: Worin liegen die ausdrucksmäßig wichtigsten phonetischen Gegensätze zwischen dem Deutschen und dem Italienischen? Das Deutsche ermöglicht durch die Dominanz des Konsonantismus eine betont ergotrope Sprechweise; dazu kommen die traditionelle Bindung motorischer Energien an das Sinnwichtige durch die Stammsilbenbetonung und durch die sprachlogische Akzentuierung und endlich die Charakteristik einer späten und reifen Sprache durch die Geschlossensilbigkeit und die durch die Lautverschiebungen bewirkte Häufigkeit «später» Laute im Sinne von *Jakobson*. In fast allen Punkten stellt das Italienische dazu einen Gegenpol dar. Es kennt im Wortton nicht das Prinzip der Stammsilbenbetonung und im Satzton nur die rhythmische Folge der Wortakzente. Der Mangel einer sprachlogischen Betonung hat eine geringe Schwankungsbreite zwischen den akzentuierten Worten zur Folge. Der Vokalismus und die – abgesehen von den Geminaten – vorherrschende Offensilbigkeit verleihen dem Italienischen Züge eines starken trophotropen Wesensanteiles.

Mag auch das Wesen der Sprache nach *Kainz* in der Darstellung liegen, aus der Darstellungsebene der Sprache allein wird sich eine befriedigende Typologie der Sprachlautsysteme kaum erarbeiten lassen. Dazu bedarf es der hier gemeinten Art der Sprachdiagnostik, die wie die anderen besprochenen Fragenkreise und manche hier nicht besprochenen zum Aufgabenbereich der *Biophonetik* gehören.

Adresse des Autors: Prof. Dr. Felix Trojan, Wassergasse 15, *Wien III* (Österreich).

Discussion

von Raffler Engel (Florenz): Ich möchte an Herrn *Trojans* Frage, «wieweit der stimmliche Ausdruck beim Menschen international gleichartig ist», anknüpfen.

Während der letzten Jahre habe ich mich mit dem «vorsprachlichen Ausdruck beim Kinde» beschäftigt und festgestellt, daß ich von der Art des Weinens Kinder der

verschiedensten Nationalitäten inklusiv von Negerkindern stets sofort sagen konnte, ob sie Hunger hatten oder gewickelt werden wollten usw. In diesem ersten Stadium existiert also zweifellos noch ein gemeinsames allgemein menschliches Element. Vor dem Alter von 5 Monaten scheinen Kinder die Sprache nur als Melodie zu erkennen. Ich habe mit meinem Sohn das Experiment nach *Tappolet* regelmäßig gemacht, und eines Tages plötzlich wollte er den gewohnten Satz in seiner Muttersprache (italienisch) hören. Bis dahin hatte der Satz «jetzt bekommst du sofort deine Milch» ihn in gleich welcher Sprache beruhigt, wenn er hungrig war; danach aber nur noch, wenn ich ihn auf italienisch sagte.

Was die Unmöglichkeit, «die Sprache vom stimmlichen Ausdruck» herzuleiten betrifft, so kann man auch da eben eine weitere Bestätigung des Zuerstserscheinens der rein melodischen Faktoren erblicken, für die ja, physiologisch gesehen, eine einfachere Muskulatur des «Mundraumes» genügt als für die Artikulation der Phoneme.

Höffe (Dortmund): Die Bedeutung des stimmlichen Ausdrucks möchte ich durch zwei weitere Beispiele unterstreichen. *Karl von Holtei*, der als Dramenvorleser zu Beginn des vorigen Jahrhunderts mehrere Länder Europas besuchte, wird nach einem Rezitationsabend von Paganini angesprochen. Begeistert äußert sich der italienische Meister über die Vortragskunst *Holteis;* dabei sagt er ihm etwa folgendes: Ich habe die Dichtung durch ihren Vortrag verstanden, obwohl ich nicht deutsch verstehe.

Und das andere Beispiel: Ungefähr bis zum ersten Lebensjahr gab eines meiner Kinder alle seine Regungen, Empfindungen und Wünsche mit dem Wörtchen «Ja» kund. Allein durch Veränderung des Stimmklanges drückte es das von ihm Gemeinte aus. Der Eindruck war so verblüffend, daß ich diesem akustischen Phänomen in einer wissenschaftlichen Arbeit nachging.

Merlingen (Wien): Eine Bemerkung zu den Beziehungen der *Sprachwissenschaft* und in ihrem Rahmen insonderheit der *Phonologie* zur Biophonetik. Es ist dies meines Erachtens ein ganz bestimmter Punkt, nämlich die *Bindung an den Menschen selbst*, die auch für die Phonologie (und Sprachwissenschaft) bestehen muß. Gemeint ist hier die Beschäftigung mit der *Artikulation* (der organgenetischen Seite), im Gegensatz etwa zur physikalischen Seite. Wir erleben jetzt eine Zeit, in der auch in der Phonologie vorwiegend von Audition und Akustik gesprochen wird (vgl. in letzter Zeit *Pilch*), ja es fehlt nicht an Stimmen, die betonen, daß wir für viele Dinge nur artikulatorische und nicht akustische Bezeichnungen haben. Es gibt, wie Sie wissen, sogar Versuche einer Sprachwissenschaft ohne Laute (und ohne Sprache). Ich meine aber, daß es in der Sprache in erster Linie und zuerst darauf ankommt, *was der Mensch dabei macht, nachher* folgt erst die Audition; und fassen können wir diese Dinge am direktesten – wenn auch noch nicht bis ins allerletzte – an der *Artikulation*. Und im Bereiche der engeren Bindung an den Menschen selbst, wie sie in der von Herrn *Trojan* initiierten Biophonetik gegeben ist, dürfen wir uns meines Erachtens auch für die Phonologie Nutzen erhoffen.

Trost (Prag): Es gibt Beziehungen vom biophonetischen Bereich (mit seinen «anthropologischen Konstanten») zum Bereich der Darstellungsphonologie. Man kann wohl von einer Quasi-Universalität der syntaktischen Intonation (mit Kadenz und Antikadenz) sprechen. Aber nicht alles ist natürlich (lautliche) Kundgabe, was dafür gehalten werden will und naiverweise dafür gehalten wird: vieles davon ist Mache, gehört mit anderen Worten in den Bereich der Appell- (und nicht der Ausdrucks-) Funktion. Ich habe seinerzeit die Beispiele in *Trubetzkoys* «Lautstilistik» durch ein weiteres vermehrt, das mir recht interessant scheint: Der österreichische Aristokrat pflegte mit hoher (und schwacher) Stimme näselnd vorgestellt zu werden. Die hohe Stimme ist eine «phonetische Metapher» («man spricht von oben herab»). Das Näseln ist auch eine «phonetische Metapher», aber ein Artikulationssymbol: die «Lässigkeit» der Artikulation (unvollkommene Hebung des Gaumensegels); der schwache Stimmaufwand ist relevant und «natürliches Zeichen».

Verh. 5. int. Kongr. Phon. Wiss., Münster 1964, pp. 548–550
(S. Karger, Basel/New York 1965).

Zur phonematischen Wertung der deutschen Diphthonge

Von PAVEL TROST, Prag

Während sich *Trubetzkoy* für monophonematische Wertung der Diphthonge in der deutschen Hochsprache erklärt hatte, sind seither einige Forscher für biphonematische Wertung eingetreten. Die biphonematische Wertung wird eingehend von *N. Morciniec* begründet, dabei als erster Bestandteil der Diphthonge /a/ und /ɔ/, als zweiter /i/ und /u/ «in der Stellung nach akzentuiertem Vokal» bestimmt. Ebenso liegt nach *W. Merlingen* wenn nicht im Süddeutschen, so doch im Nord- und Ostdeutschen Zweiphonemigkeit der Diphthonge vor.

Die biphonematische Wertung der deutschen Diphthonge wird aber auch nicht allgemein anerkannt. Kurios ist die «Widerlegung» *O. Zachers*, der ein Spektrogramm des deutschen Diphthongs *ai* beibringt, wo sich deutlich zwei Bestandteile unterscheiden lassen, aber dennoch behauptet, die Spektralanalyse lehre, «daß wir es hier nicht mit einer einfachen(?) Verbindung zweier Laute zu tun haben». Andere halten an der Argumentation *Trubetzkoys* fest und legen noch immer Wert auf die Frage, ob sich die Bestandteile der Diphthonge auf zwei Silben verteilen können oder nicht. Es ist nach *Trubetzkoy* bekanntlich eine Voraussetzung für monophonematische Wertung, daß sie sich niemals auf zwei Silben verteilen.

Tatsächlich kann man dieses Moment geltend machen: Die unsilbischen /i/ und /u/ kommen ausschließlich in den Diphthongen vor; sie bilden mit /a/ und /o/ eine Silbe unabhängig davon, ob ihnen ein Konsonant oder ein Vokal folgt. Darin liegt eine einseitige Bindung des unsilbischen /i/ und /u/ an den vorhergehenden Laut, d. h. den ersten Bestandteil des Diphthongs. Diesem gebundenen Verhalten entgegengesetzt ist das selbständige Verhalten von /i/ und /u/ nach Vokal, wenn sie vor Vokal als j und w erscheinen.

Ein solches Verhalten gibt es in deutschen Mundarten (wobei der vorhergehende silbenschließende Vokal natürlich lang ist), aber eben nicht in der Hochsprache. Das gebundene Verhalten ist ein Argument für monophonematische Wertung der Diphthonge in der deutschen Hochsprache.

Literaturverzeichnis

Merlingen, W.: Über Ein- und Zweiphonemigkeit. Z. Phon. allg. Sprachw. *13:* 98–176 (1960).
Morciniec, N.: Zur phonologischen Wertung der deutschen Affrikaten und Diphthonge. Z. Phon. allg. Sprachw. *11:* 49–66 (1958).
Trubetzkoy, N. S.: Grundzüge der Phonologie. TCLP *VII* (1939).
Vachek, J.: The phonemic status of modern English long vowels and diphthongs. Philologica Pragensia *6:* 59–71 (1963).
Zacher, O.: Deutsche Phonetik. Leningrad 1960.

Adresse des Autors: Prof. Dr. Pavel Trost, Na Micance 20, *Praha 6* (ČSSR).

Discussion

Merlingen (Wien): In einem großen Teil des deutschen Sprachgebiets sind die Diphthonge sicherlich als zweiphonemig zu rechnen. Jedoch glaube ich nicht, daß bei dieser Frage distributionalistische Gesichtspunkte anwendbar sind. Wenn aber beide Bestandteile eines Diphthongs auch in anderen Stellungen stehen können, so sind sie natürlich auch im Diphthong Phoneme für sich, und der ganze Diphthong ist zweiphonemig. Kommt jedoch einer der beiden Bestandteile in anderen Stellungen nicht vor, so ist dieser eine Bestandteil eben kein Phonem für sich, und die ganze Vokalverbindung ist einphonemig. Dies ist nun der Fall wenigstens in einem Teil der deutschen «Hochsprachen», wenn ich so sagen darf. Es gibt meines Erachtens ein paar solcher Hochsprachen, die in verschiedenen Gebieten praktisch existieren – und zwar vor allem in der mir gut bekannten österreichischen Hochsprache (die wenigstens bei den Diphthongen *nicht* von den bayrisch-österreichischen Substratdialekten beeinflußt ist, deren Diphthonge nämlich zweiphonemig sind!). Am deutlichsten zu erkennen bei dem Diphthong, der *eu, äu* geschrieben wird und dessen erster Bestandteil – etwa [å] – außerhalb dieser Verbindung *nicht* vorkommt und auch nicht etwa als Allophon eines /o/ erklärbar ist. Im Gegensatz z.B. zu der sog. «norddeutschen» Hochsprache, wie sie vor allem im Rundfunk zu hören ist, in der offenkundig unser *eu* phonologisch in /o+i/ zerfällt.

Martens (Hamburg): Herr *Trost* wollte die deutschen Diphthonge unter phonematischem und nicht unter phonetischem Gesichtspunkt betrachten. Deshalb möchte ich gegen einige Diskussionsbeiträge grundsätzliche Einwände vorbringen:

Wenn man uns sagt, daß es eine Unmenge phonetischer Realisationen gibt (Herr *Merlingen*), so ist das kein Argument für *Morčiniec* oder für *Trost.* Es ging nicht um phonetische Erscheinungen, sondern um die phonematische Wertung dieser phonetischen Realitäten. Herr *Merlingen* hätte an *einer* Stelle tatsächlich Gründe gehabt, für die monophonematische Wertung einzutreten; als er sagte, daß ein Diphthong sogar monophthongisch gesprochen wird (in einem bestimmten Gebiet). Nun, da wäre zum mindesten ein eindeutiges «phonetisches Argument» gegen eine biphonematische Wertung.

Vielleicht sollte man auch vorsichtig sein in der Verwendung solcher Begriffe wie «süddeutsche Hochsprache», «norddeutsche Hochsprache» und gar «österreichische Hochsprache». Das muß besonders verwirren, wenn bei «österreichisch» das (politisch österreichische) Gebiet von Vorarlberg ausgeklammert werden soll. Wir wollen gern zugeben, daß man von landschaftsgebundenen «Hochsprachen» reden könnte, aber man kann die Landschaft nicht «österreichisch» nennen, ohne den Begriff genauer zu umreißen.

E. Beyer (Straßburg): Deutsch ei, au, äu (eu) sind biphonematisch. Ihre Entsprechungen in den Mundarten sind es ebenfalls. So im Elsässischen und im Schweizerdeutschen. – Die Halbkonsonanten i, u, können sich vollständig konsonifizieren.

Proc. 5th int. Congr. phon. Sci., Münster 1964, pp. 551–555
(S. Karger, Basel/New York 1965).

IBM San Jose Research Laboratory, San Jose, Calif.

Duration as an Alternate Synthesis-Parameter for Intensity and Vowel-Quality

By H. M. Truby

As I indicated at the Thursday morning plenary session, in my discussion of the paper of *Fant*, the specific purpose of a given phonetic evaluation *must* be made clear! It is rarely possible to determine with any validity the *acoustic* correlates of particular *physiological* activities, and as a result the phonetician plays many rôles in present day speech research.

My own long-range purpose as a committed phonetician is to identify as many particulars relating, in whatever way, to whichever aspect of speech, as circumstance permits. Therefore I am interested in anatomic details, in articulatory details, in acoustic details, and in the ways in which these details relate to the production, transmission, and perception of speech.

This afternoon I come to you as a bearer of acoustic tidings from a machine. Please keep in mind the restrictions, limitations, and objectives that such a mission indicates.

The phonetician studies the sounds of speech – but rarely to the exclusion of context. The so-called "isolated vowel" appears in a context of silence and with a specific reference to other linguistic planes. The phonetician who deals with *synthesized* speech, operates with examples whose identification is absolutely dependent upon the effectiveness of the *caricature* nature of the presented information.

Allow me the pleasure of an analogy: The following little sketch is, in a basic way, a caricature

of a tulip...  of a man...  of a woman... 

A synthesized version "of a vowel" is a *caricature*, not only of "the vowel", but as well of a particular phonetic context, semantic context, grammatical context.

I ask your attention to a few such examples of speech caricature: It is not difficult to demonstrate the intelligibility of synthesized-speech-using-familiar-context-alignments, e.g., /[ˌgutn̩ 'tak vi 'getɪs ˌinɛn]?/, /[ˌbɔ̃ 'žuR 'kɔ̃mɔ̃n, ta le vu]?/, /[ˌd*a*s ˌdoɪ 'k*a*po sa v*a*-ɪš]/, /['hɛlo ˌhaʊ '*a*r ju]?/. Also readily demonstrable is the contribution of particular parameters to this intelligibility by a now familiar technique especially adaptable to the IBM Terminal Analog Speech Synthesizer: the first sample comprises only the acoustic information contributed by a monotoned F_0 with controlled timing ((1)); for the next sample F_1 information is added: ((2)); next we hear the sample with F_2 information added: ((3)); then with the addition of F_3 information: ((4)); then with F_0 controlled so as to simulate one of many possible intonation patterns: ((5)).

Now, it is, clearly, optimistic to expect consistently positive identification, by listeners naïve to the voicelike characteristics of a particular speech-synthesizer, of isolated simulated "words", but by the process of elimination and/or with a little practice, identifications do approach consistency. And then, too, there is the helpful – and at the same time frustrating – *perceptual* phenomenon of "knowing what the word *is*", thereby providing cues which serve as well as – or even better than – what might be termed "valid acoustic cues". For example, here are three synthesized words from our inventory which have, under ideal and even less-than-ideal listening conditions, enjoyed consistent identifications:

/[wɪnd]/ × 3 /[did]/ × 3 /[dɪd]/ × 3

It is a commonplace of phonetics that in the process identified by the term "vowel gradation", the reduced vowel *loses* its "original" vowel color and tends toward a "more neutral" vowel. Traditional observations in this regard posit four specific "neutral vowel regions" for unstressed vowels: in the "front-vowel region" the tendency is toward [ɪ], for example, /ri-/ is perceived as [rɪ-] in [rɪ'pit] (*repeat*); in the so-called "back-vowel region" the tendency is toward [ʊ], e.g. /tu-/ is perceived as [tʊ-] in [tʊ'de] (*today*); and the so-called "central-vowel region" is either retroflexed schwa [ə̢] or schwa [ə] depending upon the circumstances, e.g., note how /pə̢-/ is manifested in [pə̢'siv] (*perceive*), and how /ʌ/ becomes schwa in [ə'lon] (*alone*).

Schwa is, of course, an extremely popular linguistic notion, but the phonetician should not be so willing to accept this phonetically-broad generalization. It is certainly not new to phoneticians that, in a manner of speaking, schwa manifests itself in a wide variety of differing phonetic forms, and my own proposed title for this phenomenon is "And a Little Schwa Shall Lead Them".

In any case, certain gradations *seem* obvious, and I single out one of these for this report:

The identity of a reduced vowel is contingent upon the degree of stress of the carrier syllable as well as upon the degree of stress of the contrastive or tonic syllable. This degree of stress – or "unstress" – of the *carrier* syllable influences the phonetic character – the spectral shape in time – of all members *of* the syllable, and similarly for the members of the *tonic* syllable.

Let us take advantage of the fact that certain controls *can* be effected in speech synthesis which can*not* be effected in natural speech due to interrelated, involuntary, compensatory adjustments throughout the phonetic environment as a whole. My objective is to manufacture the following caricatures: Begin with /['kæn'did]/ (*Candide*, a play by G. B. Shaw), and systematically reduce the duration of the /-i-/-vowel until /['kændid]/ (*candied*) is heard, and eventually /['kændɪd]/ (*candid*). (Since it is clear that the minimal acoustic cues provided by speech-synthesis relate only by perceptual similarity to speech, the transcription feature /[--]/ presented in my *Acoustico-Cineradiographic Analysis Considerations*, Acta Radiologica Supplementum 182, Stockholm, 1959, is employed here).

Synthetic "utterances" are contrived with our present system by assembling sequences of what I have termed *diphones*. For example: kæ + æn + nd + di + id. Essentially, *diphone nuclei*, representing transitions, are stored in the computer, and absolutely steady-state formants are extrapolated by the computer from the relevant end-point of the particular nucleus. The systematic exploitation of each of all possible control parameters is then only a matter of computer programming. The evaluation is up to the phonetician: For example,

[kæ–	–æn–	–nd–	–di–	–id]	
180	140	70	50	80	= kændid
				85	= kændi·d
				90	= kændi:d
				95	= kændi:·d
	145				= kæn·di:·d
				90	= kæn·di:d
				85	= kæn·di·d
				80	= kæn·did
	150				= kæn:did
				85	= kæn:di·d
				etc.	
	etc.				with kæn:·
etc.					with kæ·
					with kæ:
					with kæ:·

((etc., etc.))

The figures cite diphone durations in milliseconds.

In conclusion may I offer the observation that, being phoneticians, we welcome the opportunity to "hear for ourselves" exempla ordinarily "listener-tested" and presented as statistics. As Dr. *Cooper* indicated, we can now call on the computer to provide us not only with the acoustic samples for our listening pleasure but with correlated print-outs such as these six-foot sheets of data I enfold before you.

Author's address: Dr. H. M. Truby, Communication Research Institute, *Miami, Florida* (USA).

Discussion

Isačenko (Berlin): Sie haben simulierte Rede als «Karikatur» bezeichnet. Wenn man ein Symbol, welches nur die relevanten Züge eines Objekts darstellt, als Karikatur bezeichnet, so mögen Sie recht haben. Ist aber der Klavierauszug eine «Karikatur» einer Oper oder einer Symphonie? Ist die Wiedergabe der Rede durch ein schlechtes Telephon die «Karikatur» der Rede? Ich glaube, daß die Darstellung der *relevanten* Züge eines Objekts dieses Objekt erschöpfend darstellt.

J. E. Damman (New York): Do you think the results achieved in this case reflect at all the constraints imposed by having a choice of only two words in English: "candied" and "candid"? That is, do you feel the seeming phonetic shift will carry over when a clear cut but restricted choice of this sort is not present?

Answer Truby to Mr. *Isačenko's* reluctance to accept synthesis speech within my definition of *caricature*, I pointed out that by selecting minimal spectral components and emphazing those particular components which insure phonemic identification of a particular phonetic segment, I am, certainly, producing a caricature of some Hochsprache segment.

To *Dammann:* The demonstrated "phonetic shift" will operate whenever the conditions of contrast indicated are operative.

Verh. 5. int. Kongr. Phon. Wiss., Münster 1964, pp. 556–560
(S. Karger, Basel/New York 1965).

Extensional-paradigmatische Bestimmung auditiver Qualitäten phonetischer Signale

Von Gerold Ungeheuer, Bonn

1. Vorbemerkungen

Für die moderne phonetische Forschung ist Voraussetzung nicht nur die genaue Unterscheidung zwischen den akustischen Signalen und den Wahrnehmungsdaten, sondern auch die allgemeinere Erkenntnis, daß der Fluß der phonetischen Information vom Sprecher zum Hörer eine Reihe von gegeneinander abgegrenzten Gebieten durchläuft, die von verschiedenartigen Gesetzmäßigkeiten beherrscht werden und daher die Gesamtheit der sprachlich relevanten, phonetischen Eigenschaften in verschiedener Weise ausprägen. Dies ist von besonderer Bedeutung für die Probleme der Phonemrealisationen. Die spezielle Schwierigkeit der psychologischen Phonetik liegt darin, daß die Qualitäten des Sprachschalls als psychische Gegebenheit mit Methoden festzustellen sind, für die die Psychometrie bisher nur die mathematisch-theoretischen Grundlagen bereitgestellt hat. Experimentelle Ergebnisse fehlen bis auf wenige Einzelheiten. Dies gilt allerdings nicht für psycho-akustische Experimente, deren es bereits eine große Anzahl gibt. Was ins Auge gefaßt wird, ist die Ableitung und Bestimmung der auditiven Qualitäten selbst, nicht die Abhängigkeit der auditiven Gesamtwahrnehmung phonetischer Signale von den akustischen Stimuli.

2. Das extensional-paradigmatische Verfahren

Gegeben sei eine Menge von Wahrnehmungsgrößen $g_1 \ldots g_n$, an denen sich auditive Attribute $A_1 \ldots A_k$ feststellen lassen. Zu jedem Attribut kann diejenige Teilmenge der g_i gebildet werden, die dieses Attribut besitzen. Dies ist die bekannte Zuordnung von Klassen zu Prädikaten, die jede Einführung in die Logik erwähnt. Es ist evident, daß so, wie Attributen Klassen von Elementen zugeordnet werden können, die diese Attribute besitzen, nach gegebenen Klassen umgekehrt auch Attribute definiert werden können. Wenn also ein klassenbildendes Verfahren zur Verfügung stünde, so wäre es möglich, *extensional* jeweils das zugehörige Attribut zu definieren. Auf diese Weise ist dieses Attribut nur durch einige der an ihm teilhabenden Wahrnehmungsgrößen bestimmt, die als *Paradigmata* oder Repräsentanten der gesamten Attributsklasse angesehen werden können.

Klassen von Wahrnehmungsgrößen können aus Ähnlichkeitsrelationen hergeleitet werden. Beispielsweise können Tests so ausgeführt werden, daß man Klassen aller derjenigen Elemente erhält, die für die Vpn. mit hoher Wahrscheinlichkeit alle untereinander ähnlich sind. Diesen Ähnlichkeitsklassen können dann definitorisch auditive Qualitäten zugeordnet werden, die demnach *extensional-paradigmatisch* bestimmt sind. Tests dieser Art sind der Quasianalyse *Carnaps* verwandt. (Zu Veröffentlichungen auf diesem Gebiet siehe Literaturverzeichnis.)

3. Die Tests

Die durchgeführten Tests hatten den Charakter von Erkundungsexperimenten. Insgesamt wurden 5 Tests realisiert:

a) mit vokalischen Lauten, 10 Vpn., Psychologiestudenten;
b) mit konsonantischen Lauten, 11 Vpn., Psychologiestudenten;
c) mit vokalischen Lauten, 10 Vpn., Phonetikstudenten;
d) mit vokalischen Lauten, 20 Vpn., Arbeiter, Handwerker, Angestellte;
e) mit konsonantischen Lauten, 20 Vpn., Arbeiter, Handwerker, Angestellte.

Die Laute waren als isolierte Segmente auf Tonband gesprochen, aus denen dann Testbänder hergestellt wurden, die alle möglichen Lautpaare enthielten. Die benutzten Laute sind aus den beigefügten Abbildungen ersichtlich.

Zwei Testversionen wurden ausprobiert: 1. die Vpn. sollten über die abgehörten Lautpaare entscheiden, ob sie sie als ähnlich oder unähnlich beschreiben würden; 2. die Vpn. hatten die Möglichkeit, nach einer vorgegebenen Ähnlichkeitsskala Grade von Ähnlichkeit bzw. Unähnlichkeit anzugeben.

4. Die Testergebnisse

Die Ähnlichkeitsklassen, die sich aus den Tests ableiten ließen, sind in den folgenden Abbildungen wiedergegeben. Test a) und b) sowie d) und e) ergaben eine sehr große Übereinstimmung, so daß das Ergebnis beider Serien zusammengefaßt werden konnte. Die Phonetikstudenten in Test c) waren nicht in der Lage, auf die Testanforderungen einzugehen. Sie nahmen nahezu alle Lautpaare als unähnlich wahr.

Zur Auswertung der Tests wurden statistische Hilfsmittel herangezogen. Die Testvoraussetzungen müssen diskutiert werden. Das Testverfahren selbst kann sicherlich verbessert werden. Die Tests, über die berichtet wurde, sollten lediglich dazu dienen, erste Erfahrungen auf diesem Felde der Psychophonetik zu sammeln.

5. Graphische Darstellungen der Testergebnisse

Jede der in den Abbildungen eingezeichneten Klassen von untereinander ähnlichen Lauten bestimmt eine auditive Qualität.

a) Vokalische Laute:

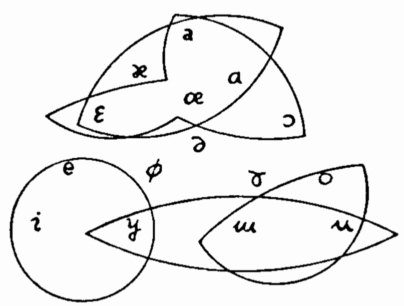


Abb. 1. Polare Vokalqualitäten.

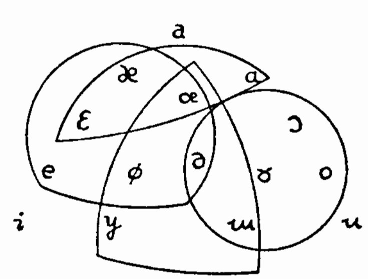


Abb. 2. Periphere Vokalqualitäten.

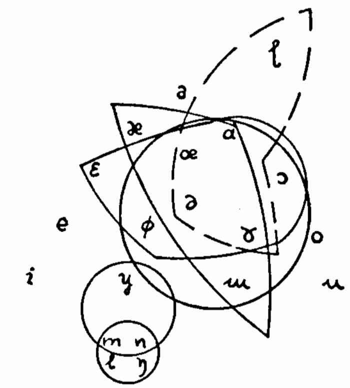


Abb. 3. Zentrale Vokalqualitäten.

b) Konsonantische Laute:

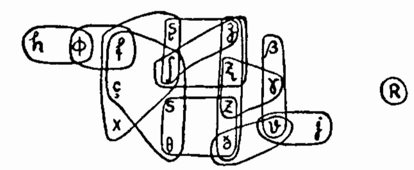


Abb. 4. Konsonantische Qualitäten.

Literatur

Carnap, R.: Der logische Aufbau der Welt, 2. Aufl., Hamburg 1961.
Hanley, C. M.: Factorial analysis of speech perception, JSHD *21:* 76–87 (1956).
Hanson, G.: Phoneme perception, a factorial investigation. Acta Univ. Uppsaliensis *11:* 109–147 (1960).
Osgood, Ch. E. and *Sebeok, T. A.* (ed.): Psycholinguistics; J. abnorm. soc. psychol. Suppl. *49* (1954).
Ungeheuer, G.: Phonetische Aspekte beim Sprachverstehen. Habilitationsschrift. Universität Bonn (1962).

Adresse des Autors: Dr. G. Ungeheuer, Institut für Phonetik und Kommunikationsforschung der Universität, Koblenzer Straße 98a, *Bonn* (Deutschland).

Discussion

Eli Fischer-Jørgensen (Virum): Zu den interessanten Ausführungen von Herrn *Ungeheuer* möchte ich hinzufügen, daß ich einige ähnliche Versuche nach verschiedenen Methoden durchgeführt habe. Die Methode, die mir am meisten zusagt, mit der ich aber nur einige vorläufige Versuche gemacht habe, ist die Vorführung von Doppelpaaren, z. B. i–e, i–a, wobei die Versuchspersonen entscheiden sollen, ob im ersten oder im zweiten Paar die Ähnlichkeit größer ist. Diese Methode hat den Vorteil, daß die Versuchspersonen nur die jedesmal dargebotenen Paare zu beobachten brauchten, ohne sich um die übrigen Laute der Versuchsreihe zu kümmern, und daß die Entscheidung in vielen Fällen sehr leicht ist. Dafür ist aber die statistische Bearbeitung relativ kompliziert.

Was die Interpretation der Ergebnisse betrifft, so möchte ich darauf aufmerksam machen, daß die Antworten von vielen Faktoren beeinflußt werden können: durch den physikalischen Stimulus, die motorische Reaktion, das phonologische System der Muttersprache, die Schrift, die Buchstabennamen usw. Die Tatsache, daß Herr *Ungeheuer* eine Ähnlichkeit zwischen u und y gefunden hat, die für dänische Versuchspersonen nicht vorzuliegen scheint, könnte z. B. dadurch erklärt werden, 1. daß es im Deutschen eine recht häufige grammatische Alternation u–ü gibt, 2. daß die Buchstaben ähnlich sind (deutsch u–ü, dänisch u–y), 3. daß ü im deutschen Schulunterricht oft als u-Umlaut bezeichnet wird.

Proc. 5th int. Congr. phon. Sci., Münster 1964, pp. 561–564
(S. Karger, Basel/New York 1965).

On Peripheral Phonemes

By Josef Vachek, Praha

It is often erroneously supposed that language is a closed system, i.e. that all its elements are rooted in it with equal firmness. The actual situation is, of course, different: any language level contains elements of transitory status which cannot be well classed with the standard clear-cut categories prevailing on that level. This is also true of the phonic level of language where, apart from the firmly rooted elements of the pattern, one can also find elements having peripheral status. Such elements are phonemes which are either not "fully integrated" in the phonemic pattern (to use *A. Martinet's* term) or exhibit a very low functional yield. A couple of instances of such peripheral phonemes, taken from Modern English, will show that a more detailed analysis of such cases may throw some light on a number of tendencies of development which otherwise do not stand out with particular prominence.

The first of such instances is that of the ModE phoneme /h/. While the Old English /h/, with its allophone [χ], could occur in a relatively large number of positions in the word, its ModE descendant /h/ is a peripheral element of this language. It virtually occurs in one single position only, viz. at the beginning of a stem-morpheme before a following vowel (or, semi-vowel). In the course of the development of English the phoneme *h/χ* was gradually ousted from all other positions in the word. The reason for this process was, in part, its isolation in the pattern of English consonant phonemes. This isolation was brought about by the loosening of the tie originally binding the *h/χ*-phoneme to the phoneme *g/ʒ*, especially after the loss of the ME phoneme /ʒ/, which had become independent during the OE period. After the phoneme *h/χ* had thus been confined to a relatively small number of word-positions, a quantitative handicap was added to the already existing qualitative one, i.e. the slight functional yield of the phoneme /h/. As is well known, in popular dialects, uncharged with cultural and civilizational tasks, so copiously imposed upon the standard language, [*h*] has been, as a rule, totally discarded as a phoneme. Thus, in the Cockney of London it is only found as a phonostylistic feature, acting, at the same time, as a signal of the beginning of words (cf. the notorious droppings and misplacements of *h*'s).

In the course of the process discarding the phoneme /h/ also several other phonemes came to be dismissed from the phonematic pattern of English, the reason being again the small functional yield of such phonemes. The concerned phonemes were the EME voiceless liquids /R, L, N/ and – somewhat later – /W/, at least in some parts of the English speaking teritorry.

The process leading to the disappearance of /h/ was motivated, in its earliest stages, by the regularities governing the activities of the speech organs (see changes like *fōhan* > *fōn*, *hræfen* > *Raven*, *hlūd* > *Lūd*, etc.). Later changes were motivated purely linguistically, i.e. by the tendency to abolish a phoneme which had become non-lucrative on account of its very slight functional yield. It should be stressed that the whole discarding process has sprung out from purely domestic roots. Some accelerating influence of French on the process, however, cannot be flatly dismissed (though it is very difficult to prove it). It is certainly remarkable that the distribution of the sounds *h* and χ in 13th century French (both positive and negative) was identical with the one that was to become characteristic of English in the centuries to come.

The other ModE phoneme to recede in the course of the development of English is /r/. Compared with its Old English ancestor, the ModE /r/ has had to give up a large number of its original positions of occurrence (cf. e.g., OE *heorte*, *wiersa*, *feor*, *þǣr*, *steorra*, *hræfen*, *wrītan* with ModE [hɑ:t, wə:s, fɑ:, ðɛə, stɑ:, reivn, rait]. Although a relatively large number of the positions of occurrence has been left to ModE /r/, some phonetic phenomena clearly reveal that the position of that phoneme in the Southern British standard of ModE has been appreciably shaken.

First, there is the rise of the "intrusive r" (as in *India-r office*), and, conversely, the abolishment of the "linking r" even in those places where it is phonetically legitimate (e.g. [fɑ: əwei]), found especially with the younger generations. Both these tendencies are probably called forth by the functional ambiguity of ModE /r/

which serves both as a means of distinctive function (as in *red*, *crab*) and as a means of purely delimitative function (as in *India-r office*). It appears that both categories of speakers tend to make the status of ModE /r/ less ambiguous: those who abolish the linking *r* emphasize the distinctive function of /r/, while those who introduce the intrusive *r* rather tend to stress its delimitative function.

The latter tendency is underlined by some other trends attacking the existence of the distinctive /r/. One of them tends to merge the initial clusters *tr–*, *dr–* with the affricates [tʃ, dʒ], so that, e.g., in childrens speech, words like *true* and *chew* sound almost as homophones. Should this trend assert itself on a wider scale (which, for the present, it does not), it might dispose of a fairly large number of the instances of distinctive /r/. – The other trend consists in a strong labialization of /r/, noted by the British phoneticians since the thirties, but reflected in literary spellings like *wough*, *dwiver*, etc. already in the 18th century. This trend, merging /r/ with /w/, is apt to dispose, at a single blow, of the distinctive function of /r/, relegating it to the status of a mere delimitative signal.

Here again, the discarding tendency is motivated both quantitatively (by the small functional yield) and qualitatively. One of the qualitative motives was the above noted functional ambiguity of /r/, serving both distinctive and delimitative functions. Besides, and this is even more important, /r/, too, is structurally isolated in the phonematic pattern of ModE consonants. This isolation became a fact when the trilled articulation of [r] had been replaced by a fricative one, and thus the tie linking the /r/ to the other liquid /l/ had been loosened (for the link tying up the two liquids see *R. Jakobson*, Proceedings, Ghent 1939).

The trends attempting to do away with the ambiguous status of ModE /r/ have not obtained universal acceptance, obviously because cultural and civilizational factors (especially the orthoepic norm) are opposed to such radical solutions. Thus the ModE /r/ remains one of the peripheral elements of its phonematic pattern.

Author's address: Prof. J. Vachek, Nám. Jiriho z Podebrad 18, *Praha 3* (ČSSR).

Discussion

Pilch (Freiburg i. Br.): Meiner Meinung nach sind altenglisch [χ] und [γ] Varianten des gleichen Phonems, und das altenglische *r* war, soweit ich weiß, retroflex.

Pike (Ann Arbor): Mr. *Váchek* – How would the transformationalist treat the problem you mentioned – and what would be your reaction to their view?

Answer *Váchek:* Answering Mr. *Pike's* questions, I must say that the transformationalists have so far – despite all their merits – shown regrettable lack of attention to the dynamic character of language (tensions within its system, including historical development). – Methodological difficulties in the establishment and differentiation of the centre and periphery of the language do exist but they must be tackled. Language being a system in which all levels are interdependent, linguistic methodology must respect this fact. First of all, one should proceed by method of trial and error.

C.R. 5e Congr. int. Sci. phon., Münster 1964, pp. 565–568
(S. Karger, Basel/New York 1965).

Réalisation du phonème dans la voix œsophagienne

Par B. Vallancien et Cl. Dinville, Paris

Les observations concernant l'émission de la voix œsophagienne dans la parole doivent êtres faites en référence avec l'étude du comportement des organes phonateurs en voix normale.

Conditions de la voix œsophagienne

C'est une voix sans larynx qui présente donc anatomiquement, physiologiquement et acoustiquement, des particularités.

Anatomiquement, le courant d'air pulmonaire qui s'échappe de la trachée, est dirigé vers l'air extérieur par un orifice pratiqué à la partie antéro-inférieure du cou. Le larynx ayant été enlevé dans sa totalité, l'hypo-pharynx se trouve prolongé par une cavité qui représente la place du larynx enlevé. Cette cavité se termine en bas au niveau de la bouche œsophagienne par un repli musculaire pouvant former sphincter, qui est le faisceau inférieur du muscle constricteur inférieur du pharynx ou muscle crico-pharyngien en regard de la sixième cervicale.

Physiologiquement, l'organe de la phonation ayant été supprimé et la respiration se trouvant directement abouchée à l'extérieur, il n'existe plus à l'endroit du carrefour aérodigestif que l'embouchure d'une voie uniquement digestive.

C'est à l'aide des replis muqueux de la sangle musculaire qui forment les parois de cette nouvelle cavité, que le sujet pourra tenter une voix de remplacement. Pour cela il devra trouver une provision d'air nécessaire pour mettre en mouvement les différentes structures musculo-muqueuses. Une petite quantité d'air doit donc être préalablement ingérée, puis comprimée et refoulée pour permettre l'écoulement qui fait entrer en vibration cette néo-glotte. Mais cette manœuvre sera d'autant plus aisée que les parois seront plus souples et que la mobilité de la musculature supérieure aura été conservée. C'est pourquoi les cicatrisations vicieuses et l'induration des tissus par la radiothérapie rendront plus difficile l'acquisition de la voix œsophagienne.

Techniques de l'approvisionnement d'air

Il s'agit donc pour le sujet d'emmagasiner une provision d'air qui, lors de son expulsion, servira de fourniture aux sons œsophagiens. Trois mécanismes sont possibles:

la déglutition
l'injection
la succion

La déglutition reste la méthode de choix, plus facile à enseigner. Elle donne l'accès à une voix œsophagienne plus timbrée puisque l'approvisionnement d'air se fait plus bas et dégage ainsi les muscles de l'articulation, contribuant à une meilleure intelligibilité. La technique est la suivante: le sujet déglutit un petit peu d'air, aidé au besoin dans les débuts par l'ingestion d'une gorgée d'eau gazeuse, et au moment où la poche d'air est suffisante, à l'aide d'une contraction de la sangle abdominale et d'une décontraction des muscles cervicaux, l'air est brusquement expulsé et servira à la fourniture du son.

L'injection se réalise aussi en inspiration, mais à l'aide des phonèmes explosifs tels que pa, ta, ka. L'air se trouve comprimé en arrière de la partie inférieure du pharynx au moment de l'implosion, puis il est brusquement rejeté et la voyelle qui suit peut ainsi être formée grâce à l'explosion. Par contre, les sifflantes se, fe, che, auront plus de difficulté à être émises car elles nécessitent une provision d'air plus grande.

Dans la succion, c'est la traction de la musculature supra laryngée qui conditionne l'ouverture de la bouche œsophagienne. Le thorax demeure en inspiration forcée. La brusque interruption de cette contraction, aidée d'une légère poussée abdominale, provoque une éructation qui pourra être modulée par les voies respiratoires supérieures.

Mécanisme de l'expulsion d'air

Une technique d'enregistrement radiocinématographique télévisée nous a permis d'étudier ce mécanisme.

Deux sondes jumelées de petit diamètre, dont les orifices sont espacés de 5 cm, sont introduites dans l'œsophage à travers les fosses nasales et reliées à deux manographes électroniques. Ceux-ci enregistrent les variations de pression qui sont d'autre part objectivées sur une échelle le long de laquelle deux spots de galvanomètre se déplacent. L'image de cette échelle, placée sur le bord de l'écran de télévision sur lequel se projette la vue radioscopique de l'œsophage, est enregistrée par une caméra de cinéma synchronisée. On peut ainsi constater que l'effort moteur déterminant l'expulsion d'air est produit par la paroi abdominale qui refoule le diaphragme.

Phonétique

D'un point de vue phonétique, les consonnes explosives faciliteront les sonorités des voyelles qui suivent. Les nasales ne présenteront de difficulté que dans la mesure ou le voile du palais reste abaissé pendant la phase de compression, ce qui crée une fuite d'air par les fosses nasales. Les liquides et vibrantes seront plus difficiles à exécuter à cause de la fuite d'air qu'elles exigent, ainsi que les sifflantes. Les voyelles seront utilisées pour donner plus d'intonation à la voix et pour dégager l'articulation de l'appui sonore qu'elles engendrent.

Etude acoustique

Les enregistrements qui ont été faits à l'aide du sonagraphe par *J. C. Lafon*, montrent une superposition remarquable des tracés des différents phonèmes dans la succession des ouvertures aussi bien laryngées qu'œsophagiennes. On note en outre une netteté consonnantique beaucoup plus accentuée, dans la voix sans larynx.

Les enregistrements de pression au niveau du pharynx et de l'œsophage comparés à la pression sonore extérieure permettent d'étudier le rendement de l'acquisition de la voix œsophagienne et de montrer l'intérêt d'impulsions brusques par décompression de l'air emmagasiné (*Damsté*).

Articulation

Ceci prouve l'intérêt d'une articulation d'autant plus parfaite que la fourniture du son est plus rudimentaire et l'écoulement d'air forcément limité.

Dans un certain nombre de cas malheureusement, l'articulation est entravée soit par la détérioration de la denture à laquelle il faudra palier par une prothèse correcte, soit parce que le système nerveux moteur des muscles de la langue et du pharynx a été plus ou moins lésé dans l'intervention ou que l'irradiation a produit secondairement des rétractions fibreuses de ces muscles. La mobilité de la langue pourra se trouver ainsi entravée ou déviée dans certains de ses mouvements.

Rythme et mélodie

Dès que le sujet sera en mesure d'émettre des sons suffisamment longs pour pouvoir prononcer des mots, il faudra lui apprendre à couper la phrase au bon endroit, à rétablir la mélodie et le rythme du discours, faute de quoi l'intelligibilité resterait très imparfaite.

Conclusion

Dans la voix œsophagienne, la réalisation du phonème doit s'apparenter d'aussi près que possible à celle de la voix normale. Les insuffisances de pression du véhicule aérien doivent être compensées par une meilleure articulation qui créera un rendement supérieur pour l'intelligibilité des phonèmes.

Un film est projeté qui montrera les variations de pression œsophagienne pendant l'expulsion à différents niveaux.

Adresse de l'auteur: Dr Bernard Vallancien, Faculté de Médecine, 16, Rue Spontini, *Paris XVIe* (France).

Proc. 5th int. Congr. phon. Sci., Münster 1964, pp. 569–573
(S. Karger, Basel/New York 1965).

Communication Sciences Laboratoy, University of Minnesota, Minneapolis

Laryngeal Analog Synthesis of Glottograms

By Ronald W. Wendahl, Minneapolis, Minn.

The subject of my paper is the use of an electrical analog of the larynx to study vocal productions.

Before we go into the operation of the analog, let me state some of the simplifying assumptions that must be made and the limitations created by these simplifying assumptions. First, one may look at a typical area analysis of the operation of the vocal folds. In figure 1 openings and closings of the folds are depicted on the ordinate, as a function of time on the abscissa. If one were interested only in areas of opening as a function of time he could plot in a single dimension as a function of time and he would have the configuration shown in figure 1-B. It would not be a great violation to translate directly to the waveform shown in figure 1-C if one were interested in preserving frequency information and were willing to accept the differences one might obtain in the harmonic spectrum.

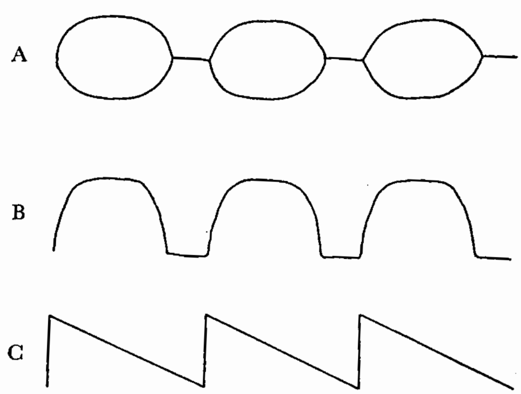


Fig. 1

In figure 2 one may see that frequency information may be preserved in several ways. Figure 2-A shows the fold area openings as a function of time complete with the closed as well as the open phases of the vocal fold actions. If one desired to preserve only frequency information he might use a waveform as is shown in figure 2-B. However, there is good reason to believe that one might have to consider the closed phase in attempting to synthesize human vocal productions and therefore he might be forced to use a waveform as is shown in figure 2-C.

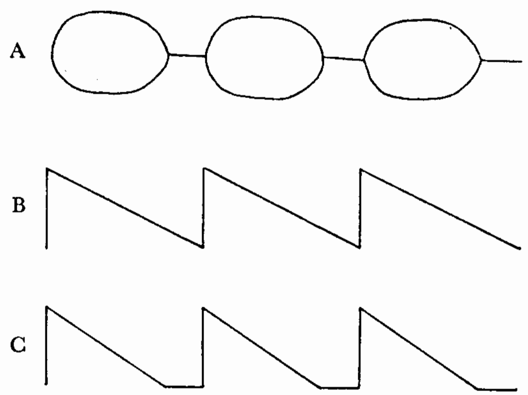


Fig. 2

Figure 3 shows a not atypical mode of vibration of the vocal folds. The second laryngeal opening in each case, is of lesser amplitude and shorter in time than the first. One may wish to be able to program amplitude as well as frequency information as in the case of figure 3-A, or he may wish to ignore amplitude information unless he suspects that the amplitude differences between adjacent wave fronts are sufficient to create a pitch change. In that case he may wish to simulate figure 3-A with figure 3-C.

In the tape recordings you will hear, all simplifying assumptions were made. The closed phase of the folds was ignored and amplitude differences between successive wave fronts also were ignored. The reason for this procedure was purely that this recording was considered to be a first approximation to the synthesis of harsh voice quality.

Before playing the tape recordings made from LADIC, may I go briefly into the operations of the device. Figure 4 shows a simplified block diagram of LADIC. The output waveform is gener-

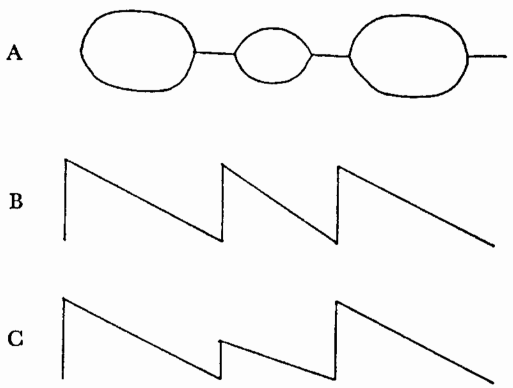


Fig. 3

ated by a master oscillator which is controlled in frequency and amplitude by a small special-purpose computer. The master oscillator is an unijunction transistor in a modified 'relaxation' circuit. This arrangement produces two waveforms; one, a slightly nonlinear sawtooth, and the other a narrow clock (synchronizing) pulse. The memory unit is binary which controls the sequence of the frequency changes. Switching from one frequency to another, or from one amplitude to another is accomplished through the use of saturated transistor gates that operate with a speed of approximately

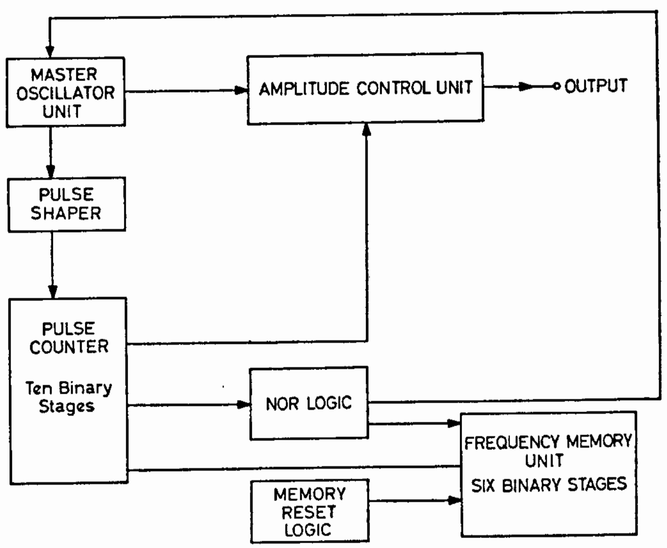


Fig. 4

500 nanoseconds. The pulse counter is also binary with the desired number of cycles of each frequency being programmed by means of toggle switches. The present counter capacity for a given frequency is 2,027 cycles. When a programmed count has been reached, NOR logic circuits cause any or all of the following events to occur in accordance with the present program.

a) The memory may be advanced to the next instruction. This causes the next sequence to begin.

b) The memory decode logic may shift the master oscillator frequency to any desired value between 50 and 1,000 cps.

c) The output amplitude may be switched to any predetermined value from zero to maximum.

d) The pulse counter may be set to zero count.

e) The count decode switches are scanned and the proper count for a new frequency is automatically selected.

The preceding sequence may be repeated, without transients, for a maximum of 63 distinct frequencies. Depending upon the purpose of the experiment, the output of LADIC either is fed directly into a tape recorder for data storage or is fed through vowel shaping filters. When it is deemed necessary for the stimuli to sound humanoid, vowel shaping circuits are used.

The procedures used to obtain the source data for programming LADIC are frequency measures obtained from phonelegrams, from ultra-high speed cinematography, and from transillumination. Each of the techniques offers advantages and has limitations. The phonelegraphic technique has the limitation of the transposition of times from points often obscure and the limited accuracy of the dividers and ruler used. A further drawback of the phonelegraphic technique of obtaining frequency information is that the open-phase time of the vocal fold vibration is indistinguishable from the closed time.

The camera technique offers the advantage of being able to separate the open from the closed phases of the vibratory cycle; however, it has the disadvantage that it is a discrete and not a continuous operation even at such speeds as 7,400 pictures per second. Hence, an error around 135 cps of plus or minus 3 cps is not unexpected.

The transillumination technique offers the advantages of being able to distinguish the open and closed phases of vibratory patterns and that it is a continuous reproduction but offers the ambiguity of divider measures.

Several subjects were chosen having deviant voice qualities. Cinematographic, photocell, and light-writing oscillographic measures were made of their phonations. From the various measures LADIC was programmed and fed to vowel filters to simulate the original phonations.

Tape recordings will present the human phonations and their synthesized counterparts. In many of the tape recordings a remarkably similar quality between the two types of stimuli can be heard. In some recordings the reverse is true. The rationale for both the successes and the failures will be described.

Author's address: Dr. Ronald W. Wendahl, Communication Sciences Laboratory, University of Minnesota *Minneapolis, Minn.* (USA).

Verh. 5. int. Kongr. Phon. Wiss., Münster 1964, pp. 574–577
(S. Karger, Basel/New York 1965).

Die Stellung der palatalisierten Velarvokale und Diphthonge in den Vokalsystemen der deutschen Mundarten

Von Peter Wiesinger, Marburg a.d. Lahn

In den deutschen Mundarten sind palatalisierte Velarvokale* und Diphthonge mehr verbreitet, als man bisher allgemein angenommen hat. Von dieser Erscheinung werden die Monophthonge U und O jeden Öffnungsgrades und jeder Quantität, die steigenden Diphthonge OU, AU und die fallenden UE, OE betroffen. Phonetisch lassen sich, grob gesprochen, 3 Typen unterscheiden. 1. Schwache Palatovelarität, bei der die Stelle der Zungenartikulation gegen den mittleren Gaumen nach vorne verlagert und die Lippenrundung aufgegeben wird. Derartige Laute klingen schwach an y und ø an, werden aber von den Sprechern noch als u und o empfunden (charakteristisch für das Obersächsische): *u*, *o*. 2. Starke Palatovelarität, bei der nebst einer intensivierten zentralen Zungenstellung häufig auch die Lippen mit leichter Rundung mitbeteiligt sind. Die Klangfarbe nähert sich stark dem y und ø, als welche die Sprecher sie auch auffassen (charakteristisch für Tiroler Hochalpentäler): ü, ö. 3. Volle Palatalität mit normaler Lippenrundung wie die hochsprachlichen Umlaute als stärkste Ausprägung: y, ø. Entsprechend verhalten sich die Diphthonge, wobei die A-Komponente des Diphthonges AU entweder zur palatalen oder velaren Seite neigt und das A nur die 1. Stufe als helles a oder dunkles *ɑ* erreicht: *ou*, *ɑu*, *ao* – öü – øy; *u*ə – üə – yə; *o*ə – öə – øə. Von den diphthongischen Grundtypen leiten sich nach einer als Lautmorphologie zu bezeichnenden, strengen, unter der Einwirkung von

* Das API-Transkriptionssystem enthält keine für dieses Thema erforderlichen Transkriptionszeichen. Wir benützen daher zur Wiedergabe der «schwachen Palatovelarität» die kursivgesetzten Zeichen *i*, *e*, *u*, *o*, *ɑ*, *a* und zur Wiedergabe der «starken Palatovelarität» den Normalsatz mit Trema ï, ë, ü, ö. Die im Text gegebene kurze Charakteristik sichert das Verstehen der einzelnen Laute.

Physiologie und Akzent stehenden Gesetzmäßigkeit folgende Diphthonge ab: 1. Von *ou*, öü bei physiologischen Veränderungen: a) velarisierte *eu*, *eo*, ëü, ëï. b) Bei Verlust der Palatovelarität durch Entvelarisierung (Physiologie): eo, ei (mit Öffnung der 1. Komponente ai). Weiterentwicklungsmöglichkeiten von eu, eo: Schließung zu iu, io. c) Bei Verlust der Palatovelarität durch Entpalatalisierung (Physiologie): oi. 2. Von *u*ə, üə: a) durch Physiologie + Akzent: ü*i*, yi; b) dessen Entpalatalisierung durch Physiologie: ui. 3. Dieselben Möglichkeiten gelten für *o*ə, öə: a) ö*i*, øi. b) oi. Eine besondere Erscheinung ist die Palatovelardiphthongierung von ü: zu üï und *o:* zu *ou*. Alle diese Lauttypen lassen sich in ihrem Wesen und Gestaltwandel mit zahllosen Varianten und Übergängen in den deutschen Mundarten belegen.

Der Phonologe stellt bei jeweils aus zweigliedrigen Reihen bestehenden Vokal- und Diphthongsystemen, bei denen der Umlaut entrundet wurde und mit dem gespreizten Palatalvokal zusammenfiel, auf Grund der korrespondierenden Bildungsweise mit Recht einem gespreizten Palatalvokal den entsprechenden gerundeten Velarvokal gegenüber. Er erhält dadurch gekoppelte Monophthongreihen i – u, ɪ – ʊ, e – o, ɛ – ɔ und gekoppelte Diphthongreihen ei – ou, ɛɪ – ɔʊ (aɪ – aʊ), iə – uə, ɪə – ʊə, eə – oə, ɛə – ɔə. Im Monophthongsystem nimmt der A-Laut seiner Bildung nach eine isolierte Stellung ein. Das Wesen der gekoppelten Reihe ist durch die korrespondierende Bildungsweise gekennzeichnet. Wirkt auf beide Glieder einer gekoppelten Reihe dieselbe veränderte Kraft ein, so vollzieht sich bei beiden ein entsprechender Wandel, der Reihenschritt genannt wird, z. B. e – o zu i – u (Hebung) oder iə – uə zu i: – u: (Monophthongierung). Die in den Vokalsystemen der deutschen Mundarten enthaltenen Reihen gehen meistens schon auf im Mittelhoch- bzw. Mittelniederdeutschen vorhandene zurück, sind aber zum Teil auch erst auf andere Weise in späterer Zeit zustande gekommen.

Die Koppelung der Laute wird durch die Einwirkung der Palatalisierung aufgehoben. Sie ist eine spezifische physiologische Veränderung, die nur das velare Glied einer Reihe betrifft und die phonetischen Bedingungen der Koppelung, nämlich die korrespondierende Bildungsweise, zerstört. Solange sich an dem palatalisierten Monophthong oder Diphthong kein Gestaltwandel vollzieht und der Grundtyp unverändert erhalten bleibt, kann auch das Reihenverhältnis, jedoch als ein gestörtes, bewahrt bleiben, z. B. hochalemannisch-urnerisch: mhd. î – û = i: – ü:; mhd. ei – ou = ɛi – ö*ü. Das ist bei der Darstellung des Vokalsystems so zu kennzeichnen, daß der palatalisierte Velarvokal mit dem entsprechenden Palatalvokal horizontal zwar in derselben Höhe stehen bleibt, aber was die Vertikale angeht, gegen die Mitte (= nach vorne zu) eingerückt wird. Eine gestörte Reihe besteht jedoch synchron nur dann, wenn nicht ein nichtpalatalisierter normaler Velarvokal oder Diphthong an seine Stelle tritt und so den palatalisierten in die Isolierung drängt, was diachron eine Verschiebung in der ursprünglichen Zusammengehörigkeit von Lauten verursacht, z. B. nordhessisch: mhd. ê – ô, mhd. ei – ou = heute: Reihe: e: – o: = ê + ei – ô; isoliert: ö: = ou. Südbairisch-ötztalerisch: mhd. ê – ô, mhd. ei – ou – öü = neue Reihen: gekoppelt: eə – oə = ê – ei, gestört: e: – *o:* = öü – ou; isoliert: *oə* = ô. Durch die Aufhebung der Koppelung kann der palatalisierte velare Monophthong anderen Weiterentwicklungen als der in der gestörten Reihe mit ihm noch gebundene Palatalvokal unterliegen. Verbreitet ist die Möglichkeit der Palatovelardiphthongierung, die eine Reihe auflöst und, wenn nicht ein korrespondierender normaler Velarvokal anderer Provenienz mit dem Palatalvokal eine neue Reihenbindung eingeht, beide Laute in die Isolierung drängt, z. B. westmittelbairisch: mhd. ê – ô, mhd. â: neue Reihe: ɛ: – ɔ: = ê – â, isoliert: *ou* = ô. Alemannische Walserkolonie Bosco-Gurin: mhd. î – û: isoliert: i:, yi. Bei gestörten Diphthongreihen verursacht ferner der Gestaltwandel des Palatovelardiphthonges (s. oben) auf jeden Fall Reihenauflösung und drängt ihn in die Isolierung. Wenn nicht der palatale Diphthong mit einem korrespondierenden normalen Velardiphthong eine neue Reihenbindung eingeht, bleibt auch er isoliert, z. B. mittelbairisch-niederösterreichisch: mhd. ie – uo, mhd. ir – ur: neue Reihe: iə – uə = ie + ir – ur, isoliert: ui = uo. Schlesisch-neiderländisch um Fraustadt: mhd. î – û: isoliert: e:, ɪo:.

Noch hinzukommende Velarisierung palataler Diphthonge und Entpalatalisierung gerundeter Umlautsdiphthonge – wesentlich seltenere Erscheinungen als die Palatalisierung velarer Diphthonge – kann beim Eintritt spezifischer Weiterentwicklungen, die jenen der palatalisierten Velardiphthonge verwandt sind, zu einer Fülle isolierter, außerhalb jeglicher Reihenbindung stehender Diphthonge führen, z. B. westfälisch-ravensbergisch: mnd. î – û – iu

* ö gibt hier stark palatovelares offenes ɔ wieder.

= ü:ï, i:u, u:i; mnd. ê⁴ – ô¹ – öe¹ = ö:ï, e:u, o:i; mnd. ê² – ô² – oe² = a:ɪ, ɛ:ɔ, ɔ:ɪ, also 9 isolierte Diphthongphoneme aus ursprünglich 3 gekoppelten Reihen.

Die Erkenntnis des Bestehens der Lautmorphologie vermag die palatalisierten Velarvokale und Diphthonge in einen organischen Entwicklungszusammenhang zu stellen. In Verbindung mit dem phonetisch-phonologischen Reihen- und Reihenschrittgesetz ergibt sich das Vorhandensein von gekoppelten und gestörten Reihen sowie isolierter Einzellaute, die durch Reihenauflösung hervorgegangen sind. Das ermöglicht die Aufstellung geordneter Monophthong- und Diphthongsysteme auch dann, wenn normale und palatalisierte Velarvokale und Diphthonge in einer Mundart nebeneinander vorkommen.

Adresse des Autors: Peter Wiesinger, Univ. Ass., Forschungsinstitut für deutsche Sprache, Deutscher Sprachatlas, Kaffweg 3, *355 Marburg a. d. Lahn* (Deutschland).

Discussion

Fourquet (Paris): M. *Beyer* a étudié la palatalisation de *u*, *uo*, *ou* comme fait articulatoire atteignant des phonèmes isolés. – M. *Wiesinger* a le mérite de poser le problème des couples de phonèmes (Reihen) tels que i–u, e–o; il este un dernier pas à faire, celui d'une étude de *tout* le système vocalique. La désorganisation d'un couple peut s'expliquer par l'intégration d'un des membres dans une autre corrélation.

Il semble bien que le déplacement vers l'avant des articulations de type u soit lié à une raison physiologique: plus la langue s'élève, plus il devient difficile de la porter vers l'arrière. – La palatalisation ne touche en alsacien que la vélaire de fermeture schême (même pas u bref, plus ouvert). Il se pourrait que la palatalisation ou semi-palatalisation de voyelles de type o, ɔ soit un fait secondaire de rétablissement de l'homogénéité du système des oppositions.

Beyer (Strasbourg): Die deutsche und österreichische Mundartforschung verwendet für die nach vorn gerückten Vokale den Terminus «palatovelare Vokale». Dieser Ausdruck scheint mir nicht geeignet zu sein. Es gibt in der Phonetik «labiovelare» oder «labiopalatale» Halbkonsonanten; zwei Organe sind dabei an der Artikulation beteiligt. Dies ist nicht der Fall bei den sogenannten «palatovelaren» Lautungen. Ich würde andere und zutreffendere Benennungen vorschlagen: Je nach der Stellung des Artikulationspunktes: «zentrale», «postpalatale», «palatale» Vokale.

Herr *Wiesinger* hat andererseits zu verstehen gegeben, daß in den deutschen palatalisierenden Mundarten alle Vokale der hinteren Reihe von dem Phänomen erfaßt worden seien. Dies trifft sicher nicht zu im Elsässischen, im Niederländischen, im Allgäu usw. In diesen Mundarten sind ausschließlich das geschlossene *u* und die *u*-haltigen Diphthonge palatalisiert worden. Alles andere ist nur der Assimilierung zuzuschreiben.

Reiffenstein (München): Gestörte Systeme haben die Tendenz, sich wieder – neu – zu «organisieren». Die Beurteilung muß vom ganzen synchronen System ausgehen, nicht von historischen Teilsystemen.

Verh. 5. int. Kongr. Phon. Wiss., Münster 1964, pp. 578–581
(S. Karger, Basel/New York 1965).

Die Tonhöhen-Unschärfe von Sprachlauten

Von F. Winckel, Berlin

Der Grund für den geringen Erfolg der Sprechmelodie-Forschung ist darin zu erblicken, daß Sprechtonhöhen absolut und auch in ihren Intervallbeziehungen in laufender Rede auditiv kaum erkannt werden, andererseits sind die psychoakustischen Zusammenhänge von Tonhöhenverläufen bisher wenig geklärt. Dies erscheint verwunderlich, da Tonhöhenmesser von hoher Meßgenauigkeit seit langem vorhanden sind.

Der Tonhöhenumfang in normaler Unterhaltung umfaßt eine Oktave, in der Emphase bis zu zwei Oktaven. Erstaunlich ist, daß der Sprechumfang von einer Oktave kaum bemerkt wird, während einfache Kinderlieder vom Umfang einer Quinte sofort eine einprägsame Melodie ergeben. Tonhöhenverläufe in der Sprache haben die Tendenz einer gewissen Stetigkeit, sprunghafte Intervalle werden möglichst vermieden. Aus einem auf Tonband gesprochenen Text hat der Verfasser Schnitte von je 100 ms angefertigt und diese stochastisch wieder zusammengesetzt. Es ist nunmehr ein deutliches Intervallspringen wahrzunehmen, das man wegen des statistischen Verhaltens natürlich nicht als «Melodie» bezeichnen darf. Auffallend ist in der stochastischen Fassung der mit wachsendem Intervall seltener werdende Anteil höherer Töne, die demnach ziemlich isoliert aus dem übrigen Verlauf hervorstechen. Aus der stochastischen Behandlung lassen sich somit eher Individualkonstanten herleiten.

Ein weiterer Grund für die schlechte Wahrnehmbarkeit der Sprachmelodie liegt darin, daß die Konsonanten, die nicht als Tonhöhen bewertet werden, in ihrem ständigen Wechsel mit den Vokalen einen hohen Prozentsatz ausmachen. Die Vokaldauern bewegen sich in der Größenordnung von 50 bis 200 ms, was grundsätzlich unterhalb der Tondauern in musikalischen Abläufen liegt.

Die Differential-Empfindlichkeit des Ohres wird festgestellt beim Vergleich von vibratoüberlagerten sinusartigen Tönen von hinreichender Dauer. Von verschiedenen Autoren werden übereinstimmend 3‰ Frequenzschwankung für Sinustöne im Bereich der Sprechgrundtonhöhen als gerade noch bemerkbar angegeben. Macht man den Versuch mit Tonimpulsen z.B. von 250 Hz, so wird man bei einer Impulsdauer von 200 ms gerade noch 3 Hz Abweichung unterscheiden können, wogegen bei einer Dauer von 5 ms das Unterscheidungsvermögen gemäß Abbildung 1 zehnmal schlechter

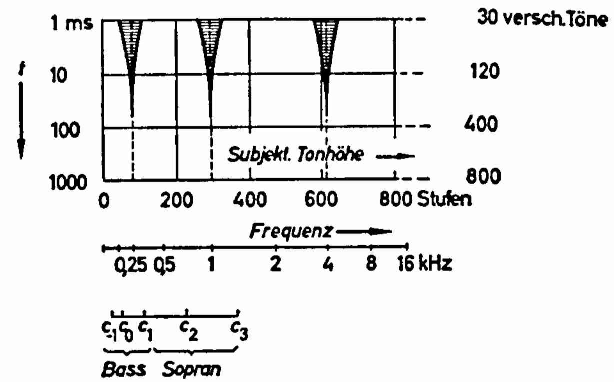


Abb. 1. Hörschärfe für Töne kurzer Dauer. Mit wachsender Dauer (Ordinate) nimmt die Hörschärfe zu, unabhängig vom Grundton, gezeigt für 250, 1000 und 4000 Hz.

wird (*Feldtkeller*). Dies ist ein Grund dafür, warum die gegenüber Musiklauten relativ kürzeren Sprachlaute im Sprechablauf so wenig melodiebildend wirken. Ferner werden Vokallaute niemals mit konstanter Tonhöhe intoniert. Anstieg und Abfall ereignen sich in großer Häufigkeit mit einer Gleitgeschwindigkeit von 1000 Hz/sec, wobei Schwankungen in den Grenzen 500 bis 2000 Hz/sec und gelegentlich auch darüber hinaus vorkommen (Frequenzmodulation). Derartige Gleitgeschwindigkeiten können trotzdem noch als konstante Tonhöhen wahrgenommen werden, wenn nur die zeitlichen Ausschnitte aus den Gleitlauten genügend klein bleiben.

Der Verfasser hat Versuche angestellt, aus Gleittönen einer Posaune, die eine Geschwindigkeit von 0,112 Oktaven/sec haben, Ausschnitte herzustellen, die als Tonhöhen konstant beurteilt werden konnten. Es ergab sich eine Dauer von etwa 0,5 sec. Eine Tonhöhenunschärfe ist zunächst aus der Heisenbergschen Unschärfe-Relation abzuleiten, woraus eine physikalische Tonhöhen-Unbestimmtheit des Glissandotones hergeleitet werden kann (vgl. Abb. 2).

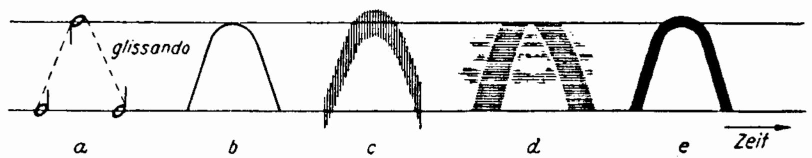


Abb. 2. Die Unbestimmtheit eines Glissandotones nach Zeit und Frequenz. a) Notierung, b) theoretischer Verlauf, c) bei exakter zeitlicher Bestimmung wird die Tonhöhenbestimmung ungenau, d) bei höchster Frequenz-Selektivität wird die zeitliche Abgrenzung ungenau, e) praktischer Fall der mittleren Zeit- und Frequenz-Ungenauigkeit (nach *F. Winckel*, Phänomene des musikalischen Hörens, Berlin 1960).

Wenn man die Frequenzmodulationsschwellen nach *Feldtkeller* und *Zwicker* heranzieht, so ergibt sich z.B. für eine Modulationsfrequenz von 16 Hz im Bereich eines 60 dB lauten Tones von 250 Hz eine Änderungsgeschwindigkeit von 80 Hz/sec. Hier handelt es sich um die periodische Wiederholung mit einem Frequenzhub von 2,5 Hz (Vibrato), also einer wesentlich geringeren Anforderung als bei der Vokalintonation mit extrem hoher Gleitgeschwindigkeit. Allerdings lagen den Messungen der Frequenzmodulationsschwellen Sinustöne zugrunde, dem Vokalmaterial dagegen obertonreiche Spektren mit 30 Teiltönen und mehr. Das gilt auch für den Posaunenversuch. In beiden Fällen sind die Teiltöne nicht exakt Harmonische, so daß an der Differenztonbildung bzw. dem Residuum sich eine Verwischung der Grundtonhöhe ergibt, während im Falle streng Harmonischer sich eine Verschärfung der Tonhöhe ergeben hätte. Anhand von Sonagrammen nach *Stevens* und *House* läßt sich der unabhängig vom Grundtonverlauf verschieden geneigte Verlauf, z.B. des 1. und 2. Formanten zeigen. Bei Vibratotönen entstehen außerdem Seitenbänder.

Die erwähnten Daten nach *Feldtkeller* und *Zwicker* können daher nur in grober Annäherung für Vokalintonation herangezogen werden. Dem Sprachcharakter wird man besser gerecht, wenn man Ausschnitte aus weißem Rauschen anstelle von Sinustönen – nach *Zwicker* – heranzieht, weil wegen der Einschwingvorgänge der Vokale zwischen den Konsonanten das Oszillogramm mehr das Aussehen permanenter Ausgleichvorgänge (transients) hat. Für die Messung mit Rauschen ergibt sich ein Unempfindlichkeitsbereich der Tonhöhenänderung – bei einer Modulationsfrequenz von 4 Hz bis zu $\pm$ 10 Hz (vgl. hierzu den Beitrag dieses Kongresses: Perzeptive Grenzen der Phonem-Unterscheidung).

Bei einem derart schlechten Unterscheidungsvermögen des Ohres für Tonhöhen-Änderung bei Sprache im Gegensatz zur Darbietung von Sinustönen sollte das Problem der Steuerung der Tonhöhe für synthetische Sprache keine Schwierigkeit sein. Es ergibt sich aber, daß eine Sprachübertragung ohne die volle Informationskapazität des Tonhöhenverlaufs unzureichend ist, d.h. für das Ohr einen leblosen maschinellen Eindruck ergibt. Bei Abschaltung der Tonhöhensteuerung – monotone Sprache – verliert die Sprache jeglichen Akzent, was durch unterschiedliche Tonlängen kaum auszugleichen ist.

Literatur

1. *Feldtkeller, R.* und *Zwicker, E.:* Das Ohr als Nachrichtenempfänger (Stuttgart 1956).
2. *Feldtkeller, R.:* Wechselbeziehungen zwischen Psychologie und Nachrichtentechnik. In: Aufnahme und Verarbeitung von Nachrichten durch Organismen (Stuttgart 1961).
3. *Kashikawa, T.* and *Sugimoto, T.:* Information rate of the pitch signal in speech. Speech Communication Seminar, Stockholm 1962.
4. *Winckel, F.:* Phänomene des musikalischen Hörens (Berlin 1960).
5. *Zwicker, E.:* Informationskapazität des Gehörs. Acustica *6:* 365–381 (1956).

Adresse des Autors: Prof. Dr.-Ing. F. Winckel, Technische Universität, Hardenbergstraße 34, *1 Berlin 12* (Deutschland).

Verh. 5. int. Kongr. Phon. Wiss., Münster 1964, pp. 582–588
(S. Karger, Basel/New York 1965).

Perzeptive Grenzen der Phonem-Unterscheidung

Von F. Winckel, Berlin

Während man anfänglich von lautanalytischen Untersuchungen ausging (Spektralanalyse), standen in den letzten Jahren die Untersuchungen mittels Segmentation im Vordergrund, die zu Klärungen führten durch die Bildung von *perceptual segments*[1]. Wichtig war die Ermittlung minimaler Phonem-Dauern, die für vokalische Laute 10 bis 15 ms beträgt[2]. Es reichen zwei Perioden zum Erkennen eines Lautes aus. Schließlich ist die digitale Synthese von Sprachlauten (Computer) gelungen[8]. Daraus geht der Aufwand an Elementarquanten für die Phonembildung hervor.

Die Problematik der Untersuchung besteht darin, daß das Phonem kein stationäres Gebilde ist, sondern in der kurzen Dauer der Intonation Änderungen in der Grundfrequenz wie auch der spektralen Zusammensetzung unterworfen ist. Die Frage ist, in welchen Grenzen solche Änderungen vom Ohr bemerkt werden. Im Sprachzusammenhang wird das komplex zusammengesetzte und zeitlich veränderliche Phonem als Ganzheit wahrgenommen. Die phonemische Folge überträgt nicht soviel Sprach-Information wie das Oszillogramm des Sprachschalls.

Nimmt man für normale Sprechgeschwindigkeit in der englischen Sprache einen Vorrat von 40 Phonemen an[4], so erhält man 5,3 bit pro Phonem, wobei strukturelle Merkmale ausgelassen sind, und mit einer sehr schnellen Folge von 10 Phonemen/sec den Wert 53 bit/sec. Hierbei kann es sich jedoch nur um einen theoretischen Wert handeln, denn die Vokale schrumpfen bei hoher Sprechgeschwindigkeit zusammen zu dem neutralen Laut ɜ[6] (Abb. 1)[9].

Die Grenzen der Sprachlauterkennung sollen im folgenden aus der Rasterschärfe der Hörfläche ermittelt werden. Tragen wir für jeden Punkt der Hörfläche das Unterscheidungsvermögen für Frequenzen und Lautstärken bei Darbietung von Sinustönen ein, so erhalten wir eine Rasteraufteilung nach *E. Zwicker* gemäß Abbil-

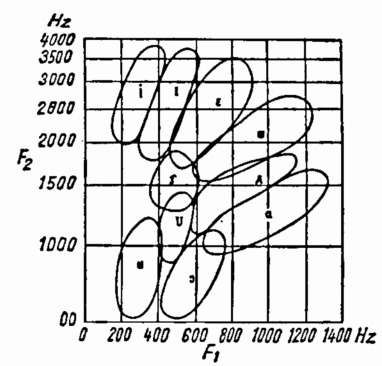


Abb. 1. Die Allophonklassen amerikanischer Vokalphoneme nach Messungen von *Peterson* und *Barney* (entnommen aus *W. Meyer-Eppler*, Informationstheorie, Berlin 1960).

dung 2. Die hörbare Änderung der Intensität Δ I und der Frequenz Δ f wurde stufenweise über die Hörfläche bestimmt durch Ermittlung der Amplitudenmodulations- und der Frequenzmodulationsschwellen. Es wurden Modulationsfrequenzen von 4 Hz gewählt, weil das Ohr in diesem Bereich die größte Empfindlichkeit besitzt. Für die Untersuchung von Sprache erscheint die Wahl günstig, soweit es die Intensität betrifft, da bei der Sprechgeschwindigkeit von

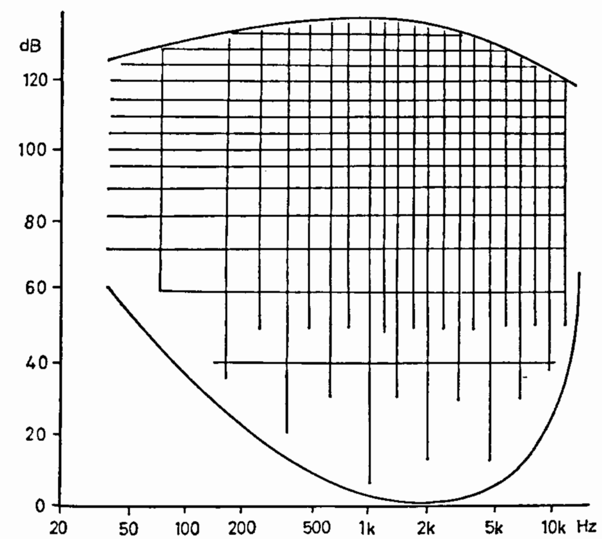


Abb. 2. Unterscheidbare Abschnitte der Hörfläche bei Sinustönen > 2 sec (je 1000 zusammengefaßt).

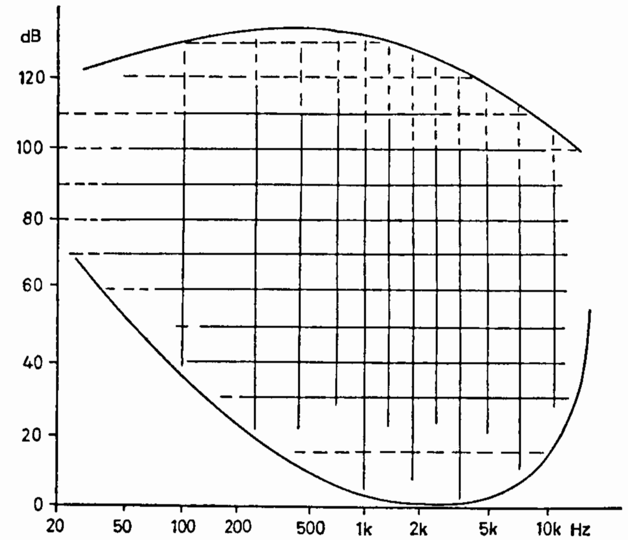


Abb. 3. Unterscheidbare Abschnitte der Hörfläche bei Breitbandrauschen (je 100 zusammengefaßt).
Muster für Reibe- und Zischlaute.

vier Silben/sec ein häufig vorkommendes Anschwellen der Sprechintensität in diesem Rhythmus erfolgt.

Dieselben Versuche wurden wiederholt bei Darbeitungen von Rauschen (Abb. 3) anstelle von Sinustönen. Aus den Messungen[11] erhält man 132 unterscheidbare Tonhöhenstufen und 120 hörbare Intensitätsunterschiede im Bereich von 10–130 dB. Somit ergeben sich aus dem Hörvergleich über die Hörfläche 15840 Valenzen von Geräuschen, dagegen 300000 von Sinustönen. Streng periodische Sinustöne kommen in der Sprache gar nicht vor, in der Musik nur selten. Vielmehr ist ein Geräuschcharakter vorherrschend (Abb. 4), so daß das dem Geräusch entsprechende Raster eher den praktischen Verhältnissen entspricht. Die Zahl der Valenzen entspricht in diesem Fall nur 5% gegenüber den von Sinustönen. Somit werden Intensitätsschwankungen von weniger als 1 dB und Tonhöhenschwankungen von $\pm$ 10 Hz bei Rauschdarbietung nicht bemerkt.

Nun gelten die angegebenen Schwellwerte für Intensitätsänderung nur für Impulsdauern oberhalb 0,1 sec. Darunter findet ein Schwellenanstieg statt[11]. Nimmt man einen noch möglichen Wert von 10 ms für das Kurzphänomen an, so gelangt man zu einem Intensitätszuwachs der Schwelle von 2 dB. Andererseits gelten die 250 unterscheidbaren Tonstufen über die ganze Frequenzskala nur

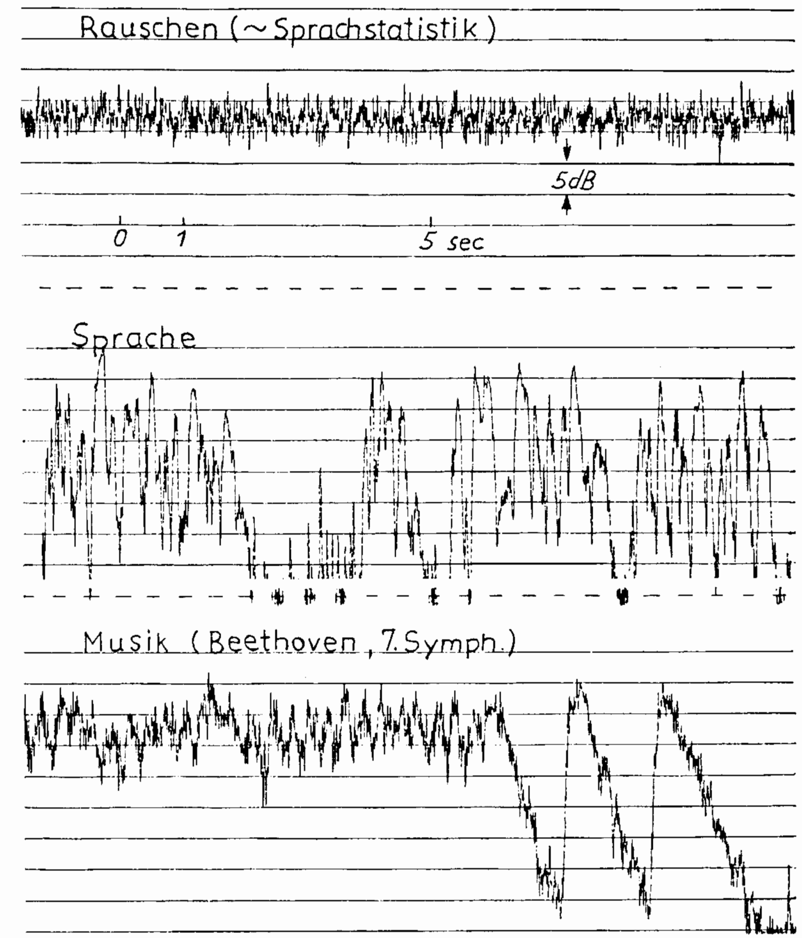


Abb. 4. Der Rauschcharakter von Sprache und Musik im Vergleich zu weißem Rauschen, das gemäß der Sprachstatistik gefiltert wurde.

für Dauertöne. Bei Darbietung von Tonimpulsen von etwa 10 ms unterscheidet das Ohr im ganzen nur 120 Tonstufen anstelle der ganzen 850 Stufen bei einer Dauer von über 250 ms[3]. Das Raster für Sinustöne muß daher 1:7 in der Frequenzskala und 1:2 in der Intensitätsskala geteilt werden, um Kurztönen von 10 ms zu entsprechen (Abb. 5). Mit abnehmender Dauer geht die Tonalität in einen geräuschähnlichen Laut über, was bei 10 ms bereits gegeben ist. Dies ist der Bereich der Explosivlaute. Andererseits bedarf die Vokalerkennung einer Mindestdauer von etwa 50 ms. Bei schnellem Sprechen werden Vokaldauern von 100 ms nur unwesentlich über-

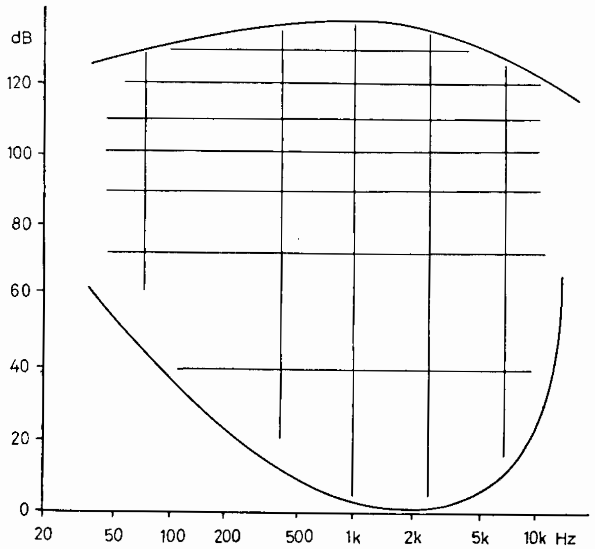


Abb. 5. Unterscheidbare Abschnitte der Hörfläche bei Sinustönen von 10 ms Dauer (je 1000 zusammengefaßt).
Muster für Explosivlaute.

schritten. Für diesen Wert gibt *Feldtkeller* bei Sinuston-Einwirkung 400 statt 850 erkennbare Tonstufen an. Das Raster für Dauertöne

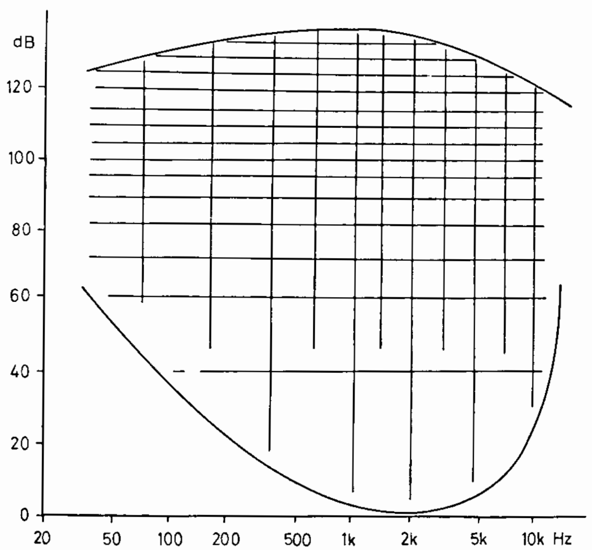


Abb. 6. Unterscheidbare Abschnitte der Hörfläche bei Sinustönen von 100 ms Dauer (je 1000 zusammengefaßt).
Muster für Vokale.

muß also in der Anwendung auf Vokale in der Frequenzskala 1:2 reduziert werden (Abb. 6). Bei strenger Vokalperiodizität der Frequenz würde eine Verschärfung der Tonhöhen-Empfindung durch Bildung des Residuums[7] eintreten. Die tatsächlich vorhandene Schwankung der Vokalfrequenzen macht die Tonhöhe-Empfindung durch das Residuum wieder unscharf.

Für die Berechnung des Informationsflusses von Sprache ist zu berücksichtigen, daß dafür nur ein kleiner Anteil der Valenzkapazität über die Hörfläche anzusetzen ist, nämlich für das angenommene ersatzweise Rauschen weniger als ein Viertel[10]. Mit einem Vorrat von nur 3400 Valenzen und 10 Phonemen/sec ergibt sich der Informationsfluß zu 115 bit/s, für den Vorrat aller Elementarlaute der Musik zu 125 bit/sec. Die Reduktion auf Telephonsprache ergibt 10,8 bit.

Die Untersuchung der Unterschiedsschwellen, die auf eine beträchtliche Unempfindlichkeit des Ohres in dieser Beziehung schließen lassen, geben den Hinweis, wie wenig die Absolutempfindung der Tonhöhe, Lautstärke, Klangfarbe, Dauer beim Erkennen von Sprache Bedeutung hat. Scheinbar sind daran noch andere Faktoren beteiligt, die sich möglicherweise in spezifischen Zusammenschaltungen des Zentralnervensystems auswirken. *W. D. Keidel* gibt den Hinweis, daß in der sekundären Rinde ganze Frequenzgruppen, wie sie für die Sprachlaute gebraucht werden, schon zusammengeschaltet sein könnten. Auch in der primären Rinde zeigt sich ein spezifisches Verhalten in der Frequenzskala, das nicht mit der Frequenzverteilung auf der Basilarmembran übereinstimmt[5].

Literatur

1. *Cohen, A.* and *'t Hart, J.:* Segmentation of the speech continuum. 4th Int. Congress Acoustics (G 51) (Kopenhagen 1962).
2. *Cramer, B.:* Über das Erkennen der Sprachlaute, in Aufnahme und Verarbeitung von Nachrichten durch Organismen (Vorträge NTG-Karlsruhe) (Stuttgart 1961).
3. *Feldtkeller, R.:* Wechselbeziehungen zwischen Psychologie, Physiologie und Nachrichtentechnik, in Aufnahme und Verarbeitung von Nachrichten durch Organismen (Vorträge NTG-Karlsruhe) (Stuttgart 1961).
4. *Fry, D. B.* and *Denes, P.:* The role of acoustics in phonetic studies in: *Richardson, E. G.* and *Meyer, E.:* Sound (Amsterdam 1962).
5. *Keidel, W. D.:* Ergebnisse der Elektrophysiologie des Hörens in: *Schubert, K.:* Theorie und Praxis der Hörgeräteanpassung (Stuttgart 1960).
6. *Lindblom, B.:* On vowel reduction. Diss. (Uppsala 1963).
7. *Schouten, I. F.:* Proc. Acad. Amsterdam *43:* 991 (1940).
8. *Strasser, B. E.* and *Matthews, M. V.:* Music from Mathematics, Bell Telephone Lab. 1961 (Beiheft zu einer Schallplatte).

9. *Ungeheuer, G.:* Anwendung der Störungsrechnung in der Eigenwerttheorie der Vokalartikulation, S. 238, 3rd Int. Congress Acoustics, Stuttgart 1959 (Amsterdam 1961).
10. *Winckel, F.:* Das Gehör in informationstheoretischer Behandlung. Arch. Ohr.-Nas.-KehlkHeilk. *182:* 456–470 (1963).
11. *Zwicker, E.:* Informationskapazität des Gehörs. Acustica *6:* 365–381 (1956).

Adresse des Autors: Prof. Dr.-Ing. F. Winckel, Technische Universität, Hardenbergstraße 34, *1 Berlin 12* (Deutschland).

Discussion

von Essen (Hamburg): Durch die Untersuchungsergebnisse ist bewiesen, daß Silbentonhöhen durch Abhören – entgegen der Behauptung von *Peters* – doch bestimmbar sind. Im übrigen wird damit erklärt, daß Gleittöne besonders in langen Silbenträgern, deren Dauer über die 0,2-Grenze hinausgeht, apperzipiert werden und dann auch phonologische Relevanz erlangen können.

Rothauser (Wien): Bei Versuchen mit synthetischer Sprache haben wir festgestellt, daß eine Rauschmodulation des Grundtons allein bereits bei einer mittleren Änderung von $\sim 3\,‰$ der ursprünglichen Grundfrequenz als Rauhigkeit des Sprachklanges bemerkbar wird. Dies scheint im Gegensatz zu dem reduzierten Tonhöhenunterscheidungsvermögen zu stehen, das vom Referenten beschrieben wird.

Janota (Prag): 1. Das Ergebnis des Experimentes mit dem Posaunengleitton stimmt gut mit der alten Erfahrung der Phonetiker überein, daß auch die Musikologen und Musiker mit einer genauen Bestimmung der Tonhöhe in zusammenhängender Rede Schwierigkeiten haben. Der Dauer des Lautes nach hören sie die gleitende Tonhöhenveränderung als eine konstante Stufe oder aber als eine schwer zu bestimmende Tonhöhe.

2. Wenn wir nun einen Gleitton zerschneiden, dann empfindet das Ohr diese Abschnitte des Gleittones als konstante Stufen. Damit ist aber keineswegs bewiesen, ob das Ohr beim Beurteilen der Sprache auf Tonhöhen*verlauf* oder Tonhöhen*stufen* reagiert.

Proc. 5th int. Congr. phon. Sci., Münster 1964, pp. 589–593
(S. Karger, Basel/New York 1965).

From the Communication Sciences Laboratory*,
University of Michigan, Ann Arbor, Michigan

Acoustic Structure of English Diphthongs and Semi-Vowels vis-à-vis Their Phonemic Symbolization

By C. M. Wise, Baton Rouge, Louisiana

What are the terminal vowel sounds of English diphthongs? Are the transient sounds between the initial and terminal vowels significant? *Chao*[1] finds that they are not. Accordingly the terminal vowel of a diphthong may be examined in isolation to discover what it is.

Consider the diphthong in *bite* (Fig. No. 1). The second spectrogram has been gated by an electronic switch to eliminate 30 ms of the transition portion of the diphthong just before the steady state of the terminal vowel. The elimination is scarcely preceptible, and

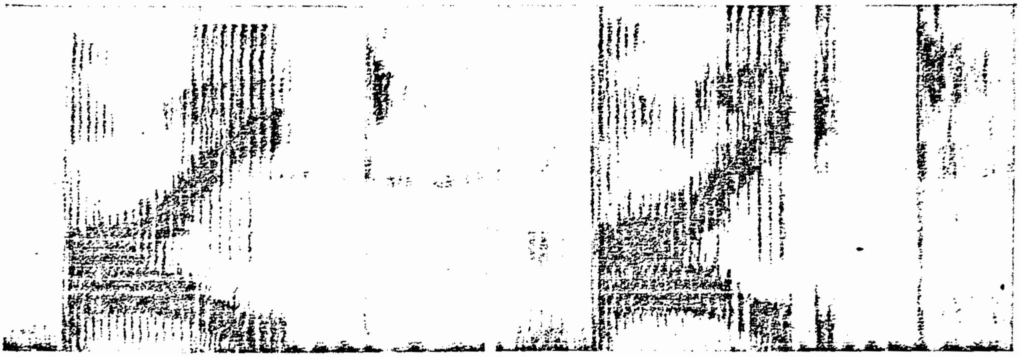
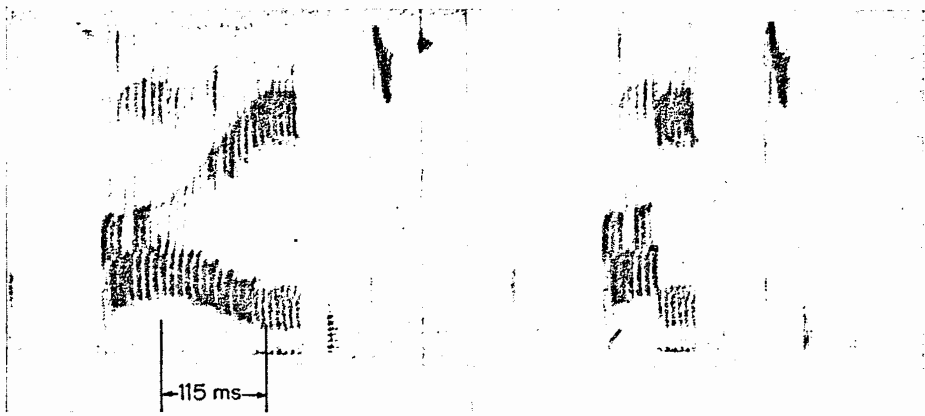

1a 1b

Fig. 1. a Spectrogram of bite without modification.
b Spectrogram of bite 'gated' with a gap of 30 ms.

* Credit due to *André-Pierre Benguerel* for laboratory work.



2a 2b

Fig. 2. *a* Spectrogram of bite. Glide of 115 ms indicated.
b Spectrogram of bite. Glide of 115 ms eliminated. Remaining parts moved together.

there is no phonemic change. It may be concluded that the portion removed is not essential to the hearer's interpretation of the word.

By means of the dual-loop recorder with associated spectrograph and gate circuit, the whole transition can be eliminated and the initial and final parts moved together (Fig. No. 2). The word is now shorter, but its meaning is the same, and the diphthong as such is not affected.

The foregoing process has been applied also to the diphthongs in *bait*, *boit*, *bout* and *boat*, with analogous results.

Speech is a phenomenon which human beings can hear as "a sequence of auditory fractions" or discrete elements, so "structured" as to "permit the exclusion of some features of the speech continuum from linguistic consideration[3]." By such exclusion of the transition within a diphthong it is again found possible to isolate end-vowels to find out what they are.

We now examine a spectrogram of the word *bite* with the end-vowel of the diphthong isolated (Fig. No. 3). With the gate circuit set to pass only this vowel, we record it. In the same way, we record the end-vowel of the diphthong in *bout*.

These brief utterances are hard to identify. Listeners often adjudge the diphthong of *bite* as ending in either /ɪ/ or /i/, and of *bout* in /ʊ/ or /u/. We now test these judgments by the measurement of the first and second formant frequencies of the two vowels and by

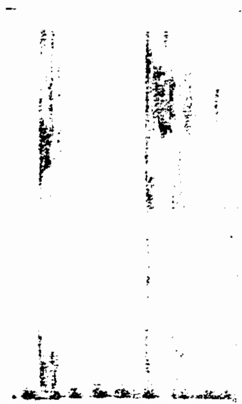


Fig. 3. Spectrogram of /ɪi/ of /baɪt/.

Table I

Comparison of End-Vowels of Diphthongs in *bite*, *bait*, *boit* with Vowels and Semi-Vowels

				Vowels which /V/ most nearly matches	
Informant CMW		F_1	F_2	F_1	F_2
bite	/baV$_1$t/	435	2550		
	/i/	275	*2550*		/i/
	/ɪ/	*450*	2350	/ɪ/	
	/j/ (in /ji/)	200	2700		
bait	/beV$_1$t/	380	2415		
	/i/	275	2550		
	/ɪ/	*450*	*2350*	/ɪ/	/ɪ/
	/j/	200	2700		
boit	/bɔV$_1$t/	410	2250		
	/i/	275	2550		
	/ɪ/	*450*	*2350*	/ɪ/	/ɪ/
	/j/	200	2700		
bout	/baV$_2$t/	445	800		
	/u/	375	975		
	/ʊ/	*450*	1000	/ʊ/	
	/o/	*450*	*825*	/o/	/o/
	/w/ (in /wi/)	300	600		
boat	/boV$_2$t/	380	870		
	/u/	*375*	975	/u/	
	/ʊ/	450	1000		
	/o/	450	*825*		/o/
	/w/ (in /wi/)	300	600		

comparing them with the formant frequencies of sustained utterances of /ɪ–i/ and /ʊ–u/, phonated by informant CMW. Here follows table I, showing the comparisons.

These comparisons show that informant CMW's end-vowels in the diphthongs of *bite, bait, boit* prove to be /ɪ/ and /i/. The end-vowels of the diphthongs in *bout, boat* range from /u/ through /ʊ/ to /o/. These findings invite measurements of the formant frequencies of many speakers.

It may be noted that the formant measurements of these end-vowels show no close relation to those of /j/ and /w/ respectively. The F_1 of /j/ is very much lower than any value of /V_1/, and F_2 much higher; and both F_1 and F_2 of /w/ are significantly lower than those of /V_2/.

Thus our evidence points toward the conclusion that the end-vowel of the diphthongs of *bite, bait, boit* lies in the range of /i–ɪ/ but far from /j/; and that the end-vowel in the diphthongs of *bout, boat* lies in the range of /u–ʊ–o/, but significantly far from /w/. These findings are supported by *Lehiste*[2].

Table II, omitted here, compares the formant frequencies of table I with those of the same vowels, as reported by earlier experimenters (*Ilse Lehiste, G. E. Peterson, H. K. Dunn*). It confirms table I.

The use of /j/ for the end-vowel of the diphthongs in *bite, boit*, and /w/ for that of *bout, boat* appears to be based on the premise that certain recorded speech sounds give similar auditory impressions if played in reverse. Steady-state vowels, with onset and terminal transition removed, do sound about the same backward, but less so with the onset and terminal transition restored. Sounds in a sequence, such as /atʃa/ and /adʒa/, sound unlike themselves when reversed.

It is also true that *yea* and *woe* sound somewhat alike when reversed; but the two versions of each syllable are only mirrored images of each other, like the right and left hands.

The phonemes /j/ and /w/ require brief steady-state vowels as beginnings, followed by rapid transitions *after* them, not *before* them, as in the terminals of *yea* and *woe;* hence /j/ and /w/ are ruled out as terminals.

However, playing utterances in reverse sometimes reveals clues for testing the validity of suspected phonemic features. The recorded word *law* played in reverse sounds considerably like *hall* because a residue of unvoiced breath escapes *after* the /ɔ/ is finished. In reverse, this rush of air attracts attention and sounds like /h/. But it has no linguistic significance.

The foregoing suggests that a similar escape of breath has been responsible for the interpretation of the terminal sound of *here* as /h/. The effect of an /h/ may appear when *here*, pronounced /hɪə/, is played backward: but then the escaping breath precedes the series of sounds, so that /həi/ results. A variation of this phenomenon can be observed when *cord*, pronounced /kɔəd/, sometimes transcribed /kɔhd/, is played in reverse and emerges as /hdəɔk/. This time the effect of /h/ is completely separated from the vowel by the /d/. Those who hit upon the transcription /hih/ for *here* must have mistaken the escape of air for the final vowel of the word, whereas it was no part of the word. Since the vowel remains /ə/, both in the forward and the reverse playing of the word, and since, when played forward, it is a clearly defined schwa, there seems no reason for symbolizing it otherwise.

Summary

The transition between the steady-state vowels of a diphthong does not appear essential to its interpretation. The end-vowel in *bite, bait, boit* is phonetically in the range of /i/ or /ɪ/. It is not /j/. The end-vowel in *bout, boat* is usually in the range of /u/ or /ʊ/ (occasionally /o/). It is not /w/.

The final phoneme of *here* /hɪə/, *there* /ðɛə/, when terminal, as well as the second vowel in *cord* /kɔəd/, is /ə/, not /h/. What has sometimes passed for /h/ appears to be the irrelevant expiration of breath after all the phonemes of a terminal word have been uttered.

Conclusion

It would seem that these phonetic facts should not be disregarded in the symbolization of the phonemes affected by them.

References

1. *Chao, Y. R.:* The non-uniqueness of phonemic solutions of phonetic systems. Bullet. of the Inst. of Hist. and Phil. (Acad. Sinica) *4:* 363–397 (1933).
2. *Lehiste, I.:* Acoustical characteristics of selected English consonants. Studies of Syllable Nuclei, No. 2, Part 3 (Ann Arbor, Mich. 1962).
3. *Lisker, L.:* Linguistic segments, acoustic segments, and synthetic speech. Language *33:* 3 (1957).

Author's address: Professor C. M. Wise, 4818 Tulane Drive, *Baton Rouge, Louisiana* (USA).

Proc. 5th int. Congr. phon. Sci., Münster 1964, pp. 594–597
(S. Karger, Basel/New York 1965).

On Preaspirated Stops in a Norwegian Dialect

By Hugo Wolter, Bergen

1. Preaspiration of voiceless stops is a well-known feature in the pronunciation of e.g. Icelandic, and descriptions of the phenomenon may be found in any handbook on that language. *Sweet* describes related phenomena in Scotch ("voiceless onglides") and *Kantner* and *West* mention a slight tendency in English "toward aspiration preceding voiceless plosives". Preaspiration in Celtic languages is described by *Marstrander*, *Bloomfield* discusses occurrences in Fox and Hopi, and in Norwegian, *Storm* and *Ross* have found it in certain dialects, in particular that of Northern Gudbrandsdal. In the dialect to be discussed in this paper, that of Northern Jæren, only *Oftedal* has mentioned occurrences of preaspiration within a rather restricted area. But, on the whole, references to preaspiration are scanty. Perhaps a word by *Haugen* may help to explain this situation: "The phoneticians have had their eyes trained too exclusively on the end of the sound where aspiration traditionally occurs."

2. I have heard preaspiration of voiceless stops on Jæren in the area from Stavanger and as far south as Bryne (30 km), and have little doubt that it may be encountered still farther south, although this remains to be investigated. The present research is confined to Stavanger and the area immediately south of that town, as far as Sandnes.

3. The investigation is based on a list of 36 words constructed as follows: I. ˟–VC:– (polysyllables with toneme 2); II. –VC: (monosyllables); and III. ˟–CC– (polysyllabic toneme 2 words containing consonant clusters with the voiceless stops /p t k/ as former cluster mates.) In the VC: dyads, the vowels /y e o a/ were used. Thus, every VC: dyad is represented twice in the material, every cluster once. The clusters in question are /pl pr ps tl tn tr ts kl kn kr ks kt/. (Note that /r/ is velar or uvular.) To find out whether stops after

long vowels are preaspirated, a few recent loanwords were included in the list (e.g. *jeep*), the stop in –V:C– being voiced in native words in this dialect.

The list was read by 21 informants, all of them students at university or high school level. They were all unaware of this feature of their pronunciation. Each informant read the list twice. Because of noise caused by imperfect recording conditions, some of the series had to be omitted, leaving a total of 26 series (936 words). Recordings of this material were made by means of a Siemens Oscillomink, at 100 mm/sec. A sample of test words was also treated by means of a sound spectrograph.

4. The auditory impression of the preaspiration is rather like an [h], although in connexion with [t] it may sound like an [f], with which it has actually been confused. The puff of breath is not equally strong with all speakers, ranging all the way from a very clear [h]-sound to the slightest voiceless transition between vowel and consonant, nearly indiscernible by ear. The auditory impression that preaspiration is stronger after rounded vowels than after unrounded ones is not borne out by the figures obtained from the *mingograms*, the difference in average duration merely amounting to 0.12 centiseconds (cs) in polysyllables and to 0.07 cs in monosyllables. The postaspiration of the preaspirated stops is usually rather weak.

Preaspiration of the stops was also registered by ear in the dyad V:C, but no measurements have been made, the material not being representative of the words known and used by some of the informants.

5. There is a significant difference between polysyllables and monosyllables containing VC: as regards the duration of the preaspiration phase. Its average duration in polysyllables is 9.09 cs, *vs* 8.53 cs in monosyllables (difference: 0.56 cs). Preaspiration amounts to an average of 28.61 % of the whole stop phase in polysyllables, to 31.73 % in monosyllables (difference: 3.12 %).

Classified after the consonant (C) of the VC: dyad, the polysyllables show the following pattern as regards average duration: /p/–8.67 cs; /t/–9.17 cs; /k/–9.41 cs, whereas the monosyllables show the inverse order: /k/–8.48 cs; /t/–8.53 cs; /p/–8.60 cs. In the latter case, however, the difference between /p/ and /k/ is so small as to seem accidental, while the corresponding difference in the polysyllable group, between /k/ and /p/, amounting to 0.74 cs, seems to indicate that the phonetic quality of C exerts a certain influence on the duration of the preaspiration phase.

Preaspiration in polysyllables accounts for an average of 29.76 % of the stop phase of /k/ (highest average) and 27.88 % of /p/ (lowest average), difference: 1.88 %. The corresponding figures for monosyllables are /k/ 33.38 % *vs* /t/ 30.36 %, difference: 2.92 %. As will be seen, averages are slightly higher in monosyllables.

It may be concluded that, on the whole, the duration of the preaspiration phase is independent on the phonetic structure of the VC: dyad in which it occurs.

6. In consonant clusters, the highest average percentage occurs for /k/ in the cluster /kt/, *viz* 40.68 %. The average duration is 6.67 cs. The corresponding figures from other cluster types are: /n/-clusters: 31.52 % ∾ 8.36 cs; /l/-clusters: 30.09 % ∾ 7.23 cs; /r/-clusters: 27.49 % ∾ 7.53 cs. No preaspiration is perceptible in /s/-clusters. This is borne out by the spectrograms.

7. On spectrograms, preaspiration manifests itself as high-frequency noise, not unlike that of [s], though considerably less strong, and poorer as regards constituent frequencies. Comparison of a large number of spectrograms might reveal a certain structure, dependent on the dyad or cluster in which the preaspiration occurs.

8. Preaspiration of voiceless stops is probably more widespread than has usually been accepted till now. Once the attention of phoneticians has been attracted to the phenomenon, the occurrence of this interesting feature will probably be detected in quite a few dialects, in which it has not yet been described.

Author's address: Gand. mag. Hugo Wolter, Fonetisk institutt, *Bergen* (Norway).

Discussion

Flifet (Bergen): Es mag erwähnt werden, daß es in Skandinavien auch eine lappische Präaspiration der Tenues gibt. Mindestens in Norwegen ist sie jedoch, soviel ich weiß, ohne Einfluß auf die lokale Artikulation der Landessprache geblieben. In engerem Kontakt mit dem Vortragsthema steht die Präaspiration in der Mundart von Nord-Gudbrandsdalen. Von den vier altnorwegischen phonematischen Silbenquantitäten haben sich hier die drei geläufigsten erhalten, nämlich die folgenden: 1. V:C, Beispiel: /ˣi:lə/; 2. VC, Beispiel: /ˣvili/; 3. VC:, Beispiel: /ˣvil:ə/. Wo bei diesen Silbengegensätzen der inter- bzw. postvokalische Konsonant ein Tenuis ist, wird dieser im Typus 3 präaspiriert, in den beiden anderen Typen hingegen nicht. In Fällen dieser Art involviert die Präaspiration somit eine perzeptorische Verdeutlichung der prosodischen Oppositionen. Andererseits muß bemerkt werden, daß die lokale Präaspiration der Tenues nach kurzem Vokal in langer Silbe auch bei heterogenen Konsonantenverbindungen geläufig ist, genau wie in der von Herrn *Wolter* behandelten Mundart.

Werner (Erlangen): Ich freue mich, daß Sie uns, Herr *Wolter*, das Phänomen der Präaspiration instrumentalphonetisch nachgewiesen und genau beschrieben haben.

Erlauben Sie mir einige Anmerkungen, die sich mehr auf die phonematische Seite der Präaspiration beziehen.

Dient die Präaspiration nicht auch in Ihrem Untersuchungsgebiet dazu, die einstige (germ.) Opposition in den Explosiven sth./Lenis – stl./Fortis (?), sagen wir /d/ – /t/, abzulösen durch /d/ – /hd/, so daß sich nur noch eine einzige Explosivreihe ergibt? Läßt sich vielleicht auch instrumentalphonetisch zeigen, daß der Explosiv sowohl mit wie ohne *h* in bezug auf Stimme und Intensität frei variiert (je nach Folgelaut...)?

Für das Neuisländische und Färingische[1] wurde dieser Systemwandel nachgewiesen.

Diese Ablösung von einstigem /d/ – /t/ zeigt sich also in weiten Räumen des Westskandinavischen, und es ergibt sich die Frage nach Entstehung und Alter der Neuerung. *Chapman*[2] konnte solche Dinge leider nur mit wenig Material andeuten.

Ich möchte daran erinnern, daß auch in weiten Räumen des übrigen Germanischen die ehemalige Opposition /d/ – /t/ in vielerlei z. T. stellungsverschiedene Oppositionen überführt oder ganz aufgegeben wurde (z.B. südnorw./dän.-Systeme, binnendeutsche Konsonantenschwächung, mittelbair. Schwächung).

Beachtenswert und vielleicht sogar sensationell erscheint mir schließlich der Befund, daß die doch sehr seltenen phonetischen Erscheinungen (Präaspiration und stl. Nasale/Liquide) – soweit ich sehe – allein hier im Nordseeraum zu finden sind: im Skandinavischen, im Keltischen und – wie uns gerade Herr *Fliflet* mitteilt – im Lappischen. Ich darf die schon 1932 von *Marstrander*[3] vorgenommene Zusammenschau nennen und meine, daß es sich lohnen würde, hier nochmals skandinavistisch-keltologisch und fennougristisch zusammenzuarbeiten – am besten in strukturalistisch-phonologischer Methodik.

[1] Z.B. *E. Haugen:* The phonemics of modern Icelandic, Language *34* (1958); *O. Werner:* Aspiration und stimmlose Nasale/Liquide im phonologischen System des Färingischen. Phonetica *9* (1963).

[2] *K.G. Chapman:* Icelandic-Norwegian linquistic relationships. NTS Suppl. VII, 1962.

[3] *C.J.S. Marstrander:* Okklusiver og Substrater. NTS *5* (1932).

Verh. 5. int. Kongr. Phon. Wiss., Münster 1964, pp. 598–602
(S. Karger, Basel/New York 1965).

Aufbau und Funktion phonologischer Einheiten

Langue und Parole

Von Ludwik Zabrocki, Poznań

Der einzige Weg, um zur Langue zu kommen, führt über Parole. So ist es beim Kinde, das die Muttersprache erlernt, so ist es auch beim Erwachsenen, der sich eine Fremdsprache mit Hilfe der direkten Methode aneignet. Die Parole-Einheiten sind dem Lernenden bewußt, die Langue-Einheiten dagegen werden ohne das Eingreifen des Bewußtseins aufgebaut. Die Langue-Einheiten, unter ihnen auch die phonologischen Einheiten, kann man sich als eine Art von Matrizen vorstellen. Wir unterscheiden hier unter anderem phonologische, Morphem-, Wort- und Satzmatrizen. Der Aufbau aller dieser Matrizen geschieht auf empirischem Wege. Grundlage und Ausgangspunkt bildet hier das als Zeichen fungierende Lautsubstanzkontinuum.

Segmentierung des Lautkontinuums

Aus den dem Menschen zur Verfügung stehenden Substanzen wurde die Lautsubstanz zum Träger der Zeichenfunktion gewählt. Daraus folgt zuerst, daß alle Substanzen, welche von uns wahrgenommen werden können, als bezeichnete Substanzen unserer Erkenntnis zugängig sind, die Lautsubstanz dagegen, soweit sie im Dienste des sprachlichen Signums steht, als bezeichnende Substanz analysiert werden kann. Sie kann aber auch in demselbem Maße als bezeichnete Substanz betrachtet werden. Also können wir letzten Endes die lautliche Substanz, die als bezeichnete Substanz erkannt worden ist, auch als *bezeichnete* Substanz analysieren. In dieser Eigenschaft als bezeichnete Substanz ist diese Substanz völlig erkennbar. Die Segmentierung eines Lautkontinuums ist grundsätzlich möglich. Es unterliegt doch denselben Segmentierungsgesetzen wie jede andere Substanz. Es besteht eben kein grundsätzlicher

Unterschied zwischen der lautlichen Substanz als bezeichneter Substanz und allen anderen bezeichneten Substanzen. Das Lautkontinuum kann hier keine Ausnahme bilden.

Bei der Erforschung des Lautkontinuums können nur naturwissenschaftliche Methoden angewandt werden. Zuständig sind hier die Akustik sowie die Experimentalphonetik als Naturwissenschaft. Die Erkennbarkeit der verschiedenen Segmente im Lautkontinuum als bezeichneter Substanz kann man mit der Erkennbarkeit aller anderen aus vielen Segmenten zusammengesetzten Substanzkontinua vergleichen. Die Erkennbarkeit des Lautkontinuums auf der Stufe der bezeichneten Substanz bildet die Grundlage für die spätere Segmentierung des Lautkontinuums als bezeichnender Substanz. Diese Segmentierung geht grundsätzlich der Segmentierung des Lautkontinuums als bezeichnender Substanz voran. Auf der Stufe der bezeichneten Substanz werden Struktureinheiten bis zur Sprechsilbe aufgebaut. Alle diese Struktureinheiten werden sowohl auf Grund von linearen (syntagmatischen) wie paradigmatischen (lateralen) Operationen gewonnen. Es wird hier sowohl mit der substanziellen Oppositionsgröße wie mit dem Verteilungsfaktor (Junktivität der Laute) gearbeitet.

Der Lautsubstanz als bezeichneter Substanz können wir grundsätzlich auch nur eine *Lautsubstanz* als bezeichnende Substanz gegenüberstellen. Diese können wir als Metalautsubstanz bezeichnen.

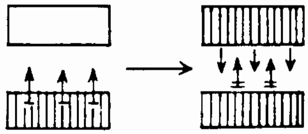
Die Lautsubstanz als bezeichnende Substanz

Der Aufbau der Strukturen der bezeichnenden Lautsubstanz erfolgt auf Grund der Einwirkung des Inhaltsplanes der Sprache. Die Einwirkung des Inhaltsplanes, der die bezeichnete Lautsubstanz zur bezeichnenden Lautsubstanz macht, verursacht eine neue Segmentierung des Lautkontinuums, und zwar sowohl in syntagmatischer wie in paradigmatischer Hinsicht. Diese Neuordnung des Lautkontinuums kann aber nicht die Grundelemente der Segmentierung des Lautkontinuums als bezeichnete Substanz zunichte machen. Die neue Segmentierung basiert auf der alten, primären Segmentierung. Die primären Grundelemente der ersten Segmentierung können aber zu neuen, höheren Einheiten organisiert werden. Die neuen Einheiten kann nur die linguistische Analyse feststellen. Die Phonetik als Naturwissenschaft ist in dieser Hinsicht hilflos.

Die phonologischen Einheiten der sprachlichen bezeichnenden Substanz werden auf empirischem Wege aufgebaut. Ausgangspunkt beim Aufbau der Langue-Einheiten bildet die Parole.

Graphisch kann man das folgendermaßen darstellen:

Matrize vor Aufbau — Matrize nach Aufbau



Konkreter Text

Erläuterung: $\uparrow\uparrow$ = Matrizenaufbau. $\updownarrow\updownarrow$ = Matrizenzuordnung.
$\downarrow\downarrow$ = Rekonstruktion nicht kongruenter Texte.

Funktion der phonologischen Einheiten

Nach dem Aufbau der phonologischen Einheiten in Gestalt von entsprechenden Matrizen verläuft der kommunikative Sprechakt in ganz anderer Art. Der konkrete Satz, also die Parole-Einheit, wird den entsprechenden phonologischen Matrizen im Abtastungsverfahren zugeordnet. Dabei wird aber die bezeichnende Substanz vorerst als bezeichnete Substanz vom Empfänger analysiert. Dann erst wird das Ergebnis dieser Analyse auf die Abstimmung mit den entsprechenden Gefüge- und Systemmatrizen überprüft, das heißt, diese Substanz wird der Reihe nach als bezeichnende Substanz analysiert und überprüft. Als letzter Akt des Vorganges kommt die entsprechende Zuordnung zustande.

Die Zuordnung zu den entsprechenden Matrizen ist von einer hundertprozentigen Übereinstimmung der Lautsubstanz als bezeichneter Substanz mit den Gefüge- und Systemmatrizen der Langue in gewissem Maße unabhängig. Gewisse Unbekannte, die die primäre Analyse mit sich bringt, können mit Hilfe einer bestimmten Zahl von bekannten Faktoren ermittelt werden. Wir haben es in dieser Hinsicht mit einer gewissen Rekonstruktion der Lautsubstanz zu tun. In einem Satz können wir diese oder jene Laute unvollkommen artikulieren, und doch wird der Satz letzten Endes verständlich. Und es kommt vor, daß der Empfänger überhaupt nicht merkt, daß er einen teilweise substanziell verstümmelten Text empfangen hat. Dieses Phänomen ist der Informationstheorie gut bekannt. Auf diese Weise kann zum Beispiel der Laut $\bar{o}$ im

Worte *Ofen* im Satze «*Der Ofen steht in der Stube*» auf der Parole-Ebene kurz ausgesprochen werden. Dennoch wird das Wort *Ofen* der richtigen Matrize zugeordnet und somit der Satz verständlich.

Linguistische Analyse und Fremdsprachenerlernen

Es besteht ein grundsätzlicher Unterschied zwischen der Analyse einer Sprache, die man spricht, und einer Sprache, die man nicht praktisch beherrscht (z. B. einen mundartlichen Text). Im ersten Falle stehen dem Sprachwissenschaftler die entsprechenden Matrizen der Langue als fertiges Werkzeug zur Verfügung, im zweiten Falle muß er diese Matrizen erst aufbauen. Die Sprachmatrizen der Langue nennt man populär Sprachgefühl.

Beim Erlernen einer neuen Sprache muß der Lernende unter anderen die Matrize der Langue aufbauen. Dabei wird dieser Aufbau von den Matrizen der Muttersprache aufs äußerste beeinflußt (Interferenz). Das Erlernen einer Fremdsprache durch Kinder ist unter anderem grundsätzlich dadurch begünstigt, daß die Kinder den Aufbau der Languematrizen der Muttersprache auch erst durchführen. Das Verfahren der Zuordnung zu den entsprechenden Languematrizen verläuft in beiden Fällen parallel zum kontinuierbaren Matrizenaufbau. Der Erwachsene dagegen kennt im Bereich der phonologischen Einheiten der Langue nur ein Zuordnungsverfahren. Dabei kann der Streuungsfaktor der sich substantiell realisierenden Phoneme und Lautklassen ziemlich groß sein, wie das die Zwirnersche Phonometrie einwandfrei bewiesen hat.

Wie aus dem oben Dargestellten hervorgeht, dürften sowohl für die Theorie und Praxis der sprachlichen Analyse wie für die Theorie des Fremdsprachenunterrichts gewisse weitere Rückschlüsse gezogen werden.

Literatur

Jakobson, R.; Fant, C. G. M. and *Halle, M.:* Preliminaries to speech analysis (Massachusets Institute of Technology, Acoustic Laboratory 1955).

Jakobson, R. and *Halle, M.:* Fundamentals of language (Mouton, 's-Gravenhage 1956).

Pike, K. L.: Phonetics (University of Michigan Press, Ann Arbor 1943). – Phonemics (University Press of Michigan, Ann Arbor 1947).

Zabrocki, L.: Sprachkode. Z. f. Phon. allg. Sprachw. *14:* 64–73 (1961). – Phon, Phonem und distinktives Morphem. Biul. Fonograf. *5:* 59–87 (1962).

Zwirner, E. und *Zwirner, K.:* Grundfragen der Phonometrie (Metten, Berlin 1936).

Adresse des Autors: Prof. L. Zabrocki, Kanclerska 11, *Poznan-Grunwald* (Polen).

Discussion

Hammarström (Uppsala): Herr *Zabrocki* hat das Wort «verstümmelte» Sprache, das etwas Herabsetzendes zu implizieren scheint, verwendet. Ich würde eher «elliptisch», «reduziert» oder etwas Ähnliches als Terminus gebrauchen. Es handelt sich ja um die Tatsache, daß der Sprecher so «undeutlich» spricht – und sprechen soll –, wie es ihm die Aufnahmemöglichkeiten des Hörers erlauben. Der Sprecher verwendet in einer uns weitgehend unbekannten Weise sehr zweckmäßige Reduktionen. (Ich habe das Wort «économie de la parole» in *Romance Philology 15:* 353 [1962] dafür vorgeschlagen.)