

HOW DETERMINABLE ARE INTONATION UNITS?

A. Wichmann and *G. Knowles

Department of Cultural Studies, University of Central Lancashire, Preston PR1 2HE

*Department of Linguistics, Lancaster University, LA1 4YT

ABSTRACT

This paper examines some of the practical and theoretical issues surrounding the assignment of intonation units in the transcription of speech corpora. We suggest that current practice may obscure some aspects of the prosodic system, and that there is a need for caution when using such data.

INTRODUCTION

There is a widespread assumption that spoken texts can be segmented into discrete prosodic units all of the same status, and known variously as *intonation units (IU)*, *tone units*, or *tone groups*. This assumption has been applied in the annotation of spoken corpora including the London-Lund Corpus (LLC) [1], and the Lancaster/IBM Spoken English Corpus, (SEC) [2]. It has most recently been applied to the Corpus of Spoken American English, (CSAE) [3]. While there are eminently good practical reasons for this approach, the resulting transcription is open to misinterpretation, especially when it is used as data by non-specialists. There are two main considerations here: the accuracy of auditory transcriptions, and the theoretical assumptions which underlie them.

THE RELIABILITY OF TRANSCRIPTIONS

Transcriber accuracy

The concept of 'accuracy' in prosodic transcription begs many questions, not

least because they are based on auditory perceptions. Human sensory perception relates to physical reality in complex ways, and is rarely discussed in terms of 'correctness', since this would presuppose that the physically measurable phenomenon in each case is criterial. It is more appropriate to consider the consistency of transcription across two dimensions. The first is consistency across transcribers, and the second is consistency within an agreed system of intonation.

The transcribers of the corpora mentioned above have undertaken various measures to ensure consistency across transcribers. The 50,000 words of the SEC were analysed by two transcribers working independently, but overlapping for about 9% of the corpus [2]. The CSAE is a much larger undertaking (200,000 words) and consequently employs many more transcribers. Consistency is sought first by intensive training of the transcribers and secondly by joint discussion of all doubtful cases [3]. In each case the transcribers are highly trained, and it is fair to assume that a high degree of internal consistency has been reached. With this degree of expertise it is possible to be relatively confident that the published spoken materials have been transcribed consistently, although this still does not make them necessarily comparable with one another.

Theoretical consistency

A more difficult issue than the reliability of transcribers is the reliability of the system itself. The biggest problem is to define the units of the system [5]. Phonological accounts differ, and different units are posited on different theoretical assumptions, see e.g. [6], [7], [8], [9], and [3]. British approaches to the IU are traditionally based on a structural definition in terms of onset or head, nucleus and tail. Other approaches assume a demarcative definition, in which the unit is defined in terms of its boundaries. Frequently adopted criteria are pause [10] and pitch contours [11].

Studies of the phonetic attributes of transcribed boundaries provide evidence for a wider range of attributes, some segmental, such as absence of assimilation and elision, others prosodic, such as syllable lengthening, changes in tempo and changes in voice quality. [12] [13]. These attributes co-occur in different ways, signalling boundaries of varying strengths, from very obvious ones to those hardly discernable. The inability of transcribers to agree on some boundaries is not evidence of lack of expertise, but an inherent problem in a phonological theory which does not allow for indeterminacy. Indeed, our experience of the analysis of prosodic data increasingly indicates that it would be desirable to incorporate such information, i.e. cases of disputed boundaries, in the transcription.

A further issue is that these various cues to boundaries may in fact belong to separate prosodic systems. We would argue that a distinction needs to be made between temporal criteria (including pause and final lengthening) and melodic criteria.

Melodic units and temporal discontinuities

The melodic and temporal systems appear to operate in different ways. Melodic patterns coincide with segments of text which can be called **melodic units**; temporal patterns divide one segment of text off from another and constitute **discontinuities**. These would also appear to have separate functions, and we would tentatively suggest that melodic units relate to the structure of the text, while temporal discontinuities are chiefly concerned with interaction management.

Melodic units have a complex structure which can be best represented by some kind of transition network [9]. The domain of these units corresponds to structural units of text. When sentences and clauses are marked off, they are given one or more melodic units. Relative clauses and appositives are marked by the repetition of melodic fragments, a pattern which is sometimes called *tone harmony*. Coordinated items and lists are assigned melodic fragments which combine to form a complete unit at a higher level.

Temporal breaks can be inserted into a tune for several reasons. The most obvious case is the hesitation pause, related to the speech planning process, but pauses can also be motivated by the text. (In the following examples (•) indicates a pause). They are used in reported speech, e.g. 'Yes' (•) *she said* where the break separates the quotation from the reporting phrase. They can draw attention to a following item, e.g. *these are called* (•) *'formants'*, or correspond to scare quotes in writing, e.g. *Mary is his* (•) *'friend'*. Skilled readers of verse may use a temporal

discontinuity within a continuing melodic unit to accommodate conflicting demands of speech rhythm and verse metre. The set of uses is not necessarily fixed, and speakers may invent idiosyncratic uses of their own.

The fitting of tunes to a text, and the insertion of temporal breaks, are of course not mutually exclusive. Indeed, temporal discontinuities commonly co-occur with the end of a tune. Consequently, a tune is prototypically - but not necessarily - bounded by temporal breaks. On the other hand there are many temporal discontinuities that are not in any way related to melodic units.

In practice, transcribers of speech corpora have used a wide range of mostly demarcative criteria to identify just one kind of unit (or two when a distinction is made between 'major' and 'minor' units). In other words a boundary has been marked variously as a response to pause, and to contour change, and to other prosodic cues.

CONCLUSION

Prosodically annotated speech corpora are valuable sources of data for the investigation of the prosodic system or systems. Such investigations will normally be based on the transcription in conjunction with the original sound recording. Corpora are, however, also intended to 'provide materials for extensive studies of all aspects of spoken English grammar and lexicon' [3]. This kind of research tends to rely on the transcription alone, taking the annotations as given.

While the original transcribers may be well aware of the different clusters of

phonetic attributes that are annotated as boundaries, the fact that they are marked in the same way makes it difficult for subsequent users of the transcription to avoid treating IU boundaries as though they were all of equal phonological status. In reality, if the distinction made in this paper is valid, then it follows that some of the stretches of speech between marked boundaries are of no phonological status at all.

REFERENCES

- [1] Svartvik, J. (ed) (1990) *The London-Lund Corpus of Spoken English: Description and Research*, Lund University Press, Lund.
- [2] Knowles, G., Williams, B., Taylor, L. (eds) (to appear) *The Lancaster/IBM Spoken English Corpus: A Corpus of Formal British English Speech*, London: Longman.
- [3] Chafe, W., Du Bois, J.W. & Thompson, S. A. (1991) "Towards a new corpus of spoken American English" in: Aijmer K. & Altenberg, B. (eds) *English Corpus Linguistics*. London: Longman.
- [5] Couper-Kuhlen E. (1986) *An Introduction to English Prosody*, London: Edward Arnold.
- [6] Crystal, D. (1969) *Prosodic Systems and Intonation in English*, Cambridge: Cambridge University Press.
- [7] Pierrehumbert, J.B. (1980) "The phonology and phonetics of English intonation". PhD thesis, MIT.
- [8] Ladd, D.R. (1986) "Intonational phrasing: the case for recursive prosodic

structure". *Phonology Yearbook* 3: 311-340.

[9] 't Hart, J., Collier, R. & Cohen, A. (1990) *A perceptual study of intonation*. Cambridge: University Press.

[10] Stenström, A-B. (1990) "Pauses in Monologue and Dialogue", in: [1].

[11] Brown, G. (1977) *Listening to Spoken English*, London: Longman.

[12] Knowles, G. (1991) "Prosodic labelling: the problem of tone group boundaries", in: Johansson, S. & Stenström, A-B. (eds) *English Computer Corpora: selected papers and research guide*. Berlin: Mouton de Gruyter.

[13] Chafe, W. (1995) "Adequacy, User-Friendliness, and practicality in Transcribing", in: Leech, G., Myers, G., Thomas, J. (eds) *Spoken English on Computer: Transcription, mark-up and application* Longman: London.