COMPUTATIONAL MODELLING AND GENERATION OF PROSODIC STRUCTURE IN SWEDISH

Merle Horne and Marcus Filipsson
Department of Linguistics and Phonetics, University of Lund

ABSTRACT

A summary of the motivation for the various levels of structure assumed in a prosodic hierarchy for Swedish and the linguistic and discourse parameters that are needed for their recognition in texts are presented.

INTRODUCTION

Current speech synthesis systems, which lack detailed prosodic structure cannot generate many of the intonational patterns that one observes in natural speech. Prosodic phenomena associated with the boundaries of clause-internal word groups constitute one problem area. More specifically, the transitions that the current Swedish text-to-speech system generates between word accents do not always correspond to those one finds in naturally occurring speech. As the F0 curve in Figure 1 (corresponding to part of the sentence in (1)) shows, the end of the focussed expression *för närvarande* 'presently' coincides with a low F0 point (L#) in the speech of the radio commentator we are modelling. This L#, we claim below, corresponds to the end of the prosodic constituent which we will define as a [+focal] Prosodic Word. In Figure 2, it is observed that the corresponding synthetic F0 curve generated using the current rule system cannot reproduce this pattern since no low point after the focal high is predicted in clause-internal position. The F0 transitions are only triggered by the positions of the word accents which can be either focal (i.e. followed by a H⁻) or nonfocal (i.e. without an additional following H⁻). Thus, after the H*L (Accent 2) word accent on the syllable *-när-*, there is a rise throughout the remainder of the word (due to an associated focal H⁻) and the first syllable of the following word, *betecknas* 'is characterized', since the underlying accent pattern of an Accent 1 word like *betecknas* is HL*, with a H on the premainstress syllable *be-* and a L* on the syllable *-teck-* [1]. Thus, the L# at the end of *för närvarande* such as in Figure 1 cannot currently be automatically generated.

(1) För närvarande betecknas tendensen som mycket svag 'At present the trend is characterized as very weak'

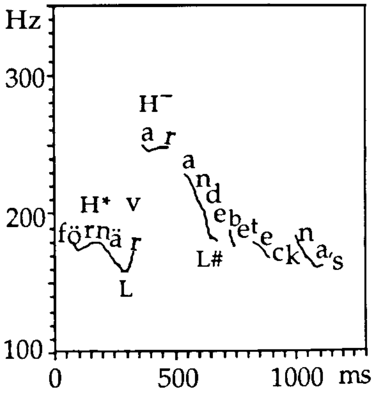

Figure 1. A partial F0 contour for the sentence in (1) uttered by a professional radio commentator.

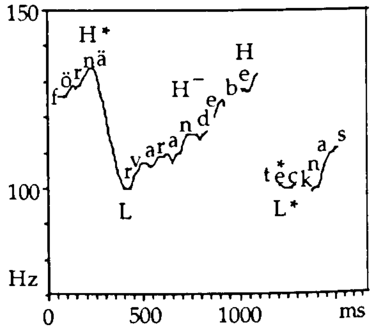

Figure 2. Synthesized F0 contour for the same sentence fragment as in Figure 1.

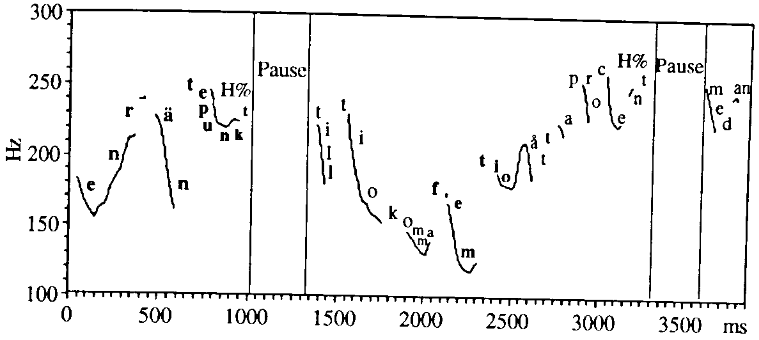

Figure 3. F0 contour for a fragment of the sentence in (2) (1 räntepunkt || till 10,58 procent || medan) with a clause-internal PP boundary after 'en räntepunkt'.

Another problem with current synthesis is that one has not been able to predict the location of clause-internal Prosodic Phrase boundaries, i.e. internal boundaries that are as strong as those which occur at the end of the majority of clauses/sentences. This is exemplified, for example, in Figure 3, which presents part of the F0 contour associated with the sentence in (2), where the internal boundary after *räntepunkt* 'interest point' has the same strength as that after *procent* 'percent'.

(2) ||*Tolvmånaders statsskuldväxlar hade gått tillbaka 1 räntepunkt || till 10,58 procent || medan sexmånadersväxlar gått upp 5 punkter till 10,50 procent*||
'Twelve-month state-debt bonds had gone back 1 point to 10.58 percent while six-month bonds had gone up 5 points to 10.50 percent.' (where || represents a Prosodic Phrase boundary)

In order to be able to recognize such internal Prosodic Phrase boundaries, one must have access to more lexico-grammatical information than is currently available in text-to-speech systems.

SWEDISH PROSODIC STRUCTURE

Prosodic Word (PW)

Three levels of prosodic structure are being assumed over the level of the syllable [2]. The smallest of these is the Prosodic Word (PW) which is defined as corresponding to a content word and any following function words up to the next content word within a given Prosodic Phrase (PPh). At the beginning of a PPh, the PW can also begin with one or more function words.

The PW is characterized by a word accent. It is also marked by a boundary tone which is realized by a final rise in the case where the content word is not focussed (i.e. contextually given) (H#) or a fall when the content word is focussed (L#). These boundary tones, we claim, play an important role in creating the transitions between consecutive PW's in a larger PPh. The unit does not necessarily correspond to a syntactic constituent; the grouping is, however, characteristic of well-planned speech. It is a rhythmic grouping with a left-headed character, where a content word can be grouped together prosodically with following function words in a manner analogous to the way the definite article and other morphological endings are attached and prosodically cliticized to the right of a lexical stem in Swedish (e.g. *bil+ar+na* 'car+pl.+the'). Thus, a PW can consist of a content word and a following preposition (e.g. *[köpt över]*, 'bought over') where the preposition is syntactically a member of a constituent that does not include the content word, as in *Lars har köpt över 100 skivor* 'Lars has bought over 100 records'.

Prosodic Phrase (PPh)

One or more PW's make up a PPh which is marked by a L% or H%

boundary tone accent, a following pause and a certain degree of Final Lengthening ([3]). It corresponds to both Pierrehumbert's 'Intonation Phrase' [4] and Lieberman's 'breath group' [5]. Factors which determine the location of PPh boundaries include the following:

a) clause/sentence boundary: A clause boundary corresponds in many cases to the end of a PPh. In an auditory analysis of a corpus of 36 radio broadcasts containing 724 clauses, where clauses also included elliptical clauses, 499 or 69% of these were characterized as ending in a boundary which was as strong or stronger than a PPh (404 were classified as PPh boundaries and 95 as Prosodic Utterance (PU) boundaries). Since we assume the Strict Layer Hypothesis in the hierarchy of prosodic constituents, this means that the end of a PU is also the end of a PPh; thus, 69% of the clauses ended in a boundary associated with a PPh at some level of analysis. Furthermore, 337 of these clauses corresponded to sentence boundaries. In the whole corpus, there were 362 sentences. This means that 93% of the sentence boundaries were assigned a prosodic boundary equal to a PPh on some level of analysis.

b) clause-internal PPh boundaries: Although PPh boundaries occur in the majority of cases in clause-final position, they may also occur optionally in clause-internal position. In our data of 724 clauses, we detected only 17 cases (2%). Although the number was extremely small, we decided, nevertheless to examine the lexico-syntactic structure of the data to determine whether one could make any generalizations concerning the environment for the insertion of these internal PPh boundaries. The following conclusions can be drawn from the investigation: in the domain-specific data-base dealing with the stock-market we investigated, 12 clause-internal PPh boundaries occurred before focussed complements (beginning with *till* 'to' or *sedan* 'since') to the verbs *gå upp* 'go up', *gå ner* 'go down', *falla* 'fall', *stiga* 'rise' IFF these complements were preceded by another focussed verb complement. Thus a PPh boundary (||) could occur before *till* in (4a) but not in (4b) where the first complement following the verb is not focussed (relevant focussed expressions are written in bold script):

(4) (a) ||*Fyra-åriga standardobligationen hade då fallit* ***4*** *punkter* || *till en ränta på* ***10,27*** *procent*||
'||The four-year standard bond had fallen **4** points|| to an interest-rate of **10,27** procent||'

(b) ||*Tolvmånaders statsskuldväxlar hade också gått tillbaka 4 punkter till* ***10,84*** *procent*||
'||Twelve-month national-debt bills had also gone back 4 points to **10.84** procent||'

The remaining 5 cases of clause-internal PPh boundaries occurred between a relatively long subject (on the average of 15 syllables) and a focussed verb.

c) length: A PPh will consist of a more or less fixed number of syllables at a given rate of speech since PPh's (as we have defined them) correspond to what is often termed 'breath groups' [5]. In our material, where the speech rate is on the average of about 5 syllables/second, PPh's contained between 7 and 63 syllables, with the mean at 24 syllables (SD=10.3). Our data also indicate that a sentence-internal clause must be of a certain length in order for it to be associated with a PPh boundary; clauses containing less than 7 syllables (12 elliptical clauses) in our database were assigned a weaker, i.e. PW boundary. PW boundaries also replace PPh boundaries in a great many other cases. The 225 clause boundaries (31%) which were associated with a PW boundary instead of a PPh boundary all arose from the linking together of clauses within a discourse segment. Such linking never occurred over discourse segment boundaries. It was observed, futhermore, that clause-linking occurred only if the first clause contained less than 30 syllables and if the resulting PPh, after the linking of two or more clauses together, did not contain (on an average of) more than 40 syllables. The linking of clauses occurred most frequently at the beginning of discourse segments and practically never between the second last and final clauses of a

discourse segment. Thus, it could be that the *non-linking* of clauses can be considered as a cue to segment finality (see also [6] for other cues).

Prosodic Utterance (PU)

One or more PPh's make up a PU, which is bounded by extended pauses [3]. These strong boundaries coincide with locations where a topic shift takes place (i.e. the end of a 'discourse segment' [7]). In our data, 95 of the 727 clauses (13%) ended in a boundary which was classified as a PU boundary. In the texts which were originally read on the radio, these correlate with the opening of a new paragraph immediately following the PU boundary (S. Haage-Palm, personal communication).

GENERATING PROSODIC STRUCTURE

In order to automatically generate prosodic structure, it is important to be able to recognize a number of different kinds of information [8]. First of all, the distinction between content words (e.g. nouns, adjectives) and function words (e.g. prepositions, conjunctions) is needed in order to define PW's. It is also crucial to be able to identify clause boundaries in a text since the clause is the basic domain over which PPh's are defined. Clause boundaries occur at certain punctuation marks, e.g. full stop, colon, semicolon, some commas (those not occurring in lists of words having the same word class), as well as before coordinate conjunctions and relative pronouns (e.g. *som* 'that') and after subordinate conjunctions (e.g. *att* 'that', *om* 'if').

In order to generate the clause-internal (optional) PPh boundaries, we have included a domain-specific module in our algorithm. This was due to the fact that the locations of clause-internal PPh boundaries seemed to be so domain-specific as regards their lexical specification. This is not the case with the module that identifies clause boundaries, which is domain-independent. Thus, the domain specific module inserts clause-internal PPh boundaries before the second focussed prepositional complement to the verbs *gå upp* 'go up', *gå ner* 'go down', *falla* 'fall 'and *stiga* 'rise'.

Moreover, in order to generate PW boundaries at the ends of clauses, i.e. in order to link two or more clauses together into a PPh, it is necessary to be able to calculate the number of syllables that a given clause consists of and the number of syllables that will result after its being linked to the following clause. This information is currently being built into the prosodic parser.

Finally, in order to generate PU boundaries, one must be able to recognize discourse segment boundaries. In the present algorithm [9] these are triggered by paragraph boundaries.

ACKNOWLEDGEMENT.

This research has been supported by a grant from the Swedish HSFR/NUTEK Language Technology Programme.

REFERENCES

[1] Bruce, G. and Granström, B. (1989). "Modelling Swedish intonation in a text-to-speech system", *STL-QPSR*, pp. 31-36.

[2] Horne, M. (1994). "Generating prosodic structure for synthesis of Swedish intonation", *Working Papers* (Dept. Ling., Univ. of Lund), pp. 43, 72-75.

[3] Horne, M., Strangert, E. and Heldner, M. (1995). "Prosodic boundary strength in Swedish: Final Lengthening and Silent Interval duration", *Proc. XIIIth ICPhS*, Stockholm.

[4] Pierrehumbert, J. (1980). *The phonetics and phonology of English intonation*. Ph.D. Diss., M.I.T.

[5] Lieberman, P. 1967. *Intonation, perception and language*, Cambridge, Mass.: MIT Press.

[6] Swerts, M. (1993). "On the prosodic prediction of discourse finality", Proc. ESCA Workshop on Prosody , *Working Papers* (Dept. Ling., U. of Lund) 41, pp. 96-99.

[7] Grosz, B. and Hirschberg, J. (1992). "Some intonational characteristics of discourse structure", *Proc. ICSLP 92*. Banff, pp. 429-432.

[8] Lindström, A., Horne, M., Svensson, T., Ljungqvist, M. and Filipsson, M. (1995), "Generating prosodic structure for restricted and "unrestricted" texts", *Proc. XIIIth ICPhS*, Stockholm.

[9] Horne, M. and Filipsson, M. (1994). "Generating prosodic structure for Swedish text-to-speech", *Proc.* ICSLP 94, Yokohama, pp. 711-714.