

FOCAL ACCENTUATION AND BOUNDARY PERCEPTION

Petra Hansson
Department of Linguistics & Phonetics,
Lund University, Sweden

ABSTRACT: In this paper, an experiment concerning focal accent distribution and phrase boundary perception in South Swedish is discussed. In previous studies on prosodic phrasing in Standard Swedish, no reliable, simple relation between focal accent distribution and boundary perception has been found. However, results from the present perception experiment show that focal accent distribution has an effect on the perceived composition of an utterance in South Swedish. Since the South Swedish focal accent is characterized by a fall (not by a rise as in Standard Swedish), the focal accent gesture can be perceived as a combined focal accent gesture and boundary signal.

INTRODUCTION

In this paper, an experiment concerning focal accent distribution and phrase boundary perception in South Swedish is discussed. We do not yet fully understand the complex interaction between phrasing and accentuation. Although phrasing and accentuation are categories that are easily separated from each other in theory, they are, nevertheless, clearly interwoven in the practical situation. Deaccentuation, for example, functions both as a prominence and as a grouping signal (for a discussion, see Bruce & Granström, 1993).

Previous studies

Bruce *et al.* (1991) and Bruce *et al.* (1992) have identified several alternative boundary-signaling strategies in Standard Swedish, e.g. a combined use of F0 (boundary tones) and preboundary lengthening. However, no reliable, simple relation between focal accent distribution and boundary perception has been found.

In Standard Swedish, the focal accent gesture is characterized by an extra H after the HL gesture for (word) accent. Bruce *et al.* (1992) have found a tendency toward perceiving a sequence of a focal and a subsequent non-focal accent as belonging to the same prosodic phrase in Standard Swedish. The focal accent has also been found to function as an indirect boundary-signal in phrase-initial position, since the initial juncture can coincide with the focal rise (instead of being realized as a separate gesture).

In South Swedish, the focal accent is characterized by a fall (HL). A focal accent has a wider range than a non-focal accent, but it is not followed by a rise as in Standard Swedish. It can therefore coincide with a final juncture (L%) and possibly be perceived as a combined focal accent gesture and low boundary tone.

Focal accent distribution

Listeners' boundary perception can be investigated by asking subjects to react on syntactically ambiguous sentences such as 'When danger threatens your children call the police' (possible clause-boundaries after 'threatens' and 'children'). The purpose of the present study is to investigate if the distribution of focal accents within a syntactically ambiguous sentence affects the perception of the internal clause-boundary's location.

Focal accentuation reflects the intended information focus of an utterance. In a so-called news sentence (Lambrecht, 1994), focus covers the entire sentence and the focal accent is placed on its last content-word. However, the heavier a constituent is, the more likely it is to constitute its own prosodic phrase and to contain its own focal accent (Ladd, 1996). The test sentences used in the experiment contain two relatively long clauses with one focal accent at the end of each clause. Other content-words were assigned non-focal accents.

Hypothesis

The experiment tests if the position of the focal accent in the first clause, influences listeners' perception of the sentence-internal clause-boundary (i.e. if the boundary is perceived at different positions depending on whether 'threatens' or 'children' is focally accented in the English example above). The hypothesis is that, given the nature of the focal accent gesture in South Swedish and the

expectation to find the focal accent on the last content-word in the clause, listeners perceive the sentence-internal clause-boundary after the first focal accent (even when no other boundary-signaling cues are synthesized).

METHOD

Stimuli

Three test sentences were chosen (1-3). Two of them have been used by Bruce and colleagues in previous studies of prosodic phrasing in Swedish (1-2). Each test sentence occurs as a minimal pair, i.e. there are two different possible interpretations of each sentence (a and b). The distinctive feature between the sentences in the minimal pairs is the location of the internal clause-boundary.

1a.

Fast man 'offrade 'bonden, och 'löparen 'hotade 'kungen.
'But we sacrificed the pawn, and the bishop threatened the king'

1b.

Fast man 'offrade 'bonden och ''löparen, 'hotade 'kungen.
'Though we sacrificed the pawn and the bishop, the king threatened'

2a.

När 'pappa ''fiskar, 'stör 'Piper ''Putte.
'When dad is fishing, Piper disturbs Putte'

2b.

När 'pappa 'fiskar ''stör, 'piper ''Putte.
'When dad is fishing sturgeon, Putte whines'

3a.

Fast man konfron'terade e''leverna, och 'rektorn 'sparkade ''studievägledaren.
'But they confronted the pupils, and the principle sacked the educational advisor'

3b.

Fast man konfron'terade e'leverna och ''rektorn, 'sparkades ''studievägledaren.
'Though they confronted the pupils and the principle, the educational advisor was sacked'

Since the distribution of focal accents within a given utterance is dependent on the utterance's information structure, it is important to control for context. In order to do so, a context was created for each of the six sentences (1A-3B). Each context was put together with both the a and the b version of the test sentence it was created for. 1A was thus combined with both 1a and 1b, 1B with both 1b and 1a and so forth.

1A.

Det fanns ett par olika alternativ. Det bästa hade kanske varit att flytta tornet. (Fast man offrade bonden, och löparen hotade kungen.)
'There were a few options. The best would have been to move the castle. (But we sacrificed the pawn, and the bishop threatened the king.)'

1B.

Det fanns inga fler alternativ. (Fast man offrade bonden och löparen, hotade kungen.)
'There were no other options. (Though we sacrificed the pawn and the bishop, the king threatened.)'

2A.

Det är alltid så här. (När pappa fiskar, stör Piper Putte.) Han låter honom inte vara i fred.
'It's always like this. (When dad is fishing, Piper disturbs Putte.) He doesn't leave him alone.'

2B.

(När pappa fiskar stör, piper Putte.) Han gillar inte stör.
'(When dad is fishing sturgeon, Putte whines.) He doesn't like sturgeon'

3A.

Det lättaste hade varit att glömma alltsammans. (Fast man konfronterade eleverna, och rektorn sparkade studievägledaren.)

'It would have been easier to forget all about it. (But they confronted the pupils, and the principle sacked the educational advisor.)'

3B.

Man hade hoppats slippa sparka någon, men så blev det inte. (Fast man konfronterade eleverna och rektorn, sparkades studievägledaren.)

'One had hoped not to have to sack anyone, but it didn't work out that way. (Though they confronted the pupils and the principle, the educational advisor was sacked.)'

Generation of stimuli

The stimuli were generated using the INTRA tool (Frid, 1999). INTRA is the result of an integration of the MBROLA-based concatenative speech synthesizer LUKAS (Filipsson & Bruce, 1997) and the PSOLA resynthesis technique.

A male speaker of South Swedish recorded the sentences in 1A, 1B, 2A, 2B, 3A and 3B (i.e. the test sentences in their contexts) three times. One fluent reading of each context was chosen. Then, the stimuli were synthesized. First, the context sentences were synthesized (i.e. all the sentences in 1A-3B except those within parenthesis). In order to make the synthesized speech as natural sounding as possible, the following method was used: The recordings were segmented and phonetically and prosodically transcribed. Synthesis was then performed preserving the original phones' duration (i.e. the duration of the phones in the recordings). Using F0-rules associating prosodic labels for accents to F0 values, an intonation contour was generated that closely resembled the intonation in the recordings.

Secondly, the a and b versions of the test sentences were synthesized. They were generated in a slightly different manner than the context sentences. The recordings of the test sentences with a late clause-boundary (i.e. the b sentences) were used to adjust the duration of the phones in the synthesized versions. However, the duration of the phones in the clause-final words (i.e. *löparen*, *stör* and *rektorn*) was adjusted to their duration in the recordings of the a sentences. By doing so, we created durationally ambiguous base versions of the test sentences, i.e. base versions in which none of the two possible clause-final words were associated with final lengthening. Pauses between the two clauses occurring in the recordings were also removed.

Finally, using the F0-rules associating prosodic labels for accents to F0 values, two versions of each base version were generated. In order to create version a from the base version, focal accents were associated with *bonden*, *fiskar* and *eleverna* (i.e. the clause-final words) as well as the utterance-final words (*kungen*, *Putte* and *studievägledaren*). All other content-words were associated with non-focal accents.

In order to generate version b from the base version, focal accents were associated with *löparen*, *stör*, and *rektorn* (the clause-final words) as well as the utterance-final words (*kungen*, *Putte* and *studievägledaren*). All other content-words were associated with non-focal accents. A low boundary tone was also generated at the end of all test sentences. Observe that no actual prosodic phrase boundaries were generated, i.e. no boundary tones, final lengthening or pauses were inserted at the location of the internal clause-boundaries in the stimuli. Figure 1 shows the two versions created of the test sentence *Fast man konfronterade eleverna och rektorn sparkade(s) studievägledaren* 'But they confronted the pupils, and the principle sacked the educational advisor' / 'Though they confronted the pupils and the principle, the educational advisor was sacked'.

Finally, the synthesized test sentences were put together with the synthesized contexts in the manner described in the previous section, i.e. stimulus pairs were created by combining context 1A with both test sentence 1a and 1b, context 1B with both test sentence 1b and 1a and so forth. All in all, six stimulus pairs were generated.

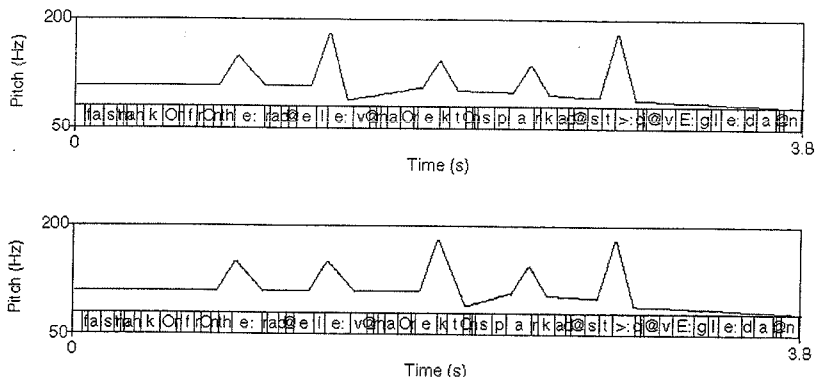


Figure 1. SAMPA transcription and F0 contours of the two versions of *Fast man konfronterade eleverna och rektorn sparkade(s) studievägledaren*. The F0 contour at the top represents 2a, the F0 contour at the bottom 2b.

Subjects and experimental procedure

Ten naïve native speakers of South Swedish were presented with 12 short texts, displayed on a computer screen (the texts in 1A-3B, each text occurring twice). They were instructed to read each text and listen to two spoken versions of it (a stimulus pair). One sentence in each text was displayed in blue. The spoken versions of the texts could be played and replayed by clicking on two buttons placed below the text on the screen. The subjects were asked to indicate for each text which of the two spoken versions they found the most natural sounding. They knew that the versions only differed with respect to the pronunciation of the blue sentence. The subjects indicated their answers on a printed answer sheet. The whole experiment lasted roughly 25 minutes. The order of presentation of the texts was randomized (but the same for all subjects).

RESULTS & DISCUSSION

The results from the perception experiment indicate that focal accent distribution has an effect on the perceived composition of an utterance in South Swedish. The subjects perceived the sentence-internal clause-boundary *after* the focally accented word in 76% of the cases (all results pooled together).

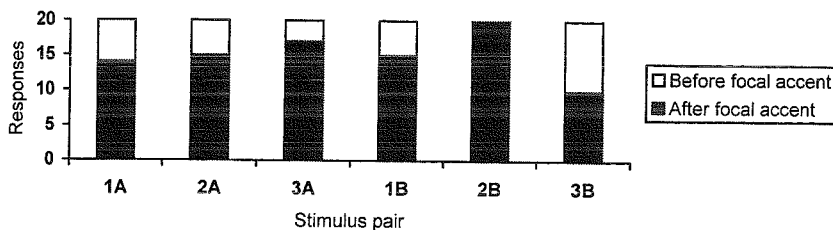


Figure 2. Subjects' responses to the different stimulus pairs (the A stimuli to the left, the B stimuli to the right). Each stimulus pair appeared twice in the test.

Figure 2 shows that the subjects demonstrated a tendency toward perceiving the internal clause-boundary after the focally accented word in all but one of the six stimulus pairs. In the 3B context, the subjects perceived the clause-boundary before the focal accent just as often as they perceived it after the focal accent. The reason for the subjects' contradictory responses to 3B is as yet unclear.

House (1990) has shown that listeners have a grouping preference or bias in the absence of phrasing cues. Subjects demonstrated a 3+2 grouping bias when asked to listen to a sequence of five fives which lacked any cues to an internal prosodic phrase boundary and determine if the sequence could best be grouped as '55-555' (2+3) or '555-55' (3+2). Such a bias is important in that it must be taken into consideration in the interpretation of the results of our perception test. However, a clear 3+2 bias can not be observed in the results of the present perception experiment, see Figure 2 above. A 3+2 grouping preference would have resulted in higher dark bars to the right than to the left in the graph.

Figure 3 shows in how many of the 12 stimulus pairs the individual subjects perceived the internal clause-boundary after the focally accented word. The mean number of stimulus pairs in which the subjects perceived the internal clause-boundary after the focal accent is 9.1.

The number of 'After focal accent' responses ranges from 7 to 12. Apparently, the subjects are not equally sensitive to focal accent distribution as a cue to prosodic phrase structure. However, observe that no subject consistently perceives the boundary *before* the focal accent.

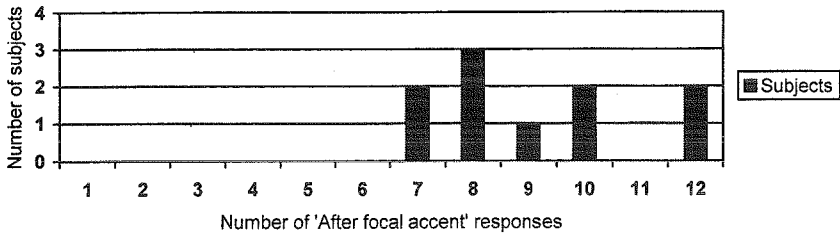


Figure 3. The number of stimulus pairs in which the individual subjects perceived the internal clause-boundary after the focal accent.

GENERAL DISCUSSION

The results from the perception experiment show that focal accent distribution has an effect on the perceived composition of an utterance in South Swedish. Subjects appeared to perceive the focal accent gesture as a combined focal accent and boundary signal. Consequently, the sequence 'focal accent + non-focal accent' does not, as in Standard Swedish, have a coherence-signaling function in South Swedish.

Cue preference

In Bruce *et al.* (1992) listeners' individual strategies or cue preferences are discussed. In their perception experiment on prosodic phrasing, subjects were able to interactively vary duration and F0 parameters by moving a point on a computer screen. Although the majority of the subjects indicated a combined use of F0 (an F0 drop to a low level) and duration (final lengthening) as cues to the location of the prosodic phrase boundary, the results also showed that individual subjects tended to have a cue preference (either F0 or duration). The difference between the subjects in the present experiment as regards how consistent they are in choosing either 'After focal accent' or 'Before focal accent' responses, may depend on their individual strategy. While e.g. final lengthening (which was lacking in the stimuli used) may be relatively unimportant for some subjects, it may be what matters most for others when determining the position of the boundary.

Context's role

In the present paper, we have argued that the relation between focal accentuation and boundary perception, observed in South Swedish but not in Standard Swedish, is possible because of different realizations of the focal accent gesture. However, the use of a context preceding and / or following the test sentence in the present perception test may also be relevant.

Bolinger, among others, has pointed out that it is methodologically unsound to try to determine how sentences are accented 'out of context' because speakers may imagine all kinds of contexts which will affect focal accent in unpredictable ways (Ladd, 1996). Following this line of argument, one could claim that (naïve) listeners are only able to make use of the focal accents' right-edge marking function, if they are given in a context in which a so-called broad focus interpretation is the only possible interpretation. Without a context, listeners may not be able to use the focal accent distribution to understand the utterance's prosodic structure.

ACKNOWLEDGEMENTS

Thanks to all the subjects who participated in the perception experiment!

REFERENCES

- Bruce, G. & Granström, B. (1993) "Prosodic modelling in Swedish speech synthesis", *Speech Communication* 13, 63-73.
- Bruce, G., Granström, B., Gustafson, K. & House, D. (1992) "Interaction of F0 and duration in the perception of prosodic phrasing in Swedish", *Nordic prosody VI. Papers from a symposium, Stockholm, August 12-14, 1992*, 7-22.
- Bruce, G., Granström, B. & House, D. (1991) "Prosodic phrasing in Swedish speech synthesis". In G. Bailly & C. Benoit (eds.), *Talking Machines: Theories, Models & Applications*, 113-125, (Elsevier Publisher).
- Filipsson, M. & Bruce, G. (1997) "LUKAS - a preliminary report on a new Swedish speech synthesis". *Working Papers* 46, 45-56.
- Frid, J. (1999) "An environment for testing prosodic and phonetic transcriptions". *Proceedings of The XVIth International Congress of Phonetic Sciences. San Francisco, 1-7 August 1999*, 2319-2322.
- House, D (1990) *Tonal Perception in Speech. Travaux de l'institut de linguistique de Lund* 24. (Lund University Press: Lund).
- Ladd, R. (1996) *Intonational Phonology. Cambridge Studies in Linguistics* 79. (Cambridge University Press: Cambridge).
- Lambrecht, K. (1994) *Information structure and sentence form. Topic, focus, and the mental representation of discourse referents. Cambridge Studies in Linguistics* 71. (Cambridge University Press: Cambridge).