

Swedish word accents in a ‘confirmation’ context

Gilbert Ambrazaitis

Linguistics and Phonetics, Centre for Languages and Literature, Lund University

Abstract

An exploratory study on the prosodic signaling of ‘confirmation’ in Swedish is presented. Pairs of subjects read short dialogs, constructed around selected target words, in a conversational style. The target utterances were produced with a rising-falling intonation, lacking any typical ‘focal accent’ (FA). Qualitative observations and acoustic measurements reveal that the signaling of the word accent contrast appears to be, to a certain degree, optional in a confirmation context. The results support the view that no tonal target needs to be assumed for accent I, and further suggest that utterance-level prominence can be realized by other means than the FA.

Introduction

An utterance of the type *ja, det var med bilen* (“Yes, (it was) by (the) car.”), may occur as a response in at least two different contexts:

(a) *Hur skulle vi egentligen åka till Helsingborg imorgon? Det var väl med bilen, eller?* (“How are we getting to Helsingborg tomorrow? We were gonna go by car, weren’t we?”), or:

(b) *Hur skulle vi egentligen åka till Helsingborg imorgon? Minns du det?* (“How are we getting to Helsingborg tomorrow? Do you remember that?”).

In (a), the speaker uttering the response confirms the truth of a given utterance, while in (b), s/he makes a new assertion. The point of departure for this study is the question of whether and how this contrast is signaled prosodically in Standard Swedish, defined as the dialect spoken roughly in the Stockholm area (*Sveamål*).

In Swedish, prosody has important functions both on the word and on the utterance level. First, Swedish has a lexical pitch accent contrast, which in the Lund account of Swedish intonation (e.g. Bruce, 1977, 2005; Bruce and Granström, 1993) is modeled in terms of the accentual pitch fall timing: In accent I the fall is ‘early’, i.e. there is a ‘high-low’ transition from the pre-stress to the lexically stressed syllable (H+L*), while in accent II it is ‘late’, i.e. the HL transition starts on the stressed syllable (H*+L).

Second, the Lund model relates prosody with two functions on the utterance level: signaling phrasing, and signaling focus. A ‘focal accent’ (FA) has been recognized (Bruce, 1977), as well as signals for coherence and boundaries (e.g. Bruce and Granström, 1993). The FA is modeled as a rising pitch gesture (H) following the word

accent gesture. It is assumed that every prosodic word is associated with a word accent (H+L* or H*+L), while only the words in ‘focus’ additionally receive a FA, resulting in a complex pitch accent (H+L* H, or H*+L H, respectively). In a broad focus condition, the last word in an utterance receives the FA (Bruce, 1977).

The difference between the dialog scenarios (a) and (b) above may be seen as a matter of the information status of the word *bilen* (“the car”), since it is ‘given’ in (a), and ‘new’ in (b), and hence, as a question of ‘focus’. In this case, an obvious hypothesis would be that a (word accent plus) FA occurs on the word *bilen* in (b), and a ‘non-focal word accent’ in (a). However, in the dialog scenario (a) above the hypothesis would predict an utterance without any FA, a situation which is not captured by the Lund model.

Both contexts (a) and (b) are currently being investigated, and the preliminary findings for (b) contexts are quite clear, since, as predicted, a FA generally occurs on the target word (*bilen* in the example). The data for (a), however, are more complex. In Ambrazaitis (2007) an exploratory study on the prosodic signaling of ‘confirmation’ in Swedish is presented, where the intonation contours of 89 responses (based on 9 speakers) in a ‘confirmation’ context were qualitatively classified. All 89 utterances differed prosodically clearly from responses in a ‘new-information’ context, since none of them were produced with a (typical) FA.

However, it was argued that the results do not confirm the simple hypothesis that the final ‘given’ word in a confirmation would be associated with a ‘non-focal word accent’: First, the falling intonation that was found in connection with the target word did not occur as a local pitch

movement, but rather as the final part of an overall rising-falling *utterance* intonation, resembling a ‘hat pattern’ as it has been described in Dutch (‘t Hart et al., 1990) and German (Kohler, 1991). Second, if the fall were merely a component of the word accent, then we would expect two clearly distinguishable gestures, one for accent I, one for accent II. However, in only 58.9% of the accent II cases, a predicted prototypically late and high starting accent II fall was observed, and 20.5% of the accent II cases exhibited an accent-I-like pattern, i.e. a low pitch level from the onset of the stressed vowel.

That is, in a ‘confirmation’ context, a falling *utterance* intonation was found in connection with the target word, and the signaling of the word accent contrast appeared to be, to a certain degree, optional. This paper presents a continuation of the Ambrazaitis (2007) study. In particular, the aim is to supplement the qualitative observations made so far by acoustic measurements.

Method

Short, constructed dialogs were read by pairs of subjects in a conversational style. This method is adopted from Kohler and Niebuhr (2007) and has two major advantages: A near-spontaneous speaking style is approached, at the same time providing highly controlled material.

Material

All dialogs had the following general structure:

A: <context-question(s)>

B: *Ja, det/den är/var* <target-phrase>.

(“Yes, it is/was <target-phrase>”)

The content of <context-question(s)> was varied systematically in order to elicit the two communicative situations of (a) ‘confirming given’ or (b) ‘introducing new’ information. The <target-phrase> consisted of a disyllabic target word with lexical stress on the first syllable, preceded by a monosyllabic function word. Ten different target words were used, five with accent I, five with accent II, cf. Table 1. The accent I class contained five nouns in definite singular form, while the accent II class contained the same words in indefinite plural form.

Table 1: The 10 target words. See text.

accent I	accent II
<i>bilen</i> /'bi:lən/ the car	<i>bilar</i> /'bi:lar/ cars
<i>boven</i> /'bu:vən/ the criminal	<i>bovar</i> /'bu:var/ criminals
<i>stigen</i> /'sti:gən/ the path	<i>stigar</i> /'sti:gar/ paths
<i>stolen</i> /'stu:lən/ the chair	<i>stolar</i> /'stu:lar/ chairs
<i>kniven</i> /'kni:vən/ the knife	<i>knivar</i> /'kni:var/ knives

A phonetic and a semantic criterion were applied simultaneously for the composition of the corpus. First, microprosodic effects were largely controlled by choosing words (i) with closed (stressed) vowels only, and (ii) such that initial-consonant perturbations should be counterbalanced in the corpus (cf. the consonants preceding the stressed vowel). Second, the chosen words are rather common, i.e., they can be expected to occur frequently in every-day conversation (*bov* “criminal” being perhaps an exception). The constituent <context-question(s)> was designed individually for each target word, in order to provide a situational context that was as natural as possible. Two examples have been presented already in the introduction (a & b).

Subjects and Procedure

Nine native speakers of Standard Swedish, 5 female and 4 male, aged 22-50, took part in the recordings. The speakers have slightly different regional backgrounds, but all can be classified as belonging to the same prosodic dialect type EAST as defined in Bruce (2005). The subjects were recorded in pairs, each one sitting in a sound-treated experimental studio, communicating via the recording microphones (*Shure BG 4.0*) and headphones. A 10th (female) speaker from another dialect area took part in one of the recordings, but was not included in the analysis. The recording equipment and the investigator were located in a separate room. The recordings were made digitally at 44.1 kHz and 24 bit.

The speakers received the dialogs printed on paper and were instructed to read them in a conversational style, however, without being too theatrical. They were encouraged to discuss their readings and, if necessary, to repeat any dialog until they were satisfied. Generally, this self-monitoring procedure worked successfully.

The 20 test dialogs (10 target words; 2 situations) were randomized and mixed with 43 other dialogs (not reported on here), yielding a corpus of 63 dialogs in total. The dialogs were arranged such that each speaker would read the A-part in every second dialog. The speakers read the whole corpus twice, with interchanged parts on the second run, such that effectively, each speaker read the whole corpus once. One dialog session consisting of instructions, the two runs, and a break in between took approximately 1h 15min.

Corpus

Recall that only the confirmation context is discussed in this paper. The total number of utterances that were analyzed in Ambrazaitis (2007)

for this context was 89 (9 speakers \times 10 utterances – 1 missing). The data inspection reported there revealed that in some cases, a confirmation was signaled by an intonation contour containing a ‘reduced focal accent’. These cases have to be treated separately, since they include another prosodic strategy for the expression of confirmation than the remaining cases, which are characterized by what can be described as a complete lack of a FA (cf. Introduction). Therefore, the clearest cases of utterances with a ‘reduced FA’ were excluded from the measurements in the present paper.

For the remaining 77 utterances by 8 speakers the measurements defined in the next subsection were carried out as far as possible. Due to the frequent occurrence of creaky voice, however, several values could not be determined.

Measurements

In order to get an impression of the pitch change from the pre-stress syllable (e.g. *med* in *med bilen*) to the stressed vowel (e.g. *i* in *bilen*), as well as during that vowel, four approximated pitch values were obtained in each utterance, defined as the following mean f_0 variables. All values were expressed in semitones (re 100Hz).

$\overline{f_{0pre}}$: mean f_0 over the last 30ms of the sonorant portion of the pre-stress syllable

$\overline{f_{0V_1}}, \overline{f_{0V_2}}, \overline{f_{0V_3}}$: mean f_0 over the first, second, and third 33.3% of the stressed vowel

Results and discussion

A number of measures were derived from the variables defined in the preceding section, two of which are presented in this paper. In a first step, we should look at those properties of the pitch contour that, on the basis of the Lund model,

should be most relevant for the realization of the word accent distinction: In a ‘non-focal’ condition, we should expect to find a high-low transition from the pre-stress to the stressed syllable for accent I, but not for accent II. Instead, for accent II, the vowel onset should be high, and a fall is expected throughout the vowel.

The difference measure $\overline{f_{0pre}} - \overline{f_{0V_1}}$ describes the transition from the pre-stress to the onset of the stressed syllable. Table 2 displays the mean values for this measure per speaker and word accent condition. Note that the maximum sample size per condition is $n = 5$ (5 target words per condition), and that the actual sample size is often lower due to missing values (creaky voice, see above).

All mean values for $\overline{f_{0pre}} - \overline{f_{0V_1}}$ are positive, regardless of the word accent. That is, even for accent II, there is a pitch drop from the pre-stress to the stressed syllable. This finding supports the previous observation that the utterances in the confirmation context were produced with a hat-pattern-like utterance intonation with the fall at the right edge of the hat associated with the stressed syllable of the target word.

Nevertheless, for speakers S2L, S5L, S5R, and to a smaller extent for S1L and S3L, there is a salient difference between the word accents: As would be predicted by the Lund model, the step from the pre-stress to the stressed syllable is larger for accent I than for accent II; and for some speakers (S2L, S5L, S3L, S5R), it is in fact below 1 semitone (*st*) for accent II, indicating that the stressed vowel has a relatively high-pitched onset. However, for two speakers (S1R, S4R) there seems to be no difference in pitch drop between accent I and accent II. For speaker S4L, the data for accent I are insufficient, and those for accent II are very variable (cf. $s = 5.76st$). In fact, for S4L the raw values range from 0.19*st* to 13.29*st*.

In order to test whether (only) accent II

Table 2: Means (\bar{x}) and standard deviations (s) for $\overline{f_{0pre}} - \overline{f_{0V_1}}$ in semitones, and sample sizes (n) for female (f) and male (m) speakers.

speaker (sex)	accent I			accent II		
	\bar{x}	s	n	\bar{x}	s	n
S1L (f)	4.94	1.92	5	2.15	1.37	5
S2L (f)	9.51	1.23	5	0.80	1.11	5
S4L (f)	(6.56)	.	1	5.63	5.76	4
S5L (f)	9.02	5.13	5	0.61	2.30	5
S1R (m)	1.31	0.95	5	2.00	1.69	5
S3L (m)	2.63	1.23	4	0.87	1.55	4
S4R (m)	1.96	2.15	3	2.07	0.92	3
S5R (m)	4.59	1.74	4	0.27	1.18	5

Table 3: Means (\bar{x}) and standard deviations (s) for $\overline{f_{0V_1}} - \overline{f_{0V_3}}$ in semitones, and sample sizes (n) for female (f) and male (m) speakers.

speaker (sex)	accent I			accent II		
	\bar{x}	s	n	\bar{x}	s	n
S1L (f)	1.30	0.72	4	3.86	1.15	5
S2L (f)	1.49	0.71	5	8.07	1.38	5
S4L (f)	.	.	0	2.64	1.96	4
S5L (f)	0.94	1.54	5	7.15	1.51	5
S1R (m)	1.89	0.41	4	2.49	0.81	5
S3L (m)	1.80	0.48	4	2.67	0.78	4
S4R (m)	(2.30)	.	1	(2.30)	.	1
S5R (m)	1.34	0.68	4	5.67	1.92	5

exhibits the expected fall through the stressed vowel, the fall range within the stressed vowel was calculated by $f_0V_1 - f_0V_3$; the results are displayed in Table 3. Generally, f_0 is slightly falling during the vowel even for accent I, but (for the speakers with no or few missing values) the mean fall range is larger for accent II than for accent I. However, this mean difference is rather small for some speakers, indicating that there are a number of accent II cases with accent-I-like patterns, in line with the earlier observations.

It should be added that the cases that were classified as having an ‘‘accent-I-like pattern’’ in Ambrazaitis (2007) in the first place received this judgment due to their perceptually low vowel. In a large proportion of these cases, this low pitch was due to, or accompanied by, creaky voice. Since cases with creaky voice have not been taken into account in the present measurements, the results weaken the earlier observations that the word accent contrast was – to a certain degree – neutralized.

Conclusions and outlook

Previous work on Swedish prosody has concentrated on a limited number of speech act conditions, i.e. first of all assertions and to some extent questions. The ongoing research, a part of which has been presented in this paper, concentrates on the expression of another speech act, namely ‘confirmation’, and the phonetic manifestation of the word accents in this condition. The results presented in Ambrazaitis (2007), now supplemented by acoustic measurements, have implications for the understanding of Swedish word and utterance prosody, as well as their interplay.

First, in the Lund model, the FA is the only means of realizing utterance-level prominence. However, what has become clear is that an utterance without any FA – a phenomenon that has not been treated systematically yet – can be consistently elicited by means of a confirmation context. Note that the target word still received strong intonational prominence due to the falling utterance intonation. That is, the present findings suggest that utterance-level prominence can be realized by other means than the FA.

Second, in the Lund model, a tonal target is assumed both for accent I and accent II. However, for the present data involving confirmations, there is no need to assume such a target for accent I, since the pitch step from the pre-stress to the stressed syllable can be argued to be part of the utterance prosody. The word accent contrast may still be encoded, but for that, it is sufficient to assume a local adjustment in the timing of the

falling utterance intonation for accent II, or possibly a high tonal target on the stressed vowel. This analysis is in line with, and extends the arguments by e.g. Engstrand (1995) and Riad (1998).

Third, in a confirmation context, the signaling of the word accent appears to be optional to a certain degree. This can perhaps be expected, considering the low informational load of the final (given) word in the confirming response – if uttered at all in an authentic situation. A conditioning factor might be the speaking style, influenced by, e.g., the communicative situation and its needs or simply individual speaker habits.

In general, it might be expected that a number of utterance-level mechanisms will be found, as soon as different speech act conditions are explored. More knowledge on Swedish utterance prosody is required before a complete understanding of the word accents and their phonetic manifestation can emerge.

Of course, more research is needed in order to confirm the conclusions drawn so far. In particular, more acoustic measurements on larger corpora, as well as perceptual experiments must be undertaken in order to investigate to what extent the word accent contrast is in fact neutralized, or maintained, in the expression of confirmation. So far, only one acoustic dimension (f_0) has been taken into account, but other dimensions should be included as well, such as duration and voice quality (e.g. Kohler and Niebuhr, 2007).

References

- Ambrazaitis G (2007). Expressing ‘confirmation’ in Swedish: the interplay of word and utterance prosody. Submitted.
- Bruce G (1977). *Swedish Word Accents in Sentence Perspective*. Lund: Gleerup.
- Bruce G (2005). Intonational prominence in varieties of Swedish revisited. In S A Jun, ed., *Prosodic Typology. The Phonology of Intonation and Phrasing*, 410–429. Oxford: Oxford University Press.
- Bruce G and Granström B (1993). Prosodic modeling in Swedish speech synthesis. *Speech Communication*, 13:63–73.
- Engstrand O (1995). Phonetic interpretation of the word accent contrast in Swedish. *Phonetica*, 52:171–179.
- Kohler K (1991). A model of German intonation. *AIPUK*, 25:295–360.
- Kohler K and Niebuhr O (2007). The phonetics of emphasis. In *Proc. ICPHS XVI*. Submitted.
- Riad T (1998). Towards a Scandinavian accent typology. In W Kehrein and R Wiese, eds., *Phonology and Morphology of the Germanic Languages*, 77–109. Tübingen: Niemeyer.
- ’t Hart J, Collier R and Cohen A (1990). *A Perceptual Study of Intonation. An Experimental-phonetic Approach to Speech Melody*. Cambridge: Cambridge University Press.