LUND UNIVERSITY

Expressing ‘confirmation’ in Swedish: the interplay of word and utterance prosody

Ambrazaitis, Gilbert

Published in:
Proceedings of the 16th International Congress of Phonetic Sciences

2007

Link to publication

Citation for published version (APA):
Ambrazaitis, G. (2007). Expressing ‘confirmation’ in Swedish: the interplay of word and utterance prosody. In J. Trouvain, & W. J. Barry (Eds.), *Proceedings of the 16th International Congress of Phonetic Sciences* (pp. 1093-1096). Universität des Saarlandes. http://www.icphs2007.de/conference/Papers/1442/

Total number of authors:
1

General rights
Unless other specific re-use rights are stated the following general rights apply:
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.
• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: https://creativecommons.org/licenses/

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

EXPRESSING 'CONFIRMATION' IN SWEDISH: THE INTERPLAY OF WORD AND UTTERANCE PROSODY

Gilbert Ambrazaitis

Linguistics and Phonetics, Centre for Languages and Literature, Lund University, Sweden
Gilbert.Ambrazaitis@ling.lu.se

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. A falling utterance intonation was found on the target word, and the signaling of word prosody (lexical pitch accent) appeared to be, to a certain degree, optional.

1. 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' a piece of 'given' information while in (b) s/he makes an assertion and introduces 'new' information. The point of departure for this study is the question of if and how this contrast is signaled prosodically in Standard Swedish, defined as the dialect spoken roughly in the Stockholm area (*Sveamål*/ East Swedish).

In Swedish, prosody has important functions both on the word and on the utterance level. As for word prosody, Swedish has a lexical pitch accent contrast, which in the Lund account of Swedish intonation (e.g., [2], [3], [4]) is modeled in terms of the accentual pitch fall timing: In accent I it 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).

As for the utterance level, the Lund model relates prosody with two functions: signaling phrasing and signaling focus in assertive utterances. A 'focal accent' (FA) has been recognized [2], as well as signals for coherence and boundaries [4]. 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 narrow focus additionally receive a FA, resulting in a complex, two-peaked pitch accent (H+L* H, or H*+L H, respectively). In a broad focus condition, the last word in a phrase receives the FA [2].

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 scenario (a) the hypothesis would predict an utterance without any FA, a situation that is not captured by the Lund model.

Both contexts (a) and (b) are currently investigated, and the preliminary findings for (b) 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. Therefore, this paper will focus on the expression of 'confirmation' in Swedish. An exploratory study is presented, and the results are discussed with reference to the background on Swedish word and utterance intonation as summarized above.

2. METHOD

Short, constructed dialogs were read by pairs of subjects in a conversational style. This method is adopted from [8] and has two major advantages: A near-spontaneous speaking style is approached, at the same time providing highly controlled material.

2.1. Material

All dialogs were of 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 scenarios of (a) 'confirmation' and (b) 'new information'. The <target-phrase> consisted of a target word preceded

Table 1: The 10 target words. See text.

accent I	accent II
bilen /ˈbiːlɛn/ the car	*bilar* /ˈbiːlar/ cars
boven /ˈbuːvɛn/ the criminal	*bovar* /ˈbuːvar/ criminals
stigen /ˈstiːgɛn/ the path	*stigar* /ˈstiːgar/ paths
stolen /ˈstuːlɛn/ the chair	*stolar* /ˈstuːlar/ chairs
kniven /ˈkniːvɛn/ the knife	*knivar* /ˈkniːvar/ knifes

by a monosyllabic function word. Twenty different target words were used, 10 of which are reported on in this paper, cf. Table 1. These were disyllabic words with lexical stress on the first syllable, grouped into two classes: Class I contained 5 nouns in definite singular form, ending in the suffix *-en*, while class II contained the same words in indefinite plural form, ending in the suffix *-ar*. The words of class I have accent I while the words of class II have accent II. That is, the corpus consists of 5 near-minimal pairs concerning word accent. Among the 10 words not reported on in this paper are the monosyllabic root forms of the I/II-words.

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 counter-balanced 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 everyday conversation (*bov* 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).

2.2. Subjects and Procedure

Nine native speakers of Standard Swedish, 5 female and 4 male, aged 22-50, were included in this study. The speakers have slightly different regional backgrounds, but all can be classified as belonging to the same prosodic dialect type EAST as defined in [3]. 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; it was hardly necessary for the investigator to interrupt the subjects.

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.

2.3. Data inspection

The data analysis consisted of (i) qualitatively judging and classifying the intonation patterns of the B-responses and (ii) counting the occurrences of the patterns found under (i). For step (i) the recordings were carefully inspected both auditorily and visually in [9], using mainly spectrograms and F0 curves. As mentioned in the introduction, only the confirmation context is presented in this paper. The total number of utterances that were analyzed in this context is 89 (9 speakers $\times$ 10 utterances $-$ 1 missing due to a technical problem). The judgment and classification of the intonation patterns was not a trivial task, e.g., due to the occurrence of creaky voice. Therefore, the frequencies for certain classes given in the next section must be regarded as approximate.

3. RESULTS

All 89 B-responses in the confirmation context sound convincingly confirming, and they differ prosodically clearly from the corresponding responses in the new-information context. As mentioned in the introduction, the latter are generally produced with a prototypical FA, as exemplified in Fig. 1. No single confirming response was produced with such a pattern.

Instead, the confirming responses were generally produced with an overall rising-falling intonation, with the initial *(ja)* and the final (target) word receiving intonational prominence associated with the rising or the falling pitch movement, respectively. Two typical examples are given in Fig. 2. In general, this global intonation reminds of a ‘hat pattern’ as it has been found for Dutch [6] and German [7].

Several variations of the overall confirmation pattern occurred in the data. First, a short pause be-

Figure 1: F0 courses and SAMPA transcriptions of 2 examples (female speaker S5L) signaling 'new information'; FA on final word; additional accent on *var* [vA:] "was". **Left:** *bilen* (accent I) "the car". **Right:** *bilar* (accent II) "cars". **Audio**: cf. audio file 1 and 2.

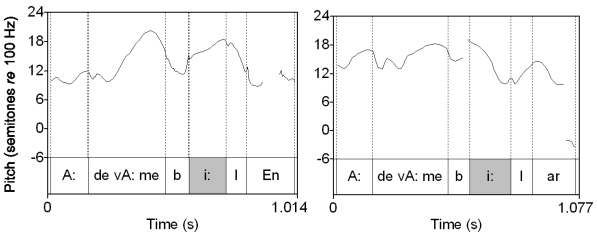


Figure 2: F0 courses and SAMPA transcriptions of 2 examples signaling 'confirmation' where pitch is falling during stressed-syllable vowel in accent II. **Left:** *bovar* "criminals" by male speaker S1R; pitch is falling from medium level. **Right:** *bilar* "cars" by female speaker S2L; pitch is falling from high. **Audio**: cf. audio file 3 and 4.

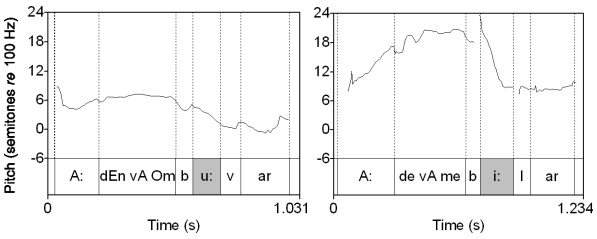
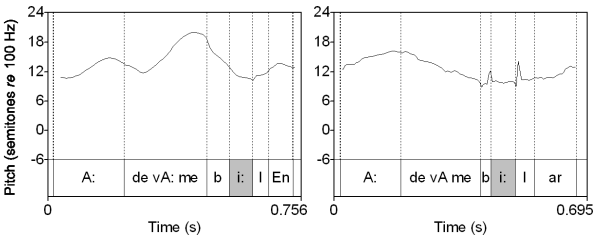


Figure 3: F0 courses and SAMPA transcriptions of 2 examples signaling 'confirmation'. **Left:** *bilen* (accent I) "the car" by female speaker S1L; additional accent on *var* [vA:] "was". **Right:** *bilar* (accent II) "cars" by female speaker S4L; pitch is low in stressed-syllable vowel, as for accent I. **Audio**: cf. audio file 5 and 6.



tween *ja* and the rest of the utterance was occasionally inserted by 5 speakers, in 19 cases in total. Second, the concatenation pattern between the two accents was variable: Often, a high plateau-like contour was found, which was usually drifting, either upwards or downwards. In several cases, however, it is more appropriate to speak of a simple rise (on *ja*) plus fall throughout the utterance, cf. Fig. 3 (right). Third, in approx. 20 cases (of 89), a function word from the intermediate part of the utterance (mostly *var* "was") received intonational prominence, as indicated by a separate pitch movement. A global, hat-pattern like contour was also found in these cases, but often rather connecting *var* and the target word, excluding the initial *ja*, cf. Fig. 3 (left).

Finally, some characteristics of the final pitch fall varied due to the word accent of the target word. For accent I a low level is usually reached already at the onset of the stressed vowel, rendering a flat, low pitch during that vowel, cf. Fig. 3 (left). For accent II pitch is often falling during the stressed vowel, either (i) from a high level at vowel onset (23 of 45 cases), cf. Fig. 2 (right), or (ii) from a level that is considerably lower than the high-plateau level (6 cases), cf. Fig. 2 (left). However, pitch may also (iii) be low and flat from the vowel onset, i.e., very like or even indistinguishable from the usual accent I pattern (8 cases), cf. Fig. 3 (right). Two cases were classified as problematic (either class ii or iii), and 6 cases were not relevant since they were produced with a FA (see below).

This general hat pattern – with an additional accent on a function word in approx. 10-50% of the cases per speaker – occurred for 7 speakers (4 female, 3 male). The 8th speaker (male) never produced an accent on a function word. The 9th speaker (female) preferred a different strategy for signaling confirmation, which could be referred to as a 'reduced FA'. That is, typical characteristics of a FA were present, e.g., the late pitch peak on the post-stress syllable for accent II, but the height of the accentual peaks was rather on a medium than a high level; the pre-stress pitch level was usually considerably higher. The 8th speaker occasionally used this strategy, too.

4. DISCUSSION

The simple hypothesis was that a confirmation would be signaled by a 'non-focal word accent'. Since in the Lund model the FA is the only means of accenting words on the utterance level, the hypothesis implicitly predicts that no utterance-level pitch accent occurs in the confirmation context, but rather a 'pure' word accent on the target word, i.e., a lexically specified pitch fall, early for accent I, late for accent II. At first sight, the hypothesis appears to be confirmed since no FA was produced by 7 of the 9 speakers. Instead, falling pitch movements on the target words were found, which could be interpreted as instances of the predicted lexical pitch accents.

However, the data may also be interpreted in another way, namely that the observed falling pitch movement on the target word is in the first place part of the utterance prosody, rather than lexically deter-

mined. The present data yield at least two preliminary arguments for this view:

1. The fall does not occur as a local pitch movement associated with the target word but constitutes the final part of a global, hat-pattern like intonation, which is likely to be an utterance-level phenomenon.

2. If the fall were merely a means of signaling word accent, then we would expect two clearly distinguishable gestures, one for accent I, one for accent II. However, in 20.5% of the 39 relevant cases, the fall for accent II appeared to be indistinguishable from the fall in the corresponding accent I word (cf. (iii) in 3.), and in only 59.0% of the cases, a predicted prototypically late and high starting accent II fall was observed (cf. (i) in 3.).

Furthermore, consider how other Germanic languages that lack any lexical use of pitch would express a confirmation in a comparable context. For example German: *Was wollten wir ihnen nochmal zur Hochzeit schenken, das war doch ein Messer, oder? – **Ja**, das war ein **Messer***. ("What were we gonna buy them for their wedding, it was a knife, wasn't it? – Yes, it was a knife.") A possible intonation here would be a hat pattern connecting *ja* and *Messer*, with an 'early peak' [7] at the right edge, i.e., basically the same pattern as the one observed here for Swedish. It would appear inappropriate to classify a phonetic expression that is used in the same function in two closely related languages as a matter of utterance prosody in the one, and as a matter of word prosody in the other language.

The alternative analysis of the falling pitch movement outlined here has implications for the understanding of Swedish word and utterance prosody, as well as their interplay. First, 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.[1] Note that the target word still received 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 classical rising 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 fall 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. This analysis is in line with and extends the arguments by, e.g., [5] and [10]. 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 confirmation context and the low functional load of the Swedish word accent contrast in general.

Of course, more research is needed in order to confirm the preliminary conclusions drawn in this paper. In particular, in order to support argument 1 above, more varied material should be investigated. In order to test argument 2, detailed measurements, as well as perceptual experiments are needed, which investigate to what extent the word accent contrast in fact is neutralized, or maintained, in the expression of confirmation. This work has been initiated in [1] where acoustic measurements are presented that support the finding of a style- or speaker-dependent word accent neutralization. So far, only one acoustic dimension (F0) has been taken into account, but others should be included as well, such as duration, voice quality, or acoustic energy [8].

Finally, more studies are needed that concentrate on utterance functions, their phonetic exponents, and the interplay of word and utterance prosody.

5. REFERENCES

[1] Ambrazaitis, G. 2007. Swedish word accents in a 'confirmation' context. *TMH-QPSR* 50 (*Proc. FONETIK 2007* Stockholm), 49–52.

[2] Bruce, G. 1977. *Swedish Word Accents in Sentence Perspective*. Lund: Gleerup.

[3] Bruce, G. 2005. Intonational prominence in varieties of Swedish revisited. In: Jun, S.-A., (ed), *Prosodic Typology. The Phonology of Intonation and Phrasing*. Oxford: OUP 410–429.

[4] Bruce, G., Granström, B. 1993. Prosodic modeling in Swedish speech synthesis. *Speech Commun.* 13, 63–73.

[5] Engstrand, O. 1995. Phonetic interpretation of the word accent contrast in Swedish. *Phonetica* 52, 171–179.

[6] 't Hart, J., Collier, R., Cohen, A. 1990. *A Perceptual Study of Intonation. An Experimental-phonetic Approach to Speech Melody*. Cambridge: CUP.

[7] Kohler, K. 1991. A model of German intonation. *AIPUK* 25, 295–360.

[8] Kohler, K., Niebuhr, O. 2007. The phonetics of emphasis. *Proc. 16th ICPhS* Saarbrücken.

[9] Praat [computer program]. http://www.praat.org.

[10] Riad, T. 1998. Towards a Scandinavian accent typology. In: Kehrein, W., Wiese, R., (eds), *Phonology and Morphology of the Germanic Languages*. Tübingen: Niemeyer 77–109.

[1] It is not entirely clear how the initial pitch rise on *ja* should be treated. For several (both functional and phonetic) reasons, however, it cannot be classified as an instance of the FA in the sense of the Lund model.