

Focal lengthening in assertions and confirmations

Gilbert Ambrazaitis

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

Abstract

This paper reports on duration measurements in a corpus of 270 utterances by 9 Standard Swedish speakers, where focus position is varied systematically in two different speech acts: assertions and confirmations. The goal is to provide information needed for the construction of a perception experiment, which will test the hypothesis that Swedish has a paradigmatic contrast between a rising and a falling utterance-level accent, which are both capable of signalling focus, the falling one being expected in confirmations. The results of the present study are in line with this hypothesis, since they show that focal lengthening occurs in both assertions and confirmations, even if the target word is produced with a falling pattern.

Introduction

This paper is concerned with temporal aspects of focus signalling in different types of speech acts – assertions and confirmations – in Standard Swedish. According to Büring (2007), most definitions of *focus* have been based on either of two ‘intuitions’: first, ‘new material is focussed, given material is not’, second, ‘the material in the answer that corresponds to the wh-constituent in the (constituent) question is focussed’ (henceforth, ‘Question-Answer’ definition). In many cases, first of all in studies treating focus in *assertions*, there is no contradiction between the two definitions; examples for usages of *focus* that are compatible with both definitions are Bruce (1977), Heldner and Strangert (2001), or Ladd (2008), where focus is defined, more or less explicitly, with reference to ‘new information’, while a question-answer paradigm is used to elicit or diagnose focus. In this study, focus is basically understood in the same sense as in, e.g. Ladd (2008). However, reference to the notion of ‘newness’ in defining focus is avoided, since it might seem inappropriate to speak of ‘new information’ in *confirmations*. Instead, the ‘Question-Answer’ definition is adopted, however, in a generalised form not restricted to wh-questions. *Focus signalling* or *focussing* is then understood as a ‘highlighting’ of the constituent in focus. Focus can refer to constituents of different size (e.g. individual words or entire

phrases), and signalled by different, e.g. morphosyntactic, means, but only *narrow focus* (i.e. focus on individual words) as signalled by *prosodic* means is of interest for this paper.

For Swedish, Bruce (1977) demonstrated that focus is signalled by a *focal accent* – a tonal rise that follows the word accent gesture. In the Lund model of Swedish intonation (e.g. Bruce et al., 2000) it is assumed that focal accent may be present or absent in a word, but there is no paradigmatic contrast of different focal accents. However, the Lund model is primarily based on the investigation of a certain type of speech act, namely assertions (Bruce, 1977). This paper is part of an attempt to systematically include further speech acts in the investigation of Swedish intonation.

In Ambrazaitis (2007), it was shown that *confirmations* may be produced without a rising focal accent (H-). It was argued, however, that the fall found in confirmations not merely reflects a ‘non-focal’ accent, but rather an utterance-level prominence, which paradigmatically contrasts with a H-. Therefore, in Ambrazaitis (in press), it is explored if and how *focus* can be signalled prosodically in confirmations. To this end, the test sentence “Wallander förlänger till november.” (*Wallander is continuing until November.*) was elicited both as an assertion and as a confirmation, with focus either on the initial, medial, or final content word. An example for a context question eliciting final focus in a confirmation is *‘Until when is Wallander continuing, actually? Until November, right?’*.

As a major result, one strategy of signalling a confirmation was by means of a *lowered H-* rise on the target word. However, another strategy was, like in Ambrazaitis (2007), to realise the target word with a *lack* of a H- rise, i.e. with falling F0 pattern (cf. Figure 1, upper panel). The initial word was always produced with a rise, irrespective of whether the initial word itself was in focus or not. Initial, pre-focal rises have been widely observed in Swedish and received different interpretations (e.g. Horne, 1991; Myrberg, in press; Roll et al., 2009). For the present paper, it is sufficient to note that an initial rise is not necessarily associated with focus.

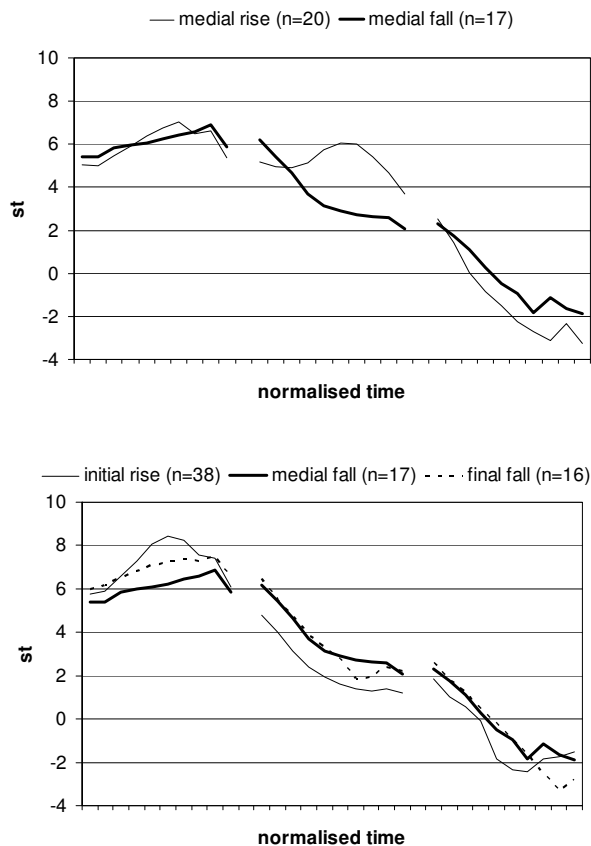


Figure 1. Mean F0 contours of the three content words in the test sentence “Wallander förlänger till november”; breaks in the curves symbolise word boundaries; time is normalised (10 measurements per word); semitones refer to an approximation of individual speakers’ base F0; adapted from Ambrazaitis (in press). Upper panel: two strategies of focus signalling on the medial word in a confirmation. Lower panel: Focus on the initial, medial, and final word in a confirmation; for medial and final focus, only the falling strategy is shown.

That is, in confirmations with intended focus on the medial or the final word, one strategy was to produce a (non-focal) rise on the initial, and two falling movements, one each on the medial and the final word. As the lower panel in Figure 1 shows, the mean curves of these two cases look very similar; moreover, they look similar to the pattern for initial focus, which was always produced with a rising focal accent. One possible reason for this similarity could be that medial or final focus, in fact, were not marked at all in these confirmations, i.e. that the entire utterance would be perceived as lacking any narrow focus. Another possibility is that all patterns displayed in Figure 1 (lower panel) would be perceived with a focal accent on the initial word. Informal listening, however, indicates that in many cases, an utterance-level prominence, indicating focus, can be perceived on the medial

or the final word. Thus, future perception experiments should test whether focus can be signalled by the falling pattern found in confirmations, and furthermore, which acoustic correlates of this fall serve as perceptual cues of focus in confirmations. Prior to that, the acoustic characteristics of the falling pattern need to be established in more detail.

It is known for a variety of languages that prosodically focussed words in assertions are not only marked tonally, i.e. by a pitch accent, but also temporally, i.e. by lengthening (e.g. Bruce, 1981, Heldner and Strangert, 2001, for Swedish; Cambier-Langeveld and Turk, 1999, for English and Dutch; Kügler, 2008, for German). Moreover, Bruce (1981) suggests that increased duration is not merely an adaptation to the more complex tonal pattern, but rather a focus cue on its own, besides the tonal rise.

The goal of this study is to examine the data from Ambrazaitis (in press) on focus realisation in assertions and confirmations in more detail as regards durational patterns. The results are expected to provide information as to whether duration should be considered as a possible cue to focus and to speech act in future perception experiments. The hypothesis is that, if focus is signalled in confirmations, and if lengthening is a focus cue independent of the tonal pattern, then focal lengthening should be found, not only in assertions, but also in confirmations. Furthermore, it could still be the case that durational patterns differ in confirmations and assertions.

Method

The following two sections on the material and the recording procedure are, slightly modified, reproduced from Ambrazaitis (in press).

Material

The test sentence used in this study was “Wallander förlänger till november” (*‘Wallander is continuing until November’*). In the case of a confirmation, the test sentence was preceded by “ja” (*‘yes’*). Dialogue contexts were constructed in order to elicit the test sentence with focus on the first, second, or third content word, in each case both as an assertion and as a confirmation. These dialogue contexts consisted of a situational frame context, which was the same for all conditions (*‘You are a police officer meeting a former colleague. You are talking about retirement and the possibility to continue working.’*), plus six different context questions, one

for each condition (cf. the example in the Introduction). While the frame context was presented to the subjects exclusively in written form, the context question was only presented auditorily. For that, the context questions were pre-recorded by a 30-year-old male native speaker of Swedish.

Recording procedure and subjects

The data collection was performed using a computer program, which both presented the contexts and test sentences to the subjects and organised the recording. First, for each trial, only the frame context was displayed on the screen in written form. The subjects had to read the context silently and to try to imagine the situation described in the context. When ready, they clicked on a button to continue with the trial. Then, the pre-recorded context question was played to them via headphones, and simultaneously, the test sentence appeared on the screen. The subject's task was to answer the question using the test sentence in a normal conversational style. The subjects were allowed to repeat each trial until they were satisfied.

Besides the material for this study, the recording session included a number of further test cases not reported on in this paper. Five repetitions of each condition were recorded, and the whole list of items was randomised. One recording session took about 15 minutes per speaker. Nine speakers of Standard Swedish were recorded (5 female) in an experimental studio at the Humanities Laboratory at Lund University. Thus, a corpus of 270 utterances relevant to this study (6 conditions, 5 repetitions per speaker, 9 speakers) was collected.

Data analysis

A first step in data analysis is reported in Ambrazaitis (in press). There, the goal was to provide an overview of the most salient characteristics of the F0 patterns produced in the different conditions. To this end, F0 contours were time and register normalised, and mean contours were calculated in order to illustrate the general characteristics of the dominant patterns found in the different conditions (cf. examples in Figure 1). The F0 patterns were classified according to the F0 movement found in connection with the stressed syllable of the target word, as either 'falling' or 'rising'.

In order to obtain duration measurements, in the present study, the recorded utterances were segmented into 10 quasi-syllables using spec-

trograms and wave form diagrams. The boundaries between the segments were set as illustrated by the following broad phonetic transcriptions: [va], ['land], [əɪ], [fœ], ['lɛŋ:], [əɪ], [tɪl], [nɔ], ['vɛmb], [əɪ]. In the case of ['land] and ['vɛmb], the final boundary was set at the time of the plosive burst, if present, or at the onset of the post-stress vowel.

It has been shown for Swedish that focal lengthening in assertions is non-linear, in that the stressed syllable is lengthened more than the unstressed syllables (Heldner and Strangert, 2001). Therefore, durational patterns were analysed on two levels, first, taking into account entire word durations, second, concentrating on stressed syllables only. In both cases, the analyses focussed on the three content words and hence disregarded the word "till".

For each word, two repeated-measures ANOVAs were calculated, one with word duration as the dependant variable, the other for stressed syllable duration. In each of the six ANOVAs, there were three factors: SPEECH ACT (with two levels: assertion, confirmation), FOCUS (POSITION) (three levels: focus on initial, medial, final word), and finally REPETITION (five repetitions, i.e. five levels).

All data were included in these six ANOVAs, irrespective of possible mispronunciations, or the intonation patterns produced (cf. the two strategies for confirmations, Figure 1), in order to obtain a general picture of the effects of focus and speech act on duration. However, the major issue is whether focus in confirmations may be signalled by a falling F0 pattern. Therefore, in a second step, durational patterns were looked at with respect to the classification of F0 patterns made in Ambrazaitis (in press).

Results

Figure 2 displays mean durations of the three test words for the six conditions (three focus positions in two speech acts). The figure only shows word durations, since, on an approximate descriptive level, the tendencies for stressed syllable durations are similar; the differences between durational patterns based on entire words and stressed syllables only will, however, be accounted for in the inferential statistics.

The figure shows that the final word ("november") is generally produced relatively long even when unfocussed, i.e. longer than medial or initial unfocussed words, reflecting the well-known phenomenon of final lengthening. Moreover, the medial word ("förlänger") is

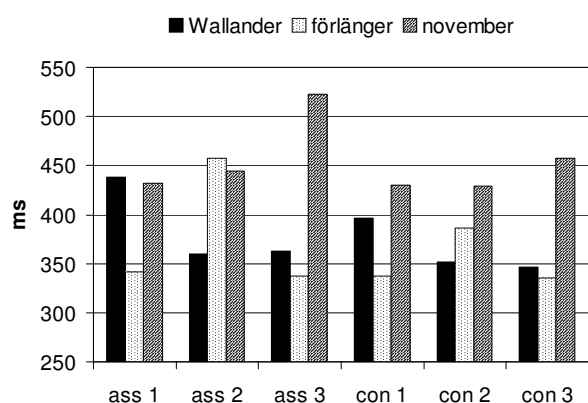


Figure 2. Mean durations of the three test words for the six conditions (ass = assertion; con = confirmation; numbers 1, 2, 3 = initial, medial, final focus), pooled across 45 repetitions by 9 speakers.

generally produced relatively short. However, influences of position or individual word characteristics are not treated in this study. The figure also shows that each word is produced longer when it is in focus than when it is pre-focal or post-focal (i.e. when another word is focussed). This focal lengthening effect can, moreover, be observed in both speech acts, although the effect appears to be smaller in confirmations than in assertions. For unfocussed words, there seem to be no duration differences between the two speech acts.

These observations are generally supported by the inferential statistics (cf. Table 1), although most clearly for the medial word: A significant effect was found for the factors SPEECH ACT and FOCUS, as well as for the interaction of the two factors, both for word duration ([fœ'leŋ:əɫ]) and stressed syllable duration ([leŋ:]); no other significant effects were found for the medial word. According to post-hoc comparisons (with Bonferroni correction),

[fœ'leŋ:əɫ] and [leŋ:] were realised with a longer duration in assertions than in confirmations ($p < .001$ in both cases) only when the medial word itself was in focus. In assertions, both the word and the stressed syllable were longer when the word was in focus than when another word was in focus ($p < .001$ for all comparisons). In confirmations, the general result was the same (entire word: focal > post-focal ($p = .016$); focal > pre-focal ($p = .003$); stressed syllable: focal > post-focal ($p = .023$); focal > pre-focal ($p = .001$)).

The situation is similar for the final word, the major difference in the test results being that the interaction of FOCUS and REPETITION was significant for word durations (cf. Table 1). Resolving this interaction shows that significant differences between repetitions only occur for final focus, and furthermore, that they seem to be restricted to confirmations. A possible explanation is that the two different strategies of focussing the final word in confirmations (rise vs. fall) are reflected in this interaction (cf. Figure 3 below). As in the case of the medial word, post-hoc comparisons reveal that both [nɔ'vembəɫ] and [vemb] were realised with a longer duration in assertions than in confirmations ($p < .001$ in both cases) only when the final word itself was in focus. Also, the entire word is longer when in focus than in the two post-focal conditions, i.e. the initial or medial word being focussed (assertions: $p < .001$ in both cases; confirmations: $p = .018$ for final vs. initial focus, $p = .007$ for final vs. medial focus). The picture is, however, different when only the stressed syllable is measured. In confirmations, no significant differences are found for [vemb] in the different focus conditions, while in assertions, the duration of [vemb] differs in all three focus conditions (final focus > medial focus ($p = .001$);

Table 1. Results of the six repeated-measures ANOVAs: degrees of freedom (Greenhouse-Geisser corrected where sphericity cannot be assumed), F-values, and p-values. Factor REPETITION was never significant; no interactions besides the one shown were significant, an exception being FOCUS*REPETITION for [nɔ'vembəɫ] ($F(8,64) = 2.21$; $p = .038$).

	[va'landəɫ]	[fœ'leŋ:əɫ]	[nɔ'vembəɫ]	[land]	[leŋ:]	[vemb]
SPEECH ACT	$F(1,8) = 9.15$ $p = .016$	$F(1,8) = 17.36$ $p = .003$	$F(1,8) = 36.33$ $p < .001$	$F(1,8) = 26.59$ $p = .001$	$F(1,8) = 73.23$ $p < .001$	$F(1,8) = 25.99$ $p = .001$
FOCUS	$F(1,13,9.07) = 19.25$ $p = .001$	$F(2,16) = 39.13$ $p < .001$	$F(2,16) = 34.57$ $p < .001$	$F(1,03,8.26) = 16.57$ $p = .003$	$F(2,16) = 36.92$ $p < .001$	$F(2,16) = 25.94$ $p < .001$
SP. ACT * FOCUS	n.s.	$F(2,16) = 20.82$ $p < .001$	$F(2,16) = 28.03$ $p < .001$	$F(2,16) = 3.75$ $p = .046$	$F(2,16) = 22.59$ $p < .001$	$F(2,16) = 31.06$ $p < .001$

final > initial ($p < .001$); medial > initial ($p = 0.19$)).

Finally, for the initial word, the interaction of FOCUS and SPEECH ACT was not significant for word duration (cf. Table 1). That is, [va'landəɹ] was produced longer in assertions than in confirmations, both when in focus and in pre-focal position (cf. also Figure 2). Post-hoc tests for FOCUS show that [va'landəɹ] is realised with a longer duration when the word is in focus than when focus is on the medial ($p = .011$) or final word ($p = .003$). However, when only the stressed syllable is taken into account, the interaction of SPEECH ACT and FOCUS is significant (cf. Table 1). As shown by post-hoc comparisons, the situation is, however, more complex than for the interactions found for the other words: First, [land] is realised longer in assertions than in confirmations not only when the initial word is in focus ($p = .002$), but also when the final word is in focus ($p = .029$). Second, in assertions, the duration of [land] differs in all three focus conditions (initial focus > medial focus ($p = .015$); initial > final ($p = .036$); final > medial ($p = .039$)), while in confirmations, [land] is significantly longer in focus than in the two pre-focal conditions only (initial > medial ($p = .005$); initial > final ($p = .016$)), i.e. no significant difference is found between the two pre-focal conditions.

In the analysis so far, all recordings have been included irrespective of the variation of F0 patterns produced within an experimental condition. As mentioned in the Introduction, confirmations were produced with either of two strategies, as classified in Ambrazaitis (in press) as either 'rising' (presence of a (lowered) H-accent on the target word), or 'falling' (absence of a H-accent on the target word), cf. Figure 1. This raises the question as to whether the focal lengthening found in confirmations (cf. Figure 2) is present in both the rising and the falling variants. Figure 3 displays the results for confirmation in a rearranged form, where the F0 pattern is taken into account.

For the medial word, Figure 3 indicates that, first, the word seems to be lengthened in focus even when it is produced with a falling pattern (cf. "förlänger" in conditions 'medial fall' vs. 'final fall', 'final rise', and 'initial rise'), and second, the focal lengthening effect still tends to be stronger when the word is produced with a rise ('medial fall' vs. 'medial rise'). However, for the final word, focal lengthening seems to be present only when the word is produced with a rise. Finally, the initial word seems to be

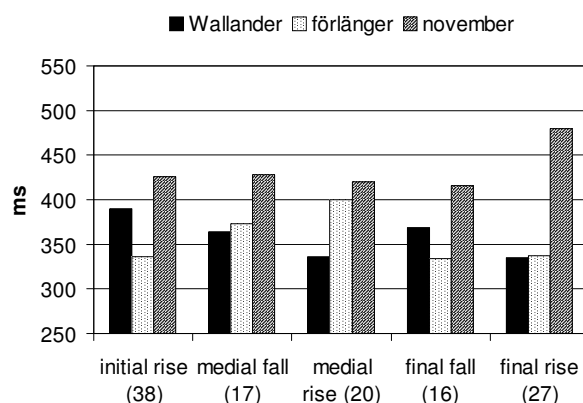


Figure 3. Mean durations of the three test words in confirmations, divided into classes according to the intended focus position (initial, medial, final word) and F0 pattern produced on the target word (rise, fall); n in parentheses.

lengthened not only when it is in focus itself, but also when medial or final focus is produced with a fall, as compared to medial or final focus produced with a rise.

Discussion

The goal of this study was to examine the durational patterns in a data corpus where focus was elicited in two different speech acts, assertions and confirmations. It is unclear from the previous F0 analysis (cf. Figure 1 and Ambrazaitis, in press) whether focus was successfully signalled in confirmations, when these were produced without a 'rising focal accent' (H-). The general hypothesis to be tested in future perception experiments is that focus in confirmations may even be signalled by a falling pattern, which would support the proposal by Ambrazaitis (2007) that there is a paradigmatic utterance-level accent contrast in Standard Swedish between a rising (H-) and a falling accent.

The present results are in line with this general hypothesis, since they have shown that focal lengthening can be found not only in assertions, but also in confirmations, although the degree of focal lengthening seems to be smaller in confirmations than in assertions. In fact, the speech act hardly affects the duration of unfocused words, meaning that speech act signalling interacts with focus signalling. Most importantly, the results also indicate that focal lengthening may even be found when the target word is produced with a falling F0 pattern, although no inferential statistics have been reported for this case. In fact, in these cases, duration differences seem to be more salient than F0 differences (cf. 'medial fall' and 'final fall' in Figures 1 and 3).

This summary of the results, however, best matches the durational patterns found for the medial word. Heldner and Strangert (2001) conclude that the medial position is least affected by factors other than the focal accent itself, e.g. final lengthening. Based on the present results, it seems obvious that even the duration of the initial word is influenced by more factors than focus, since even if the initial word is pre-focal, its duration seems to vary depending on whether the medial or the final word is focussed (when only stressed syllable is measured), or, in confirmations, whether medial or final focus is produced with a fall or a rise. More research is needed in order to reach a better understanding of these patterns. In part, durational patterns of initial words could possibly be related to the role the initial position plays in signalling phrase- or sentence prosody (Myrberg, in press; Roll et al., 2009).

Finally, for the final word, the evidence for focal lengthening in confirmations is weaker, a tendency opposite to the one found by Heldner and Strangert (2001) for assertions, where final words in focus tended to be lengthened more than words in other positions. In the present study, no focal lengthening was found for the final word in confirmations when the word was produced with a falling pattern. However, the relative difference in duration between the final and the medial word was still larger as compared to the case of intended medial focus produced with a fall (cf. the duration relations of ‘medial fall’ and ‘final fall’ in Figure 3).

Some of the duration differences found in this study are small and probably irrelevant from a perceptual point of view. However, the general tendencies indicate that duration is a possible cue to perceived focus position in confirmations and thus should be taken into account in the planned perception experiment.

Acknowledgements

Thanks to Gösta Bruce and Merle Horne for their valuable advice during the planning of the study and the preparation of the paper, to Mikael Roll for kindly recording the context questions, and, of course, to all my subjects!

References

Ambrazaitis G. (2007) Expressing ‘confirma-

- tion’ in Swedish: the interplay of word and utterance prosody. *Proceedings of the 16th ICPhS (Saarbrücken, Germany)*, 1093–96.
- Ambrazaitis G. (in press) Swedish and German intonation in confirmations and assertions. *Proceedings of Nordic Prosody X (Helsinki, Finland)*.
- Bruce G. (1977) Swedish word accents in sentence perspective. Lund: Gleerup.
- Bruce G. (1981) Tonal and temporal interplay. In Fretheim T. (ed) *Nordic Prosody II – Papers from a symposium*, 63–74. Trondheim: Tapir.
- Bruce G., Filipsson M., Frid J., Granström B., Gustafson K., Horne M., and House D. (2000) Modelling of Swedish text and discourse intonation in a speech synthesis framework. In Botinis A. (ed) *Intonation. Analysis, modelling and technology*, 291–320. Dordrecht: Kluwer.
- Büring D. (2007) Intonation, semantics and information structure. In Ramchand G. and Reiss C. (eds) *The Oxford Handbook of Linguistic Interfaces*, 445–73. Oxford: Oxford University Press.
- Cambier-Langeveld T. and Turk A.E. (1999) A cross-linguistic study of accentual lengthening: Dutch vs. English. *Journal of Phonetics* 27, 255–80.
- Heldner M. and Strangert E. (2001) Temporal effects of focus in Swedish. *Journal of Phonetics* 29, 329–361.
- Horne M. (1991) Why do speakers accent “given” information? *Proceedings of Eurospeech 91: 2nd European conference on speech communication and technology (Genoa, Italy)*, 1279–82.
- Kügler F. (2008) The role of duration as a phonetic correlate of focus. *Proceedings of the 4th Conference on Speech Prosody (Campinas, Brazil)*, 591–94.
- Ladd D.R. (2008) *Intonational phonology* 2nd ed. Cambridge: Cambridge University Press.
- Myrberg S. (in press) Initiality accents in Central Swedish. *Proceedings of Nordic Prosody X (Helsinki, Finland)*.
- Roll M., Horne M., and Lindgren, M. (2009) Left-edge boundary tone and main clause verb effects on syntactic processing in embedded clauses – An ERP study. *Journal of Neurolinguistics* 22, 55–73.