The acquisition of focal lengthening in Stockholm Swedish

Anna Sara H. Romøren[1], *Aoju Chen*[2]

[1] Oslo and Akershus University College of Applied Sciences/University of Oslo, Norway
[2] Utrecht University, The Netherlands
anna.sara.h.romoren@hioa.no, aoju.chen@uu.nl

Abstract

In order to be efficient communicators, children need to adapt their utterances to the common ground shared between themselves and their conversational partners. One way of doing this is by prosodically highlighting focal information. In this paper we look at one specific prosodic manipulation, namely word duration, asking whether Swedish-speaking children lengthen words to mark focus, as compared to adult controls. To the best of our knowledge, this is the first study on the relationship between focus and word duration in Swedish-speaking children.

Index Terms: prosody, language acquisition, focus, duration, phonetics.

1. Introduction

The notion 'focus' is perhaps one of the most studied aspects of information packaging, typically referring to 'new' or 'important' information in a sentence, contrasted with 'given' or 'established' information (e.g. [1]). Following many others we use the term 'prosody' as referring to acoustic variation in pitch, duration and intensity, but also for linguistic notions such as stress, accents, and tones.

2. Background

Traditionally, most work on phonological development in children has been concerned with segmental phonology; relatively little is known about how children develop toward adult proficiency in the production of prosodic categories such as lexical tones, pitch accents and prosodic boundaries. Stockholm Swedish (hereafter Swedish)[1] is a lexical pitch accent language, in which pitch is contrastive at the word level. In addition to marking lexical contrasts, pitch is also used for marking focus: Swedish-speaking adults add a prominence marking high (H) tone to focal constituents, generating a contour that in some ways resemble the pitch accents of German, English or Dutch, but that consists in both lexically and post-lexically assigned tones [2].

In addition to pitch manipulations, increased word duration is reported as a fairly consistent prosodic correlate of focus for both English, German, Dutch and Swedish ([4],[5],[6],[7],[8],

[9],[10],[11],[12],[13],[14],[15],[16],[17],[18],[19],[20],[21]). Whereas increased word duration is strongly associated with focus in adult Swedish, little is known about how Swedish-speaking children develop in the use of word duration for focus. As part of a larger study on the acquisition of prosodic focus marking in Dutch and Swedish ([22,23,24,25,26]), we have shown that Swedish-speaking children are remarkably adult-like in their use of accentuation for focus, that is: Swedish children add a prominence marking high (H) tone to focal constituents from either lexical accent category, and they avoid this tone post-focally.

Previous studies of the relationship between duration and focus in child speech predominantly report children at four-to-five *not* to show the focal lengthening described for adult speakers (see e.g. [27] on English [28], on German). Similarly, only spurious effects of focus have been reported in Dutch-speaking four-to-five-year-olds ([29], [30]; [26]). In terms of pitch, our own work suggests that Swedish-speaking children reach adult proficiency earlier than their Dutch-speaking peers. The question remains whether they are also ahead in the use of focal lengthening. To answer this question, we analysed semi-spontaneous sentence productions from Swedish-speaking children between four and eleven years.

Based on previous work on English, German, and Dutch, we hypothesized that our five-year-olds would not use duration in an adult-like way, but that our older child groups would. Based on previous work on accentual lengthening in adult Swedish (e.g. [21]) we expected our adult group to produce narrowly focal words with longer duration than post-focal ones, irrespective of sentence-position.

3. Method

Twenty-six Swedish-speaking children and ten Swedish-speaking adults participated in the study (Table 1).

Table 1: *Participant information*

Age group	N	Age range	Age mean	Gender (m/f)
4-5	10	4;3-5;6	5;0	6/4
7-8	8	7;6-8;8	8;3	5/3
10-11	8	10;0-11;0	10;6	4/4
Adults	10	20;0-43;10	27;2	5/5

None of the participants had any history of language disorders, and they were all native language speakers of Swedish. The children were recruited from kindergartens and schools in Stockholm, and the adults were recruited at The Royal Institute of Technology (KTH) in Stockholm. In the following

---

[1] Central Swedish or Stockholm Swedish is a regional variety of Swedish spoken around Stockholm and beyond (eastern Svealand). This variety is the closest one gets to a spoken standard in Sweden, and is by far the most well described variety where prosody is concerned (see [2]).

we will refer to the age groups as five-year-olds, eight-year-olds and eleven-year-olds, respectively.

The data collected consisted in semi-spontaneous subject-verb-object sentence productions, elicited as part of a picture-matching game (see [23] for details). In the game, the participant's task was to help the experimenter find correct picture combinations by answering the experimenter's questions about her pictures. Scripted contexts were combined with wh-questions (see examples 1-3 below), so as to make focal elements informationally new and non-focal elements informationly given in the responses produced by the participants. Six subject nouns, six transitive verbs and six object nouns (half accent 1, half accent 2) were distributed over the five sentence conditions.

1. Question: Who cooks the carrot?
   Answer: [The tiger] FOCUS cooks the carrot.

2. Question: What does the tiger do with the carrot?
   Answer: The tiger [cooks]FOCUS the carrot.

3. Question: What does the tiger cook?
   Answer: The tiger cooks [a carrot]FOCUS.

Only responses following the scripted speech context were included in the analysis. Responses were also excluded if they contained deviant word orders, deviant word choices or elided constituents, as well as self-repairs, hesitations, or background noise (see [26], for details).

The sentence conditions included in the analysis are presented in Table 2.

Table 2: *Focus conditions*

Condition	Explanation
Initial narrow	Focus on initial constituent (subject)
Medial narrow	Focus on medial constituent (verb)
Final narrow	Focus on final constituent (object)

The included responses were orthographically transcribed and segmented into words using Praat [31]). When segmenting we relied on changes in the waveform, in addition to the formant transitions shown in the spectrogram (see [32]). Conventions were established for how to segment the words at particularly challenging boundaries (e.g. onset plosives were consistently segmented right before the burst). The built in silence-detecting function in Praat ([31]) was also used as an aid when boundaries were placed, particularly in sentence-final position, where the end of a segment can be particularly challenging to determine.

All medial and final target words (i.e. verbs and object nouns) were segmented by hand, and word durations were automatically extracted from text grids using a script. All extracted values were checked for measuring and extraction errors.

4. Analysis and results

Our measures of word duration were made on sentence medial and sentence final target words, and words in either position were compared between conditions rendering them focal and post-focal, respectively.

For the statistical analysis of the effect of focus on word duration we built linear mixed effect models (LMMs) in R ([33]), with 'participant' and 'item' as random factors, and 'focus' (focal or post focal), 'group' (adult, five, eight, eleven) and 'lex' (accent one or accent 2) as independent factors. For all models the outcome variable was numeric, involving raw measures of word duration on medial or final target words. For details about the analysis, see [26]. Below we first report the results from the sentence-medial comparison, thereafter reporting the results from the sentence-final comparison.

4.1. Sentence medial position

In Figure 1 below we present a boxplot of the effect of focus on word duration sentence-medially, broken down by group and lexical accent.

Figure 1: *Word duration on sentence-medial targets*
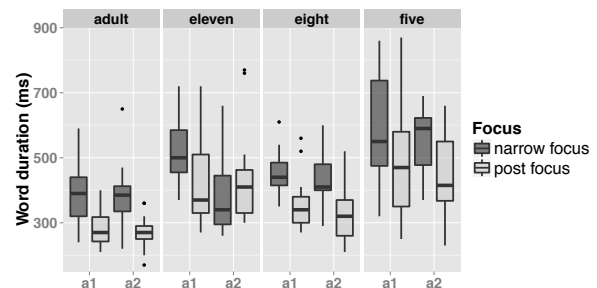
Building and comparing models on the effect of 'focus', 'group', and 'lex', on the duration of the sentence-medial target words, revealed the best model to contain main effects of 'focus' and 'group', a two-way interaction between 'group' and 'focus', and a three-way-interaction between 'focus', 'group' and 'lex'. In Table 3 we present the p-values for the model comparisons.

Table 3: *Model comparison, sentence-medial analysis*

Model comparison	p-value
0 vs 1 (adding focus)	0,000
1 vs 2 (adding group)	0,000
2 vs 3 (adding lex)	0,188
2 vs 4 (adding focus*group)	0,009
4 vs 5 (adding focus*lex)	0,156
4 vs 6 (adding group*lex)	0,419
4 vs 7 (adding focus*group*lex)	0,002

Basically, the main effect of focus consisted in focal status generally increasing the word duration on the medial targets. In addition, main effects of group consisted in generally longer word durations in the eleven and five-year-olds than the adults. Furthermore, there was an interaction between 'group' and 'lex', as accent 1-words were longer than accent 2-words in the eleven-year-olds, whereas this was not the case for the adults. Finally, the three-way-interaction was caused by the adults and eleven-year-olds differing in the effect of 'focus' by 'lex', in the sense that whereas focus lengthened the duration of both accent 1 and accent 2 words in the adults, in the eleven-year-olds this effect was only present on the accent 1 words (see Table 4 for the summary of the best model).

Table 4: *Model summary, sentence-medial analysis*

Factor	b-value	std. error	p-value
Intercept (adults, narrow focus, accent 1)	0,386	0,031	0,000
Post focus	-0,097	0,026	0,000
Group eleven	0,136	0,041	0,001
Group eight	0,060	0,041	0,146
Group five	0,206	0,039	0,000
Accent 2	-0,012	0,034	0,725
Post focus * group eleven	-0,013	0,041	0,740
Post focus *group eight	0,003	0,041	0,945
Post focus * group five	0,023	0,039	0,559
Group eleven * accent 2	-0,137	0,040	0,001
Group eight * accent 2	0,006	0,040	0,878
Group five * accent 2	-0,007	0,040	0,870
Post focus * group eleven * accent 2	0,198	0,057	0,001
Post focus * group eight * accent 2	-0,014	0,056	0,803
Post focus * group five * accent 2	-0,027	0,055	0,627

Exploring the effect of focus within each group separately showed a significant effect of focus increasing the word duration in the adults (p=.000), the eight-year-olds (p=.000) and the five-year-olds (p=.000), but not in the eleven-year-olds (p=.548). Our raw means, comparing narrowly focal to post-focal target words, showed focus to increase the duration by 106 ms (28%) in our adults, 104 ms (23%) in the eight-year-olds and by 89 ms (15%) in our five-year-olds, whereas narrowly focal words were on average 11 ms (2%) longer than post-focal ones in our eleven-year-olds, when both lexical accents were included.

4.2. Sentence final position

Figure 2 shows a boxplot of the effect of focus on word duration sentence-finally, by group and lexical accent.

Figure 2: *Word duration on sentence-final targets*
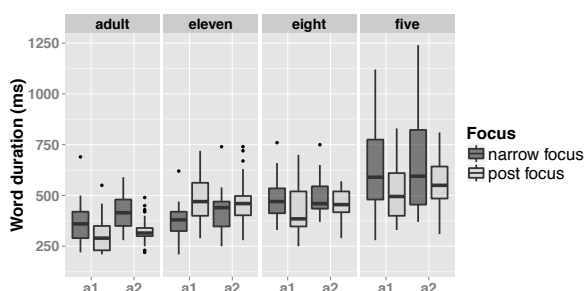
Building models for the effect of 'focus', 'group', and 'lex' on word duration sentence-finally showed the best model to include main effects of 'group' and 'focus', as well as an interaction between the two (see Table 5 for the model comparison).

Table 5: *Model comparison, sentence-final analysis*

Model comparison	p-value
0 vs 1 (adding focus)	0,000
1 vs 2 (adding group)	0,000
2 vs 3 (adding lex)	0,526
2 vs 4 (adding focus*group)	0,000
4 vs 5 (adding focus*lex)	0,680
4 vs 6 (adding group*lex)	0,652
4 vs 7 (adding focus*group*lex)	0,680

The general main effect of focus consisted in focal status significantly increasing the duration on the sentence-final target words. The group effects consisted in the eight and five-year-olds producing the sentence-final targets with longer duration than the adults. Finally, the interaction effect between focus and group consisted in focus increasing the word duration in our adult group, whereas it had the opposite effect i our eleven-year-olds. Additionally, the focal lengthening effect was stronger in our five-year-olds than in our adult controls (see Table 6 for the summary of the best model).

Table 6: *Model summary, sentence-final analysis*

Factor	b-value	std. error	p-value
Intercept (adults, narrow focus)	0,392	0,036	0,000
Post focus	-0,073	0,021	0,001
Group eleven	0,019	0,042	0,648
Group eight	0,098	0,041	0,023
Group five	0,282	0,040	0,000
Post focus * group eleven	0,140	0,034	0,000
Post focus * group eight	0,030	0,033	0,357
Post focus * group five	-0,066	0,033	0,045

Exploring the main effect of focus within each group showed that focus significantly increased the word-duration in the adults (p=.001) and the five-year-olds (p=.000), and that there was a trend in the same direction in the eight-year-olds (p=.083). A significant effect of focus was also found in our eleven-year-olds (p=.011), but in this group the effect went in the opposite direction of that found in the other groups. Comparing narrowly focal targets to post-focal ones, our raw means showed narrow focus to increase the duration by 72 ms (18%) in our adults, 50 ms (10%) in the eight-year-olds and by 140 ms (21%) in our five-year-olds, whereas narrowly focal words were on average 67 ms (16%) *shorter* than post-focal ones in our eleven-year-olds.

4.3. General discussion

Summing up our analysis of the relationship between focus and word duration in our four groups, we found general effects

of focus increasing the word duration on both medial and final target words in the adults, the eight-year-olds, and the five-year-olds, whereas the eleven-year-olds differed from the adults both sentence-medially and sentence-finally. In sentence-medial position, the eleven-year-olds differed from the adults in not showing any effect of focus on word duration in accent 2 words, whereas they behaved in line with the other groups on accent 1 words. In sentence-final position, the eleven-year-olds differed from the adults by producing post-focal target words with longer duration than focal ones across both accent types.

The patterns found in our group of adults confirm previous reports on reliable effects of focus on word duration in adult Swedish, in both sentence-medial and sentence-final position (e.g. [21]). On average, our adult participants lengthened the target words with 28% sentence medially, and 18% sentence finally. In terms of position effects, the patterns observed in this study resemble descriptions of accentual and final lengthening in English, where accentual lengthening is present both sentence-medially and sentence-finally, but where the effect of accentual lengthening is stronger in non-final than in final positions (e.g. [5],[6]). Our findings thus differ from those reported by [21], who found stronger effects of focal lengthening finally than non-finally. This study was not specifically designed to look at position effects, and we did not make direct comparisons of the degree of lengthening by position. Because of this, our observation of stronger effects of focal lengthening medially than finally needs to be confirmed by manipulating position more systematically.

Based on previous work on English, German, and Dutch, we hypothesized that our five-year-olds would not use duration in an adult-like way, but that our older child groups would. This hypothesis was disconfirmed in our data, as both the five-year-olds and the eight-year-olds performed in line with the adults. Sentence-finally, the five-year-olds showed even more robust effects of focus on word duration than the adults did. Together with the results presented in [24] and [26], these findings suggest that Swedish-speaking children reach adult proficiency in duration manipulations for focus before they reach adult proficiency in pitch-based manipulations. It also seems that Swedish-speaking children manipulate duration for focus at an earlier stage than what has been reported for children learning Dutch, German or English, indicating that Swedish-speaking children are not only ahead of Dutch-speaking children in the use of pitch-based cues to focus; they are also ahead in the use of duration manipulations.

The fact that we find robust effects of focus on duration in child Swedish, whereas this is not found in child Dutch, English or German, despite reports of adult-like accent placement in these populations, may suggest that word duration is a more reliable correlate of prominence H in child Swedish than it is of accentuation in Dutch. If the durational effect of focus or focal accentuation is generally more robust in Swedish than in it is in Dutch, the higher temporal variability characterizing child speech (e.g. [34]) may keep duration effects of accentuation from surfacing in studies on child Dutch, German and English. In order to determine whether the use of duration for focus develops differently from the use of prominence H in Swedish, younger children need to be studied, so that possible dissociations between prominence H and word duration can be systematically explored. In addition, more cross-linguistic work is needed in order to determine whether the relationship between focus, accentuation, and word duration is actually the same across different Germanic varieties.

While our duration analysis showed that both the five and the eight-year-olds performed in line with adults, our results from the eleven-year-olds are rather striking, as previous work suggests that children have largely reached adult proficiency in prosodic focus marking well before this age (e.g. [30]; [35]). Indeed, adult proficiency is also found in our measures of prominence H and pitch range in this group, thus the fact that the eleven-year-olds stand out on our duration-measures is rather surprising. The mixed results from this group are hard to explain. Tentatively, we speculate whether task effects may at least partly explain these patterns. When playing the picture-matching game with the oldest group we had the impression that the game was less interesting to these than it was to the two youngest groups. Occasionally the eleven-year-olds seemed to signal their slight disinterest in the game by slowing down their speech rate (e.g. [36]). Such lengthening seemed particularly common when the children had already provided the most important information, thus post-focal targets were more prone to such 'unengaged' prosody than focal ones. It may be the case that the tendency for the eleven-year-olds to lengthen post-focal targets may have cancelled out the effect of narrow versus post-focus that was found in the other groups. We are still working on explaining these results, and we plan to conduct more detailed analyses on the data from this group, for example by normalizing the use of duration by speaker.

5. Conclusion

In this study we have shown that Swedish-speaking children between four and eight years reliably lengthen focal words, as compared to post-focal ones. This happens across the two accentual categories, and both sentence-medially and sentence finally. Our adult data also confirms previous work, showing that adults lengthen words referring to focal information. The results from our oldest group of children were puzzling, as they showed lacking or reverse effects of focus on word duration. We speculate that task-effects may explain these patterns. Future studies should include slightly more naturalistic speech data, and should aim at studying younger children, determining at what point they master the use of duration for focus, as our youngest group already performed in line with adults.

6. References

[1] Krifka, M., & Musan, R. (2012). *The expression of information structure*. Berlin: Mouton de Gruyter.

[2] Riad, T. (2014). *The phonology of Swedish*. Oxford: Oxford University Press.

[3] Riad, T. (2006). Scandinavian accent typology. *STUF-Sprachtypologie Und Universalienforschung, 59*, pp. 36-55.

[4] Cambier-Langeveld, T., & Turk, A. E. (1999). A cross-linguistic study of accentual lengthening: Dutch vs. English. *Journal of Phonetics, 27*, pp. 255-280.

[5] Cambier-Langeveld, G. M. (2000). *Temporal marking of accents and boundaries*. PhD thesis. Leiden: Holland Institute of Generative Linguistics.

[6] Cooper, W. E., Eady, S. J., & Mueller, P. R. (1985). Acoustical aspects of contrastive stress in question–answer contexts. *Journal of the Acoustical Society of America, 77*, pp. 2142-2156.

[7] Eady, S. J., & Cooper, W. E. (1986). Speech intonation and focus location in matched statements and questions. *Journal of the Acoustical Society of America, 80*, pp. 402-415.

[8] Sityaev, D., & House, J. (2003). Phonetic and phonological correlates of broad, narrow and contrastive focus in English. In M. J. Solé, D. Recasens, & J. Romero (Eds.). *Proceedings of the 15th International Congress of Phonetic Sciences (ICPHS)*, pp. 1819-1822.

[9] Turk, A. E., & Sawusch, J. R. (1997). The domain of accentual lengthening in American English. *Journal of Phonetics, 25*, pp. 25-41.

[10] Turk, A. E., & White, L. (1997). The domain of accentual lengthening in Scottish English. *Proceedings of Eurospeech 1997*, pp. 795-798.

[11] Baumann, S., Grice, M., & Steindamm, S. (2006). Prosodic marking of focus domains-categorical or gradient. *Proceedings of Speech Prosody*, pp. 301-304.

[12] Kügler, F. (2008). The role of duration as a phonetic correlate of focus. *Proceedings of Speech Prosody*, pp. 591-594.

[13] Féry, C., & Kügler, F. (2008). Pitch accent scaling on given, new and focused constituents in German. *Journal of Phonetics, 36*, pp. 680-703.

[14] Eefting, W. (1991). The effect of ''information value'' and ''accentuation'' on the duration of Dutch words, syllables, and segments. *Journal of the Acoustical Society of America, 89*, pp. 412-424.

[15] Sluijter, A., & Van Heuven, V. J. (1995). Effects of focus distribution, pitch accent and lexical stress on the temporal organization of syllables in Dutch. *Phonetica, 52*, pp. 71-89.

[16] Sluijter, A., & Van Heuven, V. J. (1996). Acoustic correlates of linguistic stress and accent in Dutch and American English. *Proceedings of the 4th International Conference on Spoken Language Processing (ICSLP)*, pp. 630-633.

[17] Bruce, G. (1981). Tonal and temporal interplay. *Nordic prosody 2,* pp. 63-74.

[18] Bannert, R. (1979). The effect of sentence accent on quantity. *Proceedings of the 9th International Congress of Phonetic Sciences (ICPHS)*, pp. 253-259.

[19] Fant, G., Kruckenberg, A., & Nord, L. (1991). Durational correlates of stress in Swedish, French, and English. *Journal of Phonetics, 19,* pp. 351-365.

[20] Fant, G., & Kruckenberg, A. (1994). Notes on stress and word accent in Swedish. *Proceedings of the International Symposium on Prosody,* pp. 2-3.

[21] Heldner, M., & Strangert, E. (1997). To what extent is perceived focus determined by F0-cues. *Proceedings of Eurospeech 1997*, pp. 875-878.

[22] Romøren, A. S. H., & Chen, A. (2014). Accentuation, pitch and duration as cues to focus in Dutch 4-to 5-year-olds. *Proceedings of BUCLD 38*, pp. 1-12.

[23] Romøren, A. S. H., & Chen, A. (2015a). Quiet is the new loud: Pausing and focus in child and adult Dutch. *Language and Speech, 58*, pp. 8-23.

[24] Romøren, A. S. H., & Chen, A. (2015b). The acquisition of prosodic focus marking in Central Swedish: Sorting out lexical and post-lexical tones. *Proceedings of the 18th International Congress of Phonetic Sciences (ICPHS)*.

[25] Romøren, A. S. H., & Chen, A. (2016). Hunting highs and lows: Acquiring prosodic focus marking in Swedish and Dutch. *Proceedings of BUCLD 40*.

[26] Romøren, A. S. H. (2016). Hunting highs and lows: The acquisition of prosodic focus marking in Swedish and Dutch. PhD Dissertation. LOT: Utrecht University.

[27] Wonnacott, E., & Watson, D. G. (2008). Acoustic emphasis in four year olds. *Cognition, 107*, pp. 1093-1101.

[28] Sauermann, A., Höhle, B., Chen, A., & Järvikivi, J. (2011). Intonational marking of focus in different word orders in German children. *Proceedings of the 28th West Coast Conference on Formal Linguistics,* pp. 313-322.

[29] Chen, Aoju. (2009). The phonetics of sentence-initial topic and focus in adult and child Dutch. In Marina Vigário, Sonia Frota & João M. Freitas (Eds.). *Phonetics and phonology: interactions and interrelations* (pp.91-106). John Benjamins Publishing.

[30] Chen, A. (2011a). Tuning information packaging: Intonational realization of topic and focus in child Dutch. *Journal of Child Language, 38*, pp. 1055-1083.

[31] Boersma, P., & Weenink, D. (2010). Praat: Doing phonetics by computer [Computer program], available at http://www.fon.hum.uva.nl/praat/

[32] Turk, A., Nakai, S., & Sugahara, M. (2006). Acoustic segment durations in prosodic research: A practical guide. In S. Sudhoff (Ed.). *Methods in Empirical Prosody Research,* pp. 1-28. Berlin: Mouton de Gruyter.

[33] R Core Team (2014). The R project for statistical computing. R Foundation for Statistical Computing. Available at https://www.r-project.org/.

[34] Smith, B. L., Kenney, M. K., & Hussain, S. (1996). A longitudinal investigation of duration and temporal variability in children's speech production. *Journal of the Acoustical Society of America, 99*, pp. 2344-2349.

[35] Mozziconacci, S. J., & Hermes, D. J. (2000). Expression of emotion and attitude through temporal speech variations. *Proceedings of INTERSPEECH,* pp. 373-378.