

ARTICLE

# The acquisition of prosodic marking of narrow focus in Central Swedish

Anna Sara H. ROMØREN\* and Aaju CHEN\*\* 

\*Oslo Metropolitan University, Norway and \*\*Utrecht University, the Netherlands  
Address for correspondence: Anna Sara H. Romøren, Oslo Metropolitan University, Norway. E-mail: [romo@oslomet.no](mailto:romo@oslomet.no); Aaju Chen, Utrecht University, the Netherlands. E-mail: [aaju.chen@uu.nl](mailto:aaju.chen@uu.nl)

(Received 24 November 2019; revised 5 December 2020; accepted 11 December 2020;  
first published online 6 April 2021)

## Abstract

We investigated how Central Swedish-speaking four to eleven-year-old children acquire the prosodic marking of narrow focus, compared to adult controls. Three measurements were analysed: placement of the prominence-marking high tone (prominence H), pitch range effects of the prominence H, and word duration. Subject-verb-object sentences were elicited in sentence-medial and sentence-final focus conditions via a semi-spontaneous elicitation task. The children largely performed in an adult-like manner already at four to five: they predominantly added prominence H to focal words and avoided this tone post-focally in both sentence-medial and sentence-final position. The placement or avoidance of prominence H had largely the same effects on pitch range for children and adults. Finally, the four to eight-year-olds also increased the duration of the focal word, similar to adults. Hence, Central Swedish-speaking children master the use of prosody for focus marking at an earlier age, compared to children acquiring a West Germanic language.

**Keywords:** acquisition; development; focus marking; prosody; Central Swedish

## Introduction

In this study we investigate how Central Swedish-speaking children develop the ability to prosodically encode focus in their utterances. As most previous work on the acquisition of prosodic focus marking has been centred on English, German and Dutch, we expand this line of research by investigating how children learning Central Swedish, a pitch accent language, acquire prosodic focus marking, and answering the question of how the prosodic system of Central Swedish affects the way this linguistic skill is mastered by children between 4 and 11 years of age.

## Background

The notion information structure or information packaging (Halliday, 1967; Chafe, 1976; Lambrecht, 1996) concerns the adjustments speakers can make to an utterance

© The Author(s), 2021. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives licence (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided that no alterations are made and the original article is properly cited. The written permission of Cambridge University Press must be obtained prior to any commercial use and/or adaptation of the article.

in order to tailor it to the knowledge states of their listeners (Krifka & Musan, 2012). Theories of information structure share the basic idea that some parts of a sentence anchor it to previous discourse (typically given information), while other parts make a contribution to discourse (typically new information), thereby updating the common ground between the conversational partners (Vallduví & Engdahl, 1996). ‘Focus’ is perhaps one of the most extensively studied aspects of information structure. It is often defined as the new information in a sentence. However, Krifka (2008, inspired by Rooth, 1985; 1992) notes that not all instances of focus involve newness, as illustrated by the sentence *Mary only saw HIM*, where the pronoun in focus refers to a given referent. Instead, he argues that focus ‘indicates the presence of alternatives relevant for the interpretation of linguistic expressions’ (Krifka, 2008:5), and the alternatives can convey either given or new information. Focus is frequently subcategorized in terms of the *SIZE* of the focal constituent (Examples 1 and 3) and *CONTRASTIVITY* of the focal information (Examples 1 and 2) (Gussenhoven, 2004, 2007).<sup>1</sup> With regard to the size of the focal constituent, the term ‘narrow focus’ is typically used for cases where only one word of a syntactic constituent is in focus (e.g., Example 1), whereas ‘broad focus’ is used when an entire syntactic constituent or sentence is in focus (e.g., Example 3) (Ladd, 1980). *CONTRASTIVITY* concerns whether or not a contrast is explicitly evoked between the focal information and alternative candidates. A contrast is explicitly evoked in Example 2 but not in Examples 1 and 3.

(1)

Non-contrastive narrow focus (hereafter narrow focus)

Person A: Vad gör hunden med tårten? ‘What is the dog doing to the cake?’

Person B: Hunden [kastar]F tårten ‘The dog [is throwing]F the cake’.

(2)

Contrastive narrow focus (hereafter contrastive focus)

Person A: Hunden [äter]F tårten ‘The dog is eating the cake’.

Person B: Hunden [kastar]F tårten ‘The dog [is throwing]F the cake’.

(3)

Non-contrastive broad focus (hereafter broad focus)

Person A: Vad händer? ‘What’s happening?’

Person B: [Hunden kastar tårten]F ‘[The dog is throwing the cake.]F’

Across languages, different linguistic devices can be used to mark focus, such as morpho-syntactical markers (e.g., focus particles), syntactic alternations (e.g., clefting) and prosody (Vallduví & Engdahl, 1996; Krifka & Musan, 2012). In the current study we are concerned with how focus is marked using prosody, that is, acoustic variation in pitch, duration, intensity and spectral composition, giving rise to suprasegmental linguistic phenomena such as lexical stress, lexical tones and sentence accents.

<sup>1</sup>In Swedish, the simple present of a verb can be used to describe either an action that is happening now or an action at present that takes place frequently. We chose to use the present continuous in the English translation because it fits better in a game context, where the child inspects pictures and describes what’s going on in the picture (see the method section).

Research on prosodic focus marking in children is still fairly limited (for reviews, see Ito, 2014, 2018; Chen, 2018). Further, existing work has centred on children acquiring English, German and Dutch (cf. Arnhold, Chen & Järvikivi, 2016 on Finnish-speaking children; Yang & Chen, 2018 on Mandarin-speaking children; Yang, 2017 on Korean-speaking children). The present study is concerned with the acquisition of prosodic focus marking in Central Swedish, a lexical pitch accent variety of Swedish spoken in Stockholm and surrounding areas, which is considered to be the most standardised variety in Sweden. Swedish is a North Germanic language of the Indo-European language family (Riad, 2006).

Central Swedish recognises two lexical pitch accents, accent 1 (transcribed as HL\*) and accent 2 (transcribed as H\*L).<sup>2</sup> Every word has either accent 1 or accent 2 in this variety of Swedish, largely predictable from the phonological and morphological context (Gussenhoven, 2004). The starred tone is aligned with the main stressed syllable of a word, resulting in a notable difference between the two lexical pitch accents in the timing of the fall (Bruce, 1977; Myrberg, 2009; Ambrazaitis, 2009). As can be observed in the word pair *anden*<sup>1</sup> ‘the duck’ versus *anden*<sup>2</sup> ‘the spirit’ (upper panels, Figure 1), the alignment of L\* of accent 1 in a trochaic disyllabic word (i.e., with initial stress) results in low pitch in the stressed syllable and the following syllable; the alignment of H\* of accent 2 in a trochaic disyllabic word results in high pitch through most of the stressed syllable, falling to low pitch in the following syllable.<sup>3,4</sup> The accents may not be realised in certain words when they are included into larger prosodic units and consequently lose their lexical stress (see Myrberg, 2009, for a discussion).

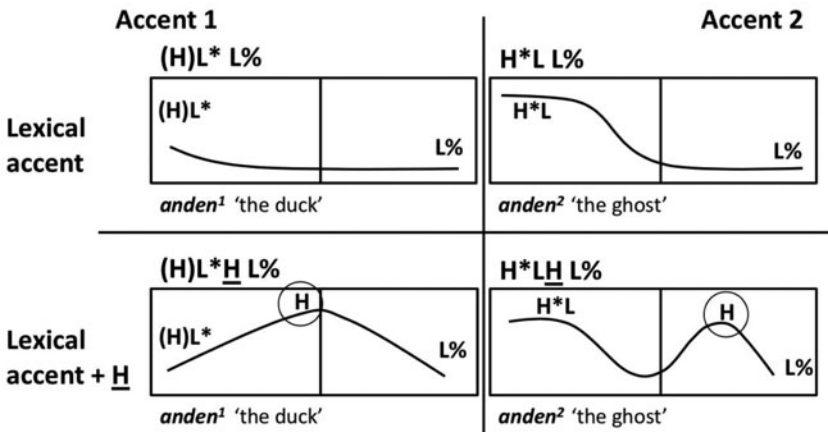
To mark focus, speakers of Central Swedish (hereafter Swedish) add a floating high (H) tone after the lexical pitch accent of the focal word (Bruce, 1977, 1987, 1998; Ambrazaitis, 2009; Myrberg, 2009; 2013). When the focus marking H tone is added to the lexical accents, the complete pattern for accent 1 is annotated as (H)L\*H and the corresponding pattern for accent 2 is H\*LH (see Figure 1). In this paper we refer to the focus-marking floating H tone as ‘prominence H’, to distinguish it from the H tones of the lexical pitch accents and other usage of a floating H tone.<sup>5</sup> As shown in the lower panels of Figure 1, the alignment of the lexical pitch accent has consequences for the timing of prominence H, typically creating a one-peaked contour in focal accent 1 words but a two-peaked contour in focal accent 2 words, at least when we only consider tonal patterns realized on the target words as in one-word intonation phrases. More specifically, after accent 1, prominence H is realised inside the stressed syllable if it is intonation phrase-final or immediately after the stressed syllable within the focal word if it is not intonation phrase-final. After accent 2, prominence H is realised on the next syllable immediately following the stressed syllable within the focal word, but it can also ‘float’ to the initial syllable of

<sup>2</sup>In recent years, researchers have questioned whether only one of the accent categories is truly lexical, and if so, which of the two that is (Riad, 2006, 2012; Lahiri et al., 2005). This ongoing theoretical debate is not relevant to the current study and will thus not be further discussed.

<sup>3</sup>We adapt the convention of marking a lexical accent with a superscripted number at the end of the relevant word, where ‘1’ refers to accent 1 words and ‘2’ refers to accent 2 words.

<sup>4</sup>The H tone of accent 1 is typically realised on the syllable preceding the stressed syllable, if there is one, and is truncated if there is not (Bruce, 1977).

<sup>5</sup>In addition to marking focus, a floating high tone can be used as a marker of a phrase-initial boundary in Swedish, referred to as ‘initiality accents’ (Myrberg, 2009, 2013) or an initial boundary tone ‘%H’ (Roll et al., 2009).



**Figure 1.** Schematized contours in Swedish without (upper two panels) and including (lower two panels) prominence H. The contours are illustrated as occurring on trochaic target words (e.g., *anden*<sup>1</sup> ('the duck') / *anden*<sup>2</sup> ('the ghost')) in sentence-final position, followed by a low boundary tone. The beginning of the contour is affected by the accentual context and may vary.

the post-focal word (Bruce, 1987; Gussenhoven, 2004). In addition, phonetic changes accompanying the adding of prominence H have been observed, including an increase in pitch range (i.e., difference between the lowest and highest pitch) and word duration in words carrying prominence H, compared to the same words without prominence H (e.g., Heldner, 2001; Myrberg, 2013). It has also been observed that listeners take account of the phonetic effects of adding prominence H in their interpretation of focus in Swedish (e.g., Heldner, 2001).

Typologically, Swedish is similar to West Germanic languages in that speakers rely primarily on prosody for marking focus, instead of relying on syntactic means (e.g., Spanish, Catalan) or on both syntactic means and prosody (e.g., Finnish). Crucially, Swedish and West Germanic languages differ in the transparency of how prosody is used for focus marking, in particular regarding the use of phonological cues (i.e., placement of prominence H, placement of a non-lexical pitch accent – accentuation or pitch accent hereafter). In Swedish, a word is realised with prominence H when focused (e.g., Bruce, 1977, 1998; Ambrazaitis, 2009; Myrberg, 2009). The mapping between the placement of prominence H and focus is thus highly transparent, especially regarding narrow focus and contrastive focus (e.g., Bruce, 1998, 2007). In contrast, in West Germanic languages deaccentuation is typically associated with post-focus but accentuation is used for both focus and pre-focus and to some degree for post-focus (in spontaneous speech). The relation between accentuation and focus is thus not transparent in West Germanic languages (Chen, 2018). On the other hand, the phonetic realisation of prominence H may be more complex in Swedish than the phonetic realisation of accentuation in a word that has no lexicon-related pitch movement in West Germanic languages, because integrating prominence H into the lexical contour requires careful timing of the tones in order to maintain the lexical contrast. Thus, the higher transparency of the mapping between prominence H and focus may make the Swedish system less complex to acquire than the West-Germanic system, but the phonetic realisation of

prominence H may be more complex than the phonetic realisation of commonly used pitch accent types such as the falling pitch accent (H\*L) in the West-Germanic system. The question that arises is how such differences between Swedish and West Germanic languages shape the acquisition of prosodic focus marking in Swedish-speaking children, compared to children acquiring a West Germanic language.

Previous work on children acquiring a West-Germanic language has shown that English-speaking children accent contrastive focus from the age of three to four (e.g., Hornby & Hass, 1970; MacWhinney & Bates, 1978; Wells, Peppé & Goulondris, 2004). Dutch-speaking children are adult-like in their use of accent placement for marking narrow focus at the age of four to five in sentence-final position but only at the age of ten to eleven in sentence-medial position (Chen, 2010, Romøren, 2016). Further, children's use of choice of accent type (e.g., accenting a focal word with a falling or a downstepped fall) is adult-like at the age of four to five in sentence-initial position but not until the age of seven to eight in sentence-medial and -final positions (e.g., Chen, 2011, Romøren, 2016). In addition, Dutch-speaking children's use of phonetic realisation to distinguish narrow focus from pre-focus is not yet adult-like even by the age of eight (Chen, 2009).

Studies on the acquisition of prosody in Swedish are limited in number and have primarily been concerned with the acquisition of the lexical accent contrasts. In these studies, children's production of isolated words of either lexical accent category has been compared to adults' production. However, isolated words are one-word utterances including sentence-level prosodic prominence, which for Swedish entails the use of prominence H, in addition to the lexical pitch accent and a boundary tone. What has actually been compared in analyses on these words in earlier studies is the entire pitch contour of a word including both word and sentence level prosody (i.e., the lower panels of Figure 1). In this sense, studies on the acquisition of lexical accents have implications not only for the production of the lexical accent contrasts but also for the way children produce the combination of each lexical accent with prominence H. For example, Engstrand, Williams and Strömquist (1991) analysed the prosody of isolated trochaic words and non-word vocalisations produced by 17-month olds. They found that accent 1 words did not differ from accent 2 words in pitch changes in the stressed syllable but accent 2 words differed from accent 1 words by carrying a rise in the post-stress syllable that was absent on the accent 1 words (Figure 1). As accent 1 and accent 2 words differ in the prosody of both the stressed and post-stress syllable in adults' production, the authors concluded that 17-month-olds do not yet differentiate the lexical pitch accents in an adult-like way. Ota (2006) reanalysed words containing a visible pitch contour for at least 150 ms in Engstrand et al.'s data and found the children distinguished accent 1 and accent 2 words in both the stressed and post-stress syllable in these words. In the context of the current study, Engstrand et al.'s results also suggest that 17-month-olds can produce prominence H, causing the rise in the post-stress syllable in accent 2 words. Kadin and Engstrand (2005) reported on accent production in 18- to 24-month-old Swedish-speaking children. Comparing on-stress falls and post-stress rises on isolated trochaic words from both lexical accent categories, the authors found that the 24-month-olds consistently produced accent 2 words with a fall on the stressed syllable and a rise on the post-stress syllable, and produced accent 1 words with a relatively small fall on the stressed syllable that kept falling toward the end of the word. Many of the 18-month-olds also produced accent 2 words with a rise in the post-stress syllable, suggesting successful production of prominence H at this age.

However, as the above-mentioned studies primarily concern words produced in isolation and lack a systematic control over information structure, they tell us little about when Swedish-learning children can assign prominence H for focus-marking purposes in the same way as adults do. Even if children may use prominence H to mark focus in isolated words at the age of 17 to 18 months, doing this in syntactically more complex constructions where focal and non-focal constituents need to be distinguished is a far more complex task. It thus remains to be investigated at what age children use prominence H for marking focus in multi-word utterances.

### Research questions and hypotheses

In this study, we have examined the prosodic marking of narrow focus in Swedish-speaking children aged four to eleven years, compared to previous findings on the acquisition of the use of pitch accent for focus marking in children acquiring a West Germanic language. Specifically, we have addressed three research questions on the use of prominence H and its effect on the pitch range of the focal word, and the use of word duration. Our research questions and hypotheses are as follows.

#### *1. Do Swedish-speaking children aged four to eleven differ from Swedish-speaking adults in their use of prominence H in narrowly focal versus non-focal target words?*

Taking into account the differences between Swedish and West Germanic languages in prosodic focus marking, we propose two plausible but opposing hypotheses:

Hypothesis 1: The transparent form-function relationship between prominence H and focus leads to earlier acquisition of the use of prominence H, compared to the acquisition of the use of pitch accent (in terms of both placement and choice of accent) for focus marking in West Germanic languages.

Hypothesis 2: The complexity in the phonetic realisation of the prominence H leads to later acquisition of the use of prominence H, compared to the acquisition of the use of pitch accent for focus marking in West Germanic languages.

#### *2. Do Swedish-speaking children aged four to eleven differ from adults in the effect that adding or avoiding prominence H has on the pitch range of target words?*

In past work on the gradient variation in pitch and duration in children's prosodic focus marking in West Germanic languages, researchers have either examined pitch-related parameters (e.g., mean pitch, maximal pitch, minimal pitch, pitch range) or word duration, often without considering whether the words involved are accented or not (e.g., Müller, Höhle, Schmitz & Weissenborn, 2006). Alternatively, researchers have conducted the analysis only on words that are accented by the same type of pitch accent regardless of focus conditions, to find out whether children nevertheless vary the phonetic realisation of certain pitch accents to distinguish focus conditions (Chen, 2009; Wonnacott & Watson, 2008). However, for Swedish, investigating the pitch characteristics of lexical pitch accents with and without prominence H can inform us whether children's production of these phonological categories is phonetically similar to that of adults, thereby shedding further light on our second hypothesis, concerning the phonetic complexity of Swedish. To our knowledge, there is only one published study on the phonetic realisation of the falling pitch accent H\*L in children aged two to six learning British English, Spanish and Catalan (Astruc, Payne, Post, Vanrell & Prieto, 2013). This study found that the children could produce adult-like peak

alignment at the youngest age in Spanish but to a lesser degree of precision in English and Catalan and they were not fully adult-like in pitch range even at the oldest age tested in English and Catalan. Based on these findings, we may hypothesise that Swedish-speaking four to eleven-year-olds differ from adults in the effect that adding or avoiding prominence H has on the pitch range of target words (Hypothesis 3).

### *3. Do Swedish-speaking children aged four to eleven differ from adults in their use of word duration to mark narrowly focal versus non-focal target words?*

Previous work on children acquiring West Germanic languages have shown that children do not use word duration to distinguish focal words from their non-focal counterparts, even at the age of seven or eight, when both the focal and non-focal words are accented (Chen, 2009). This finding would seem to suggest that Swedish-speaking children aged between four and eight may not use duration in an adult-like way, but that ten-to-eleven-year-old children might. However, research on Mandarin-speaking children showed that duration is used for focus in adult-like ways at the age of four or five, earlier than the acquisition of pitch-related cues, possibly due to the use of pitch also for lexical purposes (Yang & Chen, 2018). Based on the finding on Mandarin-speaking children, we may hypothesise that Swedish-speaking children use word duration for focus-marking purposes in an adult-like way at the age of four or five, earlier than what has previously been described for children acquiring a West Germanic language (Hypothesis 4).

On a methodological note, most previous work on the prosodic realisation of information structure in adults has been concerned with read or strictly controlled speech (for reviews of a large body of literature, see Kügler & Calhoun, 2021; Chen, 2012; but also, for counterexamples, see Bard & Aylett, 1999; Terken & Hirschberg, 1994; de Ruiter, 2010). This makes sense, as detailed prosodic analysis requires strict control over the target words under investigation. It is, however, interesting to note that several studies show that deaccenting given information is less common in spontaneous speech than what has been reported for read speech. For example, Bard and Aylett (1999) showed that second mention mostly did not lead to de-accentuation in their corpus data. This situation is also attested in cases of structural similarity across mentions of a certain referent, which was supposed to increase the likelihood of de-accentuation (Terken & Hirschberg, 1994). In another study on spontaneous speech, de Ruiter (2010) found that adult German speakers de-accented given information much less consistently in spontaneously produced narratives than in read narratives. Task effects (or effects of speech style) on the prosodic marking of information structure are particularly relevant within research on the acquisition of prosodic focus marking, as the assumed adult model should represent the natural repertoire of patterns children hear around them, not just the most 'prototypical' patterns produced by adults in highly controlled speech contexts (see also Grünloh, Lieven & Tomasello, 2015). For this reason, we address our research questions by analysing naturalistic speech from adults and children.

## **Method**

### *The picture-matching game*

Our data were elicited by means of a picture-matching game, adapted from a procedure developed by Chen (2011), also used in Romøren and Chen (2015) and Yang and Chen





Figure 2. Illustration of the experimental setup.



Figure 3. Example of picture set for a trial eliciting narrow focus on the final constituent. The target sentence is 'the dog hides THE TRAIN'.

(2018). In the picture-matching game, the participant's task was to help the experimenter find correct combinations of picture pairs by answering the experimenter's questions about her pictures.

The materials used consisted of three separate sets of pictures, two for the experimenter, and one for the participant (see Figure 2 for the setup and Figure 3 for the picture sets). The experimenter's first set (set 1) was piled face down in front of him or her. These pictures always lacked certain information, e.g., the subject, the verb, the object or all the three pieces of information. The experimenter's second set (set 2) consisted of pictures representing what was missing in set 1, but these were scrambled face up in a box located between the participant and the experimenter. The participant's set (set 3) consisted of pictures that contained all the three pieces of information, and they were piled face down in front of him/her. The pictures in Sets 1 and 3 were placed in the same order to make it easy for the participant to respond to the experimenter's question with the corresponding picture at hand each time.

Each trial was conducted as follows: the experimenter first picked up a picture from his/her set (set 1), drawing the participant's attention to it, describing the picture to establish common ground, and asking a question about the missing information or (in the case of contrastive focus) describing what he/she guessed the complete picture illustrated. The participant then inspected the corresponding picture in his/her own set (set 3) and responded to the experimenter's question or remark. The experimenter then looked for the matching picture in his/her other set (set 2) and formed a pair with his/her picture with missing information.



Prior to the game, two rules were introduced. One was that the participants should always answer in a full sentence; the other was that they should not show their own pictures to the experimenter. The experimenter was trained to speak clearly and naturally and not to deviate from what he/she was supposed to say on each picture to avoid inadvertently introducing changes to information structure. The experimenter was however encouraged to improvise outside the question-answer dialogue (e.g., when a child continued to talk about the scene in a picture after a trial was completed) if this felt natural.

Prior to the picture-matching game, a picture-naming task was conducted to ensure that the participants would use the intended words to refer to the entities in the pictures. In the case of incorrect naming (e.g., calling a lion a tiger), the experimenter explained what the relevant item should be called in this particular game, directing the participants' attention to relevant details of the depicted figure or object (e.g., it is not a tiger, it is a lion, do you see the mane?).

### Research design

This study was conducted as part of a larger project on the acquisition of prosodic focus marking across languages. For the purpose of the larger project, 30 question-answer dialogues were embedded in the picture matching game to elicit 30 SVO sentences in five focus conditions: (A) narrow focus on the subject-noun in sentence-initial position, responding to who-questions; (B) narrow focus on the object-noun in sentence-final position, responding to what-questions; (C) narrow focus on the verb in sentence-medial position, responding to what-does-X-do-to-Y questions; (D) contrastive focus in sentence-medial position, correcting the experimenter's remark about the action; (E) broad focus over the whole sentence, responding to what-happens questions, as illustrated in (1). In the current study, we analysed the data in the first three conditions, allowing comparisons between narrowly focal targets and their non-focal counterparts in sentence-medial and sentence-final positions. The sentence-medial verb was the focus in condition (C) but the pre-focus constituent in condition (B) and the post-focus constituent in condition (A). The sentence-final object-noun was the focus in condition (B) but the post-focus constituent in condition (C).

Examples of question-answer dialogues between the experimenter (E) and the participant (P) within all five conditions are presented below. For the sake of illustration, we use the same answer sentence in each example. To limit use of space, only target questions and answers are given in both Swedish and English.

- A** E: Look! The ball. The ball is in the air. It looks like someone is throwing the ball.

Vem kastar bollen? ('Who is throwing the ball?')

P: [Grodan] kastar bollen ('[The frog] is throwing the ball')

- B** E: Look! The frog. The frog stretches out its arm. It looks like the frog is throwing something.

Vad kastar grodan? ('What is the frog throwing?')

P: Grodan kastar [bollen] ('The frog is throwing [the ball]')

- C E: Look! The frog and the ball. It looks like the frog is doing something to the ball.

Vad gör grodan med bollen? ('What does the frog do to the ball?')

P: Grodan [kastar] bollen ('The frog [is throwing] the ball')

- D E: Look! The frog and the ball. It looks like the frog is doing something to the ball. I will make a guess:

Grodan kokar bollen ('The frog [is cooking] the ball.')

P: Grodan [kastar] bollen ('The frog [is throwing] the ball')

- E E: Look! My picture is very blurry. I cannot see anything clearly.

Vad händer på din bild? ('What happens in your picture?')

P: [Grodan kastar bollen] ('[The frog is throwing the ball]').

Six subject nouns, six transitive verbs and six object nouns were carefully distributed over the five focus conditions so that the answer sentences were all lexically unique. In each set of medial and final constituents, half were accent 1 words and the other half were accent 2 words. All medial and final target words were disyllabic and trochaic. The word list was constructed so that the four possible combinations of accents on medial and final targets (a1 a1, a1 a2, a2 a1, a2 a2) occurred equally frequently for both the medial and final position. We also consulted the Stanford wordbank (<http://wordbank.stanford.edu>) in order to make sure that (a) four-year-old children would know the words, (b) the words would be easy to illustrate, and (c) the words would be sufficiently flexible to combine with the other words without generating semantically odd combinations.

When ordering the stimuli, we ensured that two consecutive trials never represented the same focus condition, and they differed lexically by minimally two constituents. Following these constraints, the experimental trials were arranged into two different stimulus orders, to which the participants were randomly assigned.

## Participants

Twenty-six Swedish-speaking children aged four to eleven and ten Swedish-speaking adults participated in this study. The participants were divided into four age groups: four to five years, seven to eight years, ten to eleven years, and adults (Table 1). The choice of including children between four and eleven was based on the findings from previous work on prosodic acquisition that children at four to five are not adult-like in their production of prosodic focus marking and they undergo notable development between the age of four and eleven (see Chen, 2018, for a recent review).

The children were recruited from kindergartens and schools in Stockholm, and parents gave written consent for their children to be tested and for their speech to be recorded. Parents also filled in a form providing information about the children's language background, ensuring that all our participants were native language

**Table 1.** Background information of the participants.

Age group	N	Age range (years, months)	Age mean (years, months)	Gender
4-5	10	4;3-5;6	5;0	6 male, 4 female
7-8	8	7;6-8;8	8;3	5 male, 3 female
10-11	8	10;0-11;0	10;6	4 male, 4 female
Adults	10	20;0-43;10	27;2	5 male, 5 female

speakers of Swedish. The adult participants were recruited at the Royal Institute of Technology (KTH) in Stockholm, they were all university students and native speakers of Swedish. None of the participants reported to have had any history of language disorders, hearing problems or other known developmental disorders.

### Procedure

Each recording session included both the picture naming task and the picture-matching game. The picture-naming task took a few minutes; the picture-matching game took around 25 minutes on average, including instruction and practice trials.

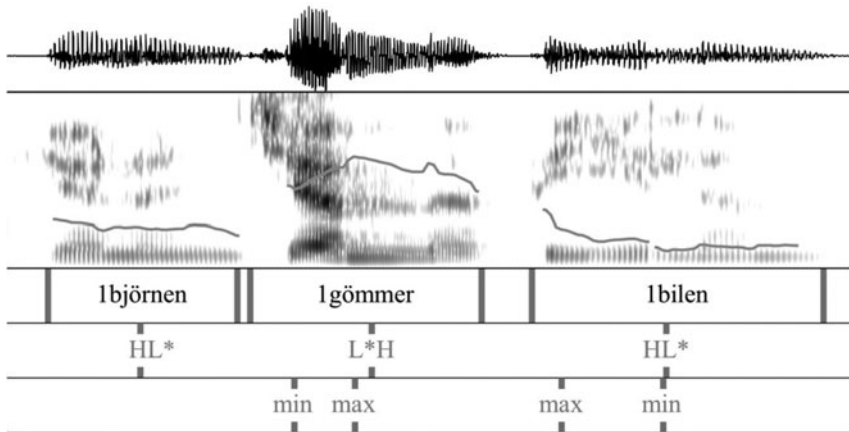
The participants were recorded individually in a quiet room, either in schools or kindergartens, or at the KTH in Stockholm, Sweden. All audio recordings were made using a portable ZOOM H1 handy recorder, with a 44.1 kHz sampling rate and 16-bit accuracy. The adults were told that the experiment was also conducted with children and was thus simple by nature, and that for the sake of consistency, the experimenter would play the game in the same way that she did with the children.

### Data selection and coding

The audio recordings were segmented into trials using Praat (Boersma & Weenink, 2010). After this, all trials were evaluated, and only responses following the scripted speech context were included in the analysis. The choice of being strict in the inclusion of responses ensured that the prosodic comparisons were made across the same target words, and that the experimental conditions were properly controlled for. In Table 2 below we report the inclusion rates for the four age groups and the number of responses excluded in each category. The category 'disfluencies' refers to cases where a response contained hesitations, repairs or filled pauses. The category 'unsuitable context' refers to cases where the speech context could not be completely controlled. For example, if a child started talking about the target focal word (e.g., the baker) before the experimenter got to ask her question (e.g., who is washing the ball?), possibly making the experimenter's question a bit artificial, since she already heard the answer (e.g., The baker is washing the ball). In such cases, it is unclear whether the baker should be assumed to be 'new' or 'informative' to the same extent as in cases where it had not already been introduced. The category 'non-target' involves responses that contained the wrong words, lacked certain constituents or had non-target constituents added to them. Finally, the category 'noise/overlap' refers to instances where a response contained noise, laughter or speech overlaps, making the recording unfit for analysis. The total number of responses analysed in the current

**Table 2.** Overview of excluded responses by group and category.

	4-5	7-8	10-11	Adults	Total
<b>Total elicited</b>	300	240	240	300	1080
<b>Total included</b>	219	203	178	249	849
<b>Disfluencies</b>	21	5	28	32	86
<b>Unsuitable context</b>	20	4	5	3	32
<b>Non-target</b>	26	20	16	8	70
<b>Noise/ overlap</b>	14	8	12	5	39
<b>Other</b>	0	0	1	3	4
<b>% included</b>	73	85	74	83	79

**Figure 4.** Illustration of the annotation procedure. The target sentence is *björnen* ('the bear') *gömmar* ('hides') *bilen* ('the car'), with focus in medial position (on the verb). Note that in our coding we conventionally annotated accent 1 words without prominence H as 'HL\*' and with prominence H as 'L\*H'.

study was 849 (79%). Since comparisons were made on both sentence-medial and sentence-final target words, this rendered a total number of 1698 target words analysed.

Included responses were orthographically transcribed and manually segmented, using the Praat software (Boersma & Weenink, 2010). When segmenting, we combined auditory perception with visual information such as changes in the waveform and formant transitions, following the standard procedure (Turk, Nakai, & Sugahara, 2006). Conventions were established for how to segment the words at particularly challenging boundaries. To illustrate, word-onset plosives following a preceding word, such as the 't' in 'kastar tårtan', were consistently segmented right before the burst, because the 'r' in 'kastar' was often elided or realized by means of changing the articulation of the 't' (see also Figure 4). Further, all segmentation was checked at least twice and revised if necessary, first during the initial round of transcription, and then during the coding for prominence H.

The medial and final target words were manually coded for whether or not the target words carried prominence H on the basis of visual inspection of the pitch contours and

auditory impression, without access to information about the focus condition. A word was considered carrying prominence H if it had a one-peak contour in the case of accent 1 and a two-peak contour in the case of accent 2, as shown in Figure 1. Prominence H was not separately labelled and singled out in the pitch contour. The coding was checked at least twice for each child. Identifying words spoken with prominence H was mostly unproblematic when both visible contours and auditory information were available, in line with Strangert and Heldner's (1995) observations.

For the analysis of pitch range, markers were manually placed on the minimum and maximum pitch points within the target word, aided by the function for detecting maximum and minimum pitch points in Praat. For sentence-medial accent 2 targets, the floating prominence H could sometimes be realized as late as on the following word. In that case, the maximum point within the boundaries of the medial target word was nevertheless used as the pitch maximum. When placing the pitch markers, we avoided the beginning, the end and the transition between segments because pitch values are subjected to micro fluctuations in these places. Furthermore, duration measures were extracted based on the word segmentation. The coding of presence/absence of prominence H and measurements of pitch range and word duration were automatically extracted from Praat using custom written scripts. The data were subsequently checked for extraction errors.

## Analysis and results

### *Analysis procedure*

Three separate analyses were conducted for this study. The first analysis addressed our first research question on the effect of narrow focus on prominence H. The second addressed our second research question, on the use of pitch range for marking focus. Finally, in the third analysis we looked at the use of word duration for marking focus. In order to control for position effects, our comparisons were always between focal and non-focal renditions of the same 6 target words in the same sentence position. The non-focal renditions could be pre-focus or post-focus.

All analyses were done by using mixed-effects modelling in the program R (R core team, 2014, including the lme4 package (Bates, Maechler, Bolker, & Walker, 2015). Fixed factors in the analyses were 'focus' (two levels: narrow focus vs. post focus or narrow focus vs. pre-focus), 'group' (four levels: four-to-five-year-olds, seven-to-eight-year-olds, ten-to-eleven-year-olds, adults) and 'lexical accent' (hereafter 'lex') (two levels: accent 1, accent 2). Random factors were 'participant' and 'item'.<sup>6</sup>

The statistical procedure used was the following: we first started out with a baseline model (hereafter model 0) in which only the intercept was included. From there we extended the model in a step-wise fashion by first adding the factor 'focus' in model 1, then adding the factor 'group' in model 2, the factor 'lex' in model 3, followed by two and three-way interactions (see Table 4). The factor 'group' included four levels (the four groups), coded into dummy variables using adults as a baseline to which the other groups were compared.

In order to assess the improvement of the model fit from models 0 through 7 we used R's 'anova' function to compare pairs of models (e.g., Quené & van den Bergh,

<sup>6</sup>The two random factors were originally added to all analyses, but for the analyses of prominence H the almost complete separation made it impossible to add them to the models.

**Table 3.** Model build-up procedure.

Model	Factor added
0	
1	Focus
2	Group
3	Lexical accent ('lex')
4	Focus x Group
5	Focus x Lex
6	Group x Lex
7	Focus x Group x Lex

**Table 4.** Model summary, prominence H, sentence-medial pre-focus vs. narrow focus comparison

Factor	b-value	std. error	p-value
Intercept (adults, narrow focus)	2.535	0.288	0.000***
Pre-focus	-4.954	0.408	0.000***

2008). When the model comparison showed a decrease in the -2 log likelihood and a p-value below 0.05, this was taken as evidence that the added parameter (main effect or interaction) significantly<sup>7</sup> improved the model fit. When the best model was established, model summaries were used in order to obtain p-values for the relevant parameters. In cases where an interaction effect was found, we re-levelled the model summary of the best model, in order to obtain estimates of the main effect of a specific factor within each level of the other factor(s), i.e., the effect of 'focus' within each 'group' or the effect of 'focus' within each 'lex'. Using the full model when re-levelling, rather than subsetting the data, made it possible for the random effects to be estimated properly for the re-levelled models.

### The use of prominence H for marking narrow focus

For the analysis of absence or presence of prominence H, we built binomial logistic regression models (GLMs) using R, following the procedure described above. The outcome variable was categorical, consisting of the binary outcome 'presence vs. absence of prominence H on the target word'. For clarity, we briefly present the results of the model comparisons, focusing on the effects found for the model with the best fit. Comparable analyses were performed for sentence-medial and sentence-final target words.

<sup>7</sup>For some of the model comparisons, complete separation (i.e. prominence H at 100% within some level of 'focus', 'group' or 'lex') made it impossible to run the analyses. In such cases we slightly manipulated the data by adding one instance of the minority pattern to each sub-level. This is not expected to have had any effect on the final results, but it made it possible to run the models in order to statistically test our hypotheses.

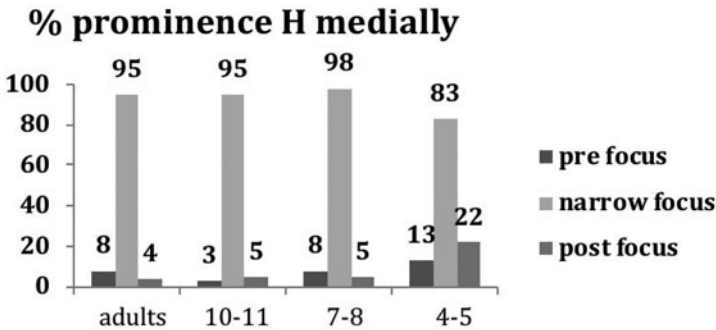


Figure 5. Percentage of prominence H on sentence-medial targets under pre-focus, narrow focus and post-focus, across groups.

Table 5. Model summary, prominence H, sentence-medial narrow focus vs. post-focus comparison.

Factor	b-value	std. error	p-value
Intercept (adults, narrow focus)	2.853	0.594	0.000***
Post-focus	-6.130	0.934	0.000***
Group 10-11	0.092	0.937	0.922
Group 7-8	0.811	1.174	0.490
Group 4-5	-1.243	0.724	0.086
Post-focus * group 10-11	0.241	1.387	0.862
Post-focus * group 7-8	-0.530	1.556	0.734
Post-focus * group 4-5	3.281	1.077	0.002**

### Sentence-medial position

The distribution of prominence H across narrow focus, pre-focus and post-focus sentence-medially in all four groups is illustrated in Figure 5. As can be observed, prominence H was the dominant pattern for focal targets, as compared to both pre and post-focal targets. The pattern was slightly less consistent for the four-to-five-year-olds.

Our first analysis compared the use of prominence H between narrow focus and pre-focus. Model comparisons showed that only the main effect 'focus' improved model fit, thus group differences observed in Figure 5 did not reach significance for this comparison. Summarizing the best model (Table 4), we observe that pre-focal status generally decreased the likelihood of prominence H ( $p < 0.001$ ) on the medial target words for all groups and both accents.

Our second analysis compared the use of prominence H between narrowly focal and post-focal targets. Comparing regression models showed the main effect 'focus' ( $p < .001$ ) and the interaction 'focus x group' ( $p < 0.005$ ) significantly improved the model fit, thus for this model the group differences did reach significance. Summarizing our best model (Table 5), we see that post-focal status significantly decreased the likelihood of prominence H on the target word ( $p < 0.001$ ). The



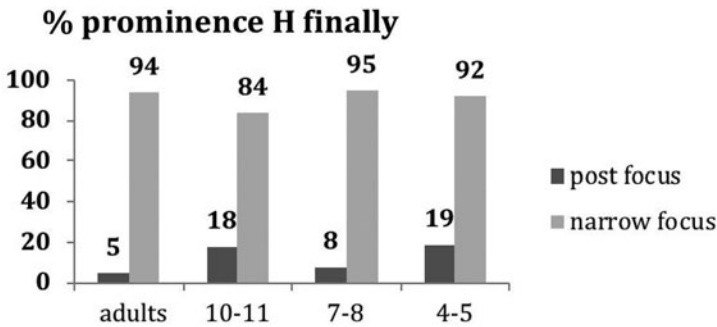


Figure 6. Percentage of prominence H on sentence-final targets under post-focus and narrow focus, across groups.

Table 6. Model summary, prominence H, sentence-final analysis.

Factor	b-value	std. error	p-value
Intercept (narrow focus)	2.535	0.288	0.000***
Post-focus	-4.954	0.408	0.000***

interaction effect showed the effect of focus differed significantly between the four-to-five-year-olds and the adults, with a weaker effect observed in the four-to-five-year-olds ( $p < 0.005$ ) than in the adults. As can be seen in Figure 5, the adults, ten-to-eleven-year-olds and seven-to-eight-year-olds produced prominence H in around 95% of the time on narrowly focal target words, and in less than 10% of the time on both pre- and post-focal target words. The four-to-five-year-olds used prominence H to a lesser degree than the other groups in the narrow focus condition (83%) and to a larger degree in the post-focus condition (22%).

### Sentence-final position

The distribution of prominence H across narrow focus and post-focus sentence-finally in all four groups is shown in Figure 6. As can be observed, prominence H was the dominant pattern for focal targets, as compared to post-focal targets. The pattern was slightly less consistent for both ten-to-eleven-year-olds and four-to-five-year-olds.

For sentence-final position, we had no pre-focal condition to which the narrow focus condition could be compared, as the final target words could not occur in a pre-focus position. Comparing regression models showed that only the main effect of focus significantly improved the model fit ( $p < 0.001$ ), thus any observable group differences did not reach significance. The summary of the best-fit model (Table 6) showed that post-focal status significantly decreased the likelihood of prominence H on our final target words, similar to what was found sentence-medially, as illustrated in Figure 5.

### Pitch range analysis of focal H for narrow focus

For the analysis on pitch range, we built and compared linear mixed effect models (LMMs) in R (R Core Team, 2014). The outcome variable was continuous, involving

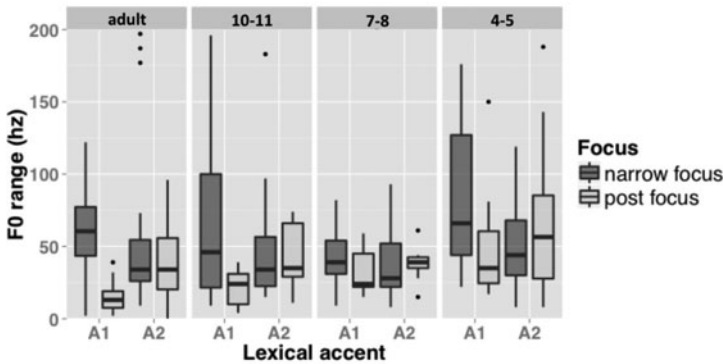


Figure 7. Pitch range on sentence-medial targets under narrow and post-focus, by group and lexical accent.

Table 7. Model summary, pitch range, sentence-medial analysis.

Factor	b-value	std. error	p-value
Intercept (narrow focus, accent 1)	66.746	6.448	0.000***
Post-focus	-34.397	6.266	0.000***
Accent 2	-18.092	5.855	0.015*
Post-focus * accent 2	34.592	8.567	0.000***

measures of pitch range in Hz (hereafter pitch range) within medial and final target words. The model comparison was done in a similar way to the GLM modelling reported in Section 3.3. In our LMMs our empty baseline model (hereafter model 0) included the crossed random effects of ‘item’ (i.e., the 6 target words appearing in each sentence position) and ‘participant’ (our 36 participants). From there we extended the model in a step wise fashion, as illustrated in Table 3.

### Sentence-medial position

The pitch range measurements in sentence-medial position are illustrated by means of a boxplot in Figure 7. Judging by this figure the difference in pitch range between focal and post-focal targets in sentence-medial position appeared to be more consistent for accent 1 words than for accent 2 words across age groups<sup>8</sup>.

Comparing mixed effect models revealed that both focus ( $p < 0.001$ ) and the interaction between focus and lex ( $p < 0.001$ ) improved the model fit. The summary in Table 7 shows that across age groups narrowly focal targets generally had a larger pitch range than their post-focal counterparts, but this effect was smaller in accent 2 words than in accent 1 words.

Exploring the effect of focus within each accent category showed that assigning prominence H significantly increased the pitch range in accent 1 words ( $p < 0.001$ ), whereas no such an effect was observed in the accent 2 words ( $p = 0.974$ ) (Figure 6).

<sup>8</sup>It should be noted that boxplots of non-normalized pitch measures may not be easy to interpret, but that our inclusion of ‘participant’ and ‘target word’ as random factors in the statistical analysis were expected to have taken care of some of the variance.

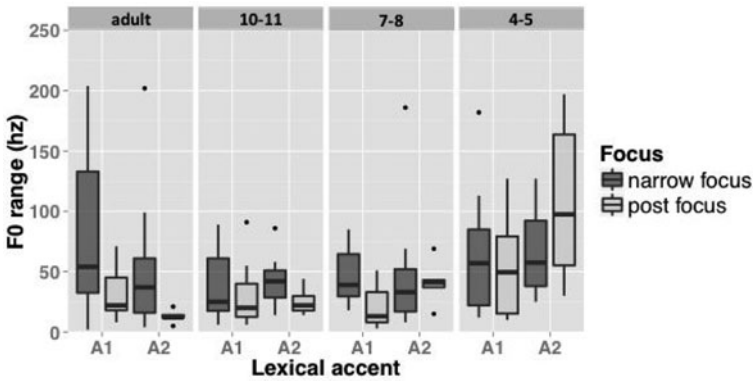


Figure 8. Pitch range on sentence-final targets under narrow and post-focus, by group and lexical accent.

Table 8. Model summary, pitch range, sentence-final analysis.

Factor	b-value	std. error	p-value
Intercept (narrow focus)	57.067	6.481	0.000***
Post-focus	-20.012	5.783	0.001**

Averaged across groups, the accent 1 targets were spoken with a mean pitch range of 29 Hz when being post-focal (lacking prominence H) and 69 Hz when being focal (carrying prominence H). The accent 2 targets were spoken with a mean pitch range of 49 Hz when being post-focal (lacking prominence H) and 48 Hz when being focal (carrying prominence H). Even if the boxplot in Figure 6 indicates a larger effect of focus on accent 1 words in adults than in children, no main effects or interactions involving group reached significance in our analyses.

### Sentence-final position

The pitch range measurements in sentence-final position are shown in Figure 8. As can be observed, the pitch range on focal versus non-focal targets in sentence-final position was rather variable, but there was a general tendency for focal targets to have a larger pitch range than non-focal targets.

Our analysis revealed that only the main effect ‘focus’ ( $p < 0.001$ ) significantly improved the model fit. The summary of the best-fit model (Table 8) showed that narrowly focal targets generally had a larger pitch range than their post-focal counterparts, confirming the pattern in Figure 8. Again, no main effects or interactions involving group were observed.

### The use of word duration for marking narrow focus

The analysis of word duration was conducted in similar ways as the analyses presented above, with comparisons between focus and post-focus in both sentence-medial and final position.

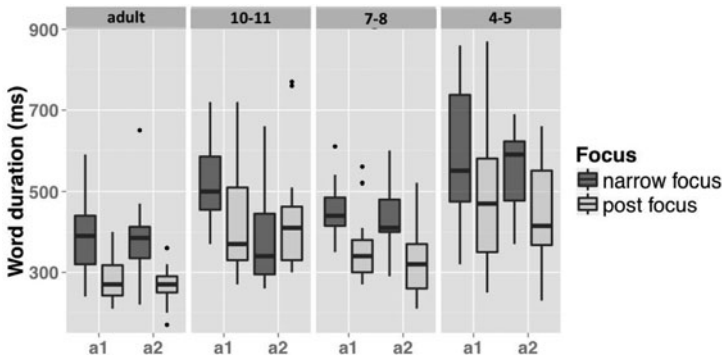


Figure 9. Word duration on sentence-medial targets.

Model comparisons between linear mixed effect models (LMMs) were made in line with the procedure used in the pitch analysis above (see also Table 3). For all models the outcome variable was continuous, involving raw measures of word duration on medial or final target words.

### Sentence-medial position

The duration measurements of the sentence-medial target words are shown in Figure 9. The focus targets had by and large a longer duration than their post-focal counterparts. The pattern appeared to be particularly consistent in the adults, four- to five-year-olds and seven- to eight-year-olds.

Model comparisons showed that the main effects of 'focus' ( $p < 0.001$ ), and 'group' ( $p < 0.001$ ), the two-way interaction between 'group' and 'focus' ( $p < 0.01$ ) and the three-way-interaction between 'focus', 'group' and 'lex' ( $p < 0.005$ ) improved the model fit. The best model is summarized in Table 9. As can be seen in Figure 9, the main effect of 'focus' consisted in an increase in word duration on the medial targets in focus than in post-focus. In addition, main effects of group consisted in generally longer word durations in the ten-to-eleven and four-to-five-year-olds than the adults. Furthermore, there was an interaction between 'group' and 'lex': accent 1-words were longer than accent 2-words in the ten-to-eleven-year-olds, whereas this was not the case for the adults. Finally, the three-way-interaction was caused by the adults and ten-to-eleven-year-olds differing in the effect of 'focus' by 'lex'; focus lengthened the duration of both accent 1 and accent 2 words in the adults and the younger children, but this effect was only present on the accent 1 words in the ten-to-eleven-year-olds.

Exploring the effect of focus within each group separately showed a main effect of focus in all but the ten-to-eleven-year-olds. Being focal led to an increase in the word duration in the adults ( $p < 0.001$ ), the seven-to-eight-year-olds ( $p < 0.001$ ) and the four-to-five-year-olds ( $p < 0.001$ ), but not in the ten-to-eleven-year-olds ( $p = 0.548$ ). Comparing the mean durations in narrowly focal to post-focal target words, we found that being focal increased the duration by 106 ms (28%) in the adults, 104 ms (23%) in the seven-to-eight-year-olds and by 89 ms (15%) in the four-to-five-year-olds, whereas narrowly focal words were on average 11 ms (2%) longer than post-focal ones in our ten-to-eleven-year-olds, when both lexical accents were included.

Table 9. Model summary, duration, 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 10–11	0.136	0.041	0.001**
Group 7–8	0.060	0.041	0.146
Group 4–5	0.206	0.039	0.000***
Accent 2	-0.012	0.034	0.725
Post-focus * group 10–11	-0.013	0.041	0.740
Post-focus *group 7–8	0.003	0.041	0.945
Post-focus * group 4–5	0.023	0.039	0.559
Group 10–11 * accent 2	-0.137	0.040	0.001***
Group 7–8 * accent 2	0.006	0.040	0.878
Group 4–5 * accent 2	-0.007	0.040	0.870
Post-focus * group 10–11 * accent 2	0.198	0.057	0.001***
Post-focus * group 7–8 * accent 2	-0.014	0.056	0.803
Post-focus * group 4–5 * accent 2	-0.027	0.055	0.627

### *Sentence-final position*

The duration measurements of the sentence-final target words are shown in [Figure 10](#). The focus targets appeared to be longer than their post-focal counterparts in the adults, four- to five-year-olds and seven- to eight-year-olds but not in the ten- to eleven-year-olds.

Model comparisons showed the best model included main effects of ‘focus’ ( $p < 0.001$ ), and ‘group’ ( $p < 0.001$ ) and an interaction between the two ( $p < 0.005$ ). The best model is summarized in [Table 10](#). The main effect of focus consisted in an increase in word duration on the sentence-final target words in focus as compared to post-focus. The group effects consisted in a longer word duration in the seven-to-eight and four-to-five-year-olds than the adults. Finally, to understand the the interaction effect between focus and group, subsequent analyses on the main effect of focus within each group showed that focus led to a significant increase in word duration in the adults ( $p < 0.005$ ), the four-to-five-year-olds ( $p < 0.001$ ), a decrease in the ten-to-eleven-year-olds ( $p < 0.05$ ) but at best a marginally significant increase in word duration in the seven-to- eight-year-olds ( $p = .083$ ). Comparing the duration of narrowly focal targets to post-focal ones, we found that narrow focus lead to an increase in word duration by 72 ms (18%) in the adults, 50 ms (10%) in the seven-to-eight-year-olds and by 140 ms (21%) in the four-to-five-year-olds, whereas narrowly focal words were on average 67 ms (16%) shorter than their post-focal counterparts in the ten-to-eleven-year-olds.

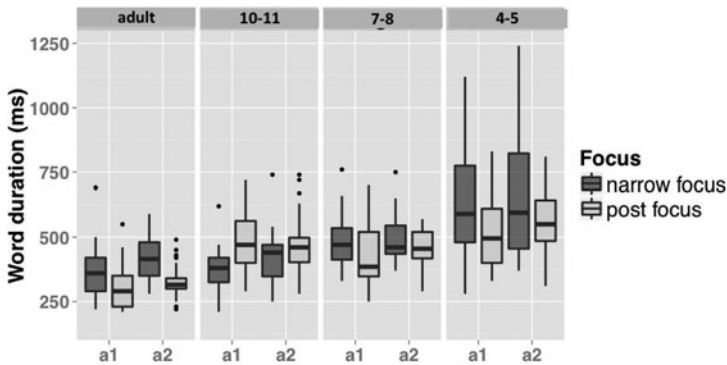


Figure 10. Word duration on sentence-final targets.

Table 10. Model summary, duration, 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 10-11	0.019	0.042	0.648
Group 7-8	0.098	0.041	0.023
Group 4-5	0.282	0.040	0.000
Post-focus * group 10-11	0.140	0.034	0.000
Post-focus * group 7-8	0.030	0.033	0.357
Post-focus * group 4-5	-0.066	0.033	0.045

## General discussion and conclusions

In this paper we have presented three separate analyses on children's use of prosody for marking focus, compared to adults' production. The first analysis concerned the question of whether Swedish-speaking children between four and eleven differ from Swedish-speaking adults in the way they use prominence H to mark narrowly focal versus non-focal target words. Our analysis revealed remarkably similar patterns across the four groups; the participants predominantly added prominence H to constituents under narrow focus, and they predominantly avoided this tone both pre- and post-focally, in line with previous descriptions of adult Swedish (i.e., Myrberg, 2013; Ambrazaitis, 2009; Riad, 2014). At the same time, interaction effects showed that whereas the seven-to-eight and ten-to-eleven-year-olds performed in line with adults in both sentence-positions, the four-to-five-year-olds differed from the adults sentence-medially by showing a slightly weaker differentiation between narrow focus and post-focus. Thus, Swedish-speaking children obtain full adult mastery of prominence H for focus by the age of four to five in sentence-final position and by the age of seven to eight in sentence-medial position, but they are fairly consistent in differentiating focus from non-focus at the age of four-to-five in sentence-medial

position. These results suggest earlier acquisition of the use of prominence H in focus marking in Swedish than the acquisition of the use of pitch accent in West Germanic languages, supporting our Hypothesis 1, not Hypothesis 2.

The earlier mastery of prominence H for marking narrow focus in sentence-final position than in sentence-medial position may be related to the fact that the IP-final position is also the default position for maximum prominence in broad focus utterances, leading to abundant exposure to the production of prosodic prominence in sentence-final position in the input. Also, this position may be particularly salient from a prosodic point of view, as important prosodic functions like turn-taking or interrogativity are typically marked sentence-finally (e.g., House, 2003, on Swedish). Alternatively, the sentence-position related-difference can also be interpreted as an effect of grammatical category, considering that the medial targets were always verbs and the final targets were always objects. As suggested by Röhr, Baumann and Grice (2015), it may be more common to mark focus on referents than on actions. Such a tendency would give children less experience with focus-marking on verbs than on nouns, and may explain why the children in our study marked focus more consistently on the sentence-final nouns than on the sentence-medial verbs.

In order to explore the hypothesis on the complexity of the contour in Swedish (Hypothesis 3), we examined whether Swedish-speaking children between four and eleven differ from adults in the effect that adding or avoiding prominence H has on the pitch range of a word. Based on the analyses presented in section 3.2, no significant differences were observed between adults and children in the way pitch range was manipulated across the relevant contour categories in either sentence position. Thus, we can reject the hypothesis that combining lexical and post lexical tones to mark focus is particularly challenging for children, at least phonetically speaking.

Nevertheless, we did find different pitch range effects for accent 1 and accent 2 words sentence-medially, where adding prominence H increased the pitch range on accent 1 words but this effect was not found for accent 2 words. This result is unexpected, as previous work has indicated that the pitch range on prominence H on accent 2-words is typically larger than on the lexical H (e.g., Ambratzitis, 2009). Our results may be related to the fact that we chose to analyse pitch range within the target words only, which may be problematic for accent 2 words where prominence H may float across word boundaries. Even if all our targets were trochaic and thus had 'sufficient space' for realizing prominence H on accent 2 words (i.e., *målar*<sup>2</sup>), the pitch might still be rising into the following word (i.e., *blomman*<sup>2</sup>). Consequently, the pitch maximum obtained for these cases might be lower than what it should be, blurring the expected pitch effects of prominence H. Sentence-finally, prominence H is expected to be realised within the target word regardless of lexical accent, and here our results showed the predicted effect of adding prominence H for both adults and children. In retrospect, our choice to include a systematic accent variance on both initial, medial and final targets may have contributed to covering possible differences between adult and child productions. With a more constrained phonetic context (keeping the verb-object combinations identical) more phonetic detail might be investigated, but with the consequence of a slightly less game-like experimental setup.

In our third analysis, we found that focal status systematically increased the word duration on both medial and final target words in the adults, the seven-to-eight-year-olds, and the four-to-five-year-olds, whereas the ten-to-eleven-year-olds differed from the adults both sentence-medially and sentence-finally. In sentence-medial



position, the ten-to-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 ten-to-eleven-year-olds differed from the adults by producing post-focal target words with longer duration than focal ones across both accent types. It can thus be concluded that four-to-five- and seven-to-eight-year-olds are adult-like in their use of word duration to distinguish narrow focus from post-focus but ten-to-eleven-year-olds are not adult-like, partially supporting Hypothesis 4.

The finding on ten-to-eleven-year-olds is unexpected. Tentatively, we suggest that task effects may at least to some degree explain the finding. The picture matching game was constructed in such a way that it would suit the youngest participants as well as older children. It could not be ruled out that some ten- to eleven-year-olds might not find the game very engaging. We noticed that occasionally some ten-to-eleven-year-olds seemed to signal their slight disinterest in the game by slowing down their speech rate on the final target words. Such lengthening seemed particularly common when the children had already provided the most important information, thus post-focal targets were more prone to such 'unengaging' prosody than focal ones. It may be the case that the tendency for the ten-to-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. Even if the task can be assumed to be similarly appealing to the adults to the ten-to-eleven-year-olds, they were told in advance that this game was constructed for children, and might thus have a different perspective on the task than the ten-to-eleven-year-olds. We also did not observe the use of slower speech rate when annotating the adult data.

To sum up, our study is the first to show that Swedish-speaking children are remarkably adult-like in their use of prominence H for focus. Their use of prominence H seems to reach adult proficiency already at four to five sentence-finally and by seven to eight sentence-medially, with robust distinction between focus and non-focus observable already at four to five in both sentence positions. In contrast, Dutch children do not reach adult proficiency in using accent placement or accent type before seven to eight, and at four to five the distribution of accentuation for focus is less clear than those found for Swedish children (Romøren & Chen, 2015; Chen, 2009).

Combining our analysis of duration and pitch, it seems 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. This suggests 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 of children acquiring a West Germanic language in the use of duration manipulations. However, it should also be acknowledged that pitch range is a rather crude measurement to be used as a diagnostic for the effect of adding prominence H because contours similar in pitch range can still differ in terms of other pitch parameters, such as timing and slope of pitch peaks and valleys. Future comprehensive prosodic analysis is needed to fully establish how adult-like Swedish-speaking children are in the phonetic realisation of prominence H at different ages.

Finally, our study underlines the importance of including children younger than four years when studying the acquisition of prosodic focus marking. For Swedish, it seems we need to study the production of children younger than five in order to see

whether the use of duration for focus marking develops differently from the use of prominence H. By simplifying the picture-matching game so that fewer conditions are elicited on simpler structures (e.g., adjective noun pairings) (Chen & Fikkert, 2007), it may be possible to use the picture-matching game with younger children. Maintaining a fairly similar setup as the one used here would make it possible to compare results from younger children to our data on children between four and eleven, thereby obtaining a clearer understanding of how children develop their ability to prosodically highlight focal information.

**Supplementary material.** For Supplementary Material accompanying this paper, visit <https://doi.org/10.1017/S0305000920000847>

**Acknowledgements.** We would like to thank David House, Jens Edlund, Sofia Strömbergsson, Christina Tärnander and their colleagues at the Royal Institute of Technology in Stockholm for kindly hosting the first author's research visits. This work was supported by a VIDI grant awarded to the second author by the Dutch Research Council (NWO) (grant number 276-89-001).

## References

- Ambrazaitis, G.** (2009). Nuclear intonation in Swedish: Evidence from experimental-phonetic studies and a comparison with German. PhD thesis, Lund University.
- Arnhold, A., Chen, A., & Järviqvist, J.** (2016). Acquiring complex focus-marking: Finnish four- to five-year-olds use prosody and word order in interaction. *Frontiers in Psychology*, 7:1886. doi: 10.3389/fpsyg.2016.01886
- Astruc, L., Payne, E., Post, B., Vanrell, M. M., & Prieto, P.** (2013). Tonal targets in early child English, Spanish, and Catalan. *Language and Speech*, 56, 229–253.
- Bard, E. G., & Aylett, M. P.** (1999). The dissociation of deaccenting, givenness, and syntactic role in spontaneous speech. In *Proceedings of the XIVth international congress of phonetic sciences* (Vol. 3, pp. 1753–6). University of California Berkeley, CA.
- Bates, D., Maechler, M., Bolker, B., & Walker, S.** (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48.
- Boersma, P., & Weenink, D.** (2010). Praat: Doing phonetics by computer [Computer program], available at <http://www.fon.hum.uva.nl/praat/>
- Bruce, G.** (1977). *Swedish word accents in sentence perspective*. Lund: Liber Läromedel.
- Bruce, G.** (1987). How Floating is Focal Accent? In: K. Gregersen and H. Basbøll (eds.), *Nordic Prosody 4*. Odense: Odense University Press. 41–49.
- Bruce, G.** (1998). *Allmän och svensk prosodi*. Lunds Universitet: Institutionen för Lingvistik.
- Bruce, G.** (2007). Components of a prosodic typology of Swedish intonation. In T. Riad & C. Gussenhoven (Eds.), *Tones and tunes. Typological and comparative studies in word and sentence prosody*, 1, pp. 113–146. Berlin: Mouton de Gruyter.
- Chafe, W. L.** (1976). Givenness, contrastiveness, definiteness, subjects, topics and point of view. In Li, N. C. (Ed.), *Subject and Topic*, pp. 27–55. Academic press.
- Chen, A.** (2009). The phonetics of sentence-initial topic and focus in adult and child Dutch. In M. C. Vigarito, S. Frota, & M. J. Freitas (Eds.), *Phonetics and phonology: Interactions and interrelations*, pp. 91–106. Amsterdam: John Benjamins.
- Chen, A.** (2010). Is there really an asymmetry in the acquisition of the focus-to-accentuation mapping? *Lingua*, 120(8), 1926–1939.
- Chen, A.** (2011). Tuning information packaging: Intonational realization of topic and focus in child Dutch. *Journal of Child Language*, 38, pp. 1055–1083.
- Chen, A.** (2012). Prosodic investigation on information structure. In M. Krifka, & R. Musan (eds.) *The expression of information structure* (pp. 251–286). Berlin: Mouton de Gruyter.
- Chen, A.** (2018). Get the focus right across languages: Acquisition of prosodic focus-marking in production. In P. Prieto & N. Esteve-Gibert. *The Development of Prosody in First Language Acquisition*. Philadelphia: John Benjamins Publishing.

- Chen, A., & Fikkert, P. (2007). "Dutch 3-year-olds' use of intonation in marking topic and focus". Poster presented at Generative Approaches to Language Acquisition (GALA), Barcelona, 6–8 September, 2007.
- De Ruiter, L. E. (2010). *Studies on intonation and information structure in child and adult German*. PhD Thesis, Max Planck Institute for Psycholinguistics.
- Engstrand, O., Williams, K., & Strömquist, S. (1991). Acquisition of the tonal word accent contrast. *Actes du XIIème Congrès International Des Science Phonétiques*, pp. 324–327.
- Grünloh, T., Lieven, E., & Tomasello, M. (2015). Young children's intonational marking of new, given and contrastive referents. *Language Learning and Development*, 11, pp. 95–127.
- Gussenhoven, C. (2004). *The phonology of tone and intonation*. Cambridge: Cambridge University Press.
- Gussenhoven, C. (2007). Types of focus in English. In C. M. Lee, M. Gordon, & D. Büring (Eds.). *Topic and focus: Cross-linguistic perspectives on meaning and intonation*, pp. 83–100. Dordrecht: Springer Publishing.
- Halliday, M. A. K. (1967). *Intonation and grammar in British English*. Berlin: Mouton de Gruyter.
- Heldner, M. (2001). *Focal accent-f0 movements and beyond*. PhD Thesis, Umeå University.
- Hornby, P. A., & Hass, W. A. (1970). Use of contrastive stress by preschool children. *Journal of Speech, Language, and Hearing Research*, 13, pp. 395–399.
- House, D. (2003). Perceiving question intonation: the role of pre-focal pause and delayed focal peak. *Proceedings of the 15th International Congress of Phonetic Sciences (ICPHS)*, pp. 755–758.
- Ito, K. (2014). Children's pragmatic use of prosodic prominence. In D. Matthews. *Pragmatic development in first language acquisition*. Philadelphia: John Benjamins Publishing.
- Ito, K. (2018). Gradual development of focus prosody and affect prosody comprehension. In P. Prieto & N. Esteve-Gibert. *The Development of Prosody in First Language Acquisition*. Philadelphia: John Benjamins Publishing.
- Kadin, G., & Engstrand, O. (2005). Tonal word accents produced by Swedish 18- and 24-month-olds. *Proceedings of Fonetik 2005*, pp. 67–70.
- Krifka, M. (2008). Basic notions of information structure. *Acta Linguistica Hungarica*, 55, pp. 243–276.
- Krifka, M., & Musan, R. (2012). *The expression of information structure*. Berlin: Mouton de Gruyter.
- Kügler, F., & Calhoun, S. (2020). Prosodic encoding of information structure: A typological perspective. In C. Gussenhoven & A. Chen (eds) *The Oxford Handbook of Language Prosody*. Oxford: OUP.
- Ladd, D. D. (1980). *The structure of intonational meaning*. Bloomington: Indiana University Press.
- Lahiri, A., Wetterlin, A., & Jönsson-Steiner, E. (2005). Lexical specification of tone in North-Germanic. *Nordic Journal of Linguistics*, 28, pp. 61–96 .
- Lambrecht, K. (1996). *Information structure and sentence form: Topic, focus, and the mental representations of discourse referents*. Cambridge University Press.
- MacWhinney, B., & Bates, E. (1978). Sentential devices for conveying givenness and newness: A cross-cultural developmental study. *Journal of Verbal Learning and Verbal Behaviour*, 17, pp. 539–555.
- Müller, A., Höhle, B., Schmitz, M., & Weissenborn, J. (2006). Focus-to-stress alignment in 4 to 5-year-old German-learning children. *Proceedings of GALA 2005*, pp. 393–407.
- Myrberg, S. (2009). *The intonational phonology of Stockholm Swedish*. PhD Thesis, Stockholm University.
- Myrberg, S. (2013). Focus type effects on focal accents and boundary tones. *Proceedings of Fonetik 2013*, pp. 53–56.
- Ota, M. (2006). Children's production of word accents in Swedish revisited. *Phonetica*, 63, pp. 230–246.
- Quené, H., & Van den Bergh, H. (2008). Examples of mixed-effects modelling with crossed random effects and with binomial data. *Journal of Memory and Language*, 58, pp. 27–51.
- Riad, T. (2006). Scandinavian accent typology. *STUF-Sprachtypologie Und Universalienforschung*, 59, pp. 36–55.
- Riad, T. (2012). Culminativity, stress and tone accent in central Swedish. *Lingua*, 122, pp. 1352–1379.
- Riad, T. (2014). *The phonology of Swedish*. Oxford: Oxford University Press.
- Röhr, C. T., Baumann, S., & Grice, M. (2015). The effect of verbs on the prosodic marking of information status: production and perception in German. *Proceedings of the 18th International Congress of Phonetic Sciences (ICPHS)*.
- Roll, M., Horne, M., & 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(1), 55–73.
- Romoren, A. S. H., & Chen, A. (2015). Quiet is the new loud: Pausing and focus in child and adult Dutch. *Language and Speech*, 58, pp. 8–23.

- Romøren, A. S. H.** (2016). *Hunting highs and lows: The acquisition of prosodic focus marking in Swedish and Dutch*. Dissertation. LOT publications, The Netherlands.
- Rooth, M.** (1985). Association with focus. PhD Thesis, University of Massachusetts at Amherst.
- Rooth, M.** (1992). A theory of focus interpretation. *Natural Language Semantics*, 1, pp. 75–116.
- R Core Team** (2014). *The R project for statistical computing*. R Foundation for Statistical Computing. Available at <https://www.r-project.org/>.
- Strangert, E., & Heldner, M.** (1995). The labelling of prominence in Swedish by phonetically experienced transcribers. *Proceedings of the 8th International Congress of Phonetic Sciences (ICPHS)*, pp. 204–207.
- Terken, J., & Hirschberg, J.** (1994). Deaccentuation of words representing ‘given’ information: Effects of persistence of grammatical function and surface position. *Language and Speech*, 37, pp. 125–145.
- 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.
- Vallduví, E., & Engdahl, E.** (1996). The linguistic realization of information packaging. *Linguistics*, 34, pp. 459–520.
- Wells, B., Peppé, S., & Goulandris, N.** (2004). Intonation development from five to thirteen. *Journal of Child Language*, 31, pp. 749–778.
- Wonnacott, E., & Watson, D. G.** (2008). Acoustic emphasis in four year olds. *Cognition*, 107, pp. 1093–1101.
- Yang, A.** (2017). *The acquisition of prosodic focus-marking in Mandarin Chinese- and Seoul Korean-speaking children*. Ph.D dissertation. University of Utrecht, The Netherlands.
- Yang, A., & Chen, A.** (2018). The developmental path to adult-like prosodic focus-marking in Mandarin Chinese-speaking children. *First Language*, 38 (1), pp. 26–46.