Hunting Highs and Lows: Acquiring Prosodic Focus Marking in Swedish and Dutch

Anna Sara H. Romøren and Aoju Chen

1. Introduction

The acquisition of prosodic focus marking has primarily been investigated for non-tonal languages; little is known about how this skill is acquired in languages where pitch is used to mark both sentence and word-level contrasts. In the current study we compare the acquisition of prosodic focus marking between a tonal language, Central Swedish (hereafter CS), and a non-tonal language, Dutch. Our hypothesis was that children learning a tonal language are more sensitive to pitch contrasts than children learning a non-tonal language, making CS-speaking children master the pitch-based cues to focus used in their language earlier than their Dutch-speaking peers.

The paper is structured as follows: In the background section (2) we introduce the terms focus and prosody, present the basics of prosodic focus marking in Dutch and Swedish, and briefly review previous work on prosody and focus in children’s speech production. In Section 3 we introduce our research questions, and in Section 4 we present our methodology. In Section 5 we present our analyses and results, and in Section 6 we discuss our findings, and what they add to our current understanding of cross-linguistic effects in the acquisition of prosodic focus marking.

2. Background

In the following we will use the term ‘prosody’ as a cover term for a number of linguistic phenomena related to acoustic variation in pitch, duration and intensity, thereby also including linguistic notions like stress, accents, and tones. The term ‘information packaging’ is often used to capture a broad range of linguistic phenomena related to the way information is presented by speakers, as opposed to the information itself. The notion ‘focus’ is perhaps one of the most studied aspects of information packaging, where ‘focus’ typically refers to ‘new’

* Anna Sara H. Romøren, Utrecht University/ HIOA Oslo, asromoren@gmail.com. Aoju Chen, Utrecht University/ Max Planck Institute for Psycholinguistics, aoju.chen@uu.nl.

1 We use the term ‘prosody’ as a cover term for a number of linguistic phenomena related to acoustic variation in pitch, duration and intensity, thereby also including linguistic notions like stress, accents, and tones.
or ‘important’ information in a sentence, contrasted with ‘given’ or ‘established’ information. Krifka (2007) maintains that even if focus is statically associated with newness or informativeness, the essential feature of focus is that it ‘highlights the presence of alternatives relevant for the interpretation of linguistic expressions’ (Krifka, 2007). In the following we will adopt Krifka’s definition of focus. Because we elicit focus by means of wh-questions (Example 1) our focal constituents can also be said to represent new and important information in the discourse.

(1)  

a. Question: Who cooks the carrot?  
Answer: [The tiger] cooks the carrot.  
b. Question: What does the tiger do with the carrot?  
Answer: The tiger [cooks]_FOCUS_ the carrot.  
c. Question: What does the tiger cook?  
Answer: The tiger cooks [a carrot]_FOCUS_.  

Across the worlds’ languages, a number of linguistic strategies are available for marking focus, like syntactic processes, morphological markers, or prosody (Vallduví & Engdahl, 1996; Krifka & Musan, 2012) but in this paper we centre on the use of prosodic means to marking focus. The most important prosodic difference between CS and Dutch is that, in the former, pitch is used for marking word-level contrasts, whereas in the latter, pitch only signals sentence-level contrasts. The presence of a lexical pitch accent contrast in CS means that the entire lexicon is split into two categories; accent 1 words and accent 2 words, and that minimal word pairs exist that are only distinguished by their associated contour (e.g. _anden_1 ‘the duck’ versus _anden_2 ‘the spirit’). Crucially, the lexical accent distinction seems to shape the way prosody is used for highlighting focal information, making prosodic focus marking work differently in CS than it does in Dutch.

In Dutch2, focus is marked by adding a pitch accent (i.e. a particular pitch movement such as a fall, rise, or level tone, aligned with the stressed syllable of a word) to the constituent in focus, and by avoiding such pitch accents post-focally (e.g. Gussenhoven, 2004; Hanssen, Peters, & Gussenhoven, 2007; Chen, 2009, 2011a). Falling pitch accents (‘H*L’) are the most common accents for marking focus, but other accents are also used (Chen, 2007, 2009). In CS, most content words carry a lexically3 defined pitch fall aligned with the main stressed syllable. Focus is marked by adding a separate prominence-marking high tone (hereafter ‘prominence H’) to these falls. Because the alignment of prominence H differs across the two lexical accent categories, two different surface contours result from

---

2 Generally, prosodic focus marking works in fairly similar ways in Dutch, German and English (see Chen, 2012)  
3 Referring to the fall of both accent 1 and accent 2 as ‘lexical’ is a simplification, as only parts of these falls seem to be truly ‘lexical’. This discussion is beyond the scope of this paper, but we refer the reader to Riad (2014), for a more elaborate discussion.
the combination of each lexical accent with prominence H. Post-focally, prominence H is avoided but lexical accents are retained and downstepped (e.g. Bruce, 1977; 1998; Heldner, 2001; Myrberg, 2009, 2013; Ambrazaitis, 2009).

On the surface, the focus-marking contours of CS (‘L*H’ for accent 1 and ‘HL*H’ for accent 2) are not all too dissimilar from the pitch accents found in Dutch; they are aligned with stressed syllables, render a word prosodically prominent, and are associated with an expanded pitch range and a longer duration on the word carrying them. Nevertheless, whereas the whole contour is assigned at the sentence-level in Dutch, it consists in both word- and sentence-level tones in CS. Further, whereas a wide number of different pitch accent types can align with stressed syllables in Dutch (Gussenhoven, Rietveld & Terken, 1999), standard descriptions of CS assume only four categorically distinct contours; two lexical accents with and without prominence H.

Previous work on the acquisition of prosodic focus marking in children’s speech production is relatively sparse, and centred on West-Germanic languages. Further, the rather limited amount of research that has been conducted is characterized by substantial methodological variability, making it hard to isolate cross-linguistic effects from effects resulting from different output measures, different ways of controlling for information packaging, or different age ranges included in the various studies.

Starting with work on English, two early studies report that contrastive stress⁴ is reliably used on contrastive information already at three to four years (Hornby & Hass, 1970; MacWhinney & Bates, 1978). Another study on English report that words involving contrast are produced with a higher mean pitch than words that are repeated across sentences (Wonnacott & Watson, 2008). At the same time, MacWhinney & Bates found development in the children’s use of contrastive stress between the age of four and six, and Wonnacott & Watson report that whereas adults increased the word duration on contrastive versus given words, the children did not, suggesting that even if some aspects of the prosodic marking of contrast are mastered already at three to four years, others develop later. This is supported by findings from Wells, Peppé & Goulandris (2011), who showed that even if children around five generally accent contrastive information, they also tend to misassign accents at this stage, primarily in terms of a preference for accenting sentence-final nouns independently of focal status. Moving on to German, Müller, Höhle, Schmitz & Weissenborn (2006) report on four- to five-year-olds producing words referring to focal information with a higher mean pitch than words representing given information, and similar patterns are also reported by Sauermann, Höhle, Chen & Järvikivi, 2011. In the latter study, the effect of focus on pitch and duration-measures depended on the sentence-position investigated, suggesting that the position investigated in a certain study may have large implications for the conclusions drawn. In a series of studies on Dutch, Chen

---

⁴ We assume the term to refer to pitch accentuation, and that the coding of contrastive stress thus involved a pitch accent judged as more prominent than neighbouring constituents (MacWhinney & Bates, 1978).
(2011a, and references therein) report that Dutch-speaking children accent focal and de-accent post-focal information in line with adults at the age of four to five, but that their use of more fine-grained phonetic manipulations for marking sentence-initial focus, as well as their accent type preferences for marking sentence-final focus, are not adult like before the age of 7 to 8, or even later. Just like the results from Sauermann et. al., (2011), Chen’s findings show that the way focus is prosodically highlighted differs across different sentence-positions, and that children’s acquisition of prosodic focus marking depends on the cues involved in the relevant position, where more fine-grained phonetic distinctions seem harder to acquire than the use of accent placement.

With the exception of Chen’s work, most previous studies on prosody and focus in children have been concerned with contrastive focus or contrast (i.e. by having children describe pictures where one element is changed from the first to the second sentence). By using wh-questions we elicit non-contrastive focus from our participating children, asking whether previous findings regarding contrastive focus can also be generalized to the marking of non-contrastive focus.

3. Research questions

The analysis presented in this paper concerns the use of accent placement for marking narrow focus in Dutch, and the use of prominence H for marking narrow focus in CS. The first research question concerns CS, the second concerns Dutch, and we asked:

1. Do CS-speaking children between 4 and 11 differ from CS-speaking adults in the way they use prominence H on narrowly focal versus post-focal target words?

2. Do Dutch-speaking children between 4 and 11 differ from Dutch-speaking adults in the way they use accent placement on narrowly focal versus post-focal target words?

Our research questions were explored for each language separately, by comparing prosodic manipulations made by three child groups to manipulations made by language-matched adult controls.

4. Method

4.1. Participants

The participants were native speakers of CS or Dutch, without any history of language or other developmental disorders (Tables 1 and 2 below). The participants were divided four age groups: 4-5 years, 7-8 years, 10-11 years and adults. In the following we will refer to the age groups as eight-year-olds, five-year-olds and eleven-year-olds.
#### Table 1: CS-speaking participants

<table>
<thead>
<tr>
<th>Age group</th>
<th>N</th>
<th>Age range (y;m)</th>
<th>Age mean (y;m)</th>
<th>Gender</th>
</tr>
</thead>
<tbody>
<tr>
<td>4-5</td>
<td>10</td>
<td>4;3-5;6</td>
<td>5;0</td>
<td>6 m, 4 f</td>
</tr>
<tr>
<td>7-8</td>
<td>8</td>
<td>7;6-8;8</td>
<td>8;3</td>
<td>5 m, 3 f</td>
</tr>
<tr>
<td>10-11</td>
<td>8</td>
<td>10;0-11;0</td>
<td>10;6</td>
<td>4 m, 4 f</td>
</tr>
<tr>
<td>Adults</td>
<td>10</td>
<td>20:00-43;10</td>
<td>27;2</td>
<td>5 m, 5 f</td>
</tr>
</tbody>
</table>

#### Table 2: Dutch-speaking participants

<table>
<thead>
<tr>
<th>Age group</th>
<th>N</th>
<th>Age range (y;m)</th>
<th>Age mean (y;m)</th>
<th>Gender</th>
</tr>
</thead>
<tbody>
<tr>
<td>4-5</td>
<td>10</td>
<td>4;4-5;8</td>
<td>5;2</td>
<td>6 m, 4 f</td>
</tr>
<tr>
<td>7-8</td>
<td>8</td>
<td>7;1-8;11</td>
<td>8;0</td>
<td>4 m, 4 f</td>
</tr>
<tr>
<td>10-11</td>
<td>8</td>
<td>10;4-12;0</td>
<td>10;7</td>
<td>4 m, 4 f</td>
</tr>
<tr>
<td>Adults</td>
<td>14</td>
<td>18;7-48;0</td>
<td>23;11</td>
<td>5 m, 9 f</td>
</tr>
</tbody>
</table>

#### 4.2. The picture-matching game

The data was gathered by means of a picture-matching game\(^5\), constructed to elicit SVO-sentences with varying information structure. The participant’s task was to help the experimenter find correct combinations of picture pairs by answering the experimenter’s questions about her or his pictures. WH-questions were constructed so as to make one constituent in the answer narrowly focal and the others non-focal. Pre-scripted short texts lead up to each question, introducing the non-focal constituents (Table 3). The materials consisted of three sets of pictures, two for the experimenter, and one for the participant (Figure 1). The experimenter’s first set (set 1) always lacked one attribute (e.g. the subject, the verb, or the object) and was piled face down in front of her. The experimenter’s second set (set 2) included pictures representing what was missing in set 1, scrambled face up in a box between the participant and the experimenter. The participant’s pictures (set 3) displayed complete actions. Sets 1 and 3 were ordered before each session, so that corresponding pictures appeared together in the same trial.

---

\(^5\) The game was adapted based on the procedure developed by Chen (2011). Different from the original setup, the robot mediator was removed, and the setup was made more game-like by introducing three picture sets that needed to be combined. New target words were also selected, and child-friendly drawings were made from scratch for all the items.
Each trial started with the experimenter picking up a picture from his/her set (set 1), uttering the corresponding context sentences (Table 3). After the target question was asked, the participant could look at his/her complete picture (set 3) and answer the question. Together they could then look for the ‘missing piece’ of the experimenter’s picture in the box (set 2), unite the pictures, and move on to the next trial. Two rules were introduced: The participants should always answer in a full sentence, and never show their own picture to the experimenter.

The game consisted both practice trials and test trials, and the total number of test trials was 30 in the Swedish version of the game and 15 in the Dutch version. Trials were originally spread over five sentence conditions; narrow focus on the initial constituent (narrow initial), narrow focus on the medial constituent (narrow medial), narrow focus on the final constituent (narrow final), in addition to contrastive focus on the medial constituent (contrastive medial) and broad focus on the whole sentence (broad focus). In the analyses presented here, only the three narrow-focus conditions were included, elicited following the structure presented in Table 3 (see Romøren & Chen, 2011, for more on the other two conditions).

Table 3: Example of context and questions implementing the three sentence conditions included in this study.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Context/ question</th>
</tr>
</thead>
<tbody>
<tr>
<td>Narrow initial</td>
<td>Look, the flower! It looks like someone is hiding the flower. Who is hiding the flower?</td>
</tr>
<tr>
<td>Narrow medial</td>
<td>Look, the flower! And there is also a frog. It looks like the frog is doing something with the flower. What is the frog doing with the flower?</td>
</tr>
<tr>
<td>Narrow final</td>
<td>Look, the frog. It looks like the frog is hiding something. What is the frog hiding?</td>
</tr>
</tbody>
</table>

In the Swedish version of the experiment, six subject nouns, six transitive verbs and six object nouns were carefully distributed over the experimental conditions; half of the target words were accent 1-words, the other half accent 2 words. In the Dutch version of the game, five subject nouns, three transitive verbs
and three object nouns were distributed over the experimental conditions. Each combination of initial, medial and final constituent only occurred once in the whole set; consecutive trials never realized the same condition, and always differed by minimally two constituents. The trials were arranged into two different stimulus orders, to which the participants were randomly assigned.

4.3. Coding

The resulting data consisted in SVO sentences where initial, medial, or final constituents were narrowly focal, whereas the other constituents were non-focal and represented contextually given information. The prosodic analyses were conducted using Praat (Boersma & Weenink, 2010). For the Dutch dataset, sentences were annotated following the ToDI system for transcribing Dutch intonation (Gussenhoven, Rietveld & Terken, 1999), coding for both accent placement and type of pitch accent. This was done by the 1st author, after a three-month training using the online ToDI tutorial. For the Swedish dataset, the sentences were hand coded for lexical accent, as well as for whether or not a word carried prominence H.

5. Analysis and results
5.1. Model comparisons

In order to statistically analyse the use of accentuation and prominence H in our eight groups, binomial logistic regression models (GLMs) were built and compared using R (R Core Team, 2014). Separate analyses were run for each language, following the same procedure. For all models, the outcome variable was categorical, consisting in the binary outcome ‘presence or absence of prominence H/pitch accent’, and our independent factors were ‘focus’ and ‘group’, as well as any interaction between them. ‘Focus’ was always a binary factor, comparing conditions rendering a target word narrowly focal to conditions rendering the same target post focal. Our factor ‘Group’ included four levels (our four groups), coded into dummy variables using adults as a baseline to which the other groups were compared.

We first built 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 and then finally adding the interaction between ‘focus’ and ‘group’ in model 3. Only factors that significantly improved the previous model were included in subsequent models. In order to assess the improvement of the model fit from models 0 through 3 we used R’s ‘anova’ function to compare pairs of models (i.e. Quené & van den Bergh, 2008). When the best model was established, model summaries were used in order to obtain estimates and p-values for the relevant parameters. In some cases additional comparisons were made on subsets of the data, for example to check for the effect of focus within specific age groups or the effect of group within specific focus conditions.
Table 4: Model build-up (added parameters only)

<table>
<thead>
<tr>
<th>Model</th>
<th>Factors added</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 0</td>
<td></td>
</tr>
<tr>
<td>Model 1</td>
<td>Focus</td>
</tr>
<tr>
<td>Model 2</td>
<td>Group</td>
</tr>
<tr>
<td>Model 3</td>
<td>Focus * Group</td>
</tr>
</tbody>
</table>

Our output measures were made on sentence-medial and sentence-final target words. Because our target sentences always had an SVO structure, medial targets were always verbs and final targets always objects. Sentence-medially, our focus condition was ‘medial narrow’ and our post focus condition was ‘initial narrow’. Sentence-finally, our focus condition was ‘final narrow’ and our post focus condition was ‘medial narrow’ (See Table 3). Below we first report on our results from CS, thereafter reporting the results from Dutch.

5.2. Prominence H for focus in CS-speaking children and adults

As can be observed in Figure 2 below, our CS-speaking adults, eleven-year-olds and eight-year-olds behaved strikingly similar in using prominence H almost at ceiling under narrow focus and almost never post-focally. The five-year-olds showed the same pattern, but used prominence H less under narrow and more under post focus compared to the other groups.

![Figure 2: Prominence H sentence-medially by focus and group](image)

Building models on the on the use of prominence H sentence-medially showed ‘focus’ (p=.000), and the interaction between ‘focus’ and ‘group’ to improve our model (p=.004), whereas ‘group’ did not (p=.671). Narrow focus lead to significantly more use of prominence H than post-focus (p=.000), and the interaction showed the effect of narrow focus to differ between the five-year-olds and the adults (p=.002). As can be observed in Figure 2, the difference between narrow and post focus was smaller in the five-year-olds (83% vs. 22%) than in the adults (95% vs. 4%).
Similar patterns of prominence H close to ceiling for narrow focus and close to the floor for post focus was also observed sentence-finally (Figure 3). This said, for the sentence-final position the patterns were slightly more variable across the three child groups than what was observed medially.

Building models on the use of prominence H sentence-finally showed focus to significantly improve our 0 model (p=.000), whereas ‘group’ (p=.639), and the interaction between ‘focus’ and ‘group’ (p=.123) did not. There was thus a general effect of narrow focus increasing the likelihood of prominence H (p=.000), but this effect did not differ significantly between the adults and any of our child groups.

5.3. The use of accent placement for focus in Dutch-speaking children and adults

As can be seen in Figure 4, the use of accentuation on narrow focus was at ceiling in all our groups, but the use of accentuation post-focally was much more common than what was the case for prominence H in our CS-speaking participants. Further, whereas the Dutch-speaking eleven-year-olds de-accented more than the adults, the Dutch-speaking eight- and five-year-olds hardly de-accented post-focally at all.
Building regression models on the use of pitch accent medially showed ‘focus’ (p=.000) and ‘group’ (p=.000) to improve our model, but the interaction between ‘focus’ and ‘group’ did not (p=.176). Focus significantly increased the likelihood of an accent (p=.000), and the group effects consisted in the eleven-year-olds accenting less (p=.043) and the five-year-olds more (p=.022) than the adults across both focus conditions. The absence of interaction effects means that observed differences between adults and children in the effect of focus did not reach significance in this dataset. Because of the way the data was collected, the Dutch dataset was half the size of the Swedish one (roughly 20 items per condition in each group). Such a small dataset makes it less likely that between-group differences like the ones observed in Figure 4 reach significance. Because of this we also chose to statistically test effect of focus in each group separately. This analysis confirmed the patterns observed in Figure 4: narrow focus significantly increased the probability of an accent in our adults (p=.000) and our eleven-year-olds (p=.000), but not in our eight year olds (p=.173) or our five-year-olds (p=.923).

In sentence-final position there was also a general pattern of more accentuation on narrow focus than on post focus in all four groups, and the effect of focus again looked stronger in the eleven-year-olds than in the adults (Figure 5).

![% accent finally](image)

**Figure 5: Accent placement sentence medially by focus and group**

Building models for the sentence-final position showed both ‘focus’ (p=.000), and the interaction between ‘focus’ and ‘group’ to improve our model (p=.041), whereas ‘group’ did not. Narrow focus increased the probability of an accent on the sentence-final target words (p=.000), and the interaction involved a stronger effect of post focus in the eleven-year-olds than in the adults (p=.029). Looking at the data from each group separately showed narrow focus to significantly increase the probability of an accent in the adults (p=.000), the eleven-year-olds (p=.000), the eight-year-olds (p=.001) and the five-year-olds (p=.002), thus in this position the effect of focus on accentuation was present in all our child groups.
5.4. Summarizing the results

Our CS-speaking children was remarkably adult-like in the use of prominence H for marking narrow focus: all groups used prominence H much more on narrowly focal targets than on post-focal targets in both sentence-positions. Sentence-medially, this effect was weaker in the five-year-olds than the adults, but the general pattern resembled that of the other groups. The results from the Dutch adults showed that even if accentuation was at ceiling for focus, they did not de-accent post-focal targets sentence-medially more than around 30% of the time, compared 60% sentence-finally. Whereas the eight- and five-year-olds showed adult-like patterns of accenting focus more than post-focus sentence-finally, they did not differentiate focal from post-focal targets using accentuation sentence-medially. Interestingly, the Dutch eleven-year-olds de-accented post-focal information more than the adults in both sentence-positions. In other words, whereas CS-speaking children predominantly use prominence H in line with adults already from the age of five, Dutch children only use accentuation in line with adults after the age of eight, at least in sentence-medial position.

6. General discussion

The finding that the CS-speaking children reach adult consistency in the use of prominence H earlier than their Dutch-speaking peers master the use of accent placement, supports our initial hypothesis that increased tonal sensitivity makes CS-speaking children faster at learning pitch-based cues to focus than Dutch-speaking children. This said we have also seen that the mapping between accent placement and focus seems less consistent in Dutch adults than the mapping between prominence H and focus in CS-speaking adults, thus initial sensitivity to pitch is not the only possible candidate for explaining our results.

Our finding that Dutch-speaking children are adult-like in using accent placement for focus by the age of four to five are in line with previous reports from Chen (2011a), where children between four and eight accented in line with adults in sentence-final position. Further, the patterns we observed sentence-medially partly resemble Chen’s (2009) sentence-initial findings, showing that both adults and children often accent both focal and non-focal targets outside sentence-final positions. Together with Chen’s work, our data suggests that de-accenting post-focal information is one of several options available for speakers of Dutch when highlighting focal information, and that de-accentuation is more common in sentence-final position than in non-final positions.

Importantly, the inconsistency with which Dutch-speaking adults de-accent post-focal information sentence-medially may explain why the Dutch-speaking eight- and five-year-olds included in this study hardly ever de-accented in this position at all. After all, the dominant pattern in our adults was accentuation. If it is the case that the assumed post-focal de-accentuation is less common in West-Germanic varieties than what is typically assumed, reports on young children’s failure to de-accent post-focal information may simply be a consequence of assuming the wrong adult model, or at least not taking the variability involved in
adult focus-to-accent mapping properly into account (see Grünloh, Lieven & Tomasello, 2015, see also De Ruiter, 2010, for similar discussions). At the same time, our finding that five and eight-year-olds did not de-accent at all, whereas the adults did, nevertheless means that Dutch children between five and eight do not yet mark focus in the same way as adults do sentence-medially, as they never make use of the option of de-accentuation.

One common developmental pattern was observed across our Dutch and CS-speaking children, in that both groups reach adult proficiency earlier in the sentence-final position than sentence-medially. This pattern may be related to the fact that the sentence-final constituent is the default position for a nuclear pitch accent in broad-focus sentences in North and West-Germanic varieties (e.g. Féry, 2013). Accenting a narrowly focal target elsewhere in the sentence, followed by de-accenting constituents following it, therefore represents the marked pattern. The fact that both Dutch and CS-speaking children master prosodic focus-marking later in non-final positions may be a result of them having less exposure to more marked accentuation patterns. More cross-linguistic work on languages with different alignment patterns for maximal prosodic prominence may shed additional light on the interplay between syntax and prosody in children’s acquisition of information structure.

References


Myrberg, Sara. (2013). Focus type effects on focal accents and boundary tones. Proceedings Fonetik 2013, Linköping, Sweden (pp. 53-56). Department of Culture and Communication, Linköping University.


