Hunting highs and lows

The acquisition of prosodic focus marking in Swedish and Dutch

The research reported here was part of the project 'Get the focus right: a cross-linguistic study on prosodic encoding of focus in children', supported by the Netherlands Organisation for Scientific Research (NWO) under project number 276-89-001.

Published by

LOT
Trans 10
3512 JK Utrecht
The Netherlands

phone: +31 30 253 6111
e-mail: lot@uu.nl
http://www.lotschool.nl

Cover illustration: Paula Cox

ISBN: 978-94-6093-208-3
NUR 616

Hunting highs and lows

The acquisition of prosodic focus marking in Swedish and Dutch

Bij hoog en bij laag

De verwerving van prosodische focusmarkering in het Nederlands en in het Zweeds

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor
aan de Universiteit Utrecht
op gezag van de rector magnificus, prof. dr. G. J. van der Zwaan,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op
vrijdag 1 juli 2016
des middags te 4.15 uur

door

Anna Sara Hexeberg Romøren

geboren 1 juni 1985
te Kristiansand, Noorwegen

Promotoren:	Prof. dr. R. W. J. Kager Prof. dr. W. Zonneveld
Copromotor:	Dr. Aoju Chen

Til mor og far

Table of contents

List of tables

List of figures

Thank you

Here it is, the easiest hardest part of a dissertation to write. It is also the last part, excluding those occasional ‘oh, that’s a good thing to put in the acknowledgements’ moments, often occurring under the influence of alcohol, jetlag, or just random tunes from that sentimental little harp player on your shoulder. Those moments which may or may not have resulted in some form of written material, which certainly is nowhere to be found when you need it. So here goes.

Aoju Chen, René Kager and Wim Zonneveld were my supervisors, and I am grateful for having had their support and guidance throughout this process. Thank you Aoju, for teaching me about prosody, for generously sending me to conferences all over the world, and for your impressive ability to answer emails at odd hours. Thank you also for your patience, and for your ambition on my behalf. René, thank you for your kindness and encouragement. You managed to get me excited about the acquisition of phonology even at the most difficult writing stages, thank you for being so engaging and knowledgeable, and for occasionally laughing at my jokes. Wim Zonneveld, thank you for giving me the best compliment of my PhD years, which I pulled out when I needed that extra kick. Thank you for challenging me by asking those difficult but important questions.

Thank you Shanley Allen, Paula Fikkert, Carlos Gussenhoven, David House and Hugo Quené, for serving as my manuscript committee.

My paranymphs, Sandrien van Ommen and Marko Simonović. Sandrien, this one is really hard, which says it all. Thank you for challenging me, for your eye for beauty in so many forms, for altering my sleeping pattern, for getting emotional, for taking me rollerblading, for being my friend. Marko/Lizzy. Thank you for asking me to marry you and politely refusing my delayed attempt at sealing the deal. For all those Lombok dates and all that picking, and your spectacular way of keeping it casual. For making me go to that sneak even if it was way past my bedtime, and for biking to work with me.

I am very grateful that I got to do my PhD in Utrecht. Thank you Maaike Schoorlemmer for coordinating, and for your genuine interest in the wellbeing of the PhDs. Thank you Yvonne van Adrichen, Martien Camphuis and the other secretaries, the cleaning staff, and the friendly faces in Trans 10. UiL OTS was a great place to be, partly because someone cleaned up the mess I made around my desk and provided me with a joke along with the office supplies. The presence of

research groups like the prosody group and the experimental phonology group (EPG) made the institute a good place to be, together with the ELiTU-talks, the Uiljtesdagen and the EMLAR. Being able to attend LOT-schools, talks, workshops and conferences both in Utrecht and at other universities in the Netherlands made the inevitably solitary process of doing a PhD feel much less so, and I learned a lot (no pun intended). Thank you UiL-OTS, and NWO, for giving me this opportunity.

Getting support on statistics was crucial for the finishing of this book. Thank you Huub and Mattis van den Bergh, Laura Boeschoten, Tom Lentz, Sandrien van Ommen and Maartje Schulpen. Marta Castella helped me clean up my references, thank you. Thank you Liza for proofreading, thank you Thomas for turning the formatting nightmare into a funky dance. Thank you Maart for help with the Dutch summary. Whatever is left of mysterious fonts, inventive spelling and ugly graphs, is all on me.

My PhD was part of a larger project, providing me with a great group of people to discuss my work with, in addition to quite a few extra hands for gathering and coding my data, for which I am very grateful. Thank you Aoju, for keeping track of it all. Thank you Anqi Yang and Zenghui Liu for soup lunches, hotpot, love poetry, for your company at conferences and institute rooftops, and for many useful discussions on prosody, prosodic analysis and statistics. Thank you Paula Cox, Frank Bijlsma, Saskia Verstegen, Mengru Han and Martine Veenendaal, for helping in collecting my Dutch data, translating the stimuli, annotating enormous amounts of data and for a bunch of other larger and smaller tasks. Thank you Joe Rodd for helping me with all kinds of technical tasks, and for your great work on the Dutch pausing data. Thank you Paula for taking me to concerts, teaching me Dutch poetry, and for all those drawings.

Thank you colleagues and friends at UiL-OTS. Office mates Assaf Toledo, Desiree Capel, Marjolein van Egmond, Marko Simonović, Anna Volkova and Anne-France Pinget, thank you for the moments we shared in 0.45. Thanks for the laughs, the occasional cry, naps on the couch (Anne-France, you gossip), and all of those conversations that you get from being around each other exactly when everything happens. Thank you Maartje Schulpen, for rooftop gardens, kitchen poetry, sjulabånn and way beyond. For your bad jokes, messy eating habits, and your face when you dance. Thank you Marta Castella, in particular for those limbo days in Utrecht where we muffled the panic of change with bio-kwark and that big-boned cat. Thank you for laminating and shipping home-made lemon curd to Oslo, and for telling me I'm fireworks. Björn 't Hart, morning coffee is nicer with you. For your kindness, your loyalty, and for making me blush. Mirjam Hachem, thanks for your excellent humour, excellent mind, and excellent vodka-orangejuice-because-

queensday-companionship. Remember the waterlilies, lady. Stavroula Alexandropoulou, thank you for letting me win in Rummikub, and for being magnificent in a way I simply can't put my finger on. Heidi Klockman, thank you for your energy and laughter, and your never-ending supply of vodka. Make sure you keep dropping by to practice your Norwegian. Brigitta Keij, huge-hearted organizational mastermind and punmachine, I'm looking forward to traveling with you again. Sophia, my competition, thank you for all those hearts, and for getting emotional. Hanna de Vries, Franca Wesseling, Jolien Scholten, I wish I had gotten to know you better. Liquan Liu, you firecracker, I still have your picture on my desk for good luck. Iris Mulders, I miss your giggle when my jokes fail. Eefje (tralalalala) Boef, you have to learn that song so that we can sing it. Liv Persson, why don't you mountain bike yourself northward soon? Alexis Dimitriadis, thanks for taking me climbing. Nadya Goldberg, that was the most elegant auction of my life. Hannah de Mulder, Marijn Struiksma, Hayo Terband, Tom Lentz, Iris Mulders, thank you for Friday drinks in Lokaal 9. Ana Aguilar, distant but present. Anna Volkova, I'm sure kayaking and folksongs also works on fjords. Thank you thank you thank you Alexia Guerra Rivera, Anja Goldschmidt, Lisa Bylinina, Loes Koring, Daria Bahtina-Jantsikene, Rob Switserlood, Carolien van den Hazelkamp, Lotte Hendriks, Eva Poortman, Choonkyu Lee, Toni Bassaganyas, Myrthe Bergstra, Dominique Blok, Anne van Leeuwen, Ao Chen, Bert le Bruyn, Xiaoli Dong, Heimir van der Feest Viðarsson, Shuangshuang Hu, Hans Rutker Bosker, Yipu Wei, and anyone I may have forgotten.

Dear Beste Eris, Maxime Barou and Altug Altay. I can't believe the luck I had living with you in Boeroehof. Thank you for cooperative shower routines, many many food feasts in the kitchen, for introducing me to your friends and family, for bike trips. I follow you, gipsy babies.

Half of my thesis is on Swedish. Luckily, that put me in touch with a bunch of great people. Thank you David House, Jens Edlund, Sofia Strömbergsson, Christina Tårnander and the others at KTH, for hosting me during my first and second visit to Stockholm, and for pretending my speaking Swedish wasn't awkward. Thanks to Sofia for lördagsgodis, and for helping me prepare my Swedish stimuli. Thank you Sara Myrberg, Tomas Riad, Mattias Heldner, for important input on Swedish prosody. Marjatta Sikström, you made me feel so at home in your beautiful house, thank you for enormous amounts of freshly pressed apple-juice, for sauna Sundays and for the wildlife in the garden. Thank you Yvonne Agatsoamedo for your friendly spirit, and for turning those improvised dinners into feasts. A big thank you to the people in Lund: Gilbert Ambrazaitis, Merle Horne, Mechtild Tronnier, Mikael Roll, Susan Sayehli and the others, for hosting me and including me in grant proposals, for flammable gluhwein and strange Swedish feminist movies. Thank you

Rebecca Sancho for being the Swedish ground crew in Utrecht, for translating my stimuli and for helping me getting my accents right.

Without the cooperation of a handful schools, kindergarten, parents, teachers, and of course, children, there would have been no data and no thesis. Thank you Svenska Skolföreningen i Wassenaar and Svenska Skolföreningen i Nederlenderna, for cooperating with me during my pilot. Thank you Råsunda Skola, Engelbrektsskolan, förskolan Sjöboden and förskolan Gräven, Montesorrischool Houten, Montessorischool Soest, and De Ontrekkingsreis Doorn, for allowing us to come by and play our game a million times. A special thank you goes to that cheerful boy with the bleeding nose who had to sit on top of four pillows to get close to the microphone, and who answered my questions with far more entertaining stories than the ones I was asking for.

I would never have come to Utrecht if it wasn't for my Master's thesis supervisor Kristian Emil Kristoffersen. Thank you for everything you taught me about linguistics and research, for encouraging me to apply for that position in Utrecht, and for regularly sending me pictures of mountains, reminding me of the world outside linguistics. Thank you also to the research group in clinical linguistics and language acquisition at ILN, for inviting me to seminars and meetings, despite my absence. Inger Moen was the one who inspired me to stick my nose into the world of prosody, and for this I am deeply grateful. Thank you Hanne Gram Simonsen, Nina Gram Garmann and Marianne Lind, for teaching me about psycolinguistics and phonology (even if I cried in the phonology class), and for involving me in your research. Thank you fellow linguists and friends in Oslo, for dropping by for Amsterdam adventures, for PhD helplines over Skype, for the occasional meetup in Oslo, and for making coming home feel good. Silje Mosgren, Stig Rognes, Idar Bergfjord, Brita Kaires, Kari Kinn, Vibeke Bø, Eirik Olsen, Pernille Hansen, Elisabeth Holm, Eli Anne Eiesland, Ingeborg Ribu, Veronica Pajaro, tusen tusen takk.

During my years as a PhD I had the opportunity to attend a number of conferences, workshops and meetings in Europe, the States, and in the Netherlands. Thank you Iberian prosodists (you know who you are) for spicing up conference dinners with singing. Thank you Jill Thorson, for hosting me in Providence and Boston, and for being excellent conference company. Thank you James Morgan for letting me visit the baby lab at Brown, and letting me crash the lab dinner at the BUCLD. Carlos Gusssenhoven, Paula Fikkert, Marilyn Vihman, Thorstein Fretheim, Gjert Kristoffersen, thank you for your advice at various stages of this project. Thank you Allison Wetterlin for sneaking me into that dinner in Oxford. Thank you NorPhLexers for input on acquisition, and for that dip into the icy cold Tromsø

waters. Thank you Jacopo Toregrossa for inviting me to Cologne, feeding me my favourite kind of potato (the fried kind) and for your valuable input on information structure.

I am incredibly grateful to my friends who are neither linguists not Utrechters. Thank you for walks, trips, hurried coffee breaks, quick breakfasts and the occasional chat. And even more so, for sticking around despite the lack thereof. Thank you Anne og Øyvind, Hege og Einar, Shani og Bengt Olav, and associated little people. Synne, Jenni og gjengen, thank you for low bar high love. Thank you Orion for staying up all night with me. Christian og Ferdinand, thank you for softening my landing in Oslo with warm winter boots, furniture, forks and a promise of bubbles on your balcony that I intend to cash in before you know it. Thank you Camilla Horne Heidum, for ever excellent life coaching. Thank you Øyunn for texts that feel like presents. Thank you Silje for your radio voice all the way to Utrecht. Thea, Johanne, Camilla, can't wait to see a bit more of you. Aline, thanks for catching me in Copenhagen. Haldis, it's always so good to see you, Tiril, thank you for walks with and without that ridiculous hat. Maren, pluggis. For all those summers in Kvernstien, and for spicing up my last month of writing/first month in Oslo with Rupaul's dragrace and smågodt.

Lastly, my family. My extraordinary parents, to whom I dedicate this book. Gro og gutta og jentene, thanks for Sunday dinners, snøba libre, pølse og Chablis. Laila, for being part of the ground crew, for skinny dipping in mountain lakes, for tussilago farfara, and for knowing where the best blueberries can be found. The Skarveien branch, Liza, Jacob, Autumn, Milly og William. For nugatti in the ceiling, tree houses, sailing trips, loud lunches and much much more. Tone, for donating half of her book collection to me, for giving me a head start at HIOA, and for getting emotional. Målfrid, Frode, Arild og Rune, for sailing trips and bonfires, coffee, waffles and good advice.

Maria, the hardest part. For your patience, for your love, thank you.

Part I:
Introducing

1 Introduction

1.1 Preliminaries

In order to be efficient communicators, children need to adapt their utterances to the common ground shared between themselves and their conversational partners. One way of doing this is by prosodically highlighting focal information. This thesis is a cross-linguistic study on how Dutch and Swedish-speaking children develop the ability to prosodically mark focal information in their speech production.

The notion 'focus' is perhaps one of the most studied aspects of information packaging, typically referring to 'new' or 'important' information in a sentence, contrasted with 'given' or 'established' information (see Chapter 2.1 for a more elaborate discussion). Across the world's languages, a number of linguistic strategies are available for marking focus, such as syntactic alternations, morphological markers, or prosody, and the studies presented in this thesis all concern prosodic means of marking focus. Following many others we use the term 'prosody' as referring to acoustic variation in pitch, duration and intensity, but also for linguistic notions such as stress, accents, and tones.[1]

The work presented in this thesis was conducted as part of the research program 'Get the focus right: a cross-linguistic study on prosodic encoding of focus across languages',[2] funded by the Netherlands Organisation of Scientific Research (NWO). In this program, the acquisition of prosodic focus marking is investigated in four languages: Standard Dutch, Central Swedish, Seoul Korean and Mandarin Chinese. The over-arching goal of the program was to develop a comprehensive theory on the acquisition of prosodic focus marking across languages, with an emphasis on language-specificity and individual variation. The larger project encompassed three different sub-projects: Two PhD projects on cross-linguistic similarities and differences in the development toward adult prosodic focus marking, and a third

---

[1] Our choice of using the term prosody rather than 'intonation', is additionally motivated by the fact that we will be dealing with both word and sentence-level prosody, and that our analyses include manipulations beyond pitch, such as word duration, pausing, and voice quality.

[2] For more information about this research program (grant number 276-89-001), see the project webpage at www.focus.wp.hum.uu.nl.

research project on individual variation in the acquisition of prosodic focus marking. Whereas Anqi Yang has compared the acquisition of prosodic focus marking between Seoul Korean and Mandarin Chinese-speaking children, I have compared the acquisition of prosodic focus marking between Central Swedish and Standard Dutch-speaking children. Anqi Yang's project will result in another dissertation, to be published in 2016 (see also Yang & Chen, 2014; Yang, Cho, Kim & Chen, 2015; Yang, to appear. See also Liu, Chen & van de Velde, 2014, 2015, for an affiliated project on the acquisition of prosodic focus marking in Mandarin L2 and Bai).

The choice of comparing children's prosodic development between Standard Dutch and Central Swedish is not random. Central Swedish (hereafter CS)[3] is a lexical pitch accent language, in which pitch is contrastive at the word level. Standard Dutch (hereafter Dutch) is a so-called 'intonation only' language, where pitch is not lexically distinctive (Yip, 2002; Hyman, 2009).[4] Prosody, and pitch in particular, is used for marking focus in both languages, but the exact prosodic means involved differs between the two. Studying the development toward adult prosodic focus marking in CS and Dutch therefore allows for establishing whether and how systematic differences between prosodic systems affect the developmental path toward adult proficiency. At the same time, comparing prosodic development between these two languages also allows for determining whether general developmental patterns are observed in children learning different prosodic systems.

This chapter starts with a few words on prosody in language acquisition (Section 1.2), before narrowing the discussion down to the prosodic systems of Dutch and CS (Section 1.3). In Section 1.4 we present our hypotheses. This is done by grouping the main differences between Dutch and Central Swedish prosodic focus marking into three domains: *Lexically determined pitch*, *general contour variability* and *prosody-focus mapping,* stressing how the two languages pattern differently within each domain. For each of these domains, specific hypotheses are presented on how these cross-linguistic differences may affect the acquisition of prosodic focus marking in the two languages. As these hypotheses are all concerned with the use of pitch for focus, we also present hypotheses on the use of word duration for marking

---

[3] Central Swedish or Stockholm Swedish is a regional variety of Swedish spoken around Stockholm and beyond (eastern Svealand). This variety is the closest one gets to a spoken standard in Sweden, and is by far the most well described variety where prosody is concerned (see Riad, 2014).

[4] Even if Standard Dutch does not have lexical tones, central Franconian dialects do, among which are the Dutch dialects spoken in the Dutch and Belgian provinces of Limburg (e.g. Gussenhoven & van der Vliet, 1999; Gussenhoven, 2004).

focus in the two languages, as duration varies systematically with focus in both CS and Dutch. In Section 1.5 we present the structure of the dissertation.

1.2 Prosody in language acquisition

Traditionally, most work on phonological development in children has been concerned with segmental phonology; compared to the development of consonant and vowel inventories, relatively little is known about how children develop toward adult proficiency in the production of prosodic categories such as lexical tones, pitch accents and prosodic boundaries. Probably helped by the vast amount of phenomena covered by the term prosody, prosodic features are reported simultaneously as being among the first aspects of language that infants are sensitive to and produce, as well as being among the last aspects of language that children master with adult proficiency. As an example, language-specific prosodic features are observable already in infant babbling (i.e. Hallé, de Boysson-Bardies, & Vihman, 1991), the production of lexical tones is in place before the age of two or three (e.g. Li & Thompson, 1977; Engstrand, Williams, & Strömqvist, 1991), and the same seems to hold for the basic repertoire of pitch accents and boundary tones in non-tonal varieties (e.g. Prieto & Vanrell, 2007; Chen & Fikkert, 2007; Snow & Balog, 2002). On the other hand, the use of prosody for discourse organization, as well as for marking contrast or focus, is not reliably in place until well into school age (e.g. Wells, Peppé & Goulandris, 2004; Chen, 2011a; De Ruiter, 2013). Further, just as for research on the development of segmental phonology, most work on prosodic development in children has been concerned with children acquiring non-tonal varieties such as English, Dutch or German, thus fairly little is known about prosodic development in children learning tone languages, and even less is known about how children speaking tone languages develop in the use of sentence-level prosody. By comparing prosodic development between Dutch and CS, this study is the first to systematically compare the acquisition of sentence-level prosody between a tonal and a non-tone language.

1.3 The languages

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 Swedish 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 (as in the word pair *anden*[1] 'the duck' versus *anden*[2] 'the

spirit'). [5] Crucially for this dissertation, the presence of a lexical accent distinction seems to shape the way prosody is used for highlighting focal information in CS, making prosodic focus marking work differently in this language than it does in Dutch (c.f. Ambrazaitis, 2009).

In Dutch, 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, 2008; Chen, 2009, 2011a). Falling pitch accents ('H*L') are reported to be the preferred accent type used for focus (Chen, 2009), but downstepped falls ('!H*L') are also common. In CS, most content words will be realized with a lexically defined pitch movement, aligned with the main stressed syllable, regardless of focal status.[6] Focus is marked by adding a separate prominence-marking high tone (hereafter 'prominence H') to the lexical contour. Since the alignment of prominence H differs across the two lexical accent categories, two different surface contours result from the combination of each lexical accent with prominence H. Post-focally, where Dutch-speakers are reported to avoid pitch accents (e.g. Chen, 2011), speakers of CS only avoid prominence H; the lexical accents are retained but downstepped (e.g. Bruce, 1977; 1998; Heldner, 2001; Myrberg, 2009, 2013; Ambrazaitis, 2009).

On the surface, the focus-marking contours of CS (a rise-fall pattern for accent 1 and a fall-rise pattern 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 (e.g. Chen, 2012). Nevertheless, whereas the whole contour is assigned at the sentence-level in Dutch, it consists in both word- and sentence-level tones in CS. Furthermore, whereas a wide number of different pitch accent types can align with stressed syllables in Dutch, standard descriptions of CS assume only four categorically distinct contours; two lexical accents with and without prominence H. Finally, literature on prosodic focus-marking in Dutch, supported by descriptions of the related systems of English and German, suggest that there is quite some variability in the way focus is prosodically marked. Firstly, accentuation

---

[5] We adapt the convention of marking lexical accent pertinence 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.

[6] Even if all words carry one of the two lexical accents when produced in isolation, and can thus be categorized as either accent 2 or accent 1 words, function words frequently lose their accent when included into larger prosodic units (Frid, 2003; see also Riad, 2014).

is not limited to focal information (i.e. Gussenhoven, 2011, and references therein). Secondly, post-focal de-accentuation seems to vary with factors like rhythm and speech style, as well as the type of information structure contrast elicited (see e.g. Horne, 1991; Bard & Aylett, 1999; Terken & Hirschberg, 1994; Chen, 2007). Thirdly, even if falling accents ('H*L') are more common on focal information than others accent types, falls are also preferred on pre-focal information in sentence-initial position, and downstepped falls ('!H*L') are fairly common on both focal and post-focal information sentence-finally (Chen, 2009). This situation is different from what is reported for CS, where prominence H is primarily associated with focus, and does not commonly occur post-focally.[7]

1.4 Hypotheses

One way to conceptualize the above-mentioned differences between CS and Dutch is to group them into three main domains, where the first concerns the presence versus absence of word level tones, the second concerns the general amount of contour variability in the prosodic system, and the third concerns variability in the mapping between certain prosodic categories and focus. We shall see that Dutch and CS pattern differently in all three domains, and we present specific hypotheses on how these differences will affect the developmental path toward adult proficiency when CS-speaking children are compared to Dutch-speaking ones.

Whereas Dutch and CS differ in the use of pitch for focus, the languages are similar in the sense that focus is reported to increase the duration of a word. In addition to our hypotheses concerning the use of pitch, we therefore also present two separate hypotheses related to the relationship between word duration, accentuation/prominence H, and focus in Dutch and CS.

Presence versus absence of lexically determined pitch

As we have seen, pitch is lexically distinctive in CS, but in Dutch it is not. Marking focus in CS requires the speaker to take the lexical accent of the focal word into account, integrating word and sentence-level tones into larger contours. Post-focally, prominence H is avoided, but the lexical accents are retained. Speakers of Dutch also align a pitch accent to the constituent in focus, but the whole pitch accent is

---

[7] In addition to marking focus, prominence H is also used as a marker of phrase initial boundaries in CS, what Myrberg (2009, 2012) refers to as 'initiality accents'.

assigned at the sentence-level, and there are no lexical restrictions on the choice of accent type. In Dutch, the whole pitch accent is avoided post-focally.

We propose two competing hypotheses on how the presence versus absence of lexically determined pitch may affect the developmental path to prosodic focus marking in these two languages.

1. The presence of lexically determined pitch contours makes CS-speaking children particularly sensitive to pitch. Early sensitivity to pitch is evidenced in that CS-speaking children distinguish the two accents in their production already before the age of two (e.g. Engstrand, Williams & Strömqvist, 1991). Such early pitch-sensitivity will help CS-speaking children discover that prominence H, a pitch-based category, is used for marking focus. Since they are used to controlling pitch gestures in their speech production, they will also produce this tone appropriately from a relatively early stage. Dutch children, for whom pitch is not a lexical cue, will not show such heightened sensitivity to pitch. This may delay the acquisition of pitch accentuation for focus in Dutch-speaking children, relative to the mastery of prominence H for focus in CS-speaking children.

2. The way prominence H interacts with the lexical accents requires speakers to combine tonal targets with different origins (i.e. word and sentence level). Further, CS-speaking children need to realize that pitch marks both word and sentence-level contrasts, whereas Dutch children can associate pitch with sentence-level contrasts only. The integration of word and sentence-level tones into one contour, and the 'double mapping' of pitch to these two different functions, makes the CS system more complex than the Dutch one. This may lead to an earlier mastery of accentuation for focus in Dutch than the mastery of prominence H for focus in CS.

Contour variability

We have seen that whereas the stress-aligned pitch contours of CS result from a combination of word and sentence-level tones, the Dutch contours (or pitch accents) consist only of sentence-level tones. This said, the standard analysis of Dutch assumes a rather extensive repertoire of pitch accent types; four basic contours, which when modifications like downstep or peak delay are added, make up a total of eight contours (Gussenhoven, 2005; Gussenhoven, Rietveld, & Terken, 1999). Conversely, only four contours are assumed for CS: two lexical accents, with and without prominence H. This situation motivates Gussenhoven's description of Swedish and Norwegian as having 'considerably simpler intonation systems than

West Germanic' (Gussenhoven, 2004: 226), at least as far as contour variability goes. That said, recent work on Swedish shows that the four contours can be considerably modified, for example by delaying the peak caused by prominence H (Ambrazaitis, 2009, see also House, 2005) or by combining the accents with different boundary categories (Myrberg, 2009). Still, standard descriptions of CS intonational phonology do not include other accentual categories than the four contours assumed here, thus for now Gussenhoven's (2004) description of CS as a 'simpler' intonation system in CS than in English, German and Dutch seems legitimate, at least in the sense of contour repertoires. This difference between Dutch and Swedish is related to our third hypothesis, presented below.

3. Learning a prosodic system with a large repertoire of pitch contours makes it harder for children to establish what the relevant accentual categories are. This makes the learning task of Dutch-speaking children harder than that of CS-speaking children, who only have to learn four different contours, which are associated with the two lexical accent categories in predictable ways. A smaller repertoire of contours may make the CS system easier to acquire than the Dutch one, causing CS-speaking children to reach adult proficiency in the use of prominence H earlier than their Dutch-speaking peers master the use of accentuation.

Mapping between focus and accentual categories

Recent literature on focus and accentuation in English and German suggests that the mapping between pitch accentuation and focus is more probabilistic than has traditionally been assumed, particularly in terms of the de-accentuation of non-focal information, which seems mediated by factors such as rhythm and speech style, as well as the type of information structure contrast elicited (see e.g. Horne, 1991; Bard & Aylett, 1999; Terken & Hirschberg, 1994; Chen, 2007; de Ruiter, 2015). Work on prosodic focus marking in Dutch has shown that post-focal material is de-accented around 80% of the time in sentence-final position, but it is not clear how common post-focal de-accentuation is in non-final sentence-positions. Even if Chen (2009) reports on falls ('H*L') being the dominant accent type used for focus, other accent types are also attested, and there is considerable overlap in the mapping between focus and accent categories, for example in that falls ('H*L') are common on both focal and non-focal constituents sentence-initially, and that downstepped falls ('!H*L') are common on both focus and post-focus sentence-finally (see also Hanssen, Peters & Gussenhoven, 2008; Terken, 1984; Nooteboom & Kruyt, 1987).

Different from the more probabilistic mapping between focus and accentuation reported for West-Germanic languages, work on CS suggests that prominence H is

used rather reliably for marking focus (e.g. Bruce, 1998; Heldner, 2001; Ambrazaitis, 2009; Myrberg, 2009, 2013), [8] and that prominence H has a more restricted use in CS than pitch accentuation has in languages like English, German and Dutch (e.g. Gussenhoven, 2004; Ladd, 2008). Based on this, a fourth hypothesis is presented below.

4. Because the association between prominence H and focus is more reliable than the association between pitch accent placement, or certain pitch accent types, and focus in Dutch, determining the prosodic categories used for marking focus may be easier for CS-speaking children than for Dutch speaking children. This may cause CS-speaking children to acquire the adult system for marking focus earlier than Dutch-speaking children.

Focus and word duration

Thus far our attention has primarily been on the use of pitch manipulations for focus, as these are typically assumed to be the most important cues to focus in Dutch and CS. We have also seen that three out of our four hypotheses concerning pitch predict a faster acquisition rate for CS-speaking children than for Dutch-speaking children. At the same time, focus is also associated with increased word duration in both Dutch and CS (e.g. Chen, 2012; Heldner & Strangert, 2001). It is unclear whether the increased word duration associated with focus should be considered a consequence of accentuation (or prominence H) or a consequence of focus, or both. Our last two hypotheses concern the use of word duration for focus in the two languages, and more specifically, the relationship between duration and accentuation/prominence H:

5. If increased word duration is associated with accentuation or prominence H, the use of word duration for focus should follow the development of accentuation/prominence H, so that the effect of focus on word duration will be present when children reliably use prominence H or accentuation for focus.

6. If increased word duration is manipulated independently of accentuation, we may observe different developmental paths for duration on the one hand and accentuation/prominence H on the other, in our two groups.

---

[8] But see Myrberg (2009), who reports that lexcal accents without prominence H are occasionally used for marking focus. In these cases, the lexical accents are ‘boosted’, and tend to be produced with expanded pitch range, as compared to cases where they do not mark focus.

The six hypotheses presented above represent the starting point for the four studies presented in this thesis. In the hypotheses, cross-linguistic differences between the two languages are primarily assumed to affect the *rate* at which the children reach adult proficiency, and comments are only occasionally made in terms of how the relevant differences may affect the *route* or developmental path toward adult proficiency in more qualitative ways. Perhaps even more interesting than the point at which prosodic focus marking is acquired, is the question of how this skill develops over time in a given language, as previous work suggests that children's ability to prosodically mark focus develops quite gradually (e.g. Chen, 2011b). Further, cross-linguistic effects may well contribute in shaping this development. By including children from a rather a wide age (four to eleven years), and by including a wide range of prosodic measures, we will investigate whether and how the language a child is acquiring affects both the rate and the route toward adult proficiency.

1.5 The structure of this thesis

This thesis is structured around four studies, presented in Chapters 4 -7. The studies all take an experimental approach to the question of how children between four and eleven years, speaking either Dutch or CS, develop toward adult proficiency in prosodic focus marking. Whereas Chapters 4 and 5 concern the acquisition of prosodic focus marking in CS, Chapters 6 and 7 concern the acquisition of prosodic focus marking in Dutch. Some of the prosodic manipulations we include have been considered in previous work on prosodic focus marking in both adults and children (e.g. pitch accentuation, prominence H, pitch maximum, minimum and range, and word duration), while others have not been systematically investigated before (pre-focal pausing in Dutch and non-modal voice quality in Swedish). When publishing the four studies together in this dissertation we have chosen to add a general background Chapter (Chapter 2), introducing the topic of information structure, as well as reviewing relevant literature on the prosodic marking of information structure in adults and children. We round off Chapter 2 with a presentation of the autosegmental-metrical approach to prosody, specifying the theoretical assumptions from this framework that are adopted in this thesis. Following Chapter 2 we use Chapter 3 to present our methodology. The same general design was used in all our four studies. At the same time, the procedure was slightly adjusted for each of the two languages, as well as to the research questions addressed in the four studies presented. In Chapter 3 we therefore limit our presentation to aspects of the methodology that were shared across our four studies, whereas adjustments specific to a certain investigation will be described in the method section of the relevant chapter.

Following the general method chapter, Chapter 4 presents the first of our two studies on CS, concerning the use of prominence H for focus in CS-speaking children and adults. In Chapter 5 we present an investigation on the use of word duration and voice quality manipulations for focus in the same population (Chapter 5). Chapter 6 presents a study on accentuation, pitch and duration manipulations for focus in Dutch-speaking children and adults, and Chapter 7 presents as study on the use of pausing for focus in the same population. In Chapter 8 we summarize the findings from our four studies, sketching the developmental path to adult prosodic focus marking in Dutch and CS, followed by a general discussion of our findings in light of the hypotheses presented in Section 1.4.

2 General theoretical background

2.1 Introduction

The aim of this chapter is to provide the reader with a useful context for the studies presented in Chapters 4 -7, making it easier to see how this work is situated in the broader area of intonation research, as well as the smaller research field where the acquisition of prosodic focus marking in children is investigated. The chapter is structured as follows. First, we present theories of information structure (Section 2.2), discussing the way focus is defined and operationalized in the studies included in this dissertation. We will also comment on the distinction between *relational* and *referential* givenness-newness (Gundel & Fretheim, 2004), as these two related dimensions of information structure are often confused in the literature, and as keeping them apart makes it easier to interpret findings from previous work. In Section 2.3 we review previous work on prosody and information structure in adult speech production, with the aim of providing a general picture of what the adult model looks like with regard to prosodic focus marking in West-Germanic languages. The specifics of prosodic focus marking in Dutch and Swedish will be treated in more detail within Chapters 4 through 7. In Section 2.4 we review past work on prosody and information structure in children's speech production, and in Section 2.5 we review previous work on the perception of prosodic focus marking in children. At the end of the background chapter (Section 2.6) we present what Ladd (1996) coined as the autosegmental-metrical (AM) approach to intonational phonology, discussing assumptions from this framework that are adopted in this thesis, together with notational conventions that we will use when presenting and discussing the data.

2.2 Information structure

The term information structure or information packaging (Halliday, 1967; Chafe, 1976; Lambrecht, 1994) encompasses aspects of language that allow for speakers to adapt the form of their utterances to the current knowledge state of their communicative partners (Krifka & Musan, 2012). In this sense, information structure concerns *how* information is presented, as opposed to the information

itself. Across the world's languages, a number of linguistic devices are available for conveying information structural contrasts, among which are morphological markers (e.g. the topic-marking morpheme *wa* in Japanese), syntactic alternations (e.g. clefting or topicalization) and prosody (e.g. accentuation) (Vallduví & Engdahl, 1996; Krifka & Musan, 2012).

Much work on information structure, regardless of the choice of terminology used, shares the basic idea that some parts of a sentence anchor it to previous discourse ('given' information) while other parts make a contribution to discourse ('new' information), updating the common ground between speaker and listener (Vallduví & Engdahl, 1996). Several (typically dichotomous) information structural primitives have been proposed in the literature, wuch as *theme-rheme*, *topic-comment*, or *focus-background* (Vallduví & Engdahl, 1996). While focus is often defined in terms of newness or informativeness, Krifka (2007, inspired by Rooth, 1985; 1992) stresses that even if focus is *statistically* associated with 'new' or 'important' information, it is not the case that all instances of focus involve such characteristics, as illustrated by the sentence *Mary only saw HIM*, where the pronominal form is focal despite referring to a contextually given referent. In order to capture all uses of focus under one definition, Krifka maintains that the essential feature of focus is that it 'indicates the presence of alternatives relevant for the interpretation of linguistic expressions' (Krifka, 2007:6). In Krifka's sense, focus serves to highlight the presence of such alternatives, thereby directing the listener's attention to them.

In the work presented in this thesis we assume Krifka's definition of focus, as it captures a wider range of focus-related phenomena than more general notions of newness or informativeness. Still, our use of wh-questions for eliciting focus means that what we define as focal information in our data also represents contextually new information. Questions indicate the communicative goal of the questioner, and specifies to the person asked in what way the common ground[9] needs to be updated (e.g. Krifka & Musan, 2012). Assuming that the questioner is asking about something she does not already know, and that the responder cooperates by providing the requested information, the focal constituent of the answer will represent new information that the responder adds to the common ground, as illustrated in Examples 1 and 2 below. In the following we will refer to the constituent in brackets as the 'focal' constituent, whereas we will refer to constituents outside the brackets as referring to 'background' information, bearing in mind that the 'focalness' is a property of the information *referred* to by the constituent, not of the form itself.

---

[9] The notion common ground refers to information that is shared between the conversational partners (e.g. Stalnaker, 1974; Karttunen,1974; Lewis, 1979).

(1)
Question: What does Mary cook?
Answer: Mary cooks [a carrot]$_F$.

(2)
Question: What does Mary do with the carrot?
Answer: Mary [cooks]$_F$ the carrot.

Related to Krifka's (2007) example of focus on given information, Gundel & Fretheim (2004) propose that information structure actually involves two separate dimensions of givenness-newness relations, and argue that the fact that these are not always systematically separated has led to some conceptual confusion in the literature. Gundel & Fretheim refer to these two dimensions as *referential* givenness-newness on the one hand, and *relational* givenness-newness on the other. Whereas the division of a sentence into 'theme-rheme' 'topic-comment' or 'focus-background' implies that some parts of the information are new, more important, or highlighted *relative* to other information in that same utterance, the study of *referential* givenness-newness is concerned with the 'relation between a linguistic expression and a corresponding non-linguistic entity in the speaker/hearer's mind, the discourse (model), or some real or possible world' (Gundel & Fretheim, 2004:2). In this sense, most of the preceding discussion, as well as the way we conceptualize focus in our work, concerns the *relational* givenness-newness dimension: We assume that focal constituents refer to information that is highlighted *relative* to other parts in the utterance. Conversely, work on the *information status* or *cognitive status* of referring expressions (e.g. Gundel, Hedberg & Zacharski, 1993), would fall under the *referential* givenness-newness dimension. In the latter case, the information status associated with a certain referring expression is determined by the status the entity referred to has in the mind of the addressee, often defined in terms of the 'cost' required for the listener to retrieve the referent from their memory.[10] In our experiment we systematically manipulate relational givenness-newness by means of questions. In terms of referential givenness-newness, we consider our target referents to be activated in the working memory of the speakers (e.g. Chafe, 1976; Lambrecht, 1994), as they are presented in a naming task prior to the picture game, and thereafter repeated semi-randomly throughout the game. Because of the way our target constituents were systematically spread across trials (see Chapter 3),

---

[10] This division may also be thought of in terms of *relational* givenness being a more 'linguistic' dimension of information structure, whereas referential-givenness-newness is a more 'psychological' one, despite both dimensions having systematic effects on the linguistic forms chosen.

we do not expect any systematic effects of referential givenness-newness on our prosodic measures.

Returning again to the notion focus, this category is frequently subcategorized along two dimensions: in terms of the *size* or *scope* of the focus (Example 3 and 5) and in terms of presence or absence of an explicit *contrast* between the focus and an alternative candidate (Example 3 and 4).

(3)
Narrow (non-contrastive) focus
Person A: What is the dog doing with the cake?
Person B: The dog $[\text{throws}]_F$ the cake.

(4)
Narrow contrastive focus
Person A: The dog eats the cake.
Person B: The dog $[\text{throws}]_F$ the cake.

(5)
Broad focus
Person A: What's happening?
Person B: $[\text{The dog throws the cake.}]_F$

Starting with differences in focus size or scope, the term 'narrow focus' is used for cases where the scope of focus is relatively small (e.g. on a word or a small phrase, as in Example 3), whereas 'broad focus' is used when the focus is on a larger domain, entailing that the focal status is shared across the entire domain (e.g. across the whole sentence, as in Example 5). Broad focus is typically elicited by means of open questions like 'What happened' or 'What is new?', but in general, sentences that are said 'out of the blue' and only contain new information are said to have broad focus. While narrow focus is typically marked by accentuation in West-Germanic languages (e.g. Ladd, 2008), the relation between accentuation and focus domain is more indirect in broad focus, where an accent on one constituent within the focus domain can mark the larger constituent as focal; a process referred to as *focus projection*[11] in the literature (see Gussenhoven, 1983, 1999; Selkirk, 1984, 1995; Büring, 2003, 2006; Welby 2003). Unlike the marked narrow focus cases, where the main prominence is moved to a non-default position, broad focus is associated with a default alignment of prosodic prominence. In West-Germanic languages this means aligning prosodic prominence with the final argument within

[11] Another term used for this phenomenon is 'integrative accent' (e.g. Fuchs,1984).

an intonation phrase (e.g. Féry, 2013), meaning that cases where the last argument of an intonation phrase is accented are ambiguous between a narrow focus and a broad focus interpretation. To illustrate, the sentences *The baker cooks a CARROT* and *the BAKER is arrested*[12] could both be either broadly focal or involve narrow focus on the accented NP (e.g. Ladd, 2008).[13]

When focus is subcategorized in terms of *contrastivity*, the distinction concerns whether or not a contrast is evoked between the focal referent and an alternative candidate. Contrastivity is often manipulated by explicitly proposing alternative candidates to the focal referent, as in Example 4 above, but contrastivity can also be implied from the context without explicit mention. Finally, contrastivity is not only relevant for focal referents; other information structural primitives can also be contrastive, as in the case of contrastive topics (e.g. Roberts, 1996; Büring, 2003). The extent to which contrastivity is to be seen as a category on its own has been a matter of debate. Whereas some researchers consider contrastive focus to be something different from non-contrastive focus (Chafe, 1976, Halliday, 1967, Molnár, 2002; Kiss, 1998), others have argued that there is no principled difference between narrow focus and contrastive focus, as a set of alternatives candidates is evoked in either case (Bolinger, 1961; Rooth, 1985; 1992). Clearly, assuming that there are differences between contrastive and non-contrastive focus does not necessarily entail that this difference is linguistically marked, either by prosody or by other manipulations. In our work we follow Kiss (1998) in assuming that contrastive focus differs from information focus with respect to the size of the set of alternatives evoked, restricted in the case of contrastive focus and not restricted in the case of narrow non-contrastive focus. By eliciting both narrow and contrastive focus, we examine whether our adult participants prosodically highlight the two focus types in similar ways. In addition, we also ask whether differences can be observed in how children develop in their ability to prosodically mark contrastive and non-contrastive focus.

An important reason for us to underline the difference between what Gundel & Fretheim (2004) coin referential and relational givenness-newness, is that previous work on prosody and information structure has targeted both dimensions, even if this has not always been made clear when different investigations are compared and reviewed. Sometimes the type of givenness-newness relation studied is linked to the

---

[12] We adopt the convention of capitalizing accented words.
[13] This kind of ambiguity has motivated researchers to study whether these accents have different phonetic realizations for broad versus narrow focus, or whether patterns of pre-nuclear accentuation differ between broad and narrow focus (e.g. Baumann et al., 2007 on German; Hanssen, et al. 2008 on Dutch).

methodological design chosen, where question-answer paradigms typically target the relational dimension, whereas experiments that do not use question-answer setups, or do so in less structured ways, often end up targeting more relational kinds of givenness-newness. Even with considerable terminological overlap, as well as relatively similar prosodic patterns associated with newness versus givenness-categories, it is not clear at this point whether the prosodic means used for marking these two dimensions of information structure are exactly the same. In the following section we will review previous work on prosody and information structure in adult speech production, and we shall see that distinguishing between relational and referential givenness-newness is helpful when interpreting the results of the reported investigations.

2.3 Prosody and information structure in adult speech production

The aim of this section is to provide the reader with an overview of the prosodic manipulations that have been shown to mark both relational and referential newness-givenness contrasts in West Germanic languages. The specifics of prosodic focus marking in Dutch and Swedish will be treated in more detail within Chapters 4 -7.

Studies of prosody and information structure frequently differ in terms of whether they take a phonetic or a phonological approach to the prosodic manipulations in question (see Chen, 2012, for a review). Whereas more phonetic approaches tend to compare acoustic measures of F0 (i.e. height of pitch peaks and valleys, pitch slope, pitch range, etc.), duration (i.e. word, syllable, segment, or pause durations) or intensity (i.e. over-all or within certain frequency bands) directly across various information structure conditions (e.g. Eady & Cooper, 1986; Xu & Xu, 2005; Breen, Fedorenco, Wagner & Gibson, 2010), more phonologically oriented approaches are based on the coding of pitch accents and boundary markers (i.e. placement and type of pitch accents and boundary tones), and the effect selected information structure contrasts have on the distribution of these phonological categories (e.g. Pierrehumbert, 1980; Chen, 2007; Hedberg & Sosa, 2008). In addition, several studies adopt a combined approach, including both phonological coding and phonetic measurements (e.g. Baumann, Grice & Steindamm, 2006; Myrberg, 2013; Braun, 2006).

Below we will first review findings related to relational newness-givenness, as those are at the centre of attention in this dissertation. Following this we will present some work on referential givenness-newness, and we shall see that the prosodic effects

associated with this dimension of information structure are not all too dissimilar from the effects of relational givenness-newness.

Starting with the use of accentuation for focus, adult speakers of English, Dutch and German are reported to assign pitch accents to narrowly focal constituents and to avoid accentuation post-focally (e.g. Ladd, 1996; Gussenhoven, 2004; Féry 1993). Whereas post-focal information tends to be de-accented, pre-focal information is often accented, at least in sentence-initial position (e.g. Chen, 2009; Baumann, Becker, Grice & Mücke, 2007). Both rhythmical factors (Horne, 1991, Terken & Hirschberg, 1994) and information status (e.g. Pierrehumbert & Hirschberg, 1990; Hirschberg et al., 2007) seem to affect the extent to which non-focal information is accented. In addition to descriptions of focal constituents being accented, some authors have also proposed that specific accent types are associated with focus. Despite differences between the notation systems used for capturing different pitch accents in Dutch, German and English (see also Section 2.6), it seems that in these languages, focus tends to be marked with a falling accent (Ladd, 2008; Gussenhoven, 1984; Féry, 1993), even if other accent types also occur (e.g. Chen, 2009).

In terms of phonetic effects, focal constituents are typically produced with an expanded pitch range, higher mean or maximum F0, increased word or syllable duration and higher intensity, as compared to non-focal constituents (Chen, 2012; see also Ladd, 2008; Gussenhoven, 2004; Jun, 2005). Most studies on the phonetics of focus marking in adult speech do not actually address differences between focal and non-focal constituents. What is typically compared in these studies are measures of pitch and duration across different focus *types* (e.g. broad, narrow, and contrastive focus). Results of such comparisons reveal differences within West-Germanic languages in the way different focus types are prosodically distinguished. Whereas Xu & Xu (2005) report on word duration being used to distinguish between broad, narrow, and contrastive focus in American English, no such effects of duration were found in a comparable study on British English (Sityaev & House, 2003). In addition, whereas contrastive focus is associated with a later peak alignment than broad focus in German (Baumann, Grice & Steindamm, 2006; Baumann, Becker, Grice & Mücke, 2007), the opposite pattern is described for Dutch, with contrastive focus involving an earlier peak alignment than broad focus (Hanssen, Peters & Gussenhoven, 2008, see also Chen, 2012). The differentiation between contrastive and narrow focus also seems to differ between languages. Whereas reports on West-Germanic varieties show limited or no prosodic differences between contrastive and narrow (non-contrastive) focus (t' Hart, Collier & Cohen, 1990; Hanssen, Peters & Gussenhoven, 2008; Baumann, Becker, Grice & Mücke, 2007; Myrberg, 2013), reports on Italian suggest that different pitch accent

types are used to distinguish contrastive from non-contrastive focus (e.g. Avesani & Varya, 2003; Krahmer & Swerts, 2001). Interestingly, in a study from Breen, Federenko, Wagner & Gibson (2010), phonetic effects of contrastivity were only found when speakers were deliberately trying to communicate the distinction between contrastive and non-contrastive focus, suggesting that the extent to which differences are observed between contrastive and narrow focus may partly depend on the elicitation method applied.

Whereas the studies reported above have predominantly been concerned with relational givenness-newness, other studies have addressed referential givenness-newness relations. For example, Röhr & Baumann (2010) showed that when new, accessible (i.e. explicitly introduced in the non-immediate preceding context or implicitly inferable from the context) and given referents were compared, the likelihood of accentuation increased from given through accessible to new information. Phonetically, going from given via accessible to new information was also associated with higher and later accentual peaks. Similar patterns of accentuation are reported by Swerts, Krahmer & Avesani (2002) who found higher accentuation rates on referentially new than on referentially given referents in Dutch. In terms of accent type preferences, Baumann (2006) and Baumann & Grice (2006) report German speakers to prefer medial peak accents ('H*') for new information, early peak accents ('H*+L') for accessible information and de-accentuation for given information. Interestingly, in another study on German, de Ruiter (2014) found far more diverse mappings between information status and particular accent types than what was reported by Baumann (2006) and Baumann & Grice (2006). De Ruiter interprets these different results as related to her dataset involving a more spontaneous speech style than the previous reports on German, suggesting that spontaneous speech involves more variable accent-to-information-status mappings than data gathered in more structured experiments. In addition, de Ruiter also points out that the fact that listeners judge certain accents as more appropriate than others (Baumann & Grice, 2006) does not necessarily entail that listeners only produce such 'appropriate' patterns when speaking.

As can be seen from this brief review, information structure systematically affects the way utterances are prosodically modified, at least in the languages discussed here. At the same time, several authors have shown that there is individual variation in the use of various prosodic manipulations (e.g. Baumann, Grice & Steindamm, 2006) and that the consistency with which certain prosodic parameters are mapped to information structure varies across different speech styles (e.g. de Ruiter, 2015). One important limitation of previous work on adult prosody is that the prosodic manipulations investigated are typically only considered in one specific sentence-position (e.g. only medial, only final, etc.), making it hard to establish whether

speakers use the same means for marking focus across different sentence-positions. Results from Chen (2009) on Dutch suggest that adult speakers actually mark focus in different ways when sentence-initial and sentence-final positions are compared. More work is also needed in order to determine how closely related the prosodic means of marking relational versus referential givenness-newness distinctions really are. Finally, work on prosody and information structure should be extended to more spontaneous speech styles, making sure that the patterns found in more controlled speech is actually representative of the way speakers prosodically highlight focal information 'in the wild', that is, under less controlled circumstances.

2.4 Prosody and information structure in children's speech production

Previous work on the acquisition of prosodic focus marking in children is rather limited, and has primarily involved West-Germanic languages. Drawing general conclusions based on the reports presented is a difficult task, as studies differ greatly in terms of the prosodic manipulation investigated (narrow phonological transcription, listener judgment, acoustic analyses of pitch and duration) and the information structural dimensions considered (contrast, focus, discourse accessibility, topic), as well as the ages and languages involved. Against a background of substantial methodological variability, cross-linguistic differences observed between studies are hard to isolate. Below we will review past production studies on children learning West-Germanic languages by taking three dimensions of information structure into account; focus or relational givenness-newness, referential givenness-newness and contrast, following Chen (2015). We will also mention some of these studies in the background sections of Chapters 4 -7, but the summaries included in those chapters will only include work particularly relevant for the study in question (e.g. centring on certain prosodic measures or on certain languages). In other words, the review presented here is meant as a point of reference for the later chapters, providing more detail than what will be found later in the thesis.

Starting with a few words on intonation development in toddlers, several studies have shown that children learning English, Dutch, Catalan and European Portuguese produce the inventory of pitch accents and boundary tones found in these languages already by the late two-word stage, even if the distributions of these patterns may be different from those of adults (Balog & Snow, 2007; Chen & Fikkert, 2007; Prieto et al., 2012; Frota & Vigário, 2008). Similar findings are reported for Swedish, where children distinguish the two lexical accent categories in their production by the age

of two (Kadin & Engstrand, 2005; Engstrand, Williams & Strömqvist, 1991). Findings like these suggest that the building blocks of the prosodic system the children are acquiring are in place relatively early. At the same time, children's ability to produce the inventory of contours at an early stage does not entail that they behave like adults in the way different contours are used to convey different meanings. Indeed, this is also what is reported in corpus studies on German and Dutch-speaking two-year-olds: Children around this age tend to accent both words in their early two-word utterances, regardless of givenness-newness relations, whereas adults typically de-accent given information (Behrens & Gut, 2005; Chen & Fikkert, 2007). This said, a re-analysis of the Dutch data presented in Chen & Fikkert (2007) showed that in the cases where the children failed to de-accent given information, they often de-voiced or downstepped their accents, indicating that *some* focus-mediated manipulations are already in place at this stage, even if the children were not yet adult-like in their use of accent placement (Chen, 2011b). A fourth study concerning effects of information structure on accentuation in young children comes from Wieman (1976), describing spontaneous two-word utterances from English-speaking two-year-olds. Wieman observed that the children's accentuation patterns were determined by semantic relationships between the two words. Interestingly, whereas noun-locative utterances were typically produced with an accent on the locative, this pattern was changed when the noun represented new information, in which case the noun was produced with an accent, whereas the locative was not. This said, Wieman's dataset consisted of seven observations only, thus the generalizability of her findings is fairly limited.

Most studies on the acquisition of prosodic focus marking have been concerned with children between three and five. In one of the earlier studies on this age range, Hornby & Hass (1970) report on the use of contrastive stress on contrastive referents in English-speaking three to four-year olds. The authors used a picture-task where picture pairs differing by one element were presented in a sequence, and the children's task was to describe the pictures, answering the question 'what happens in the picture?'. In the 12 picture pairs used, the element changing corresponded either to the subject, the verb, or the object of target subject-verb-object sentences (i.e. *the man <u>washes</u> the car - the man <u>drives</u> a car* or *<u>the girl</u> rides a bicycle- <u>the boy</u> rides a bicycle*). Hornby & Hass found that the children were more likely to use contrastive stress on the changed elements in the second pictures than on the original elements in the first pictures, even if this effect was stronger on subjects and verbs than on objects: 80% contrastive stress on contrastive subjects, 56% of the time on contrastive verbs, and 43% of the time on contrastive objects. Contrastive stress was

not explicitly defined, but it was scored both by a graduate student and by one of the authors, with 86% agreement. [14]

Using a comparable design as the one described in Hornby & Hass (1970), MacWhinney and Bates (1978) also describe the use of emphatic stress on contrastive referents in English-speaking children. In this study, 'emphatic stress' was defined as a word involving comparably more 'intonational stress' than any other item in the response, as well as more stress than what would have been assumed in a 'neutral' rendition of the sentence. Picture descriptions were elicited from children between three and six years of age, and sequences of three pictures were used, where one element (a noun, a verb, or a preposition) differed across the three pictures (i.e. *a bear is crying, a mouse is crying, a bunny is crying).* Three pictures were chosen over two in order to observe an assumed increase in newness for the contrastive element (underlined) across picture sets. The results reported by MacWhinney & Bates mirror the findings from Hornby & Hass, in that the children used more emphatic stress as newness increased. This effect was already seen at three to four, but the authors also report increased consistency in the use of emphatic stress toward the age of six.

In both Hornby & Hass (1970) and MacWhinney & Bates (1978), the information structural category involved was contrastivity, as what was referred to as 'new' or 'contrastive' elements always involved a direct contrast with a previously mentioned candidate (see Section 2.2). At the same time, the question used for eliciting picture descriptions was a typical broad-focus question (i.e. *what happens?),* thus the kind of 'newness' referred to may be considered more of a referential than a relational kind (e.g. Gundel & Fretheim, 2004; Chen, 2015). Relevant for our own work, these two studies suggest that children as young as three can mark contrastive information with prosodic emphasis, but questions remain as to whether children can also do this when no explicit contrast exists, not to mention whether the prosodic manipulations produced by children mirror those of adults, as adult controls were not included in these two studies.

In a more recent paper on English, Wells, Peppé & Goulandris (2004), report on the production of contrastive prosody in children between four and eleven. In this study, production of contrastive accentuation was examined in adjective-noun pairings, and the experiment was constructed as a picture-task where the children had to correct an experimenter's description of the pictures. The results of this study showed that

---

[14] It seems fair to assume that what is referred to as 'contrastive stress' involved pitch accentuation, possibly also accentuation with a comparably larger pitch excursion than other accents in the sentence.

children between four and eleven predominantly accented the contrastive element in their phrases, but that the four- to five-year-olds also occasionally misassigned accents, primarily in terms of a preference for accenting the sentence-final noun also in cases where only the adjective should have been accented. In other words, even if the children primarily performed in line with adults, the four- to five-year-olds were less consistent in assigning accents to contrastive information than adults and older children (Wells, Peppé & Goulandris, 2004).

In a more recent study on English-speaking children, Wonnacott & Watson (2008) used a video setup where short clips, illustrating transitive events, were presented to the children. Across two consecutive clips, the subjects could either be the same (condition 1) (*The bee hit the ladybug - the bee hit the lion),* a new subject could be introduced (condition 2) (e.g. *The bee hit the ladybug - the giraffe hit the ladybug)* or the subject of the second sentence could be shifted from direct object in the first sentence to subject in the second (condition 3) (e.g. *The bee hit the ladybug - the ladybug hit the giraffe).* Wonnacott and Watson found that their three- to five- year old children used a higher mean pitch and a higher mean intensity on subjects that appeared for the first time (i.e. *the giraffe* in 2) or were shifted from object to subject position (i.e. *the ladybug* in 3), than when they were repeated with the same grammatical role as in the preceding clip (*the bee* in 1). As argued by Chen (2015), questions were not used to render focus on certain constituents, thus this study seems to target a form of contrast combined with referential accessibility (e.g. Gundel & Fretheim, 2004) rather than focus. The results nevertheless show that children can use pitch and intensity manipulations to mark relational givenness-newness relations in sentence-initial position. Like the other studies on English, the study does not shed light on whether children at this stage can use prosodic manipulations to highlight focal referents that do *not* involve a contrast with a previously presented element. In addition, as the study involved acoustic measures only, it does not tell us anything about whether the children produced the same accentual categories as adults do, particularly as the adult data to which the children's productions were compared was taken from a different study (Watson, Arnold & Tanenhaus, 2005). Interestingly, the children participating in Wonnacott & Watson (2008) were described as adult-like in their use of pitch and intensity, but not of duration, suggesting that they had not yet mastered the complete repertoire of prosodic modifications used by English-speaking adults when differentiating between the new, accessible, or shifted elements (Wonnacott & Watson, 2008).

The relationship between information structure and prosody has also been studied in German-speaking children. In a study from Müller, Höhle, Schmitz, & Weissenborn (2006), four- to five-year-olds systematically produced words referring to contrastively focal information with a higher mean pitch than words representing

given information, independently of sentence-position and grammatical role. In this study, contrastive focus on subjects or direct objects was elicited by means of a question-answer task. In order to make the children produce full transitive sentences, a puppet mediator was introduced, and the children had to answer the questions by using the words produced by the puppet. The use of prosody was investigated on both SVO and OVS sentences, and wh-questions were used for rendering either the subject or the object contrastively focal. As the only output measure in this study was mean pitch, it is hard to say whether the contours produced by the children were categorically the same as those produced by the adult controls, even if they were comparable in mean pitch. In another study on German, Sauermann, Höhle, Chen & Järvikivi (2011) investigated the use of word order and prosody in German-speaking four-to five-year-olds, as compared to adult controls. In this study the children also answered wh-questions about pictures, re-constructing SVO-answers provided by robots, speaking with a flat intonation pattern. The conditions elicited involved broad, narrow, or contrastive focus on either the subject or the object, thus different from the study from Müller and colleagues, the comparison was between different focus types rather than between focal and non-focal conditions. The participants were allowed to produce different word orders than the robot, who produced sentences with both OVS and SVO order. The adults always reproduced the robot speech with SVO order, but the children used both orders. Measures of pitch minimum, pitch maximum, pitch range, and word duration were compared on subject and object nouns across focus type conditions. The results showed that the adult controls produced subject nouns with a larger pitch range in narrow or contrastive focus than in broad focus, and they produced contrastive focus with a lower pitch minimum than broad focus. The analysis of the object nouns showed the adults producing these with a marginally larger pitch range under contrastive focus as compared to broad focus, but no other effects on pitch or duration were found. On the SVO sentences, the children expanded the pitch range on narrowly or contrastively focal subjects, as compared to broadly focal ones, similar to what was found in the adults. In addition, the children also raised the pitch maximum for narrow and contrastive focus, compared to broad focus, unlike the adults, and no effects of focus condition were observed on the objects. On their OVS sentences, the children expanded the pitch range on narrowly focal subjects as compared to broadly focal ones, whereas the difference between contrastive and broad focus was only trending. In addition, contrastively focal subjects were marginally longer than broadly focal ones. Again, no differences between the conditions were found on the object nouns. In other words, Sauermann et al. show that children make some focus-mediated manipulations, in some sentence-positions, but that the way they do this differs from the behaviour of adults.

In addition to the work on toddlers reviewed above, Chen has also conducted a number of studies on prosodic focus marking in Dutch children between four and eight years (Chen 2007, 2009, 2011a, 2011b). In these studies, SVO-sentences with subjects as focal and objects as non-focal, or the reverse, were elicited by means of a picture-matching task. The children answered wh-questions about pictures lacking information either about the subject or about the object, thereby grouping corresponding pictures into pairs. Similar to the puppets used by Müller and colleagues (2006), robot mediators speaking with manipulated intonation[15] were used as 'mediators', making sure the children produced the target SVO structure, and that they used the intended labels for the items appearing in the pictures. Before answering the experimenter's question, the children heard the answer produced by the robot, and they had to reconstruct this answer using their own voices. Chen's results showed that both four-to five-year-olds and seven-to eight-year-olds predominantly accented focal targets in both sentence-initial and sentence-final position, but that non-focal targets tended to be accented sentence-initially and de-accented sentence-finally. Since this pattern was in line with what adults did in a previous experiment (Chen, 2009), Chen concluded that children already at four- to five are adult-like in their use of accent placement for focus (Chen, 2011a). Subsequent analyses of accent type choices among children and adults nevertheless revealed that whereas the adults showed a preference for falls ('H*L') for marking sentence-final focus, the accent type choices were more variable in the four-to five-year-olds, including a higher proportion of rises ('L*H') than what was found in the adult data. In terms of accent type preferences, the children between seven and eight years performed in line with adults, preferring falls for focus (Chen, 2011a). In a separate study, on the phonetics of sentence-initial focus, phonetic comparisons were made between focal and non-focal falling accents ('H*L'), showing that children at four-to-five did not use pitch range and duration to distinguish focal from non-focal falls in the way adults did (Chen, 2009). Running comparable analyses on the phonetics of focus-marking in children between seven and eight years showed that the seven- to eight-year-olds preferred high or rising accents ('H*') for non-focal target words, and falling accents ('H*L) for focal ones, adapting a non-adult like strategy for distinguishing focal from non-focal targets. Based on these findings, Chen concluded that whereas accent placement for focus is acquired already at four- to five, the use of accent type sentence-finally and the use of phonetic modifications

[15] The robot speech was constructed based on splicing together individual words produced by a female native speaker of Dutch, adding a 200 ms pause between each word. Further, original pitch patterns of the word productions were erased, and the pitch was set to 200 Hz for all words (see Chen, 2011a).

where phonological cues (e.g. accent placement sentence-initially) do not suffice, are not mastered before the age of seven to eight years, or even later (Chen, 2011b).

As can be seen from this review, child speakers of English, German and Dutch can make both phonological and phonetic adjustments to mark contrast, as well as differences between referential givenness and newness, by the age of four or five. Nevertheless, their ability to mark new information or contrast is still developing beyond this age, with mistakes in accent assignment and non-adult accentual patterns reported well into school age, both in terms of phonetic manipulations and in the use of phonological categories. Further, more work is needed in order to determine how children develop the ability to mark non-contrastive focus, as the majority of previous investigations has been concerned with target words that involve an explicit contrast with alternative candidates. Finally, as several of the studies reported above did not compare children and adults using the same task, it is not clear whether the patterns reported in child speech actually match the behaviour of adults, when these are faced with a similar task.

2.5 The perception of prosodic focus marking

The studies reported in this thesis are all concerned with the *production* of focus prosody in children, but work has also been carried out in order to determine when children are able to make use of prosodic cues to focus in their speech *perception*. One important challenge for the study of children's perception of focus prosody is that developing suitable tasks is difficult. In some studies, metalinguistic experiments have been used, where children hear a sentence with a particular accentuation pattern (e.g. *John's got FOUR oranges, I wanted CHOCOLATE and honey*), and are asked to judge which picture description is more 'suitable' from a set of candidate pictures (Cruttenden, 1985) or to judge what the speaker did NOT want, also based on pictures (Wells et al., 2004). In Cruttenden's study, ten-year-olds made wrong judgments more than a quarter of the time, and in the study from Wells and colleagues, five-year-olds performed at chance, and 90% accuracy was only reached as late as at thirteen years old. Reports like these are puzzling, considering the production studies reported above, showing that children as young as three to four years can use accentuation to mark contrast, even if development is also observed well beyond this age. This apparent asymmetry between production and perception in children's acquisition of prosodic focus marking has been noted and discussed by several researchers. For example, both Chen (2010) and Ito (2014) attribute the discrepancy in findings to methodological issues, claiming that metalinguistic tasks like the ones described above require much more complex reasoning than the association of accentuation to focal information, and Ito (2014)

additionally highlights the need for an appropriate context when the use of accentuation in reference resolution is examined.

In addition to discussing the alleged asymmetry between production and perception of prosodic focus marking in children, Chen (2010) presents a reaction-time experiment of children's processing of the accent-focus mapping, modelled after work by Birch & Clifton (1995). The rationale behind using reaction times was that inappropriate accent-to-focus mapping would increase the reaction time in adults, because processing an utterance with incongruent prosody is expected to take longer than processing an utterance with congruent prosody. If children were indeed not sensitive to the relationship between accents and focus, they would not be expected to show such slowed down processing speed when faced with incongruent prosody-focus mapping. The experiment was constructed as a game on a computer, showing a boy looking at pictures with his three pets. In order to determine which animal knew the pictures best, the boy asked the animals about the pictures, and the animals described them. The participant's task was to push either a red or a green button, depending on whether they judged the animal to answer 'correctly' or not. Question-answer pairings with congruent and non-congruent accentuation patterns were used as stimuli, combined with fillers involving lexico-semantic errors or mispronunciations. The results showed children's reactions to slow down with incongruent focus-prosody mapping, similarly to those found in the adults, whereas accentuation had no effect on correctness-judgments in either group. Based on these findings Chen concluded that when appropriate tasks are used, children show the same effects of accentuation on their sentence processing as adults.

Adult-like accentuation processing is also reported by Ito and colleagues (Ito & Speer, 2008; Ito, Jincho, Minai, Yamane & Mazuka, 2012). In a series of eye-tracking studies, these authors show that six-year-olds display the same facilitative effect of accentuation as adults when asked to find a correct target item in a grid. Further, infelicitous prosody (*Where is the purple squirrel? Now look at the ORANGE monkey*) led to incorrect fixations (i.e. on an orange squirrel) in both children and adults.[16] Together with the findings from Chen (2010), the work from Ito and colleagues shows that children between four and six do use prosody to guide their focus interpretation, and that this can be systematically shown using online methods. At the same time, the experiments designed by Ito and colleagues have shown that the effect of accentuation on contrast resolution is sensitive to the prosodic manipulations made, to the presence versus absence of an appropriate

[16] This effect was only found when the design included pauses between trials; without these the children showed a tendency to perseverate, fixating on the previously mentioned animal.

context and to the amount of options available in the grid, where fewer options increase the salience of the contrast, sometimes even cancelling out the effect of accentuation (Ito, 2014).

To summarize, whereas some previous work on the comprehension of focus prosody has indicated that children's acquisition of the prosody-focus mapping develops fairly late, recent studies using on-line measures have shown children between four and six to pattern close to adults in their processing of accentuation for reference resolution. At the same time, more experimental work is needed in order to determine how early children start behaving like adults in their interpretation of prosodic prominence.

2.6 Autosegmental-metrical approaches to prosody

Since most work on Dutch and Swedish prosody has been carried out within the autosegmental-metrical (hereafter AM) theory, and because this framework makes it possible to take both the phonological and the phonetic properties of prosody into account, we also adopt this approach in this thesis.

Since its first appearance in the eighties, the AM theory has become one of the most widespread approaches to intonational phonology (Bruce, 1977; Pierrehumbert, 1980; Ladd, 1983, 1996; Gussenhoven, 1984; Liberman & Pierrehumbert, 1984; Beckman & Pierrehumbert, 1986; Pierrehumbert & Hirschberg, 1990). Comprehensive introductions to this framework can be found in Ladd (2008), Gussenhoven (2004) and Jun (2005). The theory is autosegmental because it assumes that tones are represented as autonomous segmental categories (Goldsmith, 1976), and it is 'metrical' because it assumes that tones are associated with prosodic categories contained in a hierarchically organized structure of phonological constituents (Gussenhoven, 2002).

Within the AM approach, pitch contours are captured in terms of high (H) and low (L) tones, which can be combined into more complex tonal patterns. These patterns align with segmental material in ways that are specified by the phonological grammar of a language: one language may have the syllable as the so-called tone-bearing unit (TBU), whereas in another language the mora may be the TBU. Two kinds of tonal elements are recognized: pitch accents mark prominence relations within prosodic domains of various sizes (e.g. the marking of focus), and boundary tones mark the boundaries of such domains (e.g. marking the left or right edge of an intonational phrase). The number of prosodic levels (e.g. phonological phrase,

intonational phrase, intermediate phrase, prosodic word, etc.) assumed to be part of the phonological representation of a language is a matter of debate, and is likely to be shaped both by the phonological system of the language in question, as well as by the ways different researchers choose to represent these systems. As an example, Beckman & Pierrehumbert's (1986) analysis of American English assumes that both *intonational phrases* and *intermediate phrases* are tonally marked, whereas other analyses of the same language only assume the level of the intonational phrase (e.g. Gussenhoven, 2002, 2005).

Summarizing, the AM approach describes the tonal structure of an utterance as a sequential combination of tones that can be assigned to categories at different levels in the prosodic hierarchy (e.g. phonological phrase, intonational phrase and prosodic word). Ladd (2008) underlines the sequentiality assumed in AM theory as contrasting with what he refers to as superposition, or 'overlay' approaches to intonation, where the latter model the pitch contour as the result of different functions being simultaneously conveyed, so that local accentual peaks are superimposed on more global shapes (e.g. Xu & Xu, 2005; Kochanski, Shih & Jin, 2003; Ladd, 2008). Whereas overlay models tend to assume a rather direct relationship between various pragmatic functions and the surface contour, the AM theory maintains that the tonal structure of an utterance is phonological and abstract, and thus separated from its phonetic implementation (Gussenhoven, 2004).

In Section 2.3 we saw that pitch accents are used to mark focus in English, Dutch and German. Pitch accents are phonological categories that can consist of single tones (e.g. high 'H' or low 'L') or of combinations of high or low tones, resulting in more complex contours (e.g. falling 'HL', rising 'LH' or fall-rising 'HLH'). The use of the term 'accent' reflects the fact that in a number of languages, these tones or tone sequences are associated to metrically strong syllables, but importantly, pitch accents are considered independent from phonological stress (e.g. Liberman, 1975). In addition to the convention of annotating tonal events in terms of high and low tones, an asterisk (*) marks the alignment of a particular tone with the stressed syllable, so that 'H*L' means that it is the high tone that is associated, whereas in 'HL*' it is the low tone that is associated. High tones that are lowered relative to a previous high tone, typically within the same IP, are referred to as 'downstepped'. The last accent in an intonational phrase is referred to as the 'nuclear accent', whereas accents preceding this one are referred to as 'pre-nuclear'. The terms 'nuclear' and 'pre-nuclear' do not entail different phonetic realization of the same accent, but it is not uncommon for the same accent to have slightly different realizations depending on whether it appears in nuclear or pre-nuclear position. In addition, nuclear accents appear more frequently together with boundary tones than pre-nuclear accents, thus statistically they more commonly take part in more

complex tonal patterns than pre-nuclear accents. The special status of the nuclear accent is also structurally motivated, in the sense that they tend to mark the edge of a focal domain (Ladd, 2008).

Despite the dominance of the AM theory in the field of intonation research, alternative theories also exist, perhaps primarily represented by proponents of more acoustically or phonetically oriented approaches to prosody. For example, Kochanski and colleagues (Kochanski & Shih, 2003; Kochanski, Shih & Jin, 2003) argue that the global shape of the F0 is better accounted for in terms of continuous phonetic parameters that can be mathematically modelled, than by assuming abstract tonal events that are mapped onto the F0 curve as part of phonetic implementation. A similar approach is taken by Xu and colleagues (e.g. Xu & Wang, 2001; Xu & Xu, 2005). In Xu & Xu's (2005) analysis of English, the sequentiality assumed by the AM framework is replaced by parallel encoding of different prosodic functions, so that the resulting contour is seen as a combination of different functions working in tandem. Melodic primitives (pitch targets, pitch range, strength and duration) are considered as the true building blocks of intonation, and every syllable is assumed to involve a specific pitch target. As pointed out by Ladd (2008), a number of issues still remain unsolved within the field of prosody research, and it is likely that insights from different perspectives need to be combined in order to develop better accounts of how prosodic systems work and vary.

The AM framework has been used as a starting point for developing conventions for intonational transcription in a number of languages. These systems are based on the ToBI ('Tones and Break Indices') for transcribing American English (Pierrehumbert 1980; Beckman & Pierrehumbert, 1986; Silverman et al., 1992; Pitrelli et al., 1994), as well as revised versions of this system (see Jun, 2005, for an overview). Despite differences between transcription systems established for different languages, reflecting both differences between languages and differences between research traditions, the AM theory has contributed in establishing a common ground for describing cross-linguistic differences between prosodic systems, based on a limited set of assumptions about what intonational phonological structures look like and how they can be modelled (see Beckman, Hirschberg & Shattuck-Hufnagel, 2005).

In this thesis we will draw on insights from the AM framework, aiming to provide systematic descriptions of CS and Dutch-speaking children's development toward adult prosodic focus marking. In addition to studying prosodic manipulations that are well-described in adult Dutch and Swedish, we also go beyond these phenomena, adding measures of voice quality and pausing that have not yet been systematically investigated in previous accounts of Dutch or Swedish prosody. We also combine phonological coding with phonetic measurements, verifying the

categories we assume with more quantitative measures. Previous work shows that both categorical and more gradual prosodic manipulations take part in prosodic focus marking, both in the distinction between focus and non focus (Chen, 2009; Baumann et al., 2006) and in the distinction between different focus types (Baumann et al., 2007; Hanssen et. al., 2008; Myrberg, 2013). Since we are studying developing systems, combining coding of accentual categories like falls or rises with acoustic measures of pitch range and word duration allows us to capture possible intermediate steps children may take in their development toward adult proficiency. The relevance of such an approach is illustrated by previous work suggesting that children develop rather gradually in their ability to prosodically mark focal information, and that some manipulations seem to be developing earlier than others, suggesting that children resort to intermediate steps in their development of prosodic focus marking, and that these steps do not necessarily match the adult phonological categories (e.g. Chen, 2011b).

3 General method

3.1 Introduction

All the data presented in this thesis was collected by means of a picture-matching game, adapted after a procedure developed by Chen (2011a). In order to prevent too much repetition in Chapters 4 -7, we will use this chapter to describe the picture-matching game in detail. Within Chapters 4 -7 we will also comment on methodological issues, but then the presentation will be limited to aspects relevant for each particular study, as well as on methodological adaptations that were made on the Swedish or the Dutch version of the game.

Most previous work on the prosodic realization of information structure in adults has been concerned with read or strictly controlled speech. This makes sense, as detailed prosodic analysis requires strict control over the target words investigated. It is however interesting to note that several studies show that the de-accentuation (or prosodic reduction) of given material (i.e. referential givenness-newness) is less common in spontaneous speech than what has been reported for read speech. For example, Bard & 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, something Terken & Hirshberg (1994) proposed would increase the likelihood of de-accentuation. In another study on spontaneous speech, de Ruiter (2010) found that adult German speakers de-accented given information much more consistently in read than in spontaneously produced narratives. Task effects (or effects of speech style) on the prosodic marking of information structure is particularly relevant when the acquisition of prosodic focus marking is addressed, as the adult model assumed 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).

The speech children are surrounded with, and are expected to learn from, can be assumed to involve a rather variable repertoire of pitch contours and boundary tones, distributed over focal and non-focal information occurring in a wide range of sentence types, as well as at various sentence locations. Our choice of using an interactional task when gathering the data was motivated by our wish to make our methodology as ecologically valid as possible, thereby gathering speech

representative of how children aged between four and eleven speak when interacting with people in a familiar setting, i.e. that of playing a game.

In the picture-matching game, question-answer sequences are embedded in a naturalistic and child-friendly setting, whereby the task of the child is not primarily to produce sentences but to convey important information to his or her conversational partner (the experimenter), thereby helping the experimenter in finding correct picture combinations. The game is constructed so as to turn the participants' attention away from their own speech and more toward the purpose of the game, which we believe makes the speech data more spontaneous and natural than what tends to be the case in most studies of prosody and information structure.

Our picture-matching game is an adapted version of an experimental paradigm developed by Aoju Chen (e.g. Chen, 2011a). In Chen's previous version of the game, subject, verb, object - sentences (hereafter SVO) were elicited with focus either on the subject or the object of the sentence, and subjects and objects were compared between conditions rendering them focal or non-focal (topical, in Chen's terminology). In our revised version of the game the robot mediator has been removed, and new sentences and pictures[17] have been developed. Three laminated picture sets are used, making it easier for the children to see how certain pictures belong together. Thirdly, a larger set of focus conditions are included as compared to Chen's version of the game, allowing for investigating the children's prosodic manipulations in parallel ways sentence-medially and sentence-finally, in addition to exploring whether and how children prosodically differentiate different focus types (i.e. broad focus, narrow focus and contrastive focus). By using the same methodology when gathering the Dutch and the CS data, we made sure that comparable datasets were analysed for each language, addressing the hypotheses presented in Chapter 1. Additionally, this version of the picture-matching game was also used in the 'Get the focus right'- project as a whole (see Section 1.1), allowing for cross-linguistic comparisons beyond the two languages investigated here.

3.2 Procedure

The data gathered with the picture matching game consisted of semi-spontaneous subject-verb-object (SVO) or subject-verb-object-adverbial (SVOA) sentence productions. In each recording session the participants completed two tasks together

[17] All the pictures were hand drawn by Paula Cox. The pictures have been made digitally available for other researchers who want to use the game in their research. The picture database can be found at www.prosodicdevelopment.com.

with the experimenter; a picture-naming task and the picture-matching game. A complete session including both tasks typically resulted in a 20-40 minute long recording. The picture-naming task only took a few minutes; the picture-matching game took around 25 minutes on average, including instruction and practice trials. The game took longer with the four- to five-year-olds than with the older children and the adults, because the youngest children chatted more between each trial, and were more easily distracted from the task.

The participants were recorded individually in a quiet room; sessions with children took place in a designated test room at their school or kindergarten; sessions with adults took place either at The Royal Institute of Technology (KTH) in Stockholm, Sweden, or at the linguistics lab at Utrecht University, the Netherlands. All audio recordings were made using a portable ZOOM H1 handy recorder, with a 44.1 kHz sampling rate and 16-bit accuracy.

The participants first completed the picture-naming task and then played the picture-matching game. Because the experiment was designed to study prosodic focus marking in children, we included extra introductions when playing the game with adults. In these, the adults were told that the experiment they would participate in was constructed for children, and that in order to ensure consistency, the experimenter would play the game in the same way that he or she did with the children.

As the experiment was conducted by 5 different experimenters, detailed instructions were created for both the picture naming task and the picture naming game, including a script on how to explain the tasks, how to respond to unexpected situations and how to control the context for each trial of the game. We also made conventions for the intonation pattern to be used by the experimenter, making sure that each trial and each session were conducted in the same manner.

3.3 The picture-naming task

The purpose of the picture-naming task was to familiarize the participants with the nouns appearing in the picture-matching game, making sure they would use the intended target words when playing the game. In this task the participants were instructed to name figures and objects illustrated in pictures. The spoken context was scripted for each naming trial as *this is a...*, after which the participants could provide a response. In the case of incorrect naming (e.g. calling the 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?*). The target verbs were not part of the picture-naming task, but were presented, illustrated and explained, as part of the introduction to the game (e.g. *look, this is 'painting', and when someone paints there is always a brush and a bucket of blue paint)* (see Appendix for the complete sentence lists).

3.4 The picture-matching game

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 or his pictures. Questions were constructed so as to make one constituent in the answer narrowly focal and the others non-focal, representing given information (see Section 2.2). In addition, 'what happens' –questions were used to elicit broad focus, and a 'guess' from the experimenter was used to elicit contrastive (narrow) focus. Pre-scripted short texts led up to each question, introducing the non-focal constituents, following the structure in Table 1 below. As the names of all figures, objects, and actions appearing in the game were mentioned in the picture-naming task or in the introduction to the game, they can all be considered accessible in terms of information status or referential newness-givenness (e.g. Lambrecht, 1994, see also Section 2.2), but the questions served to render certain constituents focal and others non-focal. In addition, the context preceding the question served to make sure that the non-focal information was contextually given (see Chapter 2.2).

The materials used consisted of three separate sets of pictures, two for the experimenter, and one for the participant (see Figure 1 for the setup and Figure 2 for the picture set). The experimenter's first set (set 1) was piled face down in front of him or her. These pictures always lacked one attribute, e.g. the subject, the verb or the object. 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 displaying complete actions, and these were piled face down in front of him/her. Sets 1 and 3 were carefully put into the same order before each session, so that corresponding pictures always appeared together in the same trial.

Figure 1: Illustrating the experimental setup

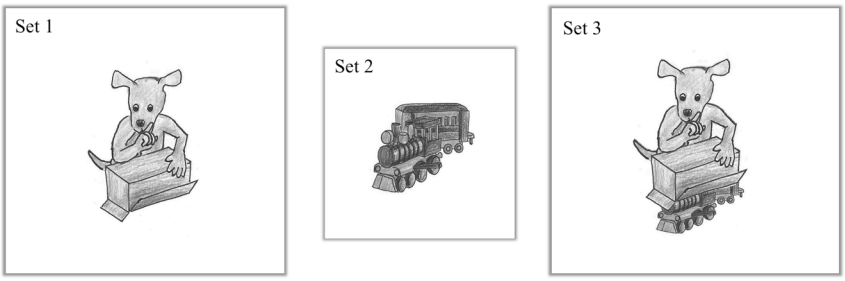

Figure 2: Example of picture set for a trial eliciting narrow focus on the final constituent. The target sentence is 'the dog hides THE TRAIN'

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, uttering the context sentences as illustrated in Table 1 below. After the target question was asked, the participant could look at his/her complete picture (set 3) in order to answer the question. Once the answer was provided, the experimenter, sometimes helped by the child, could look for the 'missing piece' of his/her picture in the box (set 2), unite the two pictures, and move on to the next trial. In the instructions 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 picture to the experimenter.

The experimenter was instructed to use a consistent intonation pattern on the scripted context and target question pertaining to each trial. This pattern consisted in maximum prominence contour ('H*LH' or 'L*H' for Swedish, 'H*L' for Dutch) on "look" as well as on the nouns and verbs, when these were introduced for the first time. In the questions, the experimenter similarly used the same maximum

prominence contours on the wh-word, and avoided accents (in the case of Dutch) and produced lexical accents only (in the case of Swedish) on the following words. In cases where the children produced elided answers, or where noise or other disfluencies was assumed to make the response unusable, the experimenter was instructed to ask the question again.

The game included one set of test trials and one set of practice trials, where the total number of test trials was 30 for both languages. Whereas all 30 trials elicited SVO sentences in the Swedish version of the game, the Dutch version was split into 15 trials eliciting SVOA and 15 trials eliciting SVO sentences. In Chapter 6, on the use of accentuation for focus in Dutch children and adults, only the SVO-sentences were included, whereas both sentence-types were included in our study on pausing for focus, presented in Chapter 7. For both languages, practice and test trials were spread over five sentence conditions, namely narrow focus on the initial constituent (narrow initial), narrow focus on the medial constituent (narrow medial), narrow focus on the final constituent (narrow final), contrastive focus on the medial constituent (contrastive medial) and broad focus on the whole sentence (broad focus). The conditions were implemented following the structure in Table 1 (see Chapter 7 for more on how the conditions were implemented in the SVOA trials).

Sentence condition	**Example context/ question**
Narrow initial	Look, the flower! It looks like someone is hiding the flower. Who is hiding the flower?
Narrow medial	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?
Narrow final	Look, the frog. It looks like the frog is hiding something. What is the frog hiding?
Contrastive medial	Look, the flower! And there is also a frog. It looks like the frog is doing something with the flower. I'll guess: The frog is PAINTING the flower. (What do you say?)
Broad focus	Look! A ghost picture, I can't see anything at all. What is happening in the picture?

Table 1: Examples of context and questions implementing the five sentence conditions

In the Swedish version of the experiment, six subject nouns, six transitive verbs and six object nouns were carefully distributed over the five sentence conditions,

rendering a total of 30 experimental trials. Among each set of initial, medial and final constituents, half were accent 1 words and the other half were accent 2 words. In the Dutch version of the game, five subject nouns, six transitive verbs and six object nouns were distributed over the same five sentence conditions. For both languages, the target words were carefully selected to make sure that four-year-old children would know them, that they would be easy to illustrate, and that they would be sufficiently flexible to combine with the other words without generating semantically odd combinations.

When creating and ordering the stimuli we ensured that each combination of initial, medial and final constituent only occurred once in the whole set. In addition, two consecutive trials never realized the same condition, and they always differed by minimally two constituents. Following these constraints, the experimental trials were arranged into two different stimulus orders, to which the participants were randomly assigned.

The procedure described above was used for collecting both the Dutch and the Swedish datasets presented in Chapters 4 -7. In addition, smaller adaptations of the methodology were made in each of the four sub-studies, particularly in terms of how the data was treated and analysed, and we will return to these issues within the relevant chapters.

Part II:
The studies

4 Adding tones to tones in Central Swedish: The acquisition of prominence H

4.1 Introduction

In this chapter we present an investigation into the development toward adult prosodic focus marking in Central Swedish-speaking children, centring on the question of how pitch-based cues to focus are acquired.[18] Even if both duration and intensity manipulations have been attested to correlate with focus in adult CS (Heldner, 2001; Heldner & Strangert, 1997), pitch is typically considered the primary cue to focus in this variety (see Riad, 2014, for a review). To the best of our knowledge, this study is the first to systematically investigate the acquisition of sentence-level prosody in Swedish.

Previous work on prosodic focus marking in children has been centred on English, German and Dutch. In these languages, focus is typically marked by accenting focal and de-accenting the post-focal constituents (e.g. Ladd, 2008; Gussenhoven, 2004; see also Chapter 2.3). Despite substantial methodological variability, previous work suggests that English, German and Dutch-speaking children accent contrastive information already from the age of three to four, but that their use of accent placement for marking non-contrastive focus only matches those of adults at around four to five. Further, in terms of more fine-grained manipulations like accent type choices, phonetic adjustments where accentuation is not sufficient, or the prosodic differentiation of different focus types, children do not reach adult proficiency before the age of seven to eight, if not later (e.g. Hornby & Hass, 1970; MacWhinney & Bates, 1978; Müller, Höhle, Schmitz & Weissenborn, 2006; Wonnacott & Watson, 2008; Wells, Peppé & Goulandris, 2004; Sauermann et al., 2011; Chen, 2011a). Because English, German and Dutch represent similar prosodic systems, studies on how children acquire prosodic focus marking need to look beyond these languages, in order to determine whether the developmental patterns described this far also generalize to prosodic systems that differ from the West-

[18] Parts of the analyses presented in Chapter 4 were published in Romøren & Chen (2015b).

Germanic ones. This study presents data on prosodic focus marking in Central Swedish, a North Germanic tonal variety that differs in systematic ways from the prosodic systems of English, German and Dutch, allowing for exploring the effect of such differences on the acquisition path.

CS is a tone language (e.g. Yip, 2002; Hyman, 2009), where pitch is contrastive at the word level, and where the contrastive pitch movement is aligned with the main stress of a word. On the surface, the lexical accent contrast is best captured as a timing difference (e.g. Bruce, 1977; see also Riad, 2014), where a falling contour is aligned *early* relative to the primary stressed syllable in accent 1 words, and *late* relative to the primary stressed syllable in accent 2 words, as illustrated in the two upper panels in Figure 1. Marking constituents as focal in CS involves assigning a separate prominence-marking high tone (hereafter 'prominence H') to the constituent in focus. When prominence H is added to the lexical accents, two different surface contours are generated as maximum-prominence variants of each lexical accent contour (e.g. the two lower panels in Figure 1).

Accent 1	Accent 2
ˈσ σ | HL*	ˈσ σ | H*L

Accent 1	Accent 2
ˈσ σ | (H)L*<u>H</u>	ˈσ σ | H*L <u>H</u>

Figure 3: Tonal characteristics of the lexical accent contrast of CS without (above) and including (below) prominence H (underlined), illustrating the alignment difference between contours of either accent category. The contrast is illustrated as it would appear on trochaic words, but similar patterns will also appear on longer words. Accent 2 requires a trochee, thus monosyllabic words or iambic words always carry accent 1.

In Chapter 1 we discussed a number of ways in which the prosodic system of CS may affect the way prosodic focus marking is acquired in this language, as compared to the development of prosodic focus marking in Dutch. Firstly, acquiring a lexical accent system may make CS-speaking children more sensitive to pitch than their Dutch-speaking peers, giving them a head start in discovering that pitch is also

used for marking focus (Hypothesis 1). Alternatively, we also proposed that the double mapping of pitch to both word and sentence-level contrasts would slow down the CS-speaking children relative to their Dutch-speaking peers, because such a double mapping may be considered more complex than the simpler one-to-one mapping between sentence-level contrasts and pitch found in Dutch (Hypothesis 2). In addition to effects based on the lexical accent contrast, we also hypothesized that generally larger contour variability, resulting from the relatively large repertoire of pitch accents used in Dutch, would make the Dutch system harder to acquire than the CS one, where only four contours are assumed (Hypothesis 3). Finally, based on what seems to be a more probabilistic mapping between accentuation and focus in Dutch, as compared to a rather consistent mapping between focus and prominence H in CS, we hypothesized an earlier mastery of prominence H for focus in CS-speaking children than the mastery of accentuation for focus in Dutch-speaking children (Hypothesis 4).

This Chapter is structured as follows: in Section 4.2.1 we present the prosodic system of CS in more detail, starting with a few words on the lexical accent contrast before moving on to the prominence marking H tone. Following this we review past work on prosodic acquisition in child Swedish (4.2.2). In Section 4.3 we present our research questions, and in Section 4.4 we briefly describe the method, with particular attention to how our methodology, described in Chapter 3, was adapted to the present study. Our analysis is presented in three separate sections, corresponding to our three research questions. In Section 4.5.3 we present an analysis on the use of prominence H for marking narrow focus in CS-speaking children and adults. In Section 4.5.4 we present an acoustic analysis of the use of pitch range for focus. In Section 4.5.5 we present an analysis of the use of prominence H across broad focus, narrow focus and contrastive focus. In Section 4.6 we summarize the results of our three analyses, and in Section 4.7 we discuss these results, comparing our findings to previous work on the acquisition of prosodic focus marking in West-Germanic languages.

4.2 Background

4.2.1 Prosodic focus marking in Central Swedish

The interaction between the lexical accents and prominence H has led many researchers to claim that CS has two separate levels of intonational prominence, where the lexical accents express the lower level and the combination of each lexical accent with prominence H expresses the higher. Below we will start by presenting

the lexical accent contrast (upper panels in Figures 1 and 2), before describing the contours resulting from adding prominence H (lower panels in Figure 1 and 2).

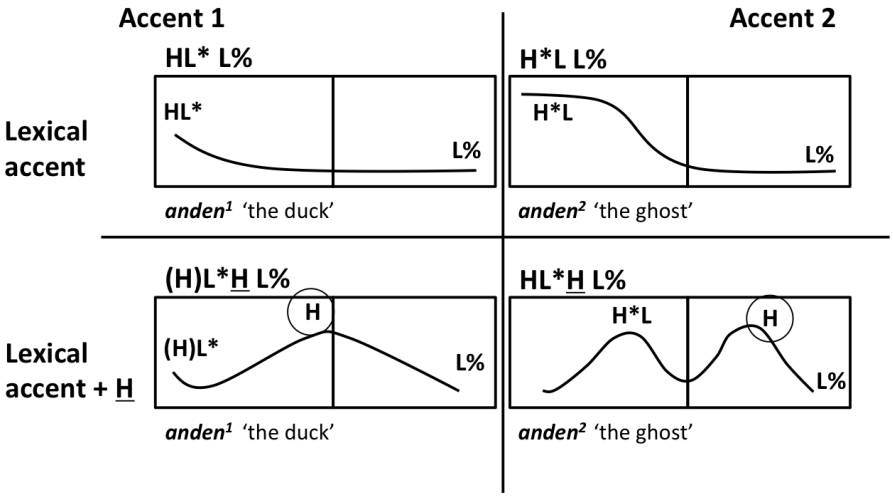

Figure 4: Schematized contours in CS assumed for lower (lexical accent only) and higher (lexical accent + prominence H) level prominence. The contours are illustrated as occurring on trochaic target words (e.g. *anden*[1] ('the duck') / *anden*[2] ('the ghost') in sentence-final position. The two higher panels illustrate post-focal productions lacking prominence H, whereas the latter illustrate focal productions including prominence H (circled).

The lexical accent contrast is present in most Swedish and Norwegian dialects, splitting the lexicon of these dialects into two categories: accent 1 words and accent 2 words, differentiated by the pitch pattern they carry on and around the stressed syllable. In CS, the lexical accent contour is described as a fall differing in timing, where accent 1 falls start *before* the main stressed syllable ('HL*') and accent 2 falls start *on* the main stressed syllable ('H*L') (Riad, 2014, and references therein). The leading H tone of accent 1 is often truncated, particularly when followed by prominence H. This means that the leading H is typically not observed in the citation form of an initially stressed accent 1 word (i.e. the lower panels in Figure 2), which is why this tone is presented in parentheses in Figures 1 and 2. Conversely, the first H of accent 2 is consistently present across both prominence levels (Engstrand, 1995,1997).

In casual speech, the pitch pattern associated with each lexical accent is mostly realized on content words, whereas function words frequently lose their accent when included into larger prosodic units (Frid, 2003; see also Riad, 2014). This means that

most content words in CS will carry a pitch fall associated with the main stress, regardless of their focal status.

Two competing accounts exist regarding how to best account for the phonological representation of the Swedish (and Norwegian) lexical accent contrast. In Gösta Bruce's original work, the contrast was treated as equipollent, and both accents were described as lexically specified. This view is maintained in what is known as the Lund model of Swedish intonation (e.g. Bruce, 1977, 1998, 2005, 2007, see also Gussenhoven & Bruce, 1999; Ambrazaitis, 2009). An alternative view of the accents considers the contrast to be privative, assuming that only one of the lexical accents contains a truly lexical tone, while the other pattern can be accounted for with phonological redundancy rules. The latter analysis is assumed by Tomas Riad, who adopts the view that only accent 2 words are specified with a H tone in the lexicon, whereas the accent 1 pattern is assigned post-lexically (see Riad 1998, 2003, 2006, 2012; see also Engstrand, 1995, 1997).[19] Regardless of what is assumed in terms of lexical specification and markedness, these two accounts do not differ in the description of the resulting surface contours; both assume an early fall to be aligned with the main stressed syllable in accent 1 and a later fall to be aligned with the main stressed syllable in accent 2, when words are produced without prominence H. Since the topic of the current study is how prominence H is acquired, we will not go further into the discussion of markedness and lexical specification here, and we will remain agnostic about whether or not a leading H should be assumed as part of the representation of accent 1 words under maximal prominence.

The lexical accents constitute the lower level of intonational prominence in CS, and the accents are not assumed to differ in terms of the prominence they involve. The highest level of prosodic prominence is achieved by aligning prominence H after the lexical accent falls, generating what Myrberg & Riad (2015) refer to as 'big accents', but what has also been referred to as a 'sentence accent' or a 'focus accent' in the literature (Bruce 1977, 1998, 2007; Heldner, 2001; Hansson 2003; Myrberg 2009, 2013; Ambrazaitis 2009; Riad 1998, 2006, Engstrand, 1995; 1997). While focal information is marked using prominence H, constituents representing non-focal information typically lack this tone, and carry lexical accents only (but see Horne, 1991; Myrberg 2009, on prominence H used on non-focal constituents). Post-focal lexical accents are described as downstepped in the literature (e.g. Bruce,

---

[19] An alternative privative description of the Norwegian and Swedish lexical accents is proposed by Allison Wetterlin and colleagues (Lahiri, Wetterlin & Jönsson-Steiner, 2005; Wetterlin 2007, 2010, see also Kristoffersen, 2006, 2007). These authors propose that it is accent 1 rather than accent 2 that involves a lexically specified tone.

1977; 1998; 2007), but multiple options seem to exist in terms of how the post-focal area is phrased, with plateau formation between prominence H and a following lexical accent being only one of several patterns attested (e.g. Myrberg, 2009). [20]

The Lund model (e.g. Bruce, 1977, 2005; Gussenhoven & Bruce, 1999; Ambrazaitis, 2009) and the model proposed by Riad (2006; 2014) both assume that the timing difference observed between the lexical falls causes the timing of prominence H to differ between the two accent categories, typically resulting in a one-peak contour on accent 1 words carrying prominence H ('L*H') and a two-peak contour ('H*LH') on accent 2 words (see Figure 2) carrying prominence H. The main difference between the two approaches relates to whether or not a leading H tone is assumed in the representation of accent 1 words when combined with prominence H. Whereas the Lund model assumes this tone to be part of the representation, even if it depends on the presence of pre-stress material to surface, Riad does not assume a leading H for accent 1 words when combined with prominence H (e.g. Bruce, 1998; Riad, 2014; Myrberg, 2009).

For the purpose of keeping prominence H separate from the contours generated lexically (or as a combination of lexical tones and post-lexical rules, according to the privative view), we will refer to the tones pertaining to the falls of either accent category (i.e. 'HL*' for accent 1 and 'H*L' for accent 2) as 'lexical', despite the possibility that only some of these tones are truly lexically specified. Further, we will use the notation 'L*H' and 'H*LH' to refer to maximum prominence contours of accent 1 and accent 2, respectively. This notation is chosen because all our target words were trochaic, and when prominence H is added to trochaic accent 1 words there is general agreement that the leading H of accent 1 does not have sufficient space to surface within that word.

4.2.2 The acquisition of prosody in Swedish

Previous work on the acquisition of prosody in Swedish is limited, and has primarily been concerned with the acquisition of the lexical accent contrast. Typically, isolated word productions of either lexical accent category have been compared, answering the question of whether children differentiate these in the same way as adults do. However, because prominence H is structurally required on isolated word productions, what has actually been compared is maximum prominence contours

[20] The observation that such plateaus end earlier in accent 1 than in accent 2 is taken as evidence that both accent 1 and accent 2 words involve falls in their lower-prominence realization, but that accent 1 falls end earlier, and close the plateaus earlier, than accent 2 words (see Riad, 2014).

across the two lexical accent categories (i.e. the lower panels of Figure 2), meaning that the results of these studies have implications not only for the production of the lexical accent contrast but also for the way children produce the combination of each lexical accent with prominence H. One of these studies comes from Engstrand, Williams & Strömqvist (1991), who analysed spontaneous trochaic isolated word productions[21] from 17-month olds. Comparing F0 changes in the main stressed syllable showed no differences between accent 1 and accent 2 words in the children's productions, 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 (compare Figure 2). As the adults phonetically distinguished the contours on both syllables, the authors concluded that 17-month-olds do not yet differentiate the lexical accents in an adult-like way, even though the contours were distinguished on the post-stress syllable. Relevant for the current study, what these results suggest is that the children produced prominence H on the post-stress syllable in accent 2, and close to the stressed syllable in accent 1, a pattern in line with the adult model. Additional phonetic evidence supporting an early mastery of the complete maximum prominence contours comes from Ota (2006), who re-analysed the data from Engstrand et al. (1991). By only including words containing visible F0 for at least 150 ms, Ota showed accent 1 words to involve a significantly smaller F0 change on the main stressed syllable than accent 2 words, and that the on-stress falls (if any) on accent 1 words were significantly later than the on-stress falls on accent 2 words. In other words, Ota showed that the children studied by Engstrand et al. (1991) *did* distinguish the two maximum prominence contours both on the stressed and the post-stress syllable (see Ota, 2006), at least when words containing a certain amount of voiced material were considered.

In a third study, Kadin & Engstrand (2005) report on accent productions in CS speaking children between 18 and 24 months. Comparing on-stress falls and post-stress rises on spontaneous trochaic target words from both lexical accent categories, the authors showed 24 -month-olds to consistently produce accent 2 words with a fall on the stress and a rise on the post-stress, whereas accent 1 words typically had a smaller fall on the stress that kept falling toward the end of the word. Many of the 18-month-olds also produced accent 2 words with a two-peak accent, but since there were very few accent 1 words in the dataset from this age group, the presence of an actual contrast between accent 1 and accent 2 could not be verified. As this study included both isolated words and words from longer stretches of speech, sentence-

---

[21] The accent 2 targets were all bisyllabic words, but due to a dominance of accent 2 words in the dataset, other trochaic non-word vocalizations were also included into the accent 1 category.

position and information structure were not systematically controlled for, thus the phonetic analysis may have been run across accents from both prominence levels, even if the phonetic results indicate that the proportion of maximum-prominence contours was fairly high.

The papers discussed above provide some evidence that CS-speaking children around two years prosodically differentiate maximally prominent accent 1 contours from maximally prominent accent 2 contours in an adult-like way, even if further development may also occur beyond this age. Nevertheless, as the reviewed studies primarily target isolated productions, and as they lack a systematic control over information structure, they shed little light on whether children around two years assign prominence H in the same way that adults do in cases where this tone is not required for structural reasons, that is: in cases where prominence H marks a constituent as focal. Even if children may get the alignment of the relevant pitch manipulations right on isolated productions, and might even mark focus[22] in similar ways as adults do on isolated word forms, doing this on syntactically more complex constructions, taking both focus placement and focus size into consideration, represents a different task. In this sense, existing work on prosodic acquisition in CS sheds *some* light on how CS-speaking children assign prominence H in their speech production, but in order to determine whether children use prominence H systematically for marking focus, a different methodological approach than the ones reviewed above is required.

4.3 Research questions

Four age groups were included in this study: four-to five-year-olds, seven- to eight-year-olds and ten- to eleven-year-olds, in addition to a group of adult controls. Our primary goal was to describe how CS-speaking children develop the ability to use prominence H for marking focus in their speech production. As most existing work on prosodic focus marking in adult Swedish comes from laboratory speech (but see Engstrand, 1997; Myrberg, 2009, for exceptions), gathering adult data also served

---

[22] Allowing for the marking of focus on single words is somewhat unintuitive based on Gundel & Fretheim's (2004) definition of focus as an instance of referential newness-givenness. At the same time, in a situation where a wh-question is answered with a one-word response (e.g. *What did you eat? Cake*), both Krifka's (2007) definition of focus as highlighting the presence of alternatives, as well as the assumption of a cooperative answer to a wh-question necessarily involving focal information, would allow for 'cake' to be interpreted as focal in this case.

the independent purpose of systematically investigating the consistency with which adult speakers use prominence H for focus in a more spontaneous speech sample.

Several focus conditions were included in this study, allowing not only for investigating the marking of narrow focus, but also for studying whether and how our participants prosodically distinguished different focus types. Comparing narrowly focal target words to pre-focal and post-focal renditions of the same word allowed for systematic comparisons of both on-focus and outside-focus manipulations in our four groups. Performing comparable analyses both sentence-medially and sentence-finally permitted looking at position effects in the acquisition of prosodic focus. Adding a broad focus condition let us explore focus *size* effects in our four groups, something that has not previously been addressed in work on prosodic acquisition (but see e.g. Hanssen et al., 2008; Baumann et al., 2007; Myrberg, 2013, for studies on focus size effects in adults). Finally, eliciting both contrastive and non-contrastive focus allowed for us to test whether the differential marking of narrow versus contrastive focus suggested by previous work could also be observed when contrastive and non-contrastive focus was elicited within the same experiment (see Chen, 2011a, 2015, for discussions). Our research questions are presented below.

Prominence H by narrow focus (research question 1)

As we saw in Section 4.2.1, marking narrow focus in CS involves modifications both on the *focal* constituent and on the *post-focal* constituent. According to Bruce (1977, 1998, 2007), post-focal lexical falls are retained as downstepped, but they lack prominence H. Avoiding prominence H post-focally can be seen as parallel to the post-focal de-accentuation patterns found in English, German and Dutch. Because Swedish has an additional level of prosodic prominence, the Swedish equivalent to de-accentuation involves avoiding the highest prominence level only, thus the lexical accents (without prominence H) are retained, but downstepped. Apart from Myrberg's (e.g. 2009) observation that the leftmost stressed syllable of an IP often carries prominence H (what she refers to as '*initiality accents'*), there are no systematic studies on whether prominence H is avoided on pre-focal constituents in CS.

The first research question to be answered in this study concerns the marking of narrow focus, and we asked:

1. Do CS-speaking children between four and eleven differ from CS-speaking adults in the way they use prominence H to mark narrowly focal versus non-focal target words?

Our study of narrow focus was conducted for each sentence position separately. Sentence-medially, the use of prominence H was compared between targets under narrow focus on the one hand and targets that were post- or pre-focal on the other. Sentence-finally, narrowly focal target words were compared to post-focal target words. Our choice of comparing narrow focus to conditions where the focus was either pre or post-focal was not only motivated by our assumption that both on-focus and post-focus manipulations are relevant for a constituent to be perceived as focal, but also by previous findings of children developing slower in their ability to de-accent (or prosodically reduce) post-focal material than in their ability to prosodically highlight focus (e.g. Grünloh, Lieven & Tomasello, 2015). For adults, reports on post-focal de-accentuation are more common than pre-focal de-accentuation, and work from Chen (2009) suggests that Dutch-speaking adults accent pre-focal subject nouns to the same extent as focal ones. Including a pre-focus comparison sentence medially thus allowed us to explore whether the pattern reported by Chen sentence-initially also holds sentence-medially, and whether our children make the same manipulations as our adults do.

Based the literature on focus marking in CS, we expected our adult group to use prominence H rather consistently on narrow focus, and to avoid this tone both pre and post-focally. If a difference would be observed between our pre- and post-focal conditions, we expected a lower proportion of prominence H post-focally than pre-focally. Based on previous descriptions of prosodic focus marking in English, German and Dutch speaking children, we expected our oldest group of ten- to eleven-year-olds to perform in line with the adults, our four- to five-year-olds to be more variable in prominence-H-to-narrow focus mapping, and that our seven to-eight-year-olds would be somewhere in between. Alternatively, if our hypotheses on a faster or slower acquisition of prominence H in Swedish than in Dutch would hold, we would expect that the use of prominence H would develop at a faster or slower rate than what has been observed in Dutch. Finally, we expected the primary differences found between the groups to be related to the avoidance of prominence H post-focally, rather than to the use of prominence H for narrow focus.

Pitch range by narrow focus (research question 2)

The output measure for research question 1 was the presence versus absence of prominence H on sentence-final and sentence-medial target words. In addition, we

also included an analysis of pitch range on our target words, adding a quantitative measure of the phonetic consequences of adding or avoiding prominence H. A closer look at the pitch patterns in our four groups is particularly interesting considering the local complexity we assume to be involved in CS focus prominence (Hypothesis 2), entailing the combination of tones with different origins, which may make the phonetic realization of the relevant contours particularly challenging for CS-speaking children (see Section 4.2.1). Our second research question was thus:

2. Do CS-speaking children between four and eleven differ from adults in the effect that adding or avoiding prominence H has on the pitch range of target words?

Our second research question was examined by comparing the pitch range between the maximum prominence contours (i.e. lexical accent + prominence H), as used on focal targets to the lower prominence contours (i.e. lexical accent only) as used on post-focal targets. Focal target words were compared to post-focal ones, allowing for parallel comparisons sentence-medially and sentence-finally. Further, we excluded misassigned accents (e.g. using a maximum prominence contour post-focally or using a lower prominence contour focally) from this analysis. This was done because we wanted a measure of what the children's manipulations were like in the cases where they used the correct manipulation, not an average of correctly and wrongly assigned contours.

Based on the shape of the two lexical accents with and without prominence 2 (Figure 2), we expected that, at least for our adults, adding prominence H to accent 1 contours would consistently cause an increase in pitch range, but that adding prominence H to accent 2 words might not cause the pitch range to increase. This is because accent 2 words carry a fall on the stress regardless of whether or not this fall is followed by prominence H, whereas accent 1 words without prominence H carry shallower falls if any, compared to a sharp rise-fall when including prominence H (Engstrand, 1995, 1997, see also Figure 2). If our assumption that the phonetic implementation of the combinations of lexical accent with prominence H represents a challenge for CS-speaking children holds, we would predict that the children would not show the same accent-mediated differences on pitch range as we expected from our adult group.

Prominence H by broad, narrow and contrastive focus (research question 3)

Previous work on prosodic focus marking in children has typically not investigated children's ability to prosodically distinguish between different *focus types* (see Sauermann et. al., 2011, for an exception, see also Chapter 2.4). In our setup we also included a broad focus condition, where the whole sentence (i.e. subject, the verb, and the direct object) represented focal information, and a contrastive narrow focus condition, where our sentence-medial targets represented contrastively focal information (see Chapter 3).

In most Germanic languages, broad focus is marked by assigning a maximal prominence (i.e. a nuclear pitch accent in languages like English) to the right edge of the last intonation phrase of an utterance (Gussenhoven, 1983, 1999; Selkirk, 1984, 1995).[23] Such a pattern is also described for Swedish, where prominence H is assigned to the right edge of the last intonation phrase in broad focus sentences (Myrberg, 2009). The extent to which prominence H also appears on pre-final constituents under broad focus is however not clear (but see Myrberg, 2009, 2013).

Most work on prosodic focus marking in younger children has been concerned with contrast or contrastive focus, whereas non-contrastive focus has only been looked at in older children. Based on the results of these studies, Chen (2011a, 2015) suggests that prosodically marking contrastive focus or contrast might be acquired earlier than narrow and non-contrastive focus. Comparing the marking of contrastive and non-contrastive focus in our child groups allows for exploring possible differences in the acquisition of these two types of focus within the same study. Our third research question addressed the marking of focus types, and we asked:

3. Do CS-speaking children between four and eleven differ from adults in the way they use prominence H to differentiate between (a) broad and narrow focus and (b) contrastive and non-contrastive focus?

The few studies systematically comparing the prosodic marking of broad versus narrow or contrastive focus in adult speech have all been conducted on constructions

---

[23] Even if an accent in this position is assumed to be sufficient for a broad focus interpretation of the utterance, pre-nuclear accents are common on constituents preceding the final one, and arguments are more likely to be accented than verbs, if they are adjacent to each other (e.g. Gussenhoven, 2011).

where the accent placement was somewhat controlled, so that narrow focus was elicited on a constituent that was also expected to carry the nuclear pitch accent in broad focus. In our setup, narrow focus was elicited in two different positions, and the extent to which the medial constituents would carry prominence H in broadly focal SVO-sentences like ours is not clear. We hypothesized that our adult participants would be more likely to use prominence H on the medial targets when these were narrowly focal than when they were part of a broad-focus utterance. In SVO sentences like ours, accenting the sentence-final object marks the whole sentence as broadly focal, thus we expected the adults to use prominence H across both broad and narrow focus sentence-finally. Because accenting the sentence-final object can be seen as the default pattern in all-new sentences, we expected the children to be reasonably good at assigning prominence H to the sentence-final targets under broad focus. At the same time, children generally seem to have more trouble with reducing the prosodic prominence (i.e. de-accentuation) than with enhancing the prosodic prominence, thus we expected that the four- to five-year-olds would be prone to use prominence H more on the medial targets under broad focus than the adults would, but that the children from our two oldest age groups would be close to adult-like.

In addition to experimentally manipulating the focus conditions, lexical accent pertinence was also included as a variable in all our analyses. This was done by balancing the number of target words pertaining to each lexical accent category (see Section 4.4.2 below), as well as by including lexical accent as a factor in our statistical analyses. For our analysis of the use of prominence H we did not expect any differences between the lexical accents in the extent to which they were used with prominence H for marking focus, but in our analysis of pitch we expected different pitch-range effects of prominence H across the two lexical accent categories.

4.4 Method

4.4.1 Participants

Twenty-six CS-speaking children and ten CS-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 2). In the following we will refer to the children as five-year-olds, eight-year-olds and eleven-year-olds.

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

Table 2: Participant information, Swedish

None of the participants were reported to have had any history of language disorders, hearing problems or other known developmental disorders. 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 detailed information about the children's language background, ensuring that all our participants were native language speakers of Central Swedish. The adult participants were recruited at The Royal Institute of Technology (KTH) in Stockholm, and were all university students.

4.4.2 Procedure

The data collected consisted of semi-spontaneous subject-verb-object (SVO) sentence productions, recorded while the participants were playing the picture-matching game together with an experimenter. The procedure is described in detail in Chapter 3. When developing the Swedish version of the game, pilot sessions were conducted with 10 Swedish-speaking children living in The Netherlands, ensuring that the task was suitable for eliciting natural speech from CS-speaking children between four and eleven.

In the Swedish version of the game, 30 test trials and 5 practice trials were included, spread over our five sentence conditions; narrow focus on the initial constituent ('initial narrow'), narrow focus on the medial constituent ('medial narrow'), narrow focus on the final constituent ('final narrow'), contrastive focus on the medial constituent ('medial contrastive') and broad focus on the whole sentence ('broad focus'), elicited following the structure in Table 1, Chapter 3.4. Furthermore, in the Swedish version of the game, one half of the target words were accent 1 words and the other half were accent 2 words. The words chosen were carefully selected, partially based on the Swedish version of the M-CDI (Berglund & Eriksson, 2000), making sure that they would be familiar to children as young as four years, that they would be easy to illustrate in pictures, and that they could be flexibly combined with each other without creating odd sentences.

4.4.3 Data selection and coding

At the first stage of analysis, 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 (i.e. Table 1, Chapter 3.4) were included in the analysis. Responses were also excluded if they contained deviant word orders, deviant word choices or elided constituents, as well as self-repairs, hesitations, or background noise.

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 3 below we report the inclusion rates for the four age groups, divided into five categories. 'Disfluencies' refers to cases where a response contained hesitations, repairs or filled pauses, and 'information structure' refers to cases where the speech context could not be completely controlled (e.g. where responses did not immediately follow the scripted context or where between-trial conversations had rendered certain constituents salient beyond the aim of the experimental setup).[24] The 'non-target' category involved responses that contained the wrong words, lacked certain constituents or had non-target constituents added to them. Finally, the category 'noise/overlap' was used for instances where a response contained noise, laughter or speech overlaps, making the recording unfit for analysis. The total inclusion rate across all groups was 79%, but the groups differed slightly both in over-all inclusion rate and in the distribution of excluded responses across the five categories. The total number of responses analysed in the current study was 849. Since comparisons were made on both sentence-medial and sentence-final target words, this rendered a total number of 1698 target words analysed.

---

[24] An example would be 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. 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.

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

Table 3: Excluded responses by group and category, Swedish

At the next step of analysis, included responses were orthographically transcribed and segmented into words using Praat (Boersma & Weenink, 2010). When segmenting, we relied on changes in the waveform in addition to the formant transitions shown in the spectrogram (see Turk, Nakai, & Sugahara, 2006).

All medial and final target words were manually coded according to whether or not they carried prominence H, resulting in a 'L*H' contour in the case of accent 1 and a 'H*LH' contour in the case of accent 2. Words lacking prominence H were conventionally annotated as 'HL*' if they were accent 1 words and 'H*L' if they were accent 2 words. Recognizing the maximum prominence contours was mostly unproblematic when both visible contours and auditory information was available (see Strangert & Heldner, 1995). The coding of presence/ absence of prominence H was extracted from Praat to Excel sheets using a script, and the data was checked for extraction errors. Figures 5-8 below illustrate our coding on medial target word productions with and without prominence H, for each lexical accent category.

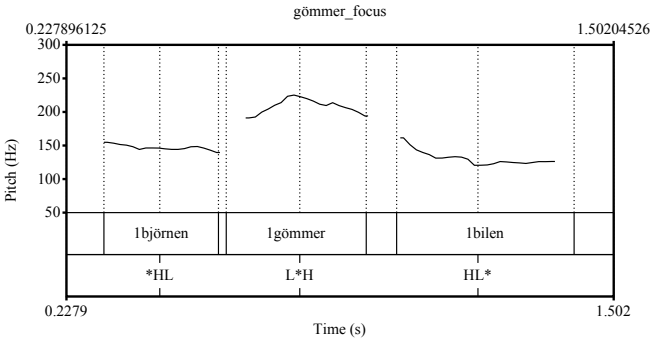

Figure 5: Example of coding of prominence H under narrow focus. The medial target word *gömmer*[1] 'hides' carries prominence H.

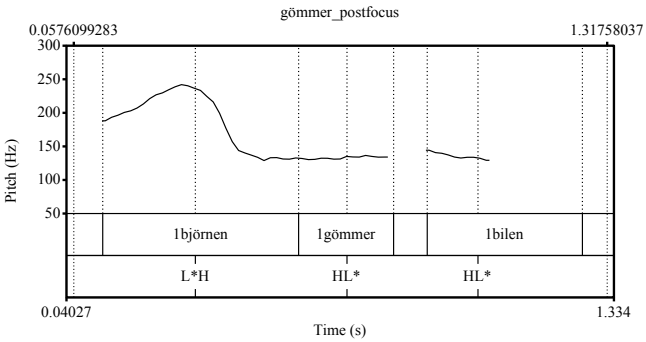

Figure 6: Example of coding of prominence H under post focus. The medial target word *gömmer*[1] 'hides' does not carry prominence H.

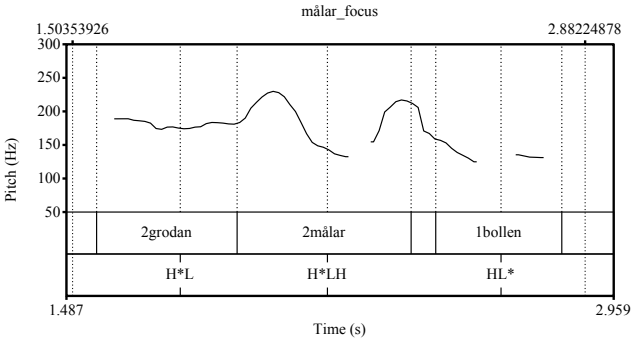

Figure 7: Example of coding of prominence H under narrow focus. The medial target word *målar*[2] 'paints' carries prominence H.

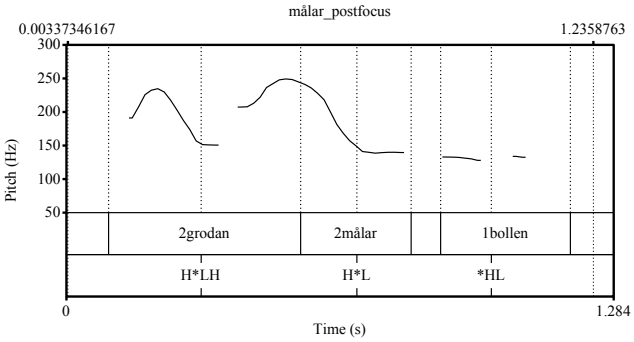

Figure 8: Example of coding of prominence H under post focus. The medial target word *målar*[2] 'paints' is post-focal and does not carry prominence H.

In addition to the manual coding of presence vs. absence of prominence H we also extracted minimum and maximum F0 measures on the target words. These measuring points were the absolute minimum and maximum points within the word, excluding micro-prosodic variation at the vicinity of plosives or at utterance edges. The maximum and minimum points were thus manually placed in areas where the pitch was reasonably stable, and where the pitch level observed in the spectrogram could be auditorily confirmed as reliably representing the F0 minimum or maximum within that word. These points were used for calculating the over-all pitch range on the target words.

4.5 Analysis and results

4.5.1 Introductory remarks

Three separate analyses are presented in Section 4.5, all addressing the question of how CS-speaking children and adults use prominence H for marking focus. The first analysis (Section 4.5.3) addresses our first research question on the effect of narrow focus on prominence H. The second addresses our second research question, on differences between adults and children in the effect prominence H has on the pitch range of focal versus non-focal words (Section 4.5.4). Finally, in Section 4.5.5, we present our analysis of focus type effects on the use of prominence H, comparing the use of prominence H between broad, narrow and contrastive focus. Before presenting the analyses we remind the reader of the conditions that were elicited in the game, explaining how these were compared in the three analyses.

4.5.2 Sentence and focus conditions

The original conditions elicited in the picture-matching game are repeated in Table 4 below.

Condition	Explanation
Initial narrow	Narrow focus on initial constituent (i.e. subject)
Medial narrow	Narrow focus on medial constituent (i.e. verb)
Final narrow	Narrow focus on final constituent (i.e. object)
Medial contrastive	Contrastive focus on medial constituent (i.e. verb)
Broad focus	Broad focus on whole sentence (i.e. subject, verb, object).

Table 4: The original sentence conditions (repeated)

Our output measures were made on sentence-medial and sentence-final target words. The target sentences always had an SVO structure, so medial targets were always verbs and final targets always objects. In the following we will use the terms medial and final targets, stressing the location at which the comparison was made, rather than the grammatical category the targets pertained to.

In order to answer our first research question as to whether our participants added prominence H to narrowly focal targets, conditions rendering a target narrowly focal ('medial narrow' for the medial position, 'final narrow' for the final position) were compared to conditions rendering the same target pre- or post-focal. Separate models were built for each sentence position. Sentence-medially, our pre focus condition was our 'initial narrow' condition, whereas our post focus condition was our 'final narrow' condition. Sentence-finally, we didn't have a pre focus condition, but 'narrow medial' was used as the post-focus condition. In order to answer our second research question on pitch range, narrowly focal targets carrying prominence H were compared to post-focal targets lacking prominence H, in parallel ways for both sentence-positions. Finally, in order to answer the third research question about focus type effects, our medial target words were compared between 'broad focus' and 'narrow medial', as well as between 'narrow medial' and 'contrastive medial'. In sentence-final position, our targets were compared between 'broad focus' and 'narrow final'.

4.5.3 Prominence H by narrow focus

Model comparisons

For the analysis of absence or presence of prominence H (Sections 4.5.3 and 4.5.5) we originally wanted to use generalized mixed-effect models for statistically analysing our data, taking the crossed random effects of item (i.e. the 6 target words on which the information structure was manipulated) and participant (our 36 participants) into account. However, because the distribution of prominence H was very skewed for most of our comparisons, our mixed effect models often failed to converge. Because of this, we decided to use binomial logistic regression models (GLMs) instead, with the side effect that random effects of participant and item could not be added to the models.

All our models were built and compared using R (R Core Team, 2014). Our outcome variable was categorical, consisting in the binary outcome 'presence or absence of prominence H', and our independent factors were 'focus' and 'group', as well as any interaction between them. When running our analyses we also added the factor 'lex', referring to whether our target word carried accent 1 or accent 2, thereby checking for main effects or interactions involving lexical accent.[25] Because this factor hardly ever improved any model, neither as a main effect nor as part of interactions, we have chosen not to include this factor when presenting our analyses on the use of prominence H, returning to the question of lexical accent effects (if any) at the end of Sections 4.5.3 and 4.5.5.

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 and then finally adding the interaction between 'focus' and 'group' in model 3. 'Focus' was always a binary factor, and for each comparison a specific combination of our original focus conditions (Table 5) was included, depending on the sentence location investigated and the research question addressed.

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. This

---

[25] An example of such an effect could be that one of the accent categories was more likely to appear with prominence H than the other, irrespective of focus or within certain focus conditions.

was done because we were primarily interested in knowing whether any of the child groups differed significantly from the adults, not whether they differed from each other. The reason for adding 'focus' before 'group' when building the model was that the primary concern of our study was whether the use of prominence H was affected by focus, and that our secondary question was whether this effect differed between our adult group on the one hand and our child groups on the other.

When building the models, 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 (e.g. Quené & van den Bergh, 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 improved the model fit. This was then taken as evidence that the parameter had a significant effect on the outcome variable. When the best model was established, model summaries were used in order to obtain 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.

Model	Factors added
Model 0	
Model 1	Focus
Model 2	Group
Model 3	Focus * Group

Table 5: GLM model build-up, showing the parameter added at each step. Added parameters (main effects or interactions) were kept for the next step if they significantly improved the previous model.

For some of the model comparisons, complete separation (i.e. prominence H at 100% within some level of 'focus', 'group' or 'lex') made the data unfit for analysis. Separation made models including the separated factor crash, making it impossible to estimate the effects we were interested in. 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.

When presenting our analysis we first report p-values for our model comparisons. After establishing the best model for a certain comparison we present the model summary of this best model, commenting on the effects found in more detail. As the

interaction model always included main effects of ‘focus’ and ‘group’, estimates of these effects were also part of the summaries from interaction models.

Sentence-medial position

As can be observed in Figure 9 below, the patterns from adults, eleven-year-olds and eight-year-olds were very similar in this position, with these three groups using prominence H around 95% of the time on narrowly focal target words, and below 10% of the time on both pre- and post-focal target words. The use of prominence H followed the same general pattern in the five-year-olds, but compared to the other groups they used prominence H *less* in the narrow focus condition and *more* in the pre and post focus conditions.

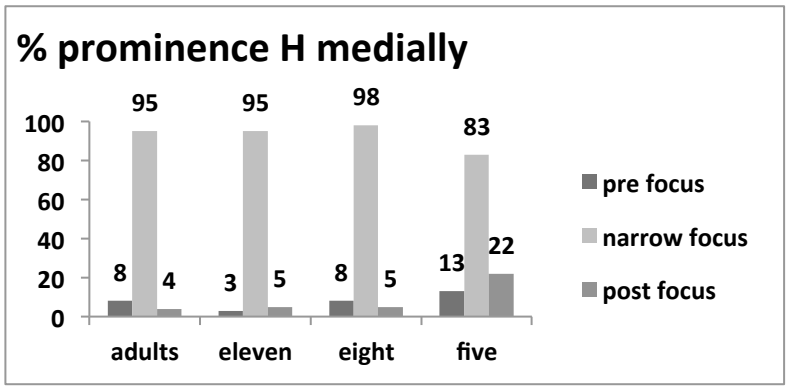

Figure 9: Percentage prominence H on sentence-medial targets under pre focus, narrow focus and post focus, across groups

We start by reporting our analysis of the use of prominence H on *narrowly focal* and *post-focal* targets. Building regression models on the outcome ‘presence of prominence H’, adding the factor ‘focus’ (‘narrow focus’ vs. ‘post focus’), and the factor ‘group’ (‘adults’, ‘eleven’, ‘eight’, ‘five’), as well as their interaction, showed ‘focus’ (p=.000) and the interaction between ‘focus’ and ‘group’ (p=.004) to improve our model, whereas ‘group’ did not (p=.671). Summarizing our best model (Table 6) we see that post-focal status significantly decreased the likelihood of prominence H on the target word (p=.000). The interaction effect showed the effect of focus to differ significantly between our five-year-olds and our adults, with a weaker effect observed in our five-year-olds (p=.002) than in our adults. As was seen in Figure 6, the difference between narrow and post focus was smaller in the five-year-olds (83% vs. 22%) than it was in the adults (95% vs. 4%).

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 eleven	0,092	0,937	0,922
Group eight	0,811	1,174	0,490
Group five	-1,243	0,724	0,086
Post focus * group eleven	0,241	1,387	0,862
Post focus *group eight	-0,530	1,556	0,734
Post focus * group five	3,281	1,077	0,002**

Table 6: Model summary, prominence H, sentence-medial post focus comparison

Comparing our medial targets between *narrow focus* and *pre focus,* similar models were built, with the factors 'focus' ('narrow focus' vs. 'pre focus'), 'group' ('adults', 'eleven', 'eight, 'five'), and their interaction. The model comparison again showed focus to improve our 0 model (p=.000), whereas 'group' did not (p=.670). For this comparison, the interaction between 'focus' and 'group' did not significantly improve our model (p=.178), despite the effect of pre focus appearing to be weaker in our five-year-olds (86% vs. 13%) than in our adults (97% vs. 8%) (Figure 6). Summarizing the best model (Table 7), which included a main effect of focus only, we observe that pre-focal status decreased the likelihood of prominence H (p=.000) on our medial target words.

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***

Table 7: Model summary, prominence H, sentence-medial pre focus comparison

Summing up our sentence-medial comparisons we find that across groups, narrow focus increased the use of prominence H, and post-focal or pre-focal status decreased it. An interaction between focus and group for our narrow focus/ post-focus comparison showed that the five-year-olds showed a weaker effect of narrow focus on prominence H than the adults did. None of the child groups differed significantly from the adults in the way they used prominence H to differentiate narrowly focal from pre-focal target words.

Sentence-final position

For our sentence final position, we had no pre-focal condition to which the narrow focus condition could be compared. Similar to what was observed sentence-medially, a general pattern of more prominence H for narrow than for post focus was also found sentence-finally, but for this position both the five and the eleven-year-olds seemed to show a weaker effect of narrow versus post focus than the adults did (Figure 10).

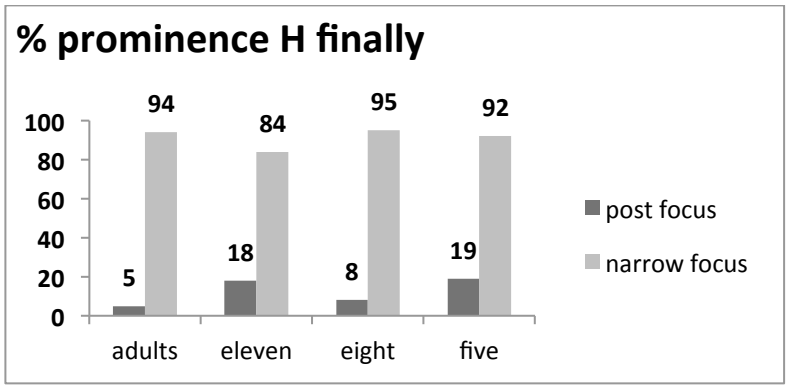

Figure 10: Percentage prominence H on sentence-final targets under post focus and narrow focus, across groups

Building models with the factors 'focus' ('narrow focus' vs. 'post focus'), and 'group' ('adults', 'eleven', 'eight', 'five'), as well as their interaction, 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. As can be seen in the summary in Table 8 below, post focal status significantly decreased the likelihood of prominence H on our final target words, similar to what was found sentence-medially.

Factor	b-value	std. error	p-value
Intercept (narrow focus)	2,535	0,288	0,000***
Post focus	-4,954	0,408	0,000***

Table 8: Model summary, prominence H, sentence-final post focus comparison

Summing up the results for our sentence-final position, we observe a general effect of narrow focus increasing the likelihood of prominence H, as compared to post

focus. Unlike what was found sentence-medially, in this position the effect did not differ significantly between the adults and any of the child groups.

Lexical accent effects

In order to check for lexical accent effects on the way prominence H was used for marking narrow focus in our four groups, comparable models to the ones described above were built, adding 'lexical accent' as a third factor, following the setup presented in Table 9.

Model	Factor added
Model 0	
Model 1	focus
Model 2	group
Model 3	lex
Model 4	focus*group
Model 5	lex*focus
Model 6	group*lex
Model 7	focus*group*lex

Table 9: GLM Model build-up, including main effects and interactions with lexical accent ('lex')

Sentence-medially, including 'lex' in our model build-up did not improve any model, neither in terms of a main effect nor as part of interactions. For the sentence-final position, our comparison between narrow and post focus showed our best model to include an interaction between 'focus' and 'lex', in addition to a main effect of focus (see Table 10 for the model summary).

Factor	b-value	std. error	p-value
Intercept (narrow focus, accent 1)	2,773	0,515	0,000***
Post focus	-5,447	0,666	0,000**
Accent 2	-0,433	0,635	0,495
Post focus * accent 2	1,736	0,809	0,032*

Table 10: Model summary, prominence H, sentence-final post focus comparison including 'lex'

As can be observed in Table 10, the interaction consisted in a weaker effect of post focus decreasing the likelihood of prominence H in accent 2 than in accent 1. In other words, there was a tendency across our groups to use prominence H more on accent 2 words than on accent 1 words when these words were post focal (see Figure 11 below).

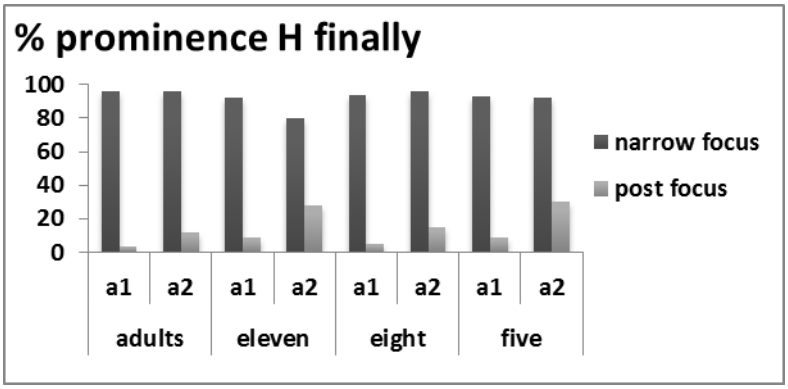

Figure 11: Percentage prominence H on sentence-final targets under narrow focus and post focus, by group and lexical accent

4.5.4 Pitch range by narrow focus

The analysis presented in this section addresses our second research question about whether prominence H, when used for marking focus, had the same effect on pitch range across our four groups. If the effect of adding or avoiding prominence H differs between our adults on the one hand and any of our child groups on the other, this could indicate that the children, even if they are recognized as producing prominence H by the coder, are not yet able to phonetically realize this tone in adult like ways. Taking a closer look at pitch range effects of adding or avoiding prominence H is thus informative to the question of whether the local complexity involved in the CS focus-marking contour (see Section 1.4) makes the phonetic realization of the four contours of CS particularly challenging for children.

Our second research question was addressed by comparing the pitch range between maximum prominence contours (i.e. lexical accent + prominence H) produced on focal targets and lower prominence contours (i.e. lexical accent only) produced on post-focal targets. In other words: misassigned accents (e.g. using a maximum prominence contour post-focally or using a lower prominence contour focally) were excluded from our analysis. Misassigned accents were generally rare in our adults

and older children, but more common in the five-year-olds, as shown by the weaker effect of focus on prominence H in this group.[26]

Model comparisons

For our statistical analysis of pitch range, we built linear mixed effect models (LMMs) in R (R Core Team, 2014). For all the models the outcome variable was numeric, involving measures of F0 range in Hz (hereafter pitch range) on medial and final target words.

The model comparison was done in a similar way as with our GLMs (Section 4.5.3), but this time 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 by first adding the factor 'focus' ('narrow focus' vs. 'post focus' on target) in model 1, then adding the factor 'group' ('adults', 'eleven', 'eight', 'five') in model 2, and then adding the factor 'lex' ('accent 1' vs. 'accent 2' on target) in model 3. After all the main effects were added, we added the three possible two-way interactions one by one, and then finally the three-way-interaction between 'focus', 'group' and 'lex', following the setup presented in Table 9, Section 4.5.1. Again, parameters (main effects or interactions) that improved a previous model were kept for the following step, and parameters that did not were excluded. 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'.[27] Using the full model when re-levelling, rather than

---

[26] Perhaps equally interesting as the question of what the children do with pitch on their correctly assigned contours is the question of what they do when they misassign prominence H, for example in terms of whether maximum-prominence contours produced post focally differ from maximum-prominence contours produced focally. Unfortunately, a systematic analysis of misassigned accents was not possible in the current dataset, as the cases of misassigned prominence H were too few for any proper statistical comparisons to be conducted.

[27] To exemplify, the default model used adults as the reference category, thus the intercept value was from the adult data. In this case, the estimated effect of focus as appearing in the model summary would be the effect of focus within adults. Re-levelling by group meant placing another group than the adults as the reference category, so that the intercept value would change from representing adults to representing for example 'group five', thereby making the estimated effect of focus shown as the effect of focus within group five.

subsetting the data, made it possible for the random effects to be estimated properly for the re-levelled models.

Below we first present the results from our sentence-medial comparisons, thereafter presenting the results from our sentence-final comparisons.

Sentence-medial position

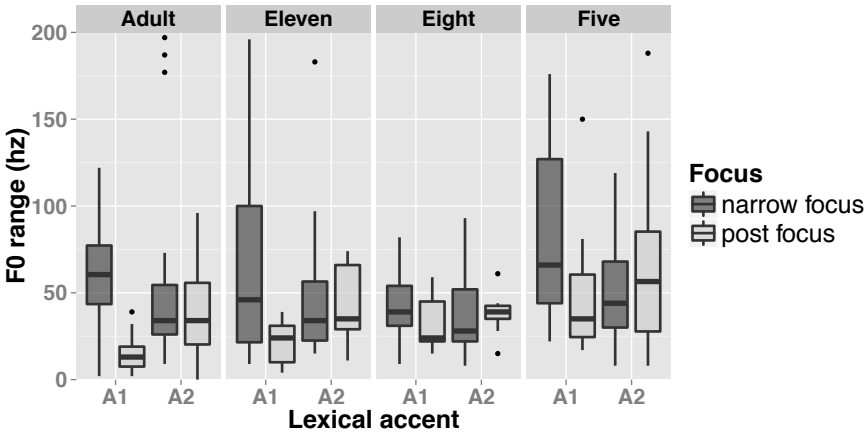

Figure 12: Pitch range on sentence-medial targets under narrow and post focus, by group and lexical accent

In Figure 12 we present a boxplot[28] of the effect of focus on pitch range by group and lexical accent, for the sentence-medial position. Building and comparing mixed effect models for the effects of 'focus', 'group' and 'lex', as well as their interactions, revealed the best model to include an interaction between 'focus' and 'lex' (see Table 11 for model comparisons and Table 12 for a summary of the best model).

[28] The band inside the box represents the median, the bottom and top of the box represent the first and third quartile. The upper whisker extends from the hinge to the highest value that is within 1.5 * inter-quartile range of the hinge. The lower whisker extends from the hinge to the lowest value within 1.5 * inter-quartile range of the hinge. Outliers are plotted as dots.

Model comparisons	p-value
0 vs. 1 (adding focus)	0,000***
1 vs. 2 (adding group)	0,332
1 vs. 3 (adding lex)	0,456
1 vs. 4 (adding focus*group)	0,339
1 vs. 5 (adding focus*lex)	0,000**
5 vs. 6 (adding group*lex)	0,112
5 vs. 7 (adding focus*group*lex)	0,462

Table 11: Model comparisons, pitch range, sentence-medial comparison

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***

Table 12: Model summary, pitch range, sentence-medial comparison

As can be seen in Table 12, narrowly focal targets had a larger pitch range than post-focal targets, and the interaction consisted in this effect being smaller in accent 2 words than in accent 1 words. Exploring the effect of focus within each accent category showed focus to significantly increase the pitch range in accent 1 words (p=.000), whereas no such effect was observed in the accent 2 words (p=.974). The accent 1 targets were spoken with a mean pitch range of 29 Hz when post focal (lacking prominence H) and 69 Hz when focal (carrying prominence H). The accent 2 targets were spoken with a mean pitch range of 49 Hz when post focal (lacking prominence H) and 48 Hz when focal (carrying prominence H). Importantly, no main effects or interactions involving group were found, thus this effect did not differ significantly between our adults and any of our child groups.

Sentence-final position

In Figure 13 we present a boxplot of the effect of focus on pitch range by group and lexical accent, for the sentence-final position. The strange presentation of post-focal accent 2 words in the data from the adults and the eight-year-olds is caused by a very low number of items in this category, which we will get back to below.

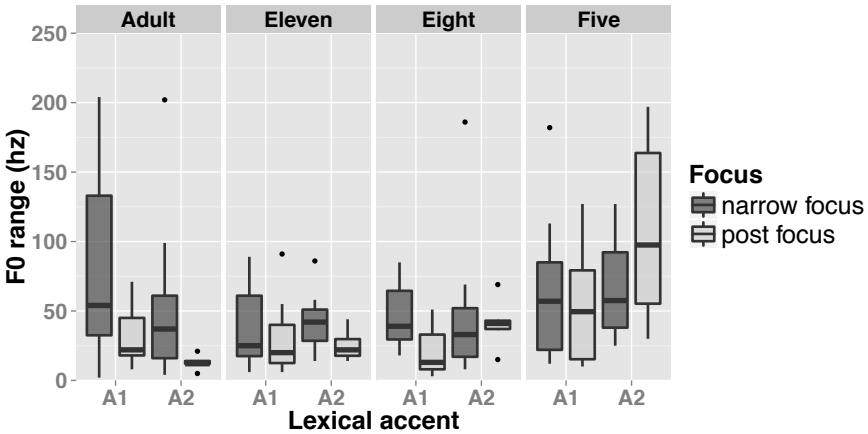

Figure 13: Pitch range on sentence-final targets under narrow and post focus, by group and lexical accent

Building and comparing models for the effect of 'focus', 'group' and 'lex', as well as their interactions, on pitch range sentence-finally, showed the best model to include a main effect of focus only (see Table 13 for the model comparison and Table 14 for a summary of the best model).

Model comparison	p-value
0 vs. 1 (adding focus)	0,000***
1 vs. 2 (adding group)	0,243
1 vs. 3 (adding lex)	0,860
1 vs. 4 (adding focus*group)	0,364
1 vs. 5 (adding focus*lex)	0,564
1 vs. 6 (adding group*lex)	0,091
6 vs. 7 (adding focus*group*lex)	0,158

Table 13: Model comparisons, pitch range, sentence-final comparisons

Factor	b-value	std. error	p-value
Intercept (narrow focus)	57,067	6,481	0,000***
Post focus	-20,012	5,783	0,001**

Table 14: Model summary, pitch range, sentence-final comparisons

As can be seen in Table 14, the main effect of focus consisted in post-focal status generally reducing the pitch range on our target words, as compared to narrow focus. Despite the lack of any interaction between 'focus' and 'lex', exploring the effect of focus within each lexical accent actually revealed a similar pattern sentence-finally as the one we observed sentence-medially; adding prominence H caused a significant pitch range increase in accent 1 (p=.001) but not in accent 2 (p=.176). This pattern was also reflected in our raw mean pitch range values: Whereas our accent 1 targets had a mean pitch range of 36 Hz when post focal (lacking prominence H) and 61 Hz when focal (carrying prominence H), our accent 2 targets had a mean pitch range of 51 Hz when post focal (lacking prominence H) and 53 Hz when focal (carrying prominence H).

The fact that the interaction between 'lex' and 'focus' did not reach significance in our final position is likely to be caused by our dataset being quite small for this position, as many words were excluded from the pitch analysis because the F0 measures could not be reliably extracted (see Section 4.4.3). This was particularly common for our post-focal target words, and often meant that only a small number of items could be included from this condition in each group. The number of items per category was further reduced when the data was broken down by 'lex', as in Figure 13 above. In Table 15 below we report the number of included items per total number of items for each group and condition. As can be seen, the number of items included from the post focus condition was often very low, possibly explaining why the interaction between 'focus' and 'lex' did not reach significance, despite the fact that the effect of focus on pitch range only reached significance in our accent 1 words.

Group	Narrow focus (included/total)	Post focus (included/total)
Adults	43/48	14/54
Eleven	22/28	12/33
Eight	28/39	14/37
Five	29/37	18/34

Table 15: Data loss, sentence-final pitch range analysis

Summing up our pitch range analysis, what we find is that the higher and lower prominence contours (Figure 2) produced by our four groups are remarkably similar in terms of the effect that prominence H had on over-all pitch range. Sentence-medially, adding prominence H to accent 1 words consistently caused the pitch range to increase, whereas no such pitch range increase accompanied prominence H

when added to accent 2 words. Sentence-finally, a similar pattern was observed, but due to the limited amount of reliable pitch data we had for this position, the interaction between 'focus' and 'lex' did not reach significance. Most importantly, no group effects or interactions involving group were found, thus our child groups were similar to our adult group in terms of the pitch range consequences of adding or avoiding prominence H.

4.5.5 Prominence H by broad, narrow and contrastive focus

For the analysis of prominence H across different focus types, the statistical procedure was exactly the same as the one described in Section 4.5.3, except that different focus conditions were included into the 'focus' factor.

Sentence-medial position

The percentage use of prominence H across contrastive focus ('medial contrastive'), narrow focus ('medial narrow') and broad focus, on our medial targets, is presented in Figure 14 below.

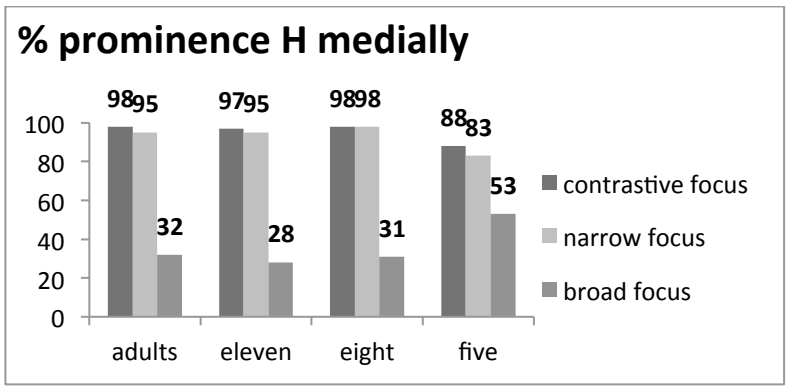

Figure 14: Percentage prominence H on sentence-medial targets under contrastive focus, narrow focus, and broad focus, across groups

As can be observed, contrastive and narrow focus patterned close together in all our groups sentence-medially. Further, the only child group that seemed to differ from the adults in the use of prominence H across broad, contrastive and narrow focus was the five-year-olds, who used prominence H more in broad focus and less in contrastive and narrow focus than the adults.

Starting with our comparison between *narrow focus* and *broad focus*, building models with the factors 'focus' ('narrow focus' vs. 'broad focus'), and 'group' ('adults', 'eleven', 'eight', 'five'), as well as their interaction, showed 'focus' to significantly improve our 0 model (p=.000), whereas 'group' (p=.842) and the interaction between 'focus' and 'group' did not (p=.090), even if the model improvement was trending for the interaction. Summarizing the interaction model showed that the trending interaction was caused by a trending difference between our adults and our five-year-olds in the effect of narrow versus broad focus, but this effect did not reach significance. As can be observed in the summary of the best model (Table 16), the main effect of focus consisted in narrow focus increasing the probability of prominence H as compared to broad focus (p=.000).

Factor	b-value	std. error	p-value
Intercept (broad focus)	-0,565	0,170	0,001**
Narrow focus	3,020	0,327	0,000***

Table 16: Model summary, prominence H, sentence-medial focus size comparison

Comparing *narrow focus* to *contrastive focus* sentence-medially, building models with the factors 'focus' ('narrow focus' vs. 'contrastive focus'), and 'group' ('adults', 'eleven', 'eight', 'five'), as well as their interaction, showed 'group' to improve our model (p=.011), whereas 'focus' (p=.273) and the interaction between 'focus' and 'group' (p=.737) did not. As can be seen in Table 17, the probability of a word being produced with prominence H was significantly lower in group five (p=.017) than in the adults when these two conditions were considered.

Factor	b-value	std. error	p-value
Intercept (adults)	3,229	0,510	0,000***
Group eleven	0,055	0,778	0,944
Group eight	-0,065	0,779	0,934
Group five	-1,412	0,591	0,017*

Table 17: Model summary, prominence H, sentence- medial contrastivity comparison

Sentence-final position

As discussed in Section 4.3, we expected prominence H to be added to the sentence-final target words under both narrow and broad focus. In Figure 15 below we see

that the dominant pattern was indeed to use prominence H across both of these conditions, even if our five-year-olds seemed to use prominence H less under broad focus than the other groups.

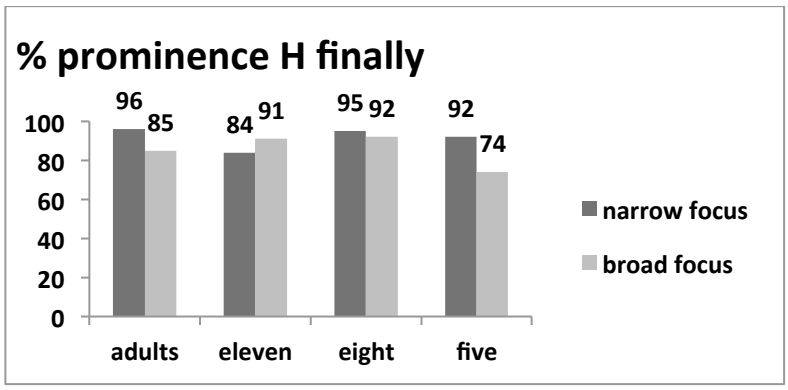

Figure 15: Percentage prominence H on sentence-final targets under narrow focus and broad focus, across groups

Building models with the factors 'focus' ('broad focus' vs. 'narrow focus'), and 'group' ('adults', 'eleven', 'eight', 'five'), as well as their interaction, showed 'focus' to significantly improve our 0 model (p=.042), whereas 'group' (p=.184) and the interaction between 'focus' and 'group' (p=.134) did not. In Table 18 below we see that broad focus significantly decreased the probability of prominence H across the groups.

Factor	b-value	std. error	p-value
Intercept (broad focus)	1.753	0.231	0.000***
Narrow focus	0.752	0.379	0.047*

Table 18: Model summary, prominence H, sentence-final focus size comparison

Summarizing our analysis of focus type effects in both sentence positions we find narrow focus to increase the likelihood of prominence H both medially and finally, as compared to broad focus. This effect was markedly stronger in the sentence-medial position than in the sentence-final position. Sentence-medially, no differences were found between contrastive and narrow focus. Importantly, no significant interactions between focus and group were found, suggesting that none of our child groups differed from the adults in the way they used prominence H across different focus types.

Lexical accent effects

We also checked for lexical accent effects in our focus type analyses, but in these 'lex' never improved any of our models, neither as a main effect nor as an interaction. Due to separation (i.e. 100% prominence H at some sublevel of 'lex') adding interactions with 'lex' to our comparison between contrastive and narrow focus was not possible. As contrastive and narrow focus were both close to ceiling we doubt that any main effects or interactions involving lexical accent would be observed if we had changed a few outcomes, following the procedure described in Section 4.5.3. The patterns observed regarding lexical accent in our focus type analysis, together with those reported at the end of Section 4.5.3, confirm previous reports showing that prominence H is used for marking focus in similar ways across the two lexical accent categories. That said, we did find one exception to this pattern, in that prominence H was more likely to appear post-focally on accent 2 words than on accent 1 words in sentence-final position.

4.6 Summarizing the results

Three separate analyses on the use of prominence H were presented in the previous sections. The first analysis concerned our first research question of whether CS-speaking children between four and eleven differ from CS-speaking adults in the way they use prominence H to mark narrowly focal versus non-focal target words. Our analysis of the use of prominence H for narrow focus revealed remarkably similar patterns across our four groups; the participants predominantly added prominence H to constituents under narrow focus, and they predominantly avoided this tone both pre- and post-focally. At the same time, interaction effects showed that whereas our eight and eleven-year-olds performed in line with adults in both sentence-positions, our five-year-olds differed from the adults sentence-medially by showing a weaker differentiation between narrow focus and post focus than the adults did, even if the effect of narrow focus was also reliably present in their data. The answer to our first research question is thus negative when our eight- and eleven-year-olds are considered, but positive for the five-year-olds, even if the five-year-olds performed in line with the adults sentence-finally.

Our second research question addressed the question of whether CS-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 our pitch range analysis, the answer to our third research question is 'no': No differences were observed between adults and children in the way the pitch range was manipulated across the relevant contour categories in either sentence position. Adding

prominence H to sentence-medial targets caused a pitch-range increase on accent 1 words, whereas adding this tone to accent 2 falls had no such effect in this position. Even if similar patterns were observed when analysing the accent categories separately, the interaction between focus and lexical accent did not reach significance sentence-finally, possibly caused by the dataset for this position being smaller than what was the case for the sentence-medial comparisons.

Our third research question concerned the question of whether CS-speaking children between four and eleven differ from adults in the way they use prominence H to differentiate between (a) broad and narrow focus and (b) contrastive and non-contrastive focus. Starting with our comparison between broad and narrow focus, the conclusion is again 'no', no significant differences were observed between our groups in how they used prominence H to differentiate broad focus from narrow focus either sentence-medially or sentence-finally: Narrowly focal targets were more likely to carry prominence H than broadly focal targets in both positions, but the effect was stronger sentence-medially than sentence-finally. Our comparison between contrastive and narrow focus sentence-medially showed none of our groups to differentiate these conditions, with prominence H at ceiling in our adults, eleven-year-olds and eight-year-olds, and slightly lower proportions for both categories in our five-year-olds.

4.7 Discussion

At a general level, our results regarding the acquisition of prominence H for narrow focus resemble descriptions of the acquisition of contrastive or focal accentuation in English, German and Dutch, where accent placement seems to be largely established by four to five, even if children occasionally misassign accents beyond this age (e.g. MacWhinney & Bates, 1978; Wells, Peppé & Goulandris, 2004; Chen, 2011a). In addition, it seems that the use of prominence H for marking narrow focus is mastered earlier in sentence-final than in sentence-medial position, similar to reports from Chen (2009) comparing sentence-final and sentence-initial positions. This finding is perhaps not so surprising, considering that in both Dutch and Swedish, the IP-final position is also the default position for maximum prominence in broad focus utterances, and it is likely that this alignment of prosodic prominence is what children have the most experience with. An additional reason that prosodic focus marking is mastered earlier utterance-finally than in other positions may be that this position may be particularly salient from a prosodic point of view, as important prosodic functions like turn-taking or interrogativity are typically marked here (e.g. House, 2003, on Swedish). Finally, our position effect could also be interpreted as an effect of grammatical category, considering that our medial targets were always

verbs and our final targets always objects. As suggested by Röhr, Baumann & 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 our children mark focus more consistently on the sentence-final nouns than on the sentence-medial verbs. Expanding work on prosodic focus marking to languages with different word orders might help in disentangling position effects on the one hand and effects of grammatical category on the other (see e.g. Yang & Chen, 2014; Yang et al., 2015, on the acquisition of prosodic focus in Korean).

The finding that CS-speaking five-year-olds are already close to adult-like in their use of prominence H for focus suggests that our investigation primarily captures the later stages of the development of prosodic focus marking in CS. The same seems to be the case for our focus type analyses, where no differences were observed between our adults and our child groups. The fact that we found no difference between contrastive and narrow focus confirms patterns described by Myrberg (2013), who reports that adult speakers of CS do not prosodically differentiate contrastive from narrow focus in their use of prominence H (see Myrberg, 2013; see also Kügler, 2008; Breen et al., 2010, for related findings). The ceiling effects found in our comparison between narrow and contrastive focus also means that we do not have evidence in favour of our hypothesis that contrastive focus may be acquired earlier than narrow non-contrastive focus (c.f. Chen, 2011a, 2015). Again, future work on the acquisition of prosodic focus marking in CS should include younger children, in order to capture the earlier stages of the developmental path toward adult proficiency. Including younger children may also make it easier to avoid ceiling effects, making it more likely to observe possible dissociations between contrastive and non-contrastive focus, but also to explore whether different developmental patterns can be observed for different prosodic manipulations.

Our analysis of lexical accent effects showed prominence H to be used in similar ways across accents 1 and 2, with one exception: Post-focal accent 2 words were more likely to carry prominence H than post-focal accent 1 words, in sentence-final position. Even if this pattern was statistically significant in our dataset, the dominant preference for both accent types was avoidance of prominence H post focally. It is possible that the double-peaked accent 2 pattern, together with it's marked status (e.g. Riad, 2014), makes accent 2 words resist being 'downgraded' to lower prominence more than accent 1 words. More data is needed in order to establish how systematic this pattern is, as the instances found in our data were limited to a few cases, even if the patterning of these cases within certain conditions was systematic enough to reach significance.

The difference we find between accent 1 and accent 2 in our analysis of pitch-range effects of (focal) prominence H, is not very surprising. Without prominence H the accent 1 contour can be flat, slightly falling or slightly rising, whereas the accent 2 contour consistently involves a fall on the stressed syllable (Engstrand, 1995; 1997). Going from a shallow fall/rise to a sharp peak in accent 1 is hence expected to have a greater effect on the pitch range than adding a second peak will have on accent 2 words, for which a certain pitch range is already present because of the lexically determined first peak (see Figure 2). The results of our pitch analysis suggest that CS-speaking children already from the age of five produce the four contours of CS in an adult-like way, indicating that integrating lexical and post-lexical tones into one complex contour does not represent a challenge for children between four and eleven. Our results thus resemble previous descriptions of adult-like contours present already at the age of two. Differently from the results reported on younger children, our findings show that five-year-olds are also able to use these complex contours for marking focus in longer stretches of speech.

Obviously, pitch range is a rather crude measure to use as a diagnostic for differences between the four contours presented in Figure 2; contours produced by our four groups could of course still differ in terms of other pitch parameters, such as timing and slope of pitch peaks and valleys. The reason why we used over-all pitch range as the outcome measure in our analyses was that our pitch data was much less 'neat' than the data typically collected in a lab. Our need to use child-friendly and semantically sensible target sentences meant that plosives or voiceless segments regularly interrupted the pitch contour. In addition, conducting our experiment in schools and kindergartens using a table-mounted microphone meant that the F0 could not always be reliably measured, as the children occasionally moved away from the microphone, and as there was occasional background noise from children playing in the background. Finally, our participants often used non-modal voice quality when producing the target words, which also contributed to lost pitch data. The above situation made it hard to target specific parts of the contour for our analysis, thus pitch range turned out to be the best candidate for a measure that we could reliably extract from our target words without excluding too many data points.

When planning this study we initially hypothesized that the interplay between the lexical accents and prominence H could represent a challenge for the CS-speaking children, due to the underlying complexity involved in the combination of lexical and post-lexical tones (e.g. Riad, 2014). Our finding that CS-speaking children already resemble adults in their pitch manipulations for focus at the age of four to five suggests that this hypothesis does not hold. In Chapter 1 we also hypothesized that early sensitivity to pitch, a comparatively smaller repertoire of pitch-accent like

contours, as well as what seems to be a more consistent mapping between prominence H and focus, could make CS a simpler system than languages like English, Dutch and German, when prosodic focus marking is considered. These hypotheses seem to be supported by our results. Comparing our findings to previous reports on prosodic focus marking in child Dutch, it indeed seems that CS-speaking children are ahead of Dutch-speaking children in the use of pitch cues to focus. In terms of tonal variability, the fact that what seems difficult for Dutch-speaking four- to five-year-olds is not only accent placement, but also the use of certain accent types, may support such an assumption (Chen, 2011a).

As previously mentioned, our findings underline the importance of including children younger than four years when studying the acquisition of prosodic focus marking. There seems to be a tendency in the literature that studies involving the acquisition of relational givenness-newness distinctions, typically studied by means of wh-questions (see Chapter 2.2), include older children, whereas studies of younger children target more relational forms of givenness-newness. This makes it hard to establish a complete developmental path to adult prosodic marking of either dimension of information structure, but also to establish whether these related functions of prosody actually develop in tandem. Some systematic methodological effort may be needed in order to develop better experimental designs for studying prosodic focus marking in younger children, maintaining a high level of experimental control while still making the task suitable for children younger than four years. Capturing earlier developmental steps in the marking of relational givenness-newness would allow for establishing a larger part of the developmental path in a uniform way, avoiding the current situation of some dimensions of information structure only being studied in younger children why others are only studied in older children. In the end it may turn out that using prosody for highlighting referentially new information is similar to prosody used for marking relationally new information, but currently this cannot be systematically determined. By simplifying the picture-matching game so that fewer conditions are elicited on simpler structures (e.g. adjective noun pairings) it may be possible to use the picture-matching game with younger children. Maintaining a 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 establishing a more complete picture of how children develop their ability to prosodically highlight focal information.

This study was only concerned with pitch cues to focus in CS. At the same time, other acoustic parameters have also been found to systematically correlate with focus in Swedish, and the literature suggests that pitch is not the only cue listeners attend to in their focus interpretation (Heldner, 2001; Heldner & Strangert, 2001).

One of these parameters is word duration, which increases systematically on focal as compared to non-focal words. In the next chapter we ask the question of how CS-speaking children acquire the use of word duration for focus. In addition, we also present an analysis of the use of non-modal voice quality for focus, as our pitch analysis suggested that speakers of CS systematically vary their voice quality to mark constituents as post focal.

5 Word duration and voice quality as cues to focus in child and adult Swedish

5.1 Introduction

In the previous chapter we saw that CS-speaking children between four and eleven are quite adult-like in their use of prominence H for focus. In this chapter we present two additional analyses of the relationship between prosody and focus in the data from our CS-speaking participants. The first analysis concerns the use of word duration for focus; the second concerns the use of non-modal voice quality for focus. Whereas increased word duration is strongly associated with focus in adult CS, little is known about how CS-speaking children develop in the use of word duration for focus. There are no previous studies on the relationship between non-modal voice quality and focus in CS, either with adults or with children. In the following we use the term non-modal voice as referring to speech produced with distortions of the normal vocal fold vibrations, such as aperiodic vibration circles or air leakage, causing both creaky and breathy voice qualities (e.g. Ní Chasaide & Gobl, 1997; Gobl & Ní Chasaide, 2010).

Whereas looking at the acquisition of duration manipulations for focus is justified by previous reports of a systematic relationship between word duration and focus adult CS, the choice of looking at non-modal voice quality was motivated by our observation, both during data collection and data annotation, that our CS participants, and predominantly the adults, tended to produce post-focal constituents with creaky or breathy voice quality. This behaviour caused considerable data-loss in our pitch analyses (Chapter 4), as creaky or breathy voice quality made it hard to obtain reliable pitch measures, particularly in sentence-final position. If the use of non-modal voice quality is systematically linked to focus in adult CS, an interesting question is how the use of voice-quality manipulations for focus develops in CS-speaking children. Further, comparing the development of non-modal voice to the use of prominence H and word duration may inform on the question of the role of non-modal voice quality in this language, in the sense that if its development

follows the development of another parameter (e.g. pitch), this may indicate that the parameters are structurally related to each other.

This chapter is structured as follows. In Section 5.2.1 we review existing literature on the relationship between focus and duration in adult speech production. In Section 5.2.2 we review past work on the relationship between duration and focus in children's speech production. Whereas effects of focus on pitch and duration are commonly described, research on the relationship between voice quality and focus in adult speech production is fairly limited. In Section 5.2.3 we review two studies that do address this relationship, also commenting on how the association between voice quality and focus is somehow implicitly acknowledged in the prosody literature, even if the role of voice quality is generally not considered in studies of prosody and information structure. In Section 5.3 we introduce our research questions, and in Section 5.4 we describe how our data was collected and analysed. In Section 5.5 we present our analyses and results, first regarding the effect of focus on word duration (5.5.1) and then on the use of non-modal voice post-focally (5.5.2). Following each analysis we will discuss the results, keeping the discussions of the findings separate for our two manipulations. In Section 5.6 we present a general discussion of both sets of results, with particular attention to the interplay between various prosodic cues to focus in CS, and how these are acquired.

5.2 Background

5.2.1 Duration and focus in adult speech

Although prominence H is traditionally considered the most important cue to focus in CS, it is not the only acoustic correlate of focus, nor the only perceptual cue listeners are attentive to in their focus interpretation (e.g. Heldner, 2001). Below we will first comment on studies of accentual lengthening in English, German and Dutch, before reviewing previous work on the relationship between focus and duration in CS.

Accentual lengthening has been the topic of a number of studies on English, German and Dutch, languages where accentuation is typically assumed to be the major prosodic cue to focus (Gussenhoven, 2004; Ladd, 2008). For these languages, accentuation has been shown to reliably increase the duration of accented words, as compared to non-accented ones (Cambier-Langeveld & Turk, 1999; Cooper et al., 1985; Eady & Cooper, 1986; Eady et al., 1986; Sityaev & House, 2003; Turk & Sawusch, 1997; Turk & White, 1999; Baumann et al., 2006; Kügler, 2008; Féry &

Kügler, 2008; Efting, 1991; Sluijter & van Heuven 1995,1996; Sluijter, 1995). It is unclear whether the domain of lengthening is the accented syllable, the accented syllable plus the syllable following it, or the word, with some studies suggesting that the domain for accentual lengthening differs between languages (Cambier-Langeveld, 2000). Another feature of accentual lengthening that seems to vary across languages is the degree to which the lengthening is affected by sentence-position. Comparing accentual lengthening in Dutch and English, Cambier-Langeveld & Turk (1999) found that whereas accentual lengthening and final lengthening were additive in English, in the sense that both finality and accentuation contributed to the lengthening of a target syllable, this was not the case for Dutch, where accentual lengthening was only observed in non-final positions (Cambier-Langeveld & Turk, 1999).

Similar patterns as those reported for English, German and Dutch have also been reported for CS, where increased word duration seems to be a reliable correlate of focal accentuation[29] (Bruce, 1981; Bannert, 1979; Fant, Kruckenberg & Nord, 1991; Fant & Kruckenberg, 1994; Heldner & Strangert, 2001). In a large study on temporal effects of accentuation in CS, Heldner & Strangert (2001) found that focally accented words (words carrying prominence H) were on average 25% longer than non-focal (either pre or post-focal) ones, and that 75% of the lengthening happened on the stressed syllable.[30] In addition, the authors report that sentence-final words were lengthened relatively more than sentence initial and sentence-medial ones. Interestingly, this effect differs from what has been reported for both Dutch and English (Cambier-Langeveld, 2000; Cooper et. al., 1985), where the lengthening effect was either similar across final and non-final positions (English), or absent sentence-finally (Dutch). The fact that accentual lengthening takes place both finally and non-finally in CS indicates accentual lengthening in this language has more in common with the additive accentual and final lengthening described for English than the pattern of non-additive lengthening effects reported for Dutch (see also Horne, Strangert & Heldner, 1995).

The traditional Swedish intonation model (e.g. Bruce, 1977, see also Chapter 4.2.1) assumes four levels of prosodic prominence: unstressed, stressed, lexically accented and focally accented, where only the latter two are assumed to involve specific pitch

---

[29] Many researchers on CS prosody refer to the maximum prominence contour consisting in lexical falls plus prominence H as 'accented'.

[30] In a second experiment the authors asked participants to read out coordinated phrases 'the man and the woman and the children', asking them to emphasize capitalized words (i.e. the 'focal' ones). In this experiment, the average duration increase for focal words was 15% (Heldner & Strangert, 2001).

movements. Lately the terms ‘focal accent’ or ‘focally accented’ are less used, and in our work we refer to the highest prominence level as the one involving prominence H, as recent work suggests that prominence H also occurs on non-focal material (e.g. Myrberg, 2009). Starting with the lower prominence level, it has been shown that increased duration is the most reliable cue in distinguishing stressed from unstressed syllables, even if intensity and vowel quality is also involved (Fant & Kruckenberg, 1994: Bruce, 1999). The distinction between a lexically accented and a non-lexically accented stress (i.e. secondary stress) is that the lexically accented stress involves an F0-fall (timed earlier for accent 1 than for accent 2), whereas a non-lexically accented stress does not. In addition to the pitch fall, syllables with stress and lexical accent tend to be longer than syllables that are stressed but not lexically accented (Bruce, 1977, 1997, 1999). Moving from the lexical accent level to the maximal prominence level, the major cue is the presence of prominence H, aligned after the lexical accent falls. As we have already seen, the use of prominence H is systematically accompanied by increased duration of the word in focus, as compared to the same word in either pre or post-focal position (Bruce, 1997, 1983, 1981; Bannert, 1979; Fant, Kruckenberg & Nord, 1991; Fant & Kruckenberg, 1994; Heldner & Strangert, 2001). In other words, even if separate cues are involved in distinguishing between the various levels of prosodic prominence assumed for Swedish, duration is relevant at all levels; somewhat similar to the situation for stress and accentuation in English, German and Dutch. The relevant levels investigated in this study are the two higher ones, that is, the maximal prominence level involving lexical accents plus prominence H, and the prominence level below that, involving lexical accents only.

One aspect making the issue of durational effects of focus challenging is determining the factor causing it, as the target words included in most studies are typically both accented and focal, and reports differ as to whether they attribute the lengthening to accentuation or to focus or to both. In some work on accentual lengthening the definition of focus is based rather directly on the presence of an accent, so that accentuation equals focus without much reference to literature on information structure. In addition, durational effects of stress and durational effects of accent have also been confounded in the literature, but work from Sluijter & van Heuven (1995, 1996) neatly shows that both stress and accent contribute to the relative duration of syllables and words in a sentence. Nevertheless, even if these authors distinguish between stress and accent, accentuation and focus are not independently manipulated, thus it is unclear whether the same lengthening effect would be observed if accentuation and focus were to be manipulated independently, to the degree that this is possible. Some evidence for independent effects of focus and accent can be found in studies on the prosodic differentiation between different focus types, which tend to show accented target words to be longer under

contrastive or narrow focus than under broad focus (e.g. Hanssen, Peters & Gussenhoven, 2008; Baumann, Becker, Grice & Mücke (2007); Myrberg, 2013). In terms of focus/ no focus, Chen (2009) reports that focal and accented subject nouns were systematically longer than their pre-focal accented counterparts (see also Arnold, 1998, for related findings). Reports such as these indicate that increased duration is not only a correlate of accentuation, but that the degree of accentual lengthening is additionally affected by focal status.

5.2.2 Duration and focus in child speech

Children's speech is characterized by longer segment durations and more durational variability than adult speech, and such characteristics remain into school age (e.g. Smith, Kenney & Hussain, 1996). Slower speech rate in children has often been attributed to children having a poorer speech-motor control than adults. Reviewing earlier work on duration in children's speech production, Smith (1992) concludes that children between three and seven years of age commonly show longer durations than older children and adults when producing comparable stimuli. The duration of the target stimuli may average as much as 50% longer for children than for adults, and differences between adults and children may remain until as late as ten to twelve years of age. Large temporal variability is reported at group and individual levels, with measures of variability often reported as much as three to four times greater than those found in adults (Smith, 1992; see also Lee, Potamianos & Narayanan (1999).

With this as a background, it is perhaps not so surprising that among the few studies that have targeted the use of duration for focus in children, none actually find systematic relationships between focus and word duration. One such study comes from Wonnacott & Watson (2008), who compared sentence-initial target words that were either (a) given (the giraffe hit the lion, *THE GIRAFFE* hit the elephant) (b) contrasting with a previously presented referent (the giraffe hit the lion, *THE ELEPHANT* hit the lion) or (c) given but shifted in grammatical role (the giraffe hit the lion, *THE LION* hit the elephant). The authors found that children aged three to four were similar to adults in producing shifted and new elements with higher mean pitch than given and non-shifted ones. Interestingly, whereas adults also produced new or shifted elements with longer duration than given ones, such effects were not found in the children, suggesting that pitch-manipulations were mastered before duration-manipulations. Similarly, Sauermann, Höhle, Chen & Järvikivi (2011), also did not find effects of different focus types on word duration in German four- to five-year-olds, whereas they did find focus effects on pitch. In this study, duration effects of focus were also rather limited in the adults, despite such effects being reported by others (e.g. Baumann et al., 2007; Kügler, 2008). It may well be that the

interactional experimental setup used by Sauermann and colleagues rendered more variable speech data than the read speech elicited by Baumann et al. (2007) and Kügler (2008), and that this variability kept systematic duration effects from surfacing[31]. At the same time, the results from Sauermann et al. (2011) point in the same direction as those from Wonnacott & Watson (2008), suggesting that children at four to five do not yet use duration-manipulations for focus in the same way as adults do. Finally, work on prosodic focus-marking in child Dutch also supports this pattern: Chen (2009, 2011a) found Dutch-speaking adults to use phonetic adjustments of pitch and duration to distinguish focal from non-focal accents sentence-initially, but neither four- to five-year-olds nor seven- to eight-year-olds used duration in this way.

Previous work on the acquisition of prosody in Swedish is limited, and has primarily been concerned with the acquisition of the lexical accent contrast, assumed to be mastered around the age of two (e.g. Engstrand, Williams & Strömqvist, 1991: Ota, 2006, see also Chapter 4.2.2). To the best of our knowledge, there is no previous work addressing the use of word duration for focus in CS-speaking children.

5.2.3 Non-modal voice in adult speech

Even if voice-quality distinctions are part of the segmental phonology in a number of languages (Gordon & Ladefoged, 2001), the use of non-modal voice quality is also widely attested at suprasegmental levels, for example as a marker of utterance and/or phrase-finality (e.g. Klatt & Klatt, 1990; Ní Chasaide & Gobl, 1997, 2010; Gordon & Ladefoged, 2001; see also Fant & Kruckenberg, 1989, on Swedish). For English, Henton & Bladon (1987) report creaky voice to be more common at the ends of utterances than elsewhere, and Pierrehumbert & Talkin (1992) report more glottalized vowels at the beginning of intonational phrases than in other locations. Accentuation also seems to affect the use of non-modal voce, as shown by Pierrehumbert (1995), who found word-initial vowels more likely to be glottalized when co-occurring with pitch accents. Similar patterns are reported by Dilley et al. (1996), who report that the use of creaky voice is positively correlated with the strength of a prosodic boundary, suggesting that non-modal voice quality is systematically affected by the prosodic structure of an utterance.

Considering that the use of non-modal voice quality is mediated by both accentuation and prosodic phrasing, and that focus is also known to affect these two (e.g. Féry, 2013), it is not unlikely that systematic relationships may exist between

---

[31] See Section 2.4 for a more detailed review of Sauermann et al. (2011).

non-modal voice quality and focus. At the same time, we have only been able to find one study explicitly addressing this relationship, reported by Arnhold (2011, 2013) and Arnhold & Féry (2014). In Arnhold & Féry (2014), results from two different experiments are presented, addressing the syntactic and prosodic marking of information structure in Finnish. In the first experiment (Arnhold, 2011), the participants produced pre-scripted SVO-sentences across seven focus conditions: broad focus on the whole sentence, narrow focus on subject, verb, or object and contrastive focus on subject, verb, or object. The conditions were implemented by means of wh-questions. The prosodic analysis showed that words referring to new information ended in creaky voice more often than words referring to given information. In addition, non-modal voice quality was more common on given information that occurred post-focally than on given information that occurred pre-focally. In the second experiment each participant described a continuously changed layout of toys to another participant, who could not see the display. This time the participants were allowed to describe the changing layout in whichever way they wanted, and the layout was changed so that target words referred to elements that were introduced (new), reintroduced (given but highlighted) or remaining across consecutive trials (given). The results from the second experiment showed that the use of non-modal voice quality was significantly more common on new or reintroduced elements than on given ones. In fact, the use of non-modal voice was the most reliable prosodic cue to focus that was found in the second experiment. Arnhold & Féry (2014) interpret the use of creaky voice as related to prosodic phrasing, whereby the participants align the focal constituent with the right edge of IP, marking this boundary with non-modal voice (Arnhold & Féry, 2014).

Even if there are some indications that creaky voice is common at utterance ends in Swedish (Fant & Kruckenberg, 1989), it is not clear whether there is a systematic relationship between focus and non-modal voice in this language, despite indications that focus does interact with phrasal structure (e.g. Myrberg, 2009, 2013).

5.3 Research questions

Two research questions were asked in this study, one concerning the use of word duration for focus and one addressing the use of voice-quality manipulations for focus. Our first research question was about word duration, and we asked:

1. Do CS-speaking children between four and eleven differ from CS-speaking adults in the way they use word duration to distinguish narrowly focal from post-focal targets?

Based on previous work on accentual lengthening in adult Swedish (e.g. Heldner & Strangert, 2001) we expected our adult group to produce narrowly focal words with longer duration than post-focal ones, irrespective of sentence-position. As was seen in Section 5.2.2, only a few studies have targeted the use of duration for focus in child speech, and the ones that have done so report no effect of focus on word duration in children between four and eight, even if children predominantly use accent placement in line with adults at this stage (see Section 2.4). Based on these reports, we expected that our five- and eight-year-olds would not use duration for focus in an adult-like way, but that our eleven-year-olds would perform in line with our adult controls.

Non-modal voice in our CS-dataset led to lost pitch data both sentence-medially and sentence-finally. As a consequence, 20% of our target words were unsuitable for pitch analysis sentence-medially, whereas the corresponding proportion was as high as 51% sentence-finally (see Section 4.5.4). Interestingly, the use of non-modal voice was markedly more frequent in our CS data than in our comparable Dutch data. Our initial observations suggested that the use of non-modal voice was particularly common on post-focal constituents, thus our analysis was aimed at testing the robustness of this pattern, asking whether non-modal voice quality was systematically associated with post-focal constituents in our data from CS. The second research question addressed in this study was thus:

2. Do CS-speaking children between four and eleven differ from CS-speaking adults in the way they use non-modal voice quality for marking constituents as post-focal?

Based Arnhold & Féry's (2014) work, we hypothesized that our CS-speaking adults would use non-modal voice quality more on post-focal constituents than on constituents that were not post-focal, and that the use of non-modal voice would be more pronounced sentence-finally than sentence-medially. Arnhold & Féry (2014) actually report increased use of creaky voice on both focal and post-focal constituents, as compared to given and pre-focal ones, suggesting that the onset of creakiness was found already at the end of the focal constituent, continuing into the post-focal area. Different from the analyses carried out by Arnhold & Féry (2014), our analysis was of a more exploratory nature, and our outcome measure was based on our coding of 'immeasurable pitch', taken from the pitch analysis presented in Chapter 4 (see Section 4.5.4.). Since this measure of non-modal voice was rather crude, and relied on non-modal voice quality being present on the whole word, we assumed that a systematic use of non-modal voice quality starting at or after the focal word would surface as a difference between post-focal target words and words

that were not post-focal (that is, either pre-focal or focal). In line with reports that that the more fine-grained prosodic modifications for focus only reach adult proficiency levels at the age of seven to eight years (e.g. Chen, 2011a), we hypothesized that children at four to five years would not make use of non-modal voice in the same way that adults did, but that this feature of adult prosody would be mastered from the age of seven to eight years.

5.4 Dataset and coding

The dataset was exactly the same as the one described in Chapter 4.4, consisting in semi-spontaneous SVO-sentences, elicited while the participants played the picture-matching game with an experimenter. Again, only responses following the scripted speech context were included in the analysis. Responses were also excluded if they contained deviant word orders, deviant word choices or elided constituents, as well as self-repairs, hesitations, or background noise (see Section 4.4.3).

As a reminder, the sentence conditions elicited in the picture-matching game are repeated in Table 19 below.

Condition	Explanation
Initial narrow	Narrow focus on initial constituent (i.e. subject)
Medial narrow	Narrow focus on medial constituent (i.e. verb)
Final narrow	Narrow focus on final constituent (i.e. object)
Medial contrastive	Contrastive focus on medial constituent (i.e. verb)
Broad focus	Broad focus on whole sentence (i.e. subject, verb, object).

Table 19: The original sentence conditions (repeated)

The included responses were orthographically transcribed and segmented into words using Praat (Boersma & Weenink, 2010). When segmenting we relied on changes in the waveform, in addition to the formant transitions shown in the spectrogram (see Turk, Nakai, & Sugahara, 2006). Conventions were established for how to segment the words at particularly challenging boundaries (e.g. onset plosives were consistently segmented right before the burst). The built in silence-detecting function in Praat (Boersma & Weenink, 2010) was also used as an aid when boundaries were placed, particularly in sentence-final position, where the end of a segment can be particularly challenging to determine.

All medial and final target words were segmented by hand, and word durations were automatically extracted from text grids using a script. Our measure of non-modal voice quality was gathered by coding cases where the pitch could not be reliably measured on a word as instances of 'non-modal voice', as opposed to instances where the pitch was regular and measurable, coded as 'modal voice'. Considering that responses including noise or other distortions to the recording were excluded prior to the coding of immeasurable pitch (Section 4.4.3), we can be reasonably certain that words lacking periodic pitch do so because of the way they were produced (i.e. because of distortions of the normal vocal fold vibrations) not because of recording factors. All extracted values were checked for measuring and extraction errors.

5.5 Analysis and results

5.5.1 The effect of focus on word duration

Our measures of word duration were made on sentence medial and sentence final target words, and, analogue to the analyses presented in Chapter 4, conditions were included into the comparisons based on the information structural status they rendered on a particular target word. Sentence-medially, 'medial narrow' represented the narrow focus condition, whereas 'initial narrow' represented the post-focus condition. For the sentence-final comparison, 'final narrow' represented the narrow focus condition, and 'medial narrow' represented the post-focus condition. The choice of using the post-focal condition as the baseline, instead of also including the pre-focal condition, was that only post-focal baselines were available across both sentence-locations. Further, the results from Heldner & Strangert (2001) showed the duration of pre- and post-focal words to be similar to each other, both differing significantly from focal ones.

For the statistical analysis of the effect of focus on word duration we built linear mixed effect models (LMMs) in R (R Core Team, 2014), with 'participant' and 'item' as random factors, and 'focus', 'group' and 'lex' as independent factors. For all models the outcome variable was numeric, involving measures of word duration on medial or final target words. The statistical procedure was exactly the same as for our analysis of focus effects on pitch range described in Chapter 4.5.4. Below we first report the results from the sentence-medial analysis, thereafter reporting the results from the sentence-final analysis.

Sentence-medial position

In Figure 16 we present a boxplot of the effect of focus on word duration sentence-medially, broken down by group and lexical accent.

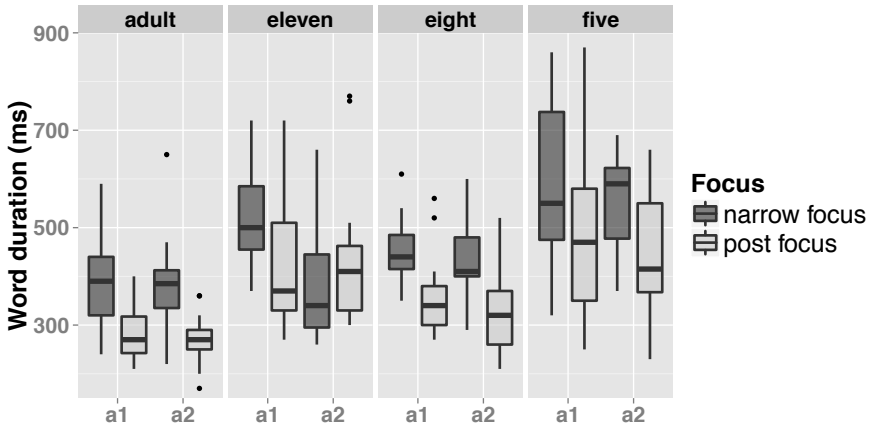

Figure 16: Word duration on sentence-medial targets under narrow and post focus, by group and lexical accent

Building and comparing models on the effect of 'focus', 'group', and 'lex', on the duration of the sentence-medial target words, revealed the best model to contain main effects of 'focus' and 'group', a two-way interaction between 'group' and 'focus', and a three-way-interaction between 'focus', 'group' and 'lex'. In Table 20 we present the p-values for the model comparisons, and in Table 21 we present the summary of the best model.

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

Table 20: Model comparisons, word duration, sentence-medial comparison

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

Table 21: Model summary, word duration, sentence-medial comparison

As can be observed in Table 21, the main effect of focus consisted in post-focus decreasing the word duration of the medial targets. In addition, main effects of group were also observed, involving generally longer word durations in the eleven- and five-year-olds than in the adults, across focus and lexical accent. Furthermore, there was also a two-way-interaction between 'group' and 'lex' for the comparison between the eleven-year-olds and the adults, involving a different effect of lexical accent in the eleven-year-olds than what was found in the adults. This difference consisted in accent 1 words being longer than accent 2 words in the eleven-year-olds, whereas the adult group produced the two lexical accent categories with comparable durations. Finally, the three-way-interaction was caused by the adults and eleven-year-olds differing in the effect of 'focus' by 'lex', in the sense that whereas focus lengthened the duration of both accent 1 and accent 2 words in the adults, in the eleven-year-olds this effect was only present on the accent 1 words (see Figure 16).

Exploring the effect of focus within each group separately showed a significant effect of focus increasing the word duration in the adults (p=.000), the eight-year-

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

Sentence-final position

Figure 17 shows a boxplot of the effect of focus on word duration sentence-finally, by group and lexical accent.

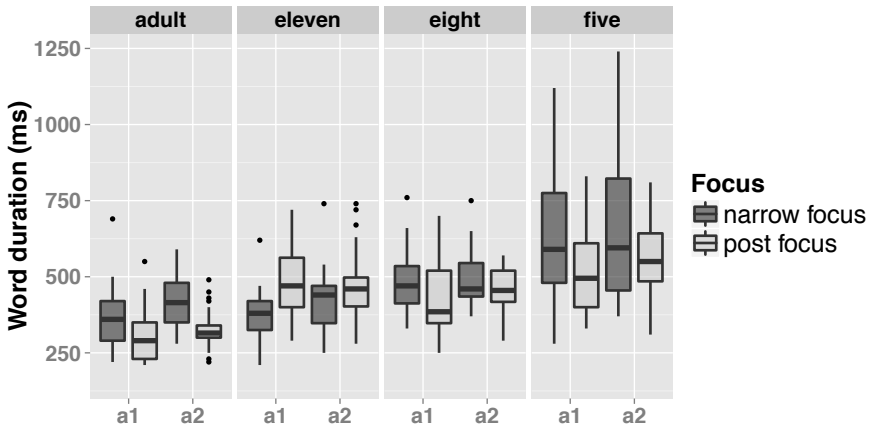

Figure 17: Word duration on sentence-final targets under narrow and post focus, by group and lexical accent

Building models for the effect of 'focus', 'group', and 'lex' on word duration sentence-finally showed the best model to include main effects of 'group' and 'focus', as well as an interaction between the two (see Table 22 for the model comparison and Table 23 for a summary of the best model).

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

Table 22: Model comparison, word duration, sentence-final comparison

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

Table 23: Model summary, word duration, sentence-final comparison

As can be observed in Table 23, the main effect of focus consisted in post-focal status significantly decreasing the duration on the sentence-final target words. The group effects consisted in the eight and five-year-olds producing the sentence-final targets with longer duration than the adults. Finally, the interaction effect between focus and group consisted in post-focal status increasing the word duration in our eleven-year-olds whereas it decreased the duration in our adults. The interaction involving the five-year-olds consisted in a stronger lengthening effect of focus in the five-year-olds than in the adults.

Exploring the main effect of focus within each group showed that focus significantly increased the word-duration in the adults (p=.001) and the five-year-olds (p=.000), and that there was a trend in the same direction in the eight-year-olds (p=.083). A significant effect of focus was also found in our eleven-year-olds (p=.011), but in this group the effect went in the opposite direction of that found in the other groups:

Whereas narrow focus increased the duration in our adults, eight- and five-year-olds, it decreased the duration in our eleven-year-olds.

Our raw means showed narrow focus to increase the duration by 72 ms (18%) in our adults, 50 ms (10%) in the eight-year-olds and by 140 ms (21%) in our five-year-olds, whereas narrowly focal words were on average 67 ms (16%) *shorter* than post-focal ones in our eleven-year-olds.

Interim discussion: Word duration and focus

Summing up our analysis of the relationship between focus and word duration in our four groups, we found general effects of focus increasing the word duration on both medial and final target words in the adults, the eight-year-olds, and the five-year-olds, whereas the eleven-year-olds differed from the adults both sentence-medially and sentence-finally. In this sense, the answer to our research question about whether the adults and children differed in their use of word duration for focus is positive when the eleven-year-olds are considered, but negative for the other groups.[32] In sentence-medial position, the eleven-year-olds differed from the adults in not showing any effect of focus on word duration in accent 2 words, whereas they behaved in line with the other groups on accent 1 words. In sentence-final position, the eleven-year-olds differed from the adults by producing post-focal target words with longer duration than focal ones across both accent types.

The patterns found in our group of adults confirm previous reports on reliable effects of focus on word duration in adult CS, in both sentence-medial and sentence-final position (e.g. Heldner & Strangert, 2001). On average, our adult participants lengthened the target words with 28% sentence medially, and 18% sentence finally. In terms of position effects, the patterns observed in this study resemble descriptions of accentual and final lengthening in English, where accentual lengthening is present both sentence-medially and sentence-finally, but where the effect of accentual lengthening is stronger in non-final than in final positions (e.g. Cambier-Langeveld, 2000; Cooper et. al., 1985). Our findings thus differ from those reported by Heldner & Strangert (2001), who found stronger effects of focal lengthening finally than non-finally. This study was not specifically designed to look at position effects, and we did not make direct comparisons of the degree of lengthening by position. Because of this, our observation of stronger effects of focal lengthening medially than finally needs to be confirmed by manipulating position more systematically.

---

[32] The lack of an interaction between focus and group for the eight-year-olds in the sentence-final analysis shows that the trending effect of focus in the eight-year-olds did not differ significantly from the significant effect of focus in the adults.

Based on previous work on English, German, and Dutch, we hypothesized that our five-year-olds would not use duration in an adult-like way, but that our older child groups would. This hypothesis was disconfirmed in our data, as both the five-year-olds and the eight-year-olds performed in line with the adults. Sentence-finally, the five-year-olds showed even more robust effects of focus on word duration than the adults did. Together with the results presented in Chapter 4, these findings suggest that CS-speaking children reach adult proficiency in duration manipulations for focus before they reach adult proficiency in pitch-based manipulations. It also seems that CS-speaking children manipulate duration for focus at an earlier stage than what has been reported for children learning Dutch, German or English, indicating that CS-speaking children are not only ahead of Dutch-speaking children in the use of pitch-based cues to focus; they are also ahead in the use of duration manipulations.

The fact that we find robust effects of focus on duration in child CS, whereas this is not found in child Dutch, English or German, despite reports of adult-like accent placement in these populations, may suggest that word duration is a more reliable correlate of prominence H in child CS than it is of accentuation in Dutch. Such a possibility is supported by the fact that Sauermann et al. (2011) failed to find significant effects of focus on word duration in German-speaking adults. If the durational effect of focus or focal accentuation is generally more robust in CS than in it is in Dutch, the higher temporal variability characterizing child speech (e.g. Smith, Kenney & Hussain, 1996) may keep duration effects of accentuation from surfacing in studies on child Dutch, German and English. In order to determine whether the use of duration for focus develops differently from the use of prominence H in CS, younger children need to be studied, so that possible dissociations between prominence H and word duration can be systematically studied. In addition, more cross-linguistic work is needed in order to determine whether the relationship between focus, accentuation, and word duration is actually the same across different Germanic varieties.

While our duration analysis showed that both the five and the eight-year-olds performed in line with adults, our results from the eleven-year-olds are rather striking, as previous work suggests that children have largely reached adult proficiency in prosodic focus marking well before this age (e.g. Chen, 2011a; Wells, Peppé & Goulandris, 2004). Indeed, adult proficiency is also found in our measures of prominence H and pitch range in this group, thus the fact that the eleven-year-olds stand out on our duration-measures is rather surprising. The opposite effect of focus on duration sentence finally, together with the different effects of focus on duration between the two lexical accent categories, is hard to explain. Tentatively, we speculate whether task effects may at least partly explain these patterns. When playing the picture-matching game with the oldest group we had the impression that

the game was less interesting to these than it was to the two youngest groups. Occasionally the eleven-year-olds seemed to signal their slight disinterest in the game by slowing down their speech rate (e.g. Mozziconacci & Hermes, 2000). Such lengthening seemed particularly common when the children had already provided the most important information, thus post-focal targets were more prone to such 'unengaged' prosody than focal ones. It may be the case that the tendency for the eleven-year-olds to lengthen post-focal targets may have cancelled out the effect of narrow versus post-focus that was found in the other groups. This said, the children did produce prominence H in line with adults, and in general they sounded natural when playing the game, thus their slight annoyance with the procedure did not keep them from marking focus in adult-like ways in terms of pitch.

5.5.2 Non-modal voice quality in the post focal area

Our hypothesis, based on the observations we made in our pitch analyses, was that non-modal voice quality was systematically associated with the post-focal area in our CS data. In order to statistically test this hypothesis, we included all our original focus conditions except broad focus (see Table 19) in the analysis, comparing the probability of non-modal voice quality between words that were post focal on the one hand and words that were not post focal on the other. For the sentence medial comparison, our 'initial narrow' condition represented the 'post focus' condition, whereas 'medial narrow[33]' and 'final narrow' were both included into the 'non-post focus' condition. Sentence finally both 'initial narrow' and 'medial narrow' were included in the 'post focus' condition, whereas 'final narrow' represented the 'non-post focus condition'.

In order to statistically explore the effect of post-focal status on non-modal voice quality we built binomial logistic regression models (GLMs) in R (R Core Team, 2014). For all models the outcome variable was categorical, and consisted in the binary option of non-modal voice quality or not on a target word. The models were built following the structure presented in Chapter 4.5.3, whereby factors were added to the model one by one until the best model was established. For the analysis of non-modal voice, lexical accent pertinence did not improve any model, thus for clarity we have left this factor out when presenting our results.

---

[33] Because our previous analyses of prominence H, pitch and duration showed no significant differences between our 'medial contrastive' and 'medial narrow' conditions (Chapter 4.5), these were grouped together in our analysis of non-modal voice, under the label 'medial narrow'.

Below we first present bar plots showing the distribution of non-modal voice across our original sentence-conditions, and then we present bar-plots of the conditions re-coded into a post-focal versus a non-post-focal condition. This is done in order to show how non-modal voice was used across the conditions that were pooled together in the statistical analysis. The pooling of the data into these two conditions was done because we wanted to know whether the post-focal area was particularly prone to being produced with non-modal voice as compared to other conditions.

Non-modal voice sentence-medially

The bar plot in Figure 18 shows the proportion of the medial target words that were produced with non-modal voice quality sentence medially, broken down by group and narrow focus location. 'Initial', 'medial' and 'final' refer to the location of narrow focus, so that for the medial comparison, 'initial' rendered the medial target post-focal, 'medial' rendered it focal and 'final' rendered it pre-focal.

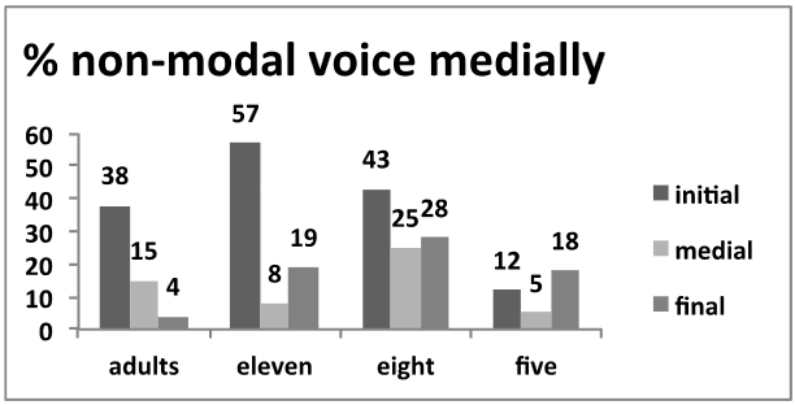

Figure 18: Percentage non-modal voice on sentence-medial targets by narrow focus location and group

For the sentence-medial position, the use of non-modal voice was relatively limited across our three sentence conditions (21% on average). However, as can be seen in Figure 18, narrow focus on the initial target led to more non-modal voice on the medial target than any of the other conditions, even if the use of non-modal voice was attested across all conditions. The distributions shown in Figure 21 thus seem to support our initial hypothesis. Further, whereas the effect of being post-focal seemed fairly systematic in the adults and the eleven-year-olds, the difference between our 'initial' condition and the other two conditions dropped in the eight-year-olds and was absent in the five-year-olds, indicating that the use of non-modal voice was

more systematically related to post-focal status in our adults and eleven-year-olds than it was in our eight and five-year-olds.

When building models on the effect of post-focal status on non-modal voice quality, our focus factor was binary, dividing our original three conditions ('initial narrow', 'medial narrow', 'final narrow') into a 'post focus' and a 'no post focus' condition, as presented in Figure 19 below.

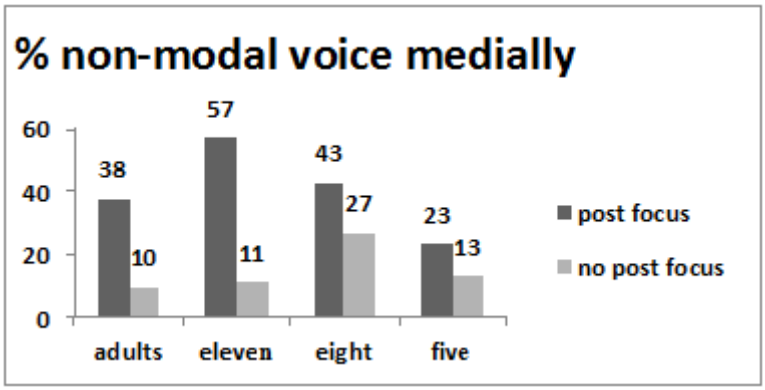

Figure 19: Percentage non-modal voice on sentence-medial targets by post focus and no post focus

Building models for the outcome 'presence versus absence of non-modal voice quality' on the sentence-medial targets, including the factors 'focus' and 'group', as well as the interaction between the two, showed the best model to include a main effect of focus (p=.000), a main effect of group (p=.000) and an interaction between post focus and group (p=.000) (see Table 24 for a summary of the best model).

	b-value	std. error	p-value
Intercept (adults, no post focus)	-2,155	0,264	0,000***
Post focus	1,673	0,383	0,000***
Group eleven	0,096	0,405	0,812
Group eight	1,163	0,334	0,000***
Group five	0,243	0,370	0,512
Post focus * group eleven	0,688	0,586	0,240
Post focus * group eight	-0,968	0,534	0,070.
Post focus * group five	-1,731	0,636	0,006**

Table 24: Model summary, voice quality, sentence-medial comparison

As can be seen in Table 24, the main effect of post focus consisted in a significantly higher probability of non-modal voice quality on post-focal as compared to non-post focal targets (p=.000). The group effect consisted in the eight-year-olds using non-modal voice significantly more than the adults across both focus conditions. Finally, the interaction consisted in a significant difference between the five-year-olds and the adults in the effect of post-focal status on non-modal voice; the effect of post-focus leading to more non-modal voice was stronger in the adults than in the five-year-olds. In addition, a trending interaction was also found for the comparison between the eight-year-olds and the adults, although this difference did not reach significance. In order to explore the effect of post-focus within each group separately, additional models were built for each group. This analysis confirmed the patterns from our larger model, showing a robust effect of post focus on non-modal voice in the adults (p=.000) and the eleven-year-olds (p=.000), but this effect was only trending in the eight-year-olds (p=.061) and absent the five-year-olds (p=.909).

Non-modal voice sentence-finally

Moving on to the sentence-final position, Figure 20 below shows the proportion final target words produced with non-modal voice, broken down by group and position of narrow focus. As can be observed, the use of non-modal voice quality was more common in the sentence final than in the sentence medial position, with 55% of the target words produced with non-modal voice quality, averaged across groups and conditions. In terms of focus effects, the pattern observed sentence-finally resembles the pattern found sentence-medially in that non-modal voice was more common in conditions rendering the target word post-focal (i.e. when the narrow focus was initial or medial) than in the condition that did not render the target post focal (i.e. final focus). In terms of group differences it seemed as though the use of non-modal voice post-focally was more pronounced in the adults, the eleven-year-olds and the eight-year-olds than it was in the in the five-year-olds, even if the effect seemed present in all four groups.

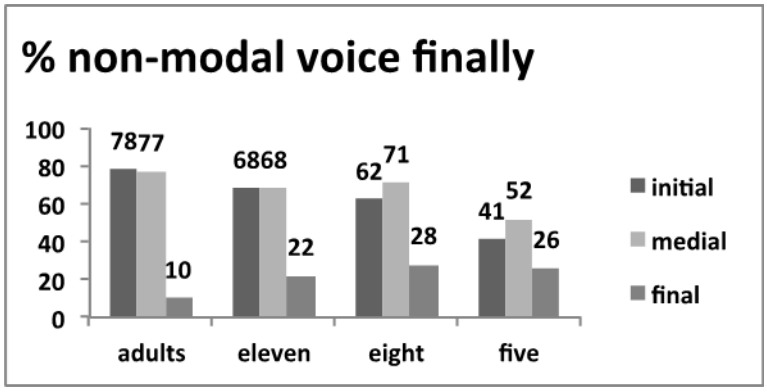

Figure 20: Percentage non-modal voice on sentence-final targets by narrow focus location and group

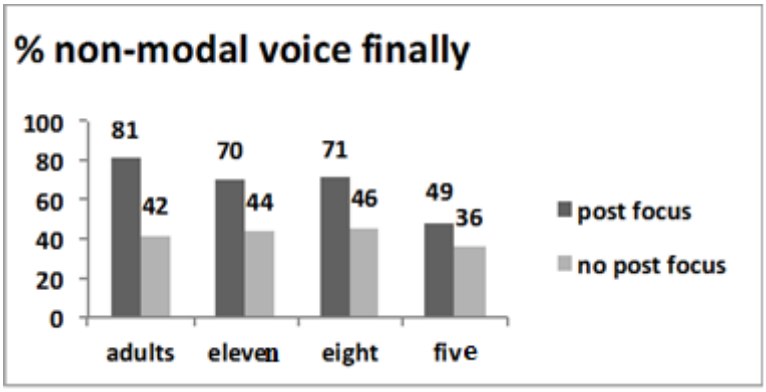

Figure 21: Percentage non-modal voice on sentence-final targets by post focus and no post focus

Building models for the outcome 'non-modal voice quality or not' on the medial targets, including the factors 'focus' and 'group', as well as the interaction between the two, showed the best model to contain a main effect of focus (p=.000), a main effect of group (p=.000) and a marginally significant interaction between post focus and group (p=.052) (see Table 25 for a summary of the best model). In Table 25 we see that the main effect of focus consisted in post-focal status increasing the probability of non-modal voice quality on the target word, and that the interaction found consisted in the five-year-olds differing significantly from the adults in the effect of post focus on non-modal voice quality.

In order to explore the effect of post-focus within each group separately, additional models were built for each group. This analysis confirmed the patterns from our

larger model, showing a robust effect of post focus on non-modal voice in the adults (p=.000), the eleven-year-olds (p=.001) and the eight-year-olds (p=.001), but that such an effect was only trending in the five-year-olds (p=.074).

	b-value	std. error	p-value
Intercept (adults, non-post focus)	-0,310	0,198	0,118
Post focus	1,757	0,318	0,000***
Group eleven	0,087	0,309	0,778
Group eight	0,132	0,301	0,660
Group five	-0,274	0,305	0,370
Post focus * group eleven	-0,674	0,471	0,153
Post focus * group eight	-0,704	0,457	0,123
Post focus * group five	-1,213	0,441	0,006**

Table 25: Model summary, sentence-final voice quality analysis

Interim discussion: Non-modal voice quality and focus

Summing up our analysis of the use of non-modal voice quality, our hypothesis is confirmed for both sentence locations: CS-speaking adults use non-modal voice quality significantly more on post-focal target words than on target words that are not post-focal. In addition, we have seen that similar uses of non-modal voice were observed both in the eleven-year-olds and the eight-year-olds, but that the five-year-olds did not show the adult pattern of using non-modal voice post-focally. The answer to our research question is thus 'yes', adults and children do differ in their use of non-modal voice quality by focus, with more systematic associations between post-focal status and non-modal voice in the adults and the eleven-year-olds, slightly weaker associations between the two in the eight-year-olds, and no significant relationship between post-focal status and non-modal voice in the five-year-olds.

The use of non-modal voice for marking constituents as post-focal is not previously described for Swedish. Even if the analysis presented here was rather rough, we believe the use of non-modal voice in CS may be interpreted in similar ways as the use of creaky voice in Finnish, namely as an indication that speakers insert prosodic boundaries after focus (Arnhold & Féry, 2014). The interaction between prominence H and intonational phrasing in CS is not fully understood, but work from Myrberg (2013) suggests that speakers of CS often insert prosodic boundaries after focal constituents, even if this is not the only option (e.g. Myrberg, 2013). More work is clearly needed in order to establish what the relationship between focus, prosodic

boundaries and voice-quality alternations is in CS. Our data provide a first indication that such a relationship may indeed be present.

5.6 Discussion

In this chapter we have seen that CS-speaking children reach adult proficiency in duration manipulations for focus earlier than what was seen for the pitch-based manipulations investigated in Chapter 4. Further, we have also seen that the use of non-modal voice quality is systematically associated with the post-focal area in adult CS, and that similar uses of non-modal voice are also observed in eleven and eight-year-olds, but not yet in five-year-olds. In this sense, the use of non-modal voice seems to follow a similar development as the use of prominence H for focus, reaching adult proficiency earlier in sentence-final than in sentence-medial positions. At the same time, whereas the use of prominence H was adult like already at five sentence-finally, this was not the case for the use of non-modal voice, thus this manipulation seems to develop slightly later than the use of prominence H.

Having established that CS-speaking children reach adult proficiency in duration manipulations for focus before they are completely adult-like in their use of pitch manipulations for the same purpose, and that the use of duration for focus is attested earlier in CS than has been reported for Dutch, English and German, an important question is how this difference can be understood. Speculating, it may be that the complexity associated with the double mapping of pitch to both word and sentence-level contrasts makes CS-speaking children rely on duration earlier than they master the use of pitch-based cues for focus. This said, as our study only captures the later stages of development, this cannot be determined. At the same time, our results show a rather robust use of duration for focus in CS-speaking adults, whereas Sauermann and colleagues (2011) actually failed to find such robust patterns in German-speaking adults. Care should be taken in order to avoid comparing child speech gathered in more interactional contexts with lab-speech elicited from adults, and future investigations on the relationship between focus, accentuation and word duration in Dutch, German and English-speaking adults should be aimed at determining whether systematic effects of focus on word duration are systematically present in more spontaneous speech samples. Finally, if it is indeed the case that duration is a more reliable cue to focus in CS than it is in non-tonal West-Germanic varieties, an interesting question is whether children learning other tone languages than CS also acquire duration for focus at an early stage. Recent findings on the acquisition of prosodic focus in Mandarin Chinese suggest that this may indeed be the case (Yang & Chen, 2014; Yang (forthcoming).

6 Focal accentuation in Dutch: The acquisition of accent placement and accent type

6.1 Introduction

In this chapter we present a study on how Dutch-speaking children between four and eleven years develop toward adult proficiency in marking focus by means of accentuation[34]. Unlike Swedish, Standard Dutch (hereafter Dutch) is a non-tone language, and pitch is not used for signalling word-level contrasts. In Dutch, prosodically highlighting focal information is achieved by means of accentuation, i.e. the placement of a particular pitch movement (such as a high tone, a low tone, or specific combinations of high or low tones) on the stressed syllable of a word (see Chapter 2.6). A number of different pitch accent types can be used for this purpose, but the most commonly used accent type used for focus is a fall ('H*L') (Chen, 2009). Constituents referring to non-focal information are treated differently depending on whether they appear before or after focal ones. Whereas pre-focal constituents are commonly accented, post-focal constituents are mostly de-accented (Gussenhoven, 1984; Chen, 2007,2009; Swerts, Krahmer & Avesani, 2002; Nooteboom & Kruyt, 1987; Rump & Collier, 1996). Previous work has shown that Dutch children's ability to prosodically highlight focal information develops rather gradually: The use of pitch accent placement is mastered between four and five years, but the adult preference for falls over other accent types, as well as the ability to use phonetic modifications of pitch and duration, in situations where accent placement is not sufficient, is not in place before seven or eight, or even later (Chen, 2007, 2009, 2011a, 2011b).

This Chapter is structured as follows. In Section 6.2.1 we review previous work on prosodic focus marking in adult Dutch, and in 6.2.2 we review what is known about prosodic focus marking in child Dutch. Following this we present our research questions (6.3). In the method section (6.4) we provide information about our

[34] An earlier analysis of some of the data presented in Chapter 6 was published as Romøren & Chen (2014)

participants, and we comment on how the picture matching game presented in Chapter 3 was adapted for this study, with particular attention to how the data was coded and analysed. In Section 6.5 we present our four analyses, corresponding to our four research questions. In Section 6.6 we summarize the results from the analyses, and in Section 7 we discuss our findings, compering our findings to previous reports on prosodic focus marking in child and adult Dutch.

6.2 Background

6.2.1 Prosodic focus marking in Standard Dutch

In West-Germanic languages, focus is predominantly marked using prosody. In Dutch this is done by assigning a pitch accent to constituents representing focal information, often associated with an expanded pitch range and an increased word duration on accented words as compared to non-accented ones (Chen, 2009, 2011a, 2011b, 2012; Gussenhoven, 1984; Hanssen, Peters & Gussenhoven, 2008). Dutch speakers can use different pitch accent types to mark focus (e.g. falls 'H*L', downstepped falls '!H*L', sustained high pitch 'H*' or rises 'L*H', to mention some), but falling accents ('H*L') are preferred for marking focus in declarative sentences, regardless of sentence position (Chen, 2007, see also Chapter 2.6). Constituents referring to non-focal information typically do not carry pitch accents (i.e. they are de-accented) but in sentence-initial position they are nearly always accented, mostly with the same fall ('H*L') that is also used for focus. In such cases, adults distinguish focal from non-focal target words by expanding the pitch range and increasing the word duration on focal as compared to non-focal ones (Chen, 2009).

Similar to the situation for English and German, the accentuation of focal information is more consistent than the de-accentuation of non-focal information, and both information status and rhythmic preferences affect the probability that non-focal information is de-accented (Horne 1991, see also Gussenhoven, 1983; Cruttenden, 2006; Röhr, Baumann & Grice, 2015). Structurally, the focal constituent tends to receive the last major prominence in the utterance (i.e. the nuclear accent), but accents also commonly occur on constituents preceding the focal one (i.e. pre-nuclear accents). The tendency for non-focal information to be both accented and de-accented is supported by perception research, showing that Dutch listeners are more lenient in accepting accentuation on non-focal information than they are in accepting absence of accentuation on focal information (Noteboom & Kruyt, 1987).

To the best of our knowledge, Chen (2007, 20011a, submitted) is the only researcher who has systematically studied the use of accent placement and different accent types for marking focus by adult speakers of Dutch. By using the picture-game setup described in Section 2.4, Chen showed that adults accent sentence-initial subject nouns regardless of focal status, whereas sentence-final objects were accented 40% of the time when non-focal, but 80% of the time when focal. Generally, the preferred accent type for focus was a fall ('H*L') regardless of sentence-position, but downstepped falls ('!H*L) were also common for marking sentence-final focus. In Chen's setup, non-focal targets in sentence-final position were always post-focal, whereas her sentence-initial non-focal targets were pre-focal, thus her findings support the pattern that non-focal information is more likely to be de-accented when following the focus then when preceding it (see Horne, 1991).

In another study on prosody and focus in Dutch, Hanssen, Peters & Gussenhoven (2008) studied the phonetic differentiation between different focus types. The authors presented acoustic analyses of accented target words in sentence-medial position, across contrastive focus, narrow focus and broad focus. The participants read scripted conversations, where the answer to the question involved different focus structures (e.g. '*What's the matter?' 'We want to stay in Manderen'/ 'Where do you want to stay?' 'We want to stay in Manderen' / 'Do you want to stay in Montfort?' 'We want to stay in Manderen').* The authors found that the participants produced the target word (*Manderen*) with a steeper F0 fall, an earlier F0 minimum, and a longer word duration, when contrastively or narrowly focal than when the word was part of a broad focus sentence.

To summarize, previous work shows that adult speakers of Dutch accent narrowly focal information and de-accent post-focal information, but that pre-focal information is predominantly accented, at least in sentence-initial position. Further, when accented targets are compared, adult speakers make use of phonetic manipulations of pitch and duration in order to differentiate focal from non-focal accents, as well as in distinguishing narrow or contrastive focus from broad focus. In terms of the learning task Dutch children are faced with, they thus need to learn that focus is accented, in what way focus is accented, and that the extent to which non-focal information is de-accented depends on where in an utterance this information is presented.

6.2.2 Prosodic focus marking in child Dutch

In a series of studies, Chen and colleagues (Chen & Fikkert, 2007, Chen 2009; 2011a; 2011b) report on Dutch children's ability to prosodically highlight focal constituents, as compared to non-focal ones. Starting with young children, Chen &

Fikkert (2007) present data showing that children around the two-word-stage accent both words in their two-word utterances, regardless of the information status of these words. Re-analysing a subset of this data, Chen (2011b) reports a tendency for the children to de-voice or use downstepped accents on non-focal words, indicating that some focus-mediated manipulations are already observable at this stage, even if the children do not yet de-accent in the way as adults. In later work, Chen used a picture-matching game for investigating prosodic focus marking in Dutch children between five and eight years (see 2.3 for a detailed description). Chen's results showed that, like adults, the four- to-five-year-olds predominantly accented focal targets in sentence-initial and sentence-final position, they predominantly de-accented post-focal targets sentence-finally, and they accented sentence-initial targets regardless of focal status. Subsequent analyses of accent type choices in the children and adults revealed that whereas the adults showed a preference for falls ('H*L') for marking sentence-final focus, the children's accent type choices where more variable, showing a higher proportion of rises ('L*H') than those found in the adult data (Chen, 2011a). In a separate study, this time on the phonetics of sentence-initial focus, phonetic comparisons were made between focal and non-focal falling accents ('H*L'), showing that whereas the adults used pitch range and duration to distinguish focal from non-focal accents in this position, children at four to five years did not (Chen, 2009).

Chen's (2011a, 2011b) data from seven- to eight-year-olds show that at this stage, the adult preference for falls for marking final focus is acquired. At the same time, non-adult patterns were also observed for this group in sentence-initial position: Whereas adults reduced the pitch range and word duration on post-focal accents as compared to focal ones, the seven- to eight-year-olds were found to prefer high or rising accents ('H*') for non-focal subjects, and falling accents ('H*L) for focal ones, thereby adapting a non-adult strategy of using different accent types when distinguishing focal from pre-focal targets.

Based on the above findings, Chen (2011a) concludes that the use of accent placement is mastered before children show the same accent type preferences as adults, and before they use more fine-grained phonetic manipulations, in cases where accentuation is not sufficient for distinguishing focal from non-focal information. Further, her results from the seven- to eight-year-olds suggest that even at this stage, non-adult like patterns are still observed in Dutch-speaking children's prosodic focus marking.

6.3 Research questions

Four age groups were included in this study: four- to five-year-olds, seven- to eight-year-olds and ten- to eleven-year-olds, in addition to a group of adult controls. By incorporating a wider set of focus conditions than what is typically done, and by including a larger age range, our aim was to add to Chen's previous studies on how Dutch children's prosodic marking of focus develops over time. Comparing narrowly focal target words to pre-focal and post-focal renditions of the same word allowed for systematic comparisons of both on-focus and outside-focus manipulations in our four groups. Performing comparable analyses both sentence-medially and sentence-finally permitted investigating position effects in the acquisition of prosodic focus marking, while keeping the comparisons the parallel in both locations. Finally, adding a contrastive focus and a broad focus condition allowed us to explore focus *type* effects in our four groups, something previous work on prosodic focus marking in children has typically not addressed (but see e.g. Hanssen et al., 2008; Baumann et al., 2007; Myrberg, 2012, for studies on focus type effects in adults, see also Sauermann et al., 2011, for a study on such effects in children).

Our analysis is divided into four parts. In the first three we look at different phonological and phonetic manipulations used for marking constituents as focal versus non-focal. We start by looking at the use of *accent placement* (research question 1), thereafter looking at the use of *accent type* (research question 2), and then finally we present phonetic measures of the *acoustic consequences* of the accentuation patterns described in the first and second analyses (research question 3). In the fourth part of the analysis we compare the use of accentuation across different focus types, that is, comparing accentuation across broad, narrow and contrastive focus (research question 4).

Accent placement by narrow focus

As we saw in Section 6.2.1, marking focus in adult Dutch involves modifications both on the focal constituent and on the post-focal constituent, thus our analysis of the marking of narrow focus concerns both on-focus and post-focus measures. Our first research question concerns the use of accent placement to mark narrow focus, and we asked:

1. Do Dutch-speaking children between four and eleven differ from Dutch-speaking adults in the way they use accent placement to mark narrowly focal versus non-focal target words?

Our study of narrow focus accentuation was conducted for each sentence position separately. Sentence-medially, the use of accentuation was compared between targets under narrow focus on the one hand and targets that were post- or pre-focal on the other. Sentence-finally, narrowly focal target words were compared to post-focal target words. For the adults we expected a consistent use of pitch accent on constituents that were narrowly focal, in addition to a fairly high proportion of accentuation on pre-focal constituents. We also expected avoidance of accentuation on post-focal constituents. Based on previous descriptions of prosodic focus marking in English-, German- and Dutch-speaking children (Chapter 2.4), as well as our own findings from CS-speaking children (Romøren & Chen, 2015, Chapters 4 and 5) we expected our oldest group of ten- to eleven-year-olds to perform in line with the adults. Based on Chen's findings we anticipated that both the four- to five-year-olds and the seven to- eight-year-olds to be adult-like in their use of accent placement sentence-finally, but that the two younger age groups would be less adult-like sentence-medially, in line with Chen's reports on the sentence-initial position. Finally, we expected the primary differences found between younger children and adults to be related to the children's ability to de-accent post-focally rather than to accent focally.

Accent type preferences by narrow focus

Our second research question addressed the use of different accent types for marking focus. Previous work from Chen (2007) has shown adults to prefer falling accents for marking focus sentence-initially and sentence-finally, but that children at four to five do not yet show this preference sentence-finally, producing more equal distributions between rising accents ('L*H'), plain falls ('H*L') and downstepped falls ('!H*L') for marking sentence-final focus (Chen, 2007; Chen, 2011a). Interestingly, non-adult distributions of certain accent types for focus are also reported in children as old as seven to eight, where children are reported to reserve falls ('H*L') for focus but high level or rising accents ('H*') for topics sentence-initially (Chen, 2009), suggesting that the use of accent type distinctions for focus changes rather dynamically through children's development. Our second research question concerns accent type preferences in our four groups, and we asked:

2. Do Dutch-speaking children between four and eleven differ from adults in the way they use certain accent types to mark narrowly focal versus non-focal target words?

Based on Chen's reports we hypothesized that our adults would show a preference for falling accents for marking focus, regardless of sentence position. We also hypothesized that differences would be found between our adult group and our four- to five-year-olds, as well as between our adults and our seven- to eight-year-olds, but that at ten to eleven our children would behave in line with the adults. Finally, we hypothesized that children who did not de-accent post-focal targets would use downstepped accents post-focally, a strategy Chen reports younger children to use at stages when they fail to de-accent completely (2011a).

Measures of pitch and duration by narrow focus

There is a tendency in the literature on prosodic focus marking in children to adapt either a phonological or a phonetic perspective on the manipulations investigated. Studies typically look at either accent placement or type, referring to assumed phonological categories and the manual coding of these, or they measure acoustic correlates of information structure without reference to a phonological level (c.f. Chen, 2007, who describes accentuation patterns vs. Wonnacott & Watson, 2008, who report on phonetic measures). Based on the literature, as well as our choice of using AM theory as our theoretical starting point, we assume that the primary cue to focus in Dutch is accentuation, and that different accent types can be used for marking focus, but that some accent categories are more common that others (see also Chapter 2.6). Typically, accenting a word is associated with increases in pitch range and duration, but there is no one-to-one relationship between accent placement and measures like pitch range, particularly when comparisons are made across a variable set of pitch accent categories.[35] This said, we do expect to see broad reflexes of our participants' accentuation patterns in our measures of pitch and duration. In addition, where differences between children and adults are observed in our accent analysis, adding phonetic measures can shed light on possible intermediate steps children take when developing their phonological system. One example could be that children, at stages where their manipulations are not sufficiently reliable to be perceived as pertaining to the adult phonological category (say, a falling accent), nevertheless systematically expand the pitch range on focal versus non-focal targets, thereby making *some* of the acoustic manipulations observed in adults, even if the children's accentuation patterns are not yet completely adult-like.

---

[35] For example, accenting a word with a rising or high level accent ('H*') does not necessarily entail a pitch excursion on the accented word, thus the word can be perceived as accented without any substantial pitch range increase.

For our phonetic analysis we chose not to include the pre-focus baseline, as our accentuation analysis showed pre-focal targets to be accented close to ceiling by all groups. Our third research question was:

3. Do Dutch-speaking children between four and eleven differ from adults in the way they manipulate maximum pitch, minimum pitch, pitch range and word duration, for distinguishing between narrowly focal and post-focal target words?

We expected the adults to accent focal and de-accent post-focal targets, and that this would lead to focal targets being produced with a larger pitch range and a higher pitch maximum than post focal ones. Further, based on the literature on accentual lengthening in Dutch we expected focal targets to be produced with a longer word duration than post-focal ones, but that this effect might not be found sentence-finally (Sluijter & Van Heuven, 1995; Cambier-Langeveld & Turk, 1999). As we expected our phonetic measurements to be related to the way the children used accent placement and accent type, we expected a weaker phonetic differentiation between focal and post-focal targets in groups shown to use accent placement and type less systematically for focus. Alternatively, we also hypothesized that the children might manipulate pitch and duration more independently of pitch accent placement and type, so that general effects of focus on pitch and duration would be observable in the children even at stages where their use of accentuation/ de-accentuation may not yet be adult-like.

Accent placement by broad, narrow and contrastive focus

Previous work on prosodic focus marking in children has typically not investigated children's ability to prosodically distinguish between different *focus types* (but see Sauermann et al., 2011, for an exception). In addition to comparing focal and non-focal renditions of our targets, we also included a broad focus condition in our design, rendering the whole target sentence (including subject, verb, and object) broadly focal, as well as contrastive narrow focus condition, rendering our sentence-medial targets contrastively focal (See Chapter 3.4). [36]

---

[36] The few studies systematically comparing the prosodic marking of broad versus narrow or contrastive focus in adult speech have all been conducted on constructions where the accentuation was controlled, so that narrow focus was elicited on a

Our fourth research question addressed the use of accent placement for marking different focus types, and we asked:

4. Do Dutch-speaking children between four and eleven differ from adults in the way they use accent placement to differentiate between (a) broad and narrow focus and (b) contrastive and non-contrastive focus?

In most Germanic languages, broad focus is marked by aligning a maximal prominence (i.e. a nuclear pitch accent in languages like Dutch) with the right edge of the last intonation phrase of an utterance. Even if an accent in this position is assumed to be sufficient for a broad focus interpretation of the utterance, pre-nuclear accents are common on constituents preceding the final one, and arguments are more likely to be accented than verbs, if they are adjacent to each other (e.g. Gussenhoven, 1983, 1999; Selkirk, 1984, 1995). Based on these descriptions, we expected the medial targets to be accented less often under broad than under narrow focus, but that final targets would be accented regardless of whether they were under narrow focus or part of a broad-focus utterance. Since we assume that aligning a pitch accent with the right edge of the intonational phrase is the default pattern in Dutch, we expected the children to be reasonably good at accenting the final targets across both narrow and broad focus. In line with the tendency of children to accent more rather than less than adults, we expected that the younger children would accent the medial targets more under broad focus than the adults would, but that their ability to de-accent the medial targets under broad-focus would develop in line with their ability to de-accent post-focally.

As was seen in Chapter 2.4, most work on prosodic focus marking in younger children has been concerned with contrast or contrastive focus, whereas non-contrastive focus has only been looked at in older children. Based on the results of these different studies, Chen (2011a) suggests that prosodically marking contrastive focus or contrast might be acquired earlier than non-contrastive focus. It is possible that a contrastive relationship between referents is more salient to children than non-contrastive cases of relational or referential newness, and that this causes children to mark contrastive focus earlier than non-contrastive focus. By including both

---

constituent that was also expected to carry the nuclear pitch accent in broad focus. In our setup, narrow focus was elicited in a slightly different way, making our approach less suitable for this kind of phonetic comparison.

contrastive and non-contrastive focus in our study we wanted to see whether children develop differently in their prosodic marking of these two types of focus. For our comparison between contrastive and narrow focus, we expected that our adults would accent contrastively and narrowly targets to the same extent, but that our younger children might be more consistent in accenting contrastively than non-contrastively focal targets.

6.4 Method

When gathering our data, a Dutch version of the picture-game presented in Chapter 3 was used, eliciting semi-spontaneous SVO sentences with varying focus structure. Compared to Chen's (e.g. 2011a) version of the game, our adapted version allowed for comparing a wider set of focus conditions, as well as for looking at a sentence-position not previously considered: the sentence medial position. Whereas Chen's (2011a) two conditions (final focus versus initial focus) involved using different non-focal baselines (i.e. pre-focus versus post focus) for the two sentence-positions investigated, our setup allowed for comparing narrowly focal targets to post focal ones both medially and finally. In addition, including a contrastive focus and a broad focus condition also allowed for investigating how prosody is used in distinguishing between different focus types, comparing accentuation patterns across broad, narrow and contrastive focus (e.g. Hanssen, Peters & Gussenhoven, 2008). Finally, compared to previous work on adult Dutch, our methodology elicits a more spontaneous speech style than what has been typical in studies of prosody and focus in adult speech (e.g. Hanssen, Peters & Gussenhoven, 2008; Sluijter & van Heuven, 1995, 1996). By incorporating question-answer sequences into a game, we hoped to gather naturalistic speech data without the speakers being particularly attentive to their own speech, something which has been shown to directly affect the way different information structure contrasts are prosodically marked (see Breen et al., 2010, for a discussion).

6.4.1 Participants

Twenty-six Dutch-speaking children and fourteen Dutch-speaking adults participated in the study. The participants were divided into four age groups; four- to five-year-olds, seven- to eight-year-olds, ten- to eleven-year-olds and adults (see Table 26). In the following we will refer to the children as five-year-olds, eight-year-olds and eleven-year-olds.

Age group	N	Age range (years, months)	Age mean (years, months)	Gender
4-5	10	4;4-5;8	5;2	6 male, 4 female
7-8	8	7;1-8;11	8;0	4 male, 4 female
10-11	8	10;4-12;0	10;7	4 male, 4 female
Adults	14	18;7-48;0	23;11	5 male, 9 female

Table 26: Participant information, Dutch

None of our Dutch-speaking participants had any history of language disorders, hearing problems or other known developmental disorders, and they were all first language speakers of standard Dutch. The children were recruited from kindergartens and schools around Utrecht, The Netherlands, 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 detailed information about the children's language background. The adult participants were recruited from the participant pool at the linguistics lab, Utrecht Institute of Linguistics, Utrecht University. They were all university students and first language speakers of standard Dutch, and none of them studied linguistics at the time of recruitment.

6.4.2 Procedure

The data analysed consisted of semi-spontaneous subject-verb-object (hereafter SVO) sentence productions, recorded while the participants were playing the picture-matching game together with an experimenter (see Chapter 3 for a detailed description). Game sessions with children took place in a designated test room at their school; sessions with adults were conducted in a sound attenuated booth at the Linguistics Lab at Utrecht University. Four experimenters were trained to carry out the experiment according to a detailed instruction, and all sessions were video recorded to control for consistency across sessions.

The Dutch version of the game consisted of 30 test trials and 5 practice trials, spread over five sentence conditions, namely narrow focus on the initial constituent ('initial narrow'), narrow focus on the medial constituent ('medial narrow'), narrow focus on the final constituent ('final narrow'), contrastive focus on the medial constituent ('medial contrastive') and broad focus on the whole sentence ('broad focus'). In the Dutch version of the game we originally elicited two different sentence-types, subject-verb-object (SVO) sentences and subject-verb-object-adverbial (SVOA) sentences. In the analyses presented in this chapter, only the SVO sentences were included, meaning that the included number of trials per participant was 15. The

SVOA data was presented in Romøren & Chen (2014) and is also included in the analyses presented in Chapter 7.

6.4.3 Data selection and coding

At the first stage of analysis, 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. Responses were also excluded if they contained deviant word orders, deviant word choices or elided constituents, as well as self-repairs, hesitations, or background noise.

In Table 27 we present an overview of the inclusion rates for the four Dutch groups. Excluded responses are sorted into five categories. 'Disfluencies' refer to cases where a response contained repairs or filled pauses, and 'information structure' refers to cases where the speech context could not be completely controlled (e.g. where responses did not immediately follow the scripted context or where between-trial conversations had rendered certain constituents salient). The 'non-target' category involved responses that contained the wrong words, lacked certain constituents or had non-target constituents added to them. Finally, the category 'noise/overlap' contained instances where a response included noise, laughter or speech overlaps, making the recordings unfit for analysis. The total inclusion rate was 83%, but the groups differed both in over-all inclusion rate and in the distribution of excluded responses across the five categories. The total number of responses included in the current dataset was 500. As two words per sentence were analysed, this renders a total dataset of 1000 words analysed.

	4-5	**7-8**	**10-11**	**Adults**	**Total**
Total elicited	150	120	120	210	600
Total included	104	96	110	190	500
Disfluencies	9	9	2	8	28
Information structure	13	4	2	6	25
Non-target	22	6	2	1	31
Noise/ overlap	1	5	4	5	15
Other	1	0	0	0	1
% included	69	80	92	91	83

Table 27. Excluded responses by group and category, Dutch

Just as for the Swedish dataset, included responses were orthographically transcribed and segmented into words using Praat (Boersma & Weenink, 2010). When segmenting we relied on changes in the waveform, in addition to the formant transitions shown in the spectrogram (see Turk, Nakai, & Sugahara, 2006). Conventions were established for how to segment the words at particularly challenging boundaries (e.g. onset plosives were consistently segmented right before the burst). The built in silence-detecting function in Praat (Boersma & Weenink, 2010) was also used as an aid when boundaries were placed, particularly in sentence-final position, where the end of a segment can be particularly challenging to determine.

All medial and final target words were hand coded for presence/ absence of pitch accent, as well as type of pitch accent. The coding was conducted according to the ToDI system (Gussenhoven, 2005; Gussenhoven, Rietveld & Terken, 1999) and was performed by the author, after a three-month period of working through the online ToDI manual, regularly discussing and comparing annotations with an experienced annotator. All the data went through three rounds of coding, and difficult cases were discussed with the more experienced ToDI trainer when necessary.

In ToDI, four basic accent types are recognized: 'H*' (high pitch stretch or weak rise), 'L*' (low pitch stretch), 'H*L' (a fall), and 'L*H' (a rise). These contours can be further modified by two phonological processes: downstep and peak delay. The downstepped counterparts of falls and sustained high accents are coded as '!H*L' and '!H*', respectively. A fall with a delayed peak is coded as 'L*HL'. Finally, the notation 'H*LH' is used for pre-nuclear steep falls that rise again after the fall, and the notation 'H*!H' is used for the vocative chant. In addition to these accent types, ToDI also assumes boundary tones at the edges of intonational phrases, which can be high ('%H'), low ('%L') or falling ('%HL') initially, and high ('H%') low ('L%') or half-completed ('%') finally. Both boundary tones and pitch accents were included in our original coding of the data, but in the analyses presented here we only consider the use of accentuation.

In addition to our ToDI annotation, four phonetic measures were added to our analysis of the sentence-medial and sentence-final target words: minimum F0 in target word (hereafter pitch minimum), maximum F0 in target word (hereafter pitch maximum), the pitch range on the target word (difference between the maximum and minimum F0) and target word duration. F0 minimum and maximum were the absolute minimum and maximum pitch levels within the target word. When placing landmarks, care was taken to avoid areas where micro-prosodic effects originating from segment or word-transitions, as well as areas with creaky voice quality. In other words, the maximum and minimum points were manually placed in areas

where the pitch was reasonably stable, and where the pitch level observed in the spectrogram could be auditorily confirmed as reliably representing the F0 minimum or maximum within that word. All measurements were printed to Excel sheets using scripts; values were checked for tracking and measuring errors, and corrected where necessary.

6.5 Analysis and results

6.5.1 Introductory remarks

In this section we present four different analyses of how our participants used prosody for distinguishing between different focus conditions. The four analyses correspond to the four research questions presented in Chapter 6.3. In the first analysis (6.5.3) we look at the use of pitch accent placement, answering our first research question on how our participants use accent placement to mark narrow focus. After that we address the second research question, concerning the use of different accent types for marking narrow focus in our four groups (6.5.4). In the third analysis we compare measures of pitch and duration between focal and post-focal target words, answering our third research question concerning the phonetic differentiation between narrowly focal and post-focal target words (6.5.5). Finally, the fourth analysis answers our fourth research question on the use of accentuation for distinguishing between different focus types, namely broad, narrow and contrastive focus (6.5.6).

6.5.2 Sentence and focus conditions

As a reminder, the original conditions elicited in the picture-matching game are repeated in Table 28 below.

Condition	Explanation
Initial narrow	Narrow focus on initial constituent (i.e. subject)
Medial narrow	Narrow focus on medial constituent (i.e. verb)
Final narrow	Narrow focus on final constituent (i.e. object)
Medial contrastive	Contrastive focus on medial constituent (i.e. verb)
Broad focus	Broad focus on whole sentence (i.e. subject, verb, object).

Table 28: The original sentence conditions (repeated)

All our output measures were gathered 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. In the following we will use the terms medial and final targets, stressing the location of the comparison rather than the grammatical category to which the targets pertained.

In order to answer our first research question on whether our participants assigned pitch accents to narrowly focal targets, conditions rendering a target narrowly focal ('medial narrow' for the medial position, 'final narrow' for the final position) were compared to conditions rendering the same target pre- or post-focal. Sentence-medially, our pre focus condition was our 'initial narrow' condition, whereas our post focus condition was our 'final narrow' condition. Sentence-finally, we didn't have a pre focus condition, but our 'narrow medial' condition served as the post focus condition. In the analysis of accent type placement (6.5.2) as well as our analysis of pitch and duration (6.5.3), the post-focal condition served as the 'no focus' condition. This was because the post focus condition allowed for parallel comparisons sentence-medially and sentence-finally.

In order to answer the fourth research question about focus type effects, our medial target words were compared between 'broad focus', 'narrow medial' and 'contrastive medial', whereas our final targets were compared between 'broad focus' and 'narrow final'.

6.5.3 Accent placement by narrow focus

Model comparisons

For our analysis of accent placement, our outcome measure was binomial, consisting in the presence or absence of a pitch accent on a target word. Originally we wanted to use generalized mixed-effect models for statistically analysing our data, taking the crossed random effects of item (i.e. the 3 target words on which the information structure was manipulated) and participant (our 40 participants) into account, but as our outcome measure was very skewed for most of our comparisons, our mixed effect models mostly failed to converge. Consequently, we decided to use binomial logistic regression models (GLMs) instead, with the side effect that random effects of participant and item could not be added to the model.

All our models were built and compared using R (R Core Team, 2014). The statistical procedure was exactly the same as the one used in the analysis of prominence H, described in Chapter 4.5.3. For some of the model comparisons,

complete separation (i.e. pitch accentuation at 100% within a level of ‘focus’ or ‘group’), made the model unfittable. Separation made models including the separated factor crash, so in such cases we slightly manipulated the data by adding one instance of the minority pattern to each sub-level where separation took place. 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.

As will be seen below, interactions between focus and group rarely reached significance in our full models on accent placement by narrow focus in our four groups, despite the raw data often suggesting that such differences were present. We believe the reason the interactions did not reach significance in the full models is that the dataset included in this study was quite small (500 sentences, spread over four groups and five conditions, rendering around 20 items per condition in our child groups and around 40 items per condition in our adult groups, in each sentence position). Lack of interaction effects in the full model forces the conclusion that the effect of focus was the same between the adults and any of the child groups, but in many cases the raw data strongly suggested that the effect was absent in one or several groups. We therefore decided to also run analyses of the effect of focus *within* each group, independent of whether interactions between focus and group reached significance in the full models. A larger dataset is required in order to confirm the preliminary conclusions we draw, based on analysing data from each group separately.

Sentence medial position

The percentage accentuation across narrow focus (‘medial narrow’), pre focus (‘final narrow’) and post focus (‘initial narrow’) on the sentence-medial target words, is presented in Figure 22 below.

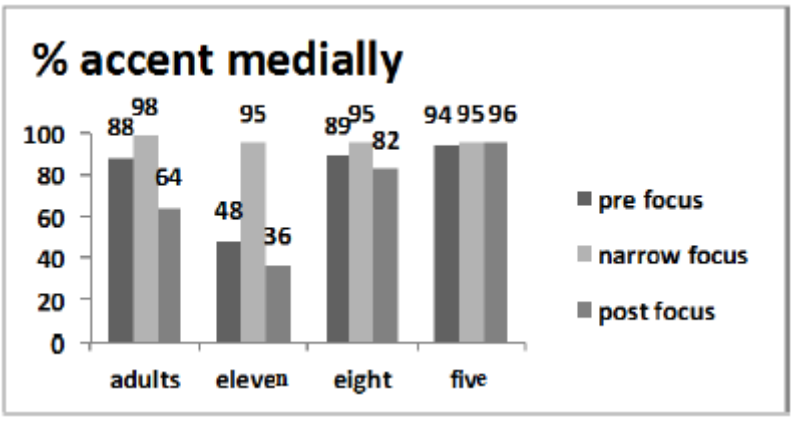

Figure 22: Percentage accentuation on sentence-medial targets under pre focus, narrow focus and post focus, across groups

As can be observed in Figure 22, the patterns from adults, eleven-year-olds and eight-year-olds were similar in the sense that narrow focus led to the most accentuation, followed by pre focus and then post focus. Still, the difference between narrow focus on the one hand and pre- and post-focus on the other seemed larger in the eleven-year-olds than in the adults, and smaller in the eight-year-olds than in the adults. Our group of five-year-olds did not seem to differentiate the conditions at all, with accentuation close to ceiling for all three conditions.

We start by reporting on models comparing the accentuation between *narrowly focal* and *post-focal* targets (the two rightmost bars in the plot). Building regression models on the outcome 'presence of pitch accent sentence medially', adding 'focus' ('narrow focus' vs. 'post focus'), 'group' ('adults', 'eleven', 'eight', 'five'), and their interaction, showed 'focus' (p=.000) and 'group' (p=.000) to improve our model, but the interaction between 'focus' and 'group' did not (p=.176). In other words, the observed between-group differences in the use of accent for focus (Figure 22) did not reach significance, thus our best model for this comparison was a model including main effects of 'focus' and 'group' only, as summarized in Table 29 below.

Factor	b-value	std. error	p-value
Intercept (adults, narrow focus)	3,285	0,584	0,000***
Post focus	-2,651	0,574	0,000***
Group eleven	-0,997	0,492	0,043*
Group eight	0,688	0,589	0,243
Group five	1,834	0,799	0,022*

Table 29: Model summary, accent placement, sentence-medial post focus comparison

In Table 29 we see that post-focal status significantly decreased the likelihood of a pitch accent on our medial targets. Further, we see that the group effects consisted in the eleven-year-olds accenting significantly less than our adult group and the five-year-olds accenting significantly more than our adult group, across both narrow and post focus.

Running analyses of the effect of focus within each group separately showed narrow focus to significantly increase the probability of a pitch 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).

We also compared our medial targets between *narrow focus* and *pre focus*, asking whether our children and adults also differentiated these two conditions. Our pre-focus comparison was carried out in exactly the same way as our post-focus comparison, but this time the factor 'focus' included the *pre focus* ('final narrow') and the *narrow focus* ('medial narrow') conditions. Building models with 'focus', 'group', and their interaction, rendered similar results as our post-focus comparison: 'focus' (p=.000) and 'group' (p=.002) improved our model, but the interaction did not (p=.311). Our best model thus contained only the main effects of 'focus' and 'group', as summarized Table 30 below. From the model summary we see that pre-focal status significantly decreased the likelihood of a pitch accent on our medial target words, and that the eleven-year-olds accented significantly less than the adults across both focus conditions.

Factor	b-value	std. error	p-value
Intercept (adults, narrow focus)	3,854	0,660	0,000***
Pre focus	-1,941	0,590	0,001**
Group eleven	-1,755	0,566	0,002**
Group eight	-0,054	0,752	0,943
Group five	0,353	0,856	0,680

Table 30: Model summary, accent placement, sentence-medial pre focus comparison

Looking at the effect of focus within each group separately showed narrow focus to significantly increase the probability of an accent (as compared to pre focus) in our eleven-year-olds (p=.009), whereas this effect was only trending in our adults (p=.072) and not significant in our eight-year-olds (p=.514) and five-year-olds (p=.911).

Sentence final position

The proportion accentuation across narrow focus ('final narrow') and post focus ('medial narrow') on our sentence-final targets is presented in Figure 23 below. For this position we had no pre focus condition.

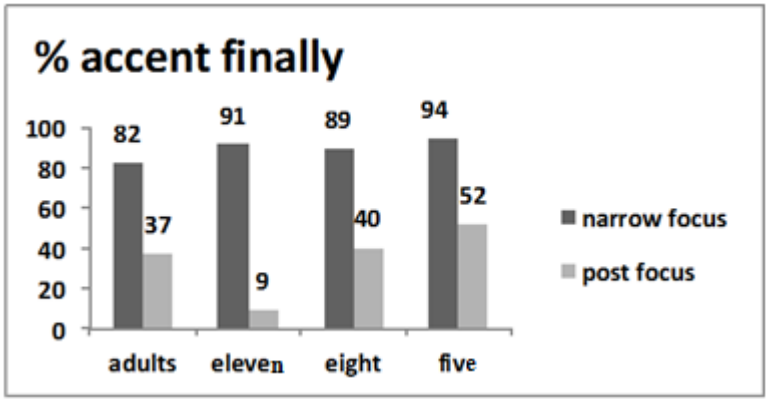

Figure 23: Percentage accentuation on sentence-final targets under narrow focus and post focus, across groups

As can be seen in Figure 23, sentence-finally there was a general pattern of more accentuation on narrow focus than on post focus in all four groups, even if the effect seemed larger in the eleven-year-olds than in the adults.

Building models comparing accent placement between *narrow focus* and *post focus*, with the factors 'focus', 'group', and their interaction, showed 'focus' (p=.000) and the interaction between 'focus' and 'group' to improve our model (p=.041), whereas 'group' did not, even if the effect was trending (p=.061). Summarizing our interaction model (Table 31), we see that post-focal status generally lowered the probability of an accent, that no group effects reached significance in the full model, and that the interaction involved a stronger effect of post focus in the eleven-year-olds than in the adults. Exploring the effect of focus in 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).

Factor	b-value	std. error	p-value
Intercept (adults, narrow focus)	1,551	0,416	0,000***
Post focus	-2,101	0,528	0,000***
Group eleven	0,801	0,849	0,346
Group eight	0,590	0,856	0,491
Group five	1,283	1,110	0,248
Post focus * group eleven	-2,553	1,173	0,029**
Post focus *group eight	-0,445	1,023	0,663
Post focus * group five	-0,637	1,236	0,606

Table 31: Model summary, accent placement, sentence-final post focus comparison

Interim summary

We find no significant between-group differences in our full models sentence-medially. Nevertheless, looking at the effect of narrow focus within each group showed our adults and eleven-year-olds to accent significantly more on narrowly focal targets than on post-focal targets, whereas this effect did not reach significance in our eight- and five-year-olds. Furthermore, only our eleven-year-olds significantly differentiated focal from pre-focal targets, even if a trend in the same direction was found for the adults. Based on these findings we conclude that the use of accentuation for marking narrow focus is only mastered between eight and eleven, when the sentence-medial position is concerned.

For the sentence-final position our full model revealed a stronger effect of narrow versus post focus in our eleven-year-olds than in our adults, but no other differences between adults and children reached significance in the full model. Looking at the

effect of focus within each group separately showed all groups to use accentuation significantly more on focal than on post-focal targets in this position. We therefore conclude that using accent placement for marking narrow focus is mastered already at the age of five sentence-finally, in line with previous reports from Chen (e.g. 2011a).

6.5.4 Accent type preferences by narrow focus

Our second analysis addressed the question of accent type choices in our four groups, asking whether our child groups differed from our adults in their accent type preferences for marking focal versus post-focal target words. Since the dataset presented here was so small we decided not to run any statistical analyses on the use of different accent types for focus. Instead we base our analysis on patterns observable from the raw proportions of each accent type, broken down by group and focus. Below we first present the distribution of different accent types by focus and group sentence-medially, and then the distribution of different accent types by focus sentence-finally.

Sentence medial position

In Table 32 below we report percentages for the most common accent types occurring sentence medially, broken down by focus and group. The 'other' category refers to accent types that occurred less than 5% of the time in this position, across groups and sentence conditions.

		H*L	**!H*L**	**!H***	**L*H**	**no accent**	**other**
adult	narrow	**56**	**37**	2	2	0	2
	post	0	**46**	10	5	**36**	3
eleven	narrow	**64**	0	0	9	5	**23**
	post	5	**14**	5	9	**64**	5
eight	narrow	**55**	10	10	0	5	**20**
	post	0	**45**	9	**23**	18	5
five	narrow	**38**	**33**	14	0	0	14
	post	21	**33**	**29**	0	4	12

Table 32: Accent type choices on sentence-medial targets under narrow and post focus. The numbers represent proportions. The most common and the second most common pattern used by each group in each condition appear in bold.

In addition to the different accent types used by our participants, the choice of ‘no accent’ is also included into the proportions, repeating the finding presented in 6.5.3, of more de-accentuation post focally in our adults and eleven-year-olds than in our eight and five-year-olds.

In Table 32 we see that in our adults, the dominant pattern for focus sentence-medially was the falling accent (‘H*L’), followed by the downstepped fall (‘!H*L’). Post-focally, the adults preferred downstepped falls (‘!H*L’), followed by de-accentuation. In other words, the adults commonly used downstepped falls both focally and post-focally. In contrast to the adults, our eleven- and eight-year-olds seemed to avoid downstepped falls for focus, reserving ‘H*L’ for focus and ‘!H*L' for post-focus, thereby implementing a more categorical accent-type-to-focus mapping than the adults did. In addition to this, both the eleven and the eight-year-olds produced a certain proportion of accents categorized as ‘other’ on focal targets, something that was not found in the adults. Whereas the eight-year-olds rarely de-accented post-focally, they did use the adult pattern of downstepped falls in this condition (‘!H*L’). In addition, the eight-year-olds also used rising (‘L*H’) accents post-focally, a pattern rarely observed in the adults.

Our five-year-olds resembled our adults in showing a preference for falls followed by downstepped falls for focus, but whereas the adults used more non-downstepped falls than downstepped falls for focus, the two patterns were about equally common for focus in the five-year-olds. Post-focally we also observe differences between the five-year-olds and the other groups in that they used a fair amount of non-downstepped falls (‘H*L’) together with downstepped level high tones (‘!H*’) post-focally, something that was uncommon in the other groups.

Sentence final position

In Table 33 below we present proportions of the most common accent types occurring sentence finally, broken down by focus and group. Again, the ‘other’ category refers to accent types that occurred less than 5% of the time in this position, across groups and sentence conditions. In this position the accent type choices were generally less variable across our four groups, with only four outcomes occurring above our 5% threshold.

		H*L	!H*L	L*H	no accent	other
adult	narrow	12	**52**	8	**18**	10
	post	0	**29**	5	**63**	2
eleven	narrow	**43**	**35**	9	9	4
	post	0	**9**	0	**91**	0
eight	narrow	**16**	**58**	11	11	5
	post	5	**20**	10	**60**	5
five	narrow	**22**	**72**	0	6	0
	post	5	**33**	10	**48**	5

Table 33: Accent type choices on sentence-final targets under narrow and post focus. The numbers represent proportions. The most common and the second most common pattern used by each group in each condition appear in bold.

As can be seen in Table 33, our adult group showed a preference for dowstepped falls ('!H*L') for marking sentence final focus, and the proportion of non-downstepped falls was fairly low. Post-focally, the adults showed a preference for de-accenting, followed by downstepped falls ('!H*L'), thus again the downstepped falls were used on both focal and post-focal target words. Our eleven-year-olds showed a more even distribution between falls ('H*L') and downstepped falls ('!H*L') for focus than the adults did, with non-downstepped falls being slightly more common than downstepped ones. As seen in Section 6.5.3, the eleven-year-olds de-accented post-focal targets more than the adults, and when they did not, they used downstepped falls ('!H*L'). Both our eight-year-olds and our five-year-olds resembled the adults in showing a preference for downstepped falls ('!H*L') for focus, and in showing a preference for de-accentuation followed by downstepped falls ('!H*L') post-focally.

Interim summary

Based on the results from our accent type analysis, the answer to our second research question is positive: Differences were again found between adults and children. Just as in the analysis of accent placement, the differences were more pronounced sentence-medially than sentence finally. Sentence-medially, the eleven and eight-year-olds produced a sharper distribution of certain accent types by focus than the adults, who used downstepped falls across both focus conditions. Additionally, our eleven and eight-year-olds also occasionally produced accent types that were not used equally often by adults, as shown by their use of 'other'

accent types for focus, and the use of post-focal rises in the eight-year-olds. Finally, the five-year-olds differed from the adults and older groups in using non-downstepped falls ('H*L') and downstepped high level accents ('!H*') post-focally, as well as in using falls ('H*L') less on focal targets.

Sentence-finally, the children differed less from the adults, and the major difference between adults and children was found in that the eleven-year-olds displayed stronger preferences for ('H*L') for focus and de-accentuation for post-focus than the adults did.

6.5.5 Measures of pitch and duration by narrow focus

Model comparisons

In order to statistically explore the effect of focus on our phonetic measures we built linear mixed effect models (LMM) in R (R Core Team, 2014), following the same procedure that was described in Chapter 4.5.4, without the factor 'lexical accent'. For our pitch analysis, the outcome measures were pitch maximum, pitch minimum and pitch range, on medial and final target words. For our word duration analysis, the outcome measure was word duration on medial and final target words. When building models, our random factors were 'participant' and 'item'; our fixed factors were 'focus', 'group', as well as interaction between the two. Focal targets were compared to post-focal targets only. Again, due to the dataset explored being so small, we also included analyses of the effect of focus within each group, as analysed separately. Below we first report the results of our sentence-medial comparisons, followed by the results of our sentence-final ones.

Sentence medial position

Pitch maximum

In Figure 24 below we present a boxplot of the effect of narrow versus post focus on pitch maximum in our four groups.

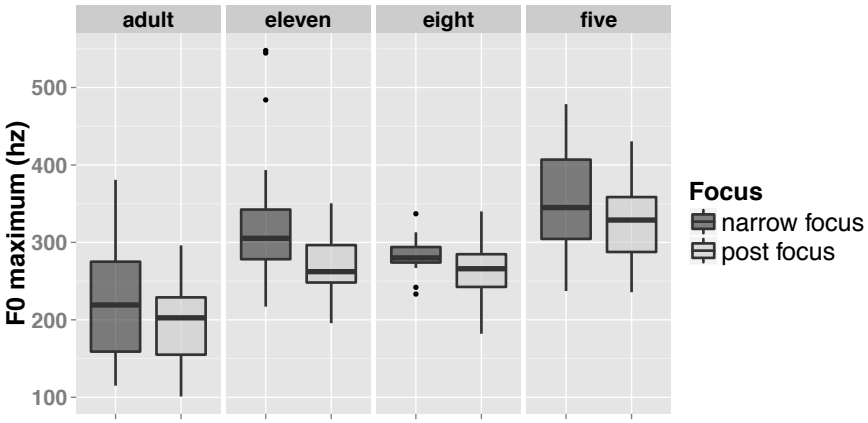

Figure 24: Pitch maximum on sentence-medial targets under narrow and post focus, across groups

Building LMMs on the effect of 'focus', 'group', and the interaction between the two, on pitch maximum medially, showed focus (p=.000) and group (p=.000) to improve the model, whereas the interaction between focus and group did not reach significance (p=.136). Summarizing our best model (Table 34) we see that post-focal status significantly decreased the pitch maximum on our medial target words, whereas the pitch maximum was generally higher in all our child groups than in our adult group.

Factor	b-value	std. error	p-value
Intercept (adults, narrow focus)	226,181	13,334	0,000***
Post focus	-31,889	5,701	0,000***
Group eleven	91,614	21,740	0,000***
Group eight	64,933	21,772	0,005**
Group five	128,820	20,612	0,000***

Table 34: Model summary, pitch maximum, sentence-medial comparison

Looking at the data from each group separately, we find a significant effect of focus in the adults (p=.000) and the eleven-year-olds (p=.000), a trending effect in the eight-year-olds (p=.073), and no significant effect in the five-year-olds (p=. 185).

Pitch minimum

In Figure 25 we present a boxplot of the effect of narrow versus post focus on pitch minimum in our four groups.

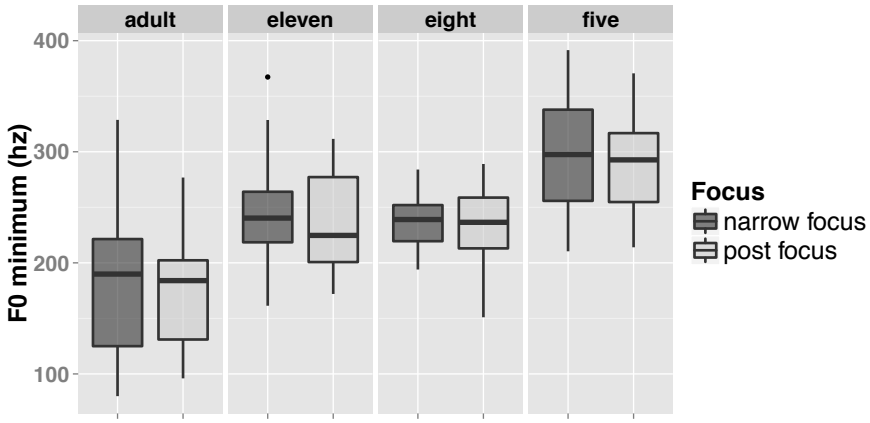

Figure 25: Pitch minimum on sentence-medial targets under narrow and post focus, across groups

Building LMMs with the factors 'focus' and 'group', as well as the interaction between the two, showed only 'group' to improve our model (p=.000), whereas 'focus' (p=.125) and the interaction between 'focus' and 'group', did not (p=.489). Summarizing our best model (Table 35), we see that the group effects consisted in a higher pitch minimum in our three child groups than in our adults.

Factor	b-value	std. error	p-value
Intercept (adults)	179,550	12,010	0,000***
Group eleven	57,500	17,050	0,002**
Group eight	56,370	17,070	0,002**
Group five	112,290	16,180	0,000***

Table 35: Model summary, pitch minimum, sentence-medial comparison

Looking at the data from each group separately, we find no significant effect of focus in the adults (p=.160), the eleven-year-olds (p=.377), the eight-year-olds (p=.431), or the five-year-olds (p=. 997).

Pitch range

In Figure 26 we present a boxplot of the effect of narrow versus post focus on pitch range in our four groups.

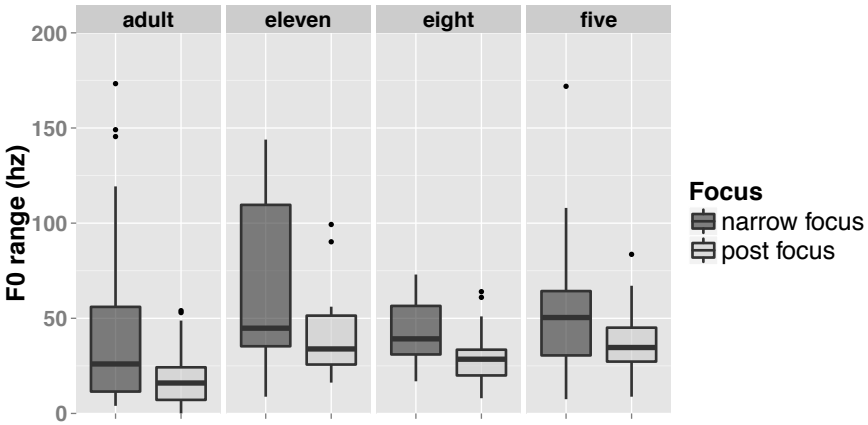

Figure 26. Pitch range on sentence-medial targets under narrow and post focus, across groups

Building LMMs with the factors 'focus' and 'group', as well as the interaction between the two, showed 'focus' (p=.000) and 'group' (p=.039), to improve the model, whereas the interaction did not (p=.123). The model summary in Table 36 shows that narrow focus significantly increased the pitch range on our target words, and that our eleven-year-olds produced the medial targets with a significantly wider pitch range than our adults across narrow and post focus.

Factor	b-value	std. error	p-value
Intercept (adults, narrow focus)	43,124	9,061	0,000***
Post focus	-24,823	4,864	0,000***
Group eleven	34,156	11,981	0,007**
Group eight	7,744	12,025	0,524
Group five	16,070	11,493	0,170

Table 36: Model summary, pitch range, sentence-medial comparison

Looking at the effect of focus in each group separately showed focus to significantly increase the pitch range in the adults (p=.004) and the eleven-year-olds (p=.000), but not in the eight-year-olds (p=.183) or the five-year-olds (p=.099).

Word duration

In Figure 27 we present a boxplot of the effect of narrow versus post focus on word duration in our four groups.

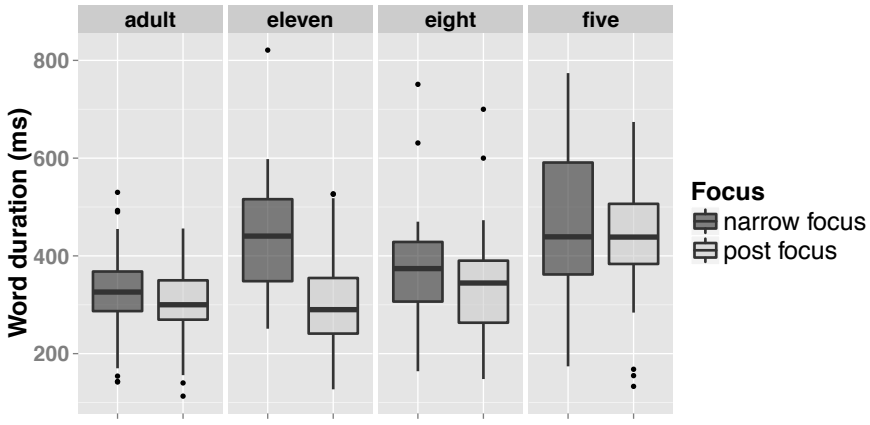

Figure 27: Word duration on sentence-medial targets under narrow and post focus, across groups

Building LMMs with the factors 'focus' and 'group', as well as the interaction between the two, showed 'focus' (p=.000) and 'group' (p=.001), as well as the interaction between them (p=.008) to improve our model. The summary of the interaction model (Table 37) shows no main effect of focus to reach significance when the interaction was added, but that the group effect consisted in the eleven-year-olds and the five-year-olds generally producing longer words than the adults. Finally, the interaction effect consisted in a stronger effect of focus in the eleven-year-olds than in the adults.

Looking at the data for each group separately showed no significant effect of focus on word duration in the adults (p=.181) or the eight-year-olds (p=.187), but significant effects of focus in the eleven-year-olds (p=.000) and the five-year-olds (p=.036).

Factor	b-value	std. error	p-value
Intercept (adults, narrow focus)	0,327	0,041	0,002**
Post focus	-0,026	0,019	0,181
Group eleven	0,122	0,037	0,002**
Group eight	0,059	0,038	0,122
Group five	0,155	0,036	0,000***
Post focus * group eleven	-0,108	0,032	0,001**
Post focus *group eight	-0,009	0,033	0,777
Post focus * group five	-0,030	0,033	0,365

Table 37: Model summary, word duration, sentence-medial comparison

Interim summary: Sentence medial position

The results of our group wise analysis on the use of pitch and duration for focus are summarized in Table 38 below.

Group	Pitch max	Pitch min	Pitch range	Duration
Adults	yes	no	yes	no
Eleven	yes	no	yes	yes
Eight	no	no	no	no
Five	no	no	no	yes

Table 38: Summary of the phonetic analyses, sentence-medial position. 'Yes' means that a significant effect of narrow versus post-focus was found on the relevant parameter for the relevant group. 'No' means that no such effect reached significance.

As can be observed in Table 38, our adults produced narrowly focal targets with a larger pitch range and a higher pitch maximum than post-focal ones, but no effects of focus were found on pitch minimum or word duration. Our eleven-year-olds behaved similarly to our adults with regard to pitch maximum and range, but in this group word duration was also used to distinguish focal from post-focal targets. None of our outcome measures were significantly correlated with focus in our eight year olds. In our five year-olds, the only phonetic measure affected by focus was word duration.

Sentence final position

Pitch maximum

In Figure 28 we present a boxplot of the effect of narrow versus post focus on pitch maximum sentence-finally in our four groups.

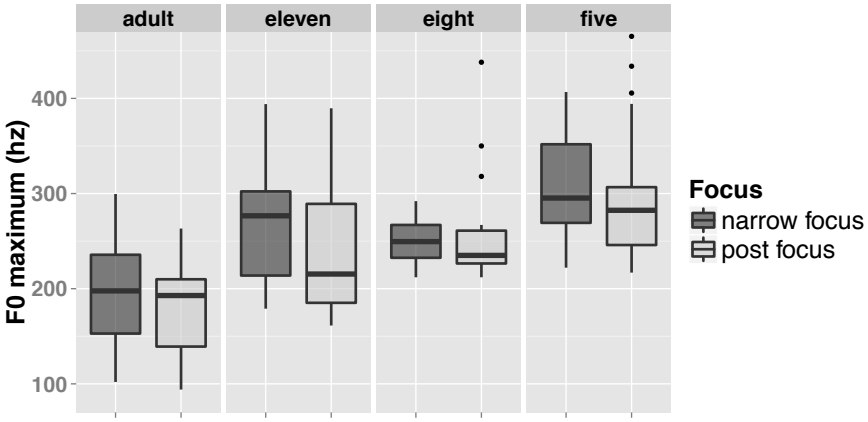

Figure 28: Pitch maximum on sentence-final targets under narrow and post focus, across groups

Building LMMs with the factors 'focus' and 'group', as well as the interaction between the two, showed both 'focus' (p=.000) and 'group' (p=.000) to improve the model, whereas the interaction did not (p=.123). Summarizing the best model (Table 39), we see that focus significantly raised the pitch maximum on our sentence-final target words, and that the group effect consisted in all our child groups producing these target words with a higher pitch maximum than our adult group, across both focus conditions.

Factor	b-value	std. error	p-value
Intercept (adults, narrow focus)	195,969	16,242	0,000***
Post focus	-24,219	6,268	0,000***
Group eleven	73,794	25,884	0,008**
Group eight	83,217	26,195	0,003**
Group five	117,922	24,581	0,000***

Table 39: Model summary, pitch maximum, sentence-final comparison

Looking at the data from each group separately, we find a significant effect of focus in the adults (p=.033), the eleven-year-olds (p=.005) and the five-year-olds (p=. 004), but not in the eight-year-olds (p=.810).

Pitch minimum

In Figure 29 we present a boxplot of the effect of narrow versus post focus on pitch minimum sentence finally, in our four groups.

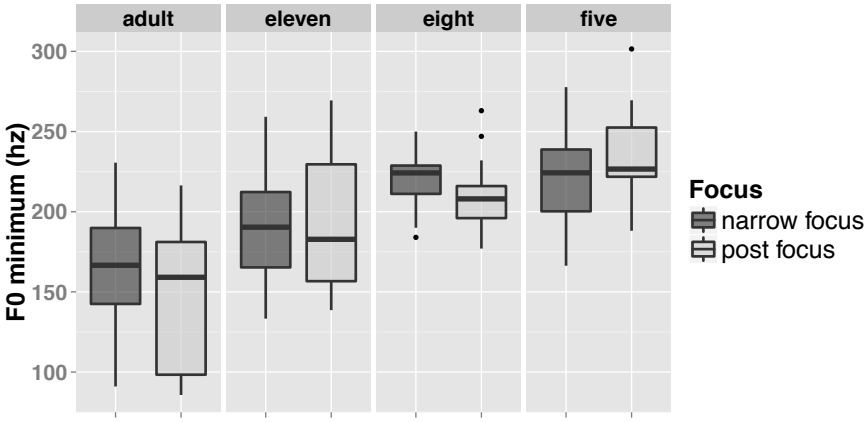

Figure 29: Pitch minimum on sentence-final targets under narrow and post focus, across groups

Building LMMs with the factors 'focus' and 'group', as well as the interaction between the two, showed only 'group' to improve the model (p=.000), whereas focus (p=.172) and the interaction between focus and group (p=.150) did not. The model summary presented in Table 40 shows the group effect to involve a significantly higher pitch minimum in all our child groups, as compared to our adults.

Factor	b-value	std. error	p-value
Intercept (adults)	150,991	8,151	0,000***
Group eleven	44,487	13,030	0,002**
Group eight	66,282	13,234	0,000***
Group five	74,613	12,433	0,000***

Table 40: Model summary, pitch minimum, sentence-final comparison

Looking at the data from each group separately, the effect of focus on pitch minimum did not reach significance in the eleven-year-olds (p=.804), the eight-year-olds (p=.157), or the five-year-olds (p=.250), but it was trending in the adults (p=.074).

Pitch range

In Figure 30 we present a boxplot of the effect of narrow versus post focus on pitch range sentence-finally in our four groups.

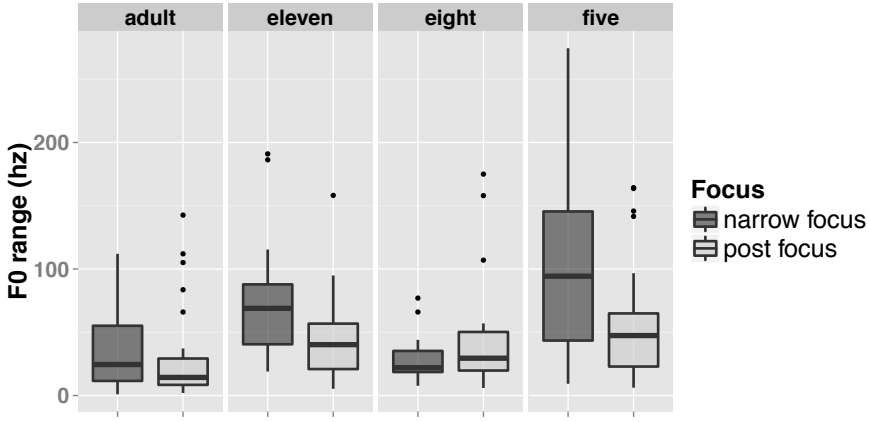

Figure 30: Pitch range on sentence-final targets under narrow and post focus, across groups

Building LMMs with the factors 'focus' and 'group', as well as the interaction between the two, showed 'focus' (p=.001), 'group' (p=.039), and the interaction between the two (p=.000) to improve the model. The best model is summarized in Table 41. Here we see that the main effect of focus is no longer significant in the full model, but that there is a group effect of significantly larger pitch range in our eleven- and five-year-olds than our adults. Finally, the interaction consisted in a stronger effect of focus in our five-year-olds than our adults.

Factor	b-value	std. error	p-value
Intercept (adults, narrow focus)	38,017	11,137	0,001**
Post focus	-10,632	8,990	0,239
Group eleven	39,884	17,977	0,032*
Group eight	-2,255	18,850	0,905
Group five	62,869	17,648	0,001**
Post focus * group eleven	-22,040	14,437	0,129
Post focus *group eight	29,490	15,690	0,062.
Post focus * group five	-33,370	14,418	0,022*

Table 41: Model summary, pitch range, sentence-final comparison

Looking at the effect of focus within each group separately showed a significant effect of focus in the eleven-year-olds (p=.005) and the five-year-olds (p=.000), but not in the adults (p=.239) or the eight-year-olds (p=.145).

Word duration

In Figure 21 we present a boxplot of the effect of narrow versus post focus on word duration sentence-finally, in our four groups.

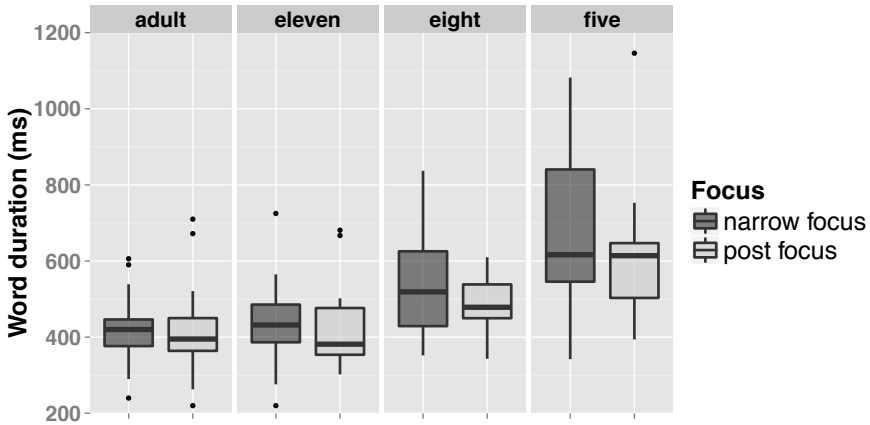

Figure 31: Word duration on sentence-final targets under narrow and post focus, across groups

Building LMMs with the factors ‘focus’ and ‘group’, as well as the interaction between the two, showed ‘group’ to improve the model (p=.000), whereas ‘focus’ (p=.579) and the interaction (p=.731) did not. In Table 42 we see that the group effect consisted in our five-year-olds producing the final targets with significantly longer duration than the adults, across both focus conditions.

Factor	b-value	std. error	p-value
Intercept (adults)	415,982	50,238	0,000***
Group eleven	-0,069	62,484	0,999
Group eight	91,195	62,886	0,156
Group five	277,764	59,652	0,000***

Table 42: Model summary, word duration, sentence-final comparison

Looking at the data for each group separately showed no significant effect of focus in either the adults (p=.507), the eleven-year-olds (p=.583), the eight-year-olds (p=.993) or the five-year-olds (p=.290).

Interim summary: Sentence final position

The results of our group-based phonetic analyses for the sentence-final position are presented in Table 43 below.

Group	F0 max	F0 min	F0 range	Duration
Adults	yes	no	no	no
Eleven	yes	no	yes	no
Eight	no	no	no	no
Five	yes	no	yes	no

Table 43: Summary of the phonetic analyses, sentence-final position. ‘Yes’ means that a significant effect of narrow versus post-focus was found on the relevant parameter for the relevant group. ‘No’ means that no such effect reached significance.

As can be seen, in the sentence-final position only pitch maximum was significantly affected by focus in the adult data, whereas both pitch maximum and pitch range was affected by focus in our eleven-year-olds. Similarly to what was found medially, no measures were related to focus in our eight-year-olds. Finally, our five-year-olds patterned with the eleven-year-olds in producing focal targets with a higher pitch maximum and a larger pitch range than non-focal ones; in fact, an

interaction showed a significantly stronger effect of focus on pitch range in this group than in the adults.

6.5.6 Accent placement by broad, narrow and contrastive focus

Model comparisons

For the analysis of accent placement across different focus types, the procedure was exactly the same as the one described in 4.5.5, except that the output measure was presence/absence of pitch accent and that the factor 'lex' was never included.

Sentence medial position

The percentage accentuation across contrastive focus ('medial contrastive'), narrow focus ('medial narrow') and broad focus, on medial targets, in our four groups, is presented in Figure 32 below. As can be observed, all three focus type conditions rendered accentuation close to ceiling.

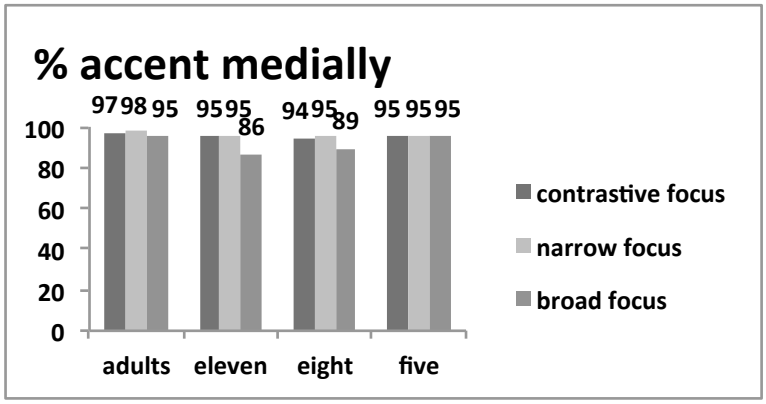

Figure 32: Percentage accentuation on sentence-medial target under contrastive, narrow and broad focus, across groups

Our first focus type comparison concerned *focus size* effects, comparing the effect *narrow focus* versus *broad focus* had on accentuation sentence-medially. For this comparison, regression models with 'focus', 'group', and their interaction, showed neither 'focus' (p=.230) nor 'group' (p=.593) to improve our empty model. Since

our five-year-olds accented at exactly the same frequency across these two conditions, the interaction could not be added for this comparison. Exploring the effect of focus in each group separately confirmed the pattern observed in Figure 32, of no effect of focus in the adults (p=.567), the eleven-year-olds (p=.284) or the eight-year-olds (p=.514). Models could not be built for the five-year-olds when this group was looked at separately, but it is clear from the proportions that the conditions did not differ in this group either. A model summary including main effects of 'focus' and 'group' is presented in Table 44, illustrating the lack of any effects.

Factor	b-value	std. error	p-value
Intercept (adults, broad focus)	2,977	0,625	0,000***
Narrow focus	0,748	0,632	0,236
Group eleven	-0,992	0,791	0,210
Group eight	-0,818	0,844	0,333
Group five	-0,294	0,936	0,754

Table 44: Model summary, accent placement, sentence-medial focus size comparison

Our second focus type comparison concerned contrastivity, comparing *narrow focus* to *contrastive focus* sentence medially. Building models with the factors 'focus' and 'group', as well as their interaction, showed neither 'focus' (p=.896), 'group' (p=.842), or the interaction between 'focus' and 'group', to improve our model (p=.991). Looking at the data from each group separately confirmed the pattern observed in Figure 45, with no significant difference between contrastive and narrow focus in our adults (p=.957), eleven-year-olds (p=.973) and eight-year-olds (p=.871). Separate models could not be run for the five-year-olds, as they accented at exactly the same proportion across these two conditions. In Table 46 we present the summary of our interaction model, showing that none of our parameters reached significance.

Factor	b-value	std. error	p-value
Intercept (adults, contrastive focus)	3,611	1,013	0,000***
Narrow focus	0,078	1,432	0,957
Group eleven	-0,615	1,441	0,669
Group eight	-0,903	1,447	0,533
Group five	-0,666	1,442	0,644
Narrow focus * group eleven	-0,029	2,037	0,989
Narrow focus *group eight	0,158	2,042	0,938
Narrow focus * group five	-0,027	2,038	0,990

Table 45: Model summary, accent placement, sentence-medial contrastivity comparison

Sentence final position

The percentage accentuation across narrow focus ('final narrow') and broad focus on final targets, in our four groups, is presented in Figure 33 below. Sentence-finally we had no contrastive focus condition, thus our comparison involved comparing final targets under *narrow focus* and *broad focus.* As discussed in Chapter 6.3, we expected broad focus to lead to accentuation on the sentence-final targets across both narrow and broad focus. In Figure 33 below we see that the dominant pattern was indeed to accent the final target word in both conditions, but the children seemed to accent slightly more than the adults across both narrow and broad focus.

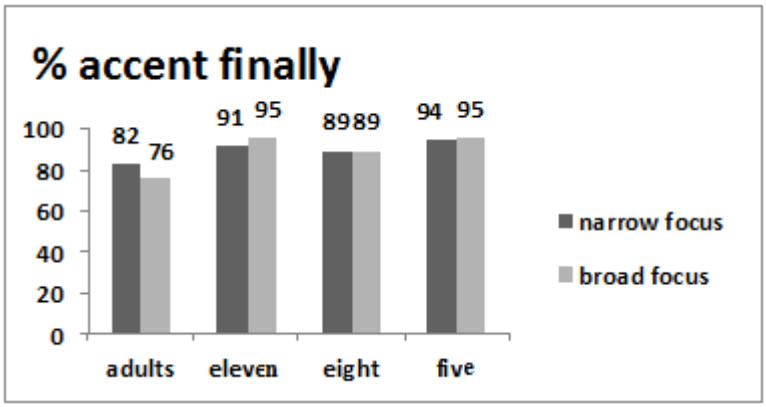

Figure 33. Percentage accentuation on sentence-final targets under narrow and broad focus, across groups.

Building models with the factors ‘focus’, ‘group’, and their interaction, showed only ‘group’ to improve our model (p=.036), whereas ‘focus’ (p=.754) and the interaction (p=.934) did not. In other words, our best model for this comparison was a model including group effects only, as presented in Table 46 below. The summary shows that our eleven and five-year-olds accented significantly more than our adults across both narrow and broad focus.

Factor	b-value	std. error	p-value
Intercept (adults)	1,341	0,272	0,000***
Group eleven	1,298	0,657	0,048*
Group eight	0,799	0,595	0,179
Group five	1,577	0,775	0,042*

Table 46: Model summary, accent placement, sentence-final focus size comparison

Looking at the data from each group separately confirmed the pattern observed in Figure 33, with no significant difference between broad and narrow focus in our adults (p=.480), eleven-year-olds (p=.573) or five-year-olds (p=.911). The models could not be run for the eight-year-olds, as they accented at exactly the same proportion across conditions.

Summarizing our analysis of focus type effects on accent placement we find accentuation at ceiling for broad, narrow and contrastive focus sentence-medially, as well as for broad and narrow focus sentence-finally, and no between-group differences. In other words, neither the adults nor the children differentiated any of our focus type categories by using accent placement.

6.6 Summarizing the results

Starting with our first research question, on whether the adults and children differed in their use of accent placement for marking narrow focus, our results answer this question positively, but only for the younger two groups. Whereas all our groups accented less on post-focal than on narrowly focal targets sentence-finally, only our adults and eleven-year-olds reliably did this sentence-medially, supporting earlier findings that marking focus in non-final sentence positions is harder than marking focus sentence-finally. In sentence-final position the eleven-year-olds de-accented post-focal targets even more than the adults, and this group was also the only group to significantly differentiate pre-focal from focal targets. This suggests that the

eleven-year-olds distinguished focal from non-focal target words even more systematically than our adult group did.

Our second analysis concerned accent type choices in our four groups, and we asked whether our child groups differed from the adults in the way they used different accent types to mark narrow focus. The results from this analysis showed that they did, primarily in the sentence-medial position. In this position the eleven and eight-year-olds tended to use non-downstepped accents more for focus and downstepped accents more for post-focus, whereas the adults used downstepped accents across both categories, even if non-downstepped ones were preferred for focus and downstepped ones were preferred for post-focus. In addition, our five-year-olds showed more non-adult patterns than the older children, for example by using non-downstepped accents and downstepped level accents post-focally. Sentence-finally the four groups were more parallel in their accent type choices by focus, but the eleven-year-olds differed from the adults in showing a stronger preference for non-downstepped falls over downstepped ones for focus, and de-accentuation over downstepped falls for post-focus.

Finally, in our third research question we asked whether our four groups differed in terms of how their accentuation patterns affected measures of pitch and duration on our target words. For the adults and eleven-year olds, the behaviour of accenting focal more than post-focal information, and of using non-downstepped accents more for focus than for post-focus, was primarily reflected in our measure of maximum pitch, which was reliably higher on focal than on post-focal targets in these groups. Sentence-medially, adults and eleven-year-olds also produced focal targets with a larger pitch range, whereas the pitch range effect of focus was only observed in the eleven-year-olds sentence-finally. Somewhat surprisingly, the accentuation patterns produced by the adults was not reflected in effects of focus on word duration in either sentence-position, whereas this was the case in the eleven-year-olds, who we saw de-accented post-focus more consistently than the adults. It may be that the adults' use of de-accentuation was not sufficiently robust for duration effects to surface in our small dataset, but it was for the eleven-year-olds, who de-accented more.

The accentuation patterns produced by our eight-year-olds rendered no significant effects of focus on any of our pitch and duration measures. This is perhaps unsurprising sentence-medially, where they accented across the board, but sentence-finally we would have expected to see similar effects as in the other groups, as the eight-year-olds resembled the adults in both accent placement and accent type choices. At the same time, the patterns from the adults only caused a significant effect of focus on pitch maximum, thus it may be that the eight-year-olds realized

their accent types in a slightly more variable way, and that this kept systematic effects on pitch from surfacing. Finally, the absence of post-focal de-accentuation observed in the five-year-olds sentence-medially may explain why none of our pitch measures were significantly associated with focus in the data from this group. Interestingly, duration was, suggesting that the five-year-olds may be using word duration for marking focus at stages when they do not yet master the accentuation/de-accentuation pattern. Sentence-finally effects of focus on pitch were also found in the five-year-olds, supporting our findings from the analysis of accent placement and type, showing that the five-year-olds were more adult-like sentence-finally than sentence-medially.

In our fourth analysis, on focus type effects, we asked whether our child groups differed from the adults in their use of accent placement for distinguishing between narrow and broad focus on the one hand and narrow and contrastive focus on the other. The results from our focus type analysis answers this question negatively: All our participants accented close to ceiling on broad, narrow, and contrastive focus sentence-medially, and on both broad and narrow focus sentence-finally. In other words, the adults did not mark focus type differences by means of accent placement, and neither did the children. Based on the literature on pre-nuclear accentuation (e.g. Gussenhoven, 2011, and references therein), as well as on our results from CS, we expected that broad focus would decrease the likelihood of an accent on the medial targets, but unlike what was found in our CS data, this was not the case for the Dutch data. It thus seems that the verbs are more likely to be accented under broad focus in Dutch than they are in CS.

6.7 Discussion

The results presented in this chapter confirm previous reports of Dutch-speaking children developing rather gradually in their ability to prosodically highlight focal information. Before sketching the developmental path as suggested by our results, and discussing these in relation to what has been found in previous work, a note on the adult model is required. Our study suggests that adult speakers of Dutch are less consistent in their accent-to-focus mapping than what is typically suggested in the literature, particularly with reference to post-focal de-accentuation. The results from our analysis of narrow focus accentuation (research question 1) showed that although the adults accented close to ceiling on narrowly focal constituents, they actually only de-accented post-focal information 36% of the time sentence-medially and 63% of the time sentence-finally. This means that the adult pattern Dutch children need to acquire is clearly more variable than a simple ‘accent focus, de-accent post-focus’ configuration, particularly when different sentence-positions are

taken into account (e.g. Gussenhoven, 2004; Ladd, 2008). With regard to our pre-focus condition, the accentuation patterns between focal and pre-focal targets sentence medially were found not to differ in our adults, similarly to Chen's reports for the sentence-initial position. Further, even if our adults preferred falls ('H*L') for marking focus sentence-medially, their preferred pattern for marking sentence-final focus was downstepped falls ('!H*L'). Downstepped falls were also very common post-focally in both sentence-positions, suggesting considerable overlap in the distribution of this accent type between focus and post-focus. The variability observed in our analysis of accentuation also seems to have effected our results regarding effects of focus on pitch and duration, where adults produced focal targets with a higher pitch maximum and range medially and higher pitch maximum finally, but where, counter to previous findings (e.g. Sluijter & van Heuven, 1995; Cambier-Langeveld & Turk, 1999), no duration effects of focus were observed in any position, and no effect of focus on pitch range was found sentence-finally.

Sketching a developmental path to adult prosodic focus marking in Dutch based on our child data suggests that children already at four to five mark sentence-final focus in adult-like ways: They accent focal more than post-focal information, they prefer '!H*L' followed by 'H*L' for marking focus, and de-accentuation followed by '!H*L' for post-focus. In addition, their accentuation patterns rendered a higher pitch maximum on focal as compared to post-focal words in final position, similar to patterns observed in the adults. Sentence-medially however, four- to-five-year-olds are not yet adult-like: They fail to de-accent post-focal targets, and even if they generally resemble adults in their accent type choices they also produce non-adult patterns, particularly on post-focal constituents. The less systematic pitch manipulations found for this group are reflected in that none of the pitch-measures from the five-year-olds varied systematically with focus in the sentence-medial position. Interestingly, in this position the five-year-olds increased the duration on focal versus post-focal target words, suggesting that when their pitch manipulations for focus are still developing, duration is used to distinguish focal from non-focal constituents.

At seven to eight years, Dutch-speaking children use accent placement and accent type in similar ways to adults sentence-finally, but they still do not de-accent in the same way as adults sentence-medially. Compared to the five-year-olds, the accent type choices of the eight-year-olds were more adult-like in terms of avoiding non-downstepped accents post-focally. Interestingly, at this stage the children seemed to rely more on accent types for marking focus than adults, avoiding the use of the same accent category ('!H*L') for focus and post-focus, something we found to be fairly common in our adult data. In addition, our seven- to eight year-olds also produced a larger proportion of post-focal rising accents ('L*H') than the adults.

Despite the pattern of using non-dowstepped accents for focus and downstepped accents for post-focus sentence-medially, these manipulations were not systematic enough to show up as significant effects of focus on pitch in this group, suggesting that the eight-year-olds are still developing toward adult proficiency in the use of pitch for focus in this position. Somewhat surprisingly, we also did not find any significant relationships between focus and pitch in sentence-final position for this group, even if in this position, the accentuation patterns of the eight-year-olds resembled those of the other groups.

Finally, at ten to eleven, Dutch children seem to master the use of accent placement for focus both sentence-finally and sentence-medially. In fact, we often observed even sharper differentiation between focus and post-focus in the eleven-year-olds than in the adults, both in terms of accent placement and in terms of accent type choices. This was also reflected in our phonetic analyses, where focal and post-focal target words were differentiated in both maximum pitch, pitch range and word duration sentence medially, and in maximum pitch and pitch range sentence-finally.

Our analysis of accent placement across focus types showed all our groups to accent close to ceiling for both broad, narrow, and contrastive focus sentence-medially, and for broad and narrow focus sentence-finally. These results suggest that the use of pre-nuclear accents on sentence-medial verbs in broad focus is common in Dutch, counter to previous descriptions of verbs being less likely to carry accents than arguments in broadly focal SVO-sentences (e.g. Gussenhoven, 2011). Comparing accent placement across different focus conditions in the way that we have done here may seem rather unintuitive, particularly as the literature suggests that both contrastive, narrow and broad focus would lead to accentuation in adult speakers of Dutch, but that the shape of the accent differs between narrow and contrastive focus on the one hand and broad focus on the other (Hanssen et al., 2008). At the same time, asking whether children also accent across focus types is another issue, and our analysis has shown that they do.

Typically, focus type effects are analysed by comparing phonetic measures on nuclearly accented (preferably with the same accent type) target words across broad, narrow and contrastive focus. The analyses are typically performed on read speech collected in the lab, and care is taken to ensure that the target word is the word that is expected to carry the nuclear accent across all three conditions (e.g. Hanssen et al., 2008; Baumann et al., 2007; Myrberg, 2013). In our setup, the only location where narrow, contrastive and broad focus was elicited on the same constituent was the sentence-medial position, but what is expected in this position under broad focus is a pre-nuclear accent (if any), as the nuclear accent is expected to be on the sentence-final object. Phonetically comparing pre-nuclear accents in broad focus to

nuclear accents on narrow focus would not serve to isolate the effect of nuclear/pre-nuclear accentuation from the effect of focus type. Whereas our setup does not make it impossible to compare the focus type conditions (as shown by us doing so for our accent placement measure), it does make the data less ideal for focus type comparisons, at least if our measures are to be compared to previous work. Additionally, considering the size of our dataset, phonetic differences between our focus type conditions are not very likely to surface, and if they do, it is questionable how generalizable these findings would be. Finally, studies on the phonetics of accentuation typically compare phonetic manipulations on one particular accent type. As the choice of accent types for focus varied considerably in our dataset, limiting our analysis to only one accent category would entail throwing out a large portion of our data.

In this study, differences found between adults and children were primarily related to post-focal manipulations, with less de-accentuation in younger than older children, and with more non-adult accent type choices post-focally than focally. This is in line with earlier descriptions of younger children being better at accenting focal than at de-accenting given information (e.g. Chen & Fikkert, 2007; Snow & Balog, 2002; Grünloh, Lieven & Tomasello, 2015), something Grünloh and colleagues attribute to a lower proportion of de-accentuation in infant directed speech (IDS) (Grünloh, Lieven and Tomasello, 2015, see also Bortfeld & Morgan, 2010). The children included in this study are older than the children that have been found to fail at de-accenting in previous studies, and the speech our participants are surrounded with is presumably not IDS. At the same time, our adult data suggests that de-accentuation of post-focal material is less common than what is typically assumed in the literature, even for non-IDS speech. The fact that our children do not de-accent post-focally may simply be because this pattern is not the only option in adults either; in fact, sentence-medially, accentuation was more common than de-accentuation, even for post-focal targets. A possible explanation why we find less de-accentuation in our adults than what is typically reported may be that our method elicits more spontaneous-like speech than what is the case for experiments using scripted or read speech (see de Ruiter, 2010; Breen et al., 2010, for related discussions). More work is clearly needed in order to establish the consistency with which adult speakers of Dutch de-accent post-focal information, as well as how different speech styles affect de-accentuation patterns.

The results from our five and eight-year-olds suggest that adult mastery of prosodic focus marking is reached earlier for the sentence-final than for the sentence-medial position. This is in line with previous work from Chen (2011a), comparing prosodic focus marking between the sentence-initial and sentence-final position. It is perhaps not so surprising that accentuation for focus is mastered earlier in this position than

in other positions, as accenting in sentence-final position can be considered the default pattern in Dutch. This said, what distinguishes focal from post-focal targets in this position is that post-focal ones are more likely to be de-accented, thus what the children master better finally than medially is the avoidance of accents when a target follows the focus. Other possible reasons why the sentence-final position is mastered earlier than the sentence-medial one is that the sentence-medial position may require more prosodic integration (e.g. the scaling of tonal targets according to adjacent tonal events on either side) than the sentence-final position, making de-accenting here more demanding from a production point of view. Another possibility is that the relative inconsistency with which adults de-accent sentence-medially makes it hard for children to discover that de-accentuation is actually an option. It is also possible that the sentence-final position is generally more salient from a perception point of view, and that children therefore detect the pattern for focus marking earlier in this position than for sentence-internal ones. The fact that a number of prosodic contrasts are marked sentence-finally (e.g. interrogativity, turn-taking) may make this position particularly important from a prosodic point of view, possibly making children attend more to patterns occurring sentence-finally than patterns occurring in other positions. Finally, in our setup, sentence position co-varied with grammatical category, thus our position effect could also be an effect of grammatical category, as our medial targets were always verbs and our final targets always objects. As suggested by Röhr, Baumann & Grice (2015), it may be more common to mark focus on referents than on actions. Such a tendency would give children less experience with marking focus on nouns than on verbs, and might explain why our children mark focus more reliably on the sentence-final nouns than on the sentence-medial verbs. Expanding work on prosodic focus marking to languages with different word orders might help in disentangling position effects on the one hand and effects of grammatical category on the other (see e.g. Yang et al., 2015; Yang, forthcoming, on the acquisition of prosodic focus in Korean).

Studies of the prosodic marking of information structure in children have typically relied on either a phonological (e.g. manual coding of accent patterns) or a phonetic (e.g. comparing measures of pitch and duration) approach to the phenomena in question. Whereas some studies report on similar patterns in children and adults when phonetic measures of pitch and duration are compared (Müller et al. 2006; Wonnacott & Watson, 2008), others have focused on accentuation patterns, often reporting non-adult accent choices well beyond this age (e.g. Wells, Peppé & Goulandris, 2004; Chen, 2011a and references therein). In the current study we combined different measures, with the benefit of getting more information about the kinds of manipulations the children make, also in cases where the children adapt non-adult strategies when differentiating focal from post-focal constituents. An important methodological insight from our analyses is that the choice of output

measure has important consequences for the conclusions that are drawn: Whereas our accent placement analysis shows that four- to five-year-olds do not distinguish focal from post-focal targets at all in sentence-medial position, our analysis of accent types, together with our analysis of pitch and duration, shows that this group does resemble adults in using downstepped falls post-focally, and that increased word duration was systematically associated with focus in this group. Similarly, describing the proficiency of the seven- to eight-year-olds based only on their accent placement would mean ignoring their rather systematic use of accent types for focus, where they are actually more consistent than the adults in reserving non-downstepped falls for focus and downstepped falls for post-focus. This pattern resembles previous descriptions from Chen (2009) on sentence-initial focus marking, suggesting that children between seven and eight go through a phase of using accent type contrasts for marking focus. Methodologically thus, performing a wider set of analyses than those typically employed has the benefit of capturing more information about the children's prosodic development than when purely relying on one measure, making it possible to see children's development more as a gradual process than as a question of mastery or not. In addition to giving a more complete picture of the children's manipulations, our combined analysis shows us that there is no one-to-one mapping between accent placement and measures of pitch maximum, minimum and range, and that accentuation versus de-accentuation need not always have consistent effects on these measurements, particularly when a small dataset is considered. Nonetheless, there is also correspondence between our different measures, seen for example in that the sharper differentiation between focus and post-focus found in the accentuation patterns of our eleven-year-olds rendering more systematic effects of focus on pitch and duration in our phonetic analysis.

A few words should be added regarding the lack of duration effects of focus in our adult data, which contrasts with previous findings of a strong relationship between focal accentuation and duration in Dutch (Sluijter & Van Heuven, 1995, 1996; Hanssen, Peters & Gussenhoven, 2008), but that also partly confirms the findings from Cambier-Langeveld & Turk (1999), that accentual lengthening does not take place sentence-finally in Dutch. Whereas studies describing reliable lengthening effects of focus, or accentuation, or both, are all conducted in the lab, using rather controlled speech material, our sentences contained a number of different target words, and they were elicited in a more interactional setting. It is possible that increased duration is less reliable as a cue to focus in more spontaneous speech styles than pitch cues are, as spontaneous speech may be more temporally variable than more controlled speech (e.g. Laan, 1997). It is also possible that focus-associated lengthening is somehow less robust in Dutch than in Swedish, as shown by the consistent effect of focus on word duration reported in Chapter 6. At the same

time, the rather inconsistent use of de-accentuation found among our Dutch adults may also mean that systematic effects of accentuation on duration are harder to find in an analysis that included both accented and de-accented target words, particularly in a small dataset like ours. Before concluding, the effect of focus on word duration in Dutch needs to be investigated in a larger dataset, preferably eliciting spontaneous-like data, in order to establish how reliably duration is correlated with focal accentuation in Dutch.

7 Quiet is the new loud: Pausing and focus in child and adult Dutch

7.1 Introduction

In this chapter we present a study on how Dutch-speaking children between four and eleven manipulate pause duration to mark focus.[37] Even if pausing is not traditionally seen as a cue to focus in Dutch, several previous studies have pointed to a possible relationship between pausing and information structure in adult speech production (Dahan & Bernard, 1996; Gee & Grosjean, 1984; Gu & Lee, 2007; Huang & Liao, 2002; Swerts & Geluykens, 1994). Considering that Dutch-speaking four- to five-year-olds are not yet completely proficient in using accentuation for focus (Chen, 2011a, 2011b; Romøren & Chen, 2014, see also Chapter 6), we asked whether pausing could be an available parameter for Dutch children at this age to manipulate for focus.

Speakers pause for various reasons, ranging from speech planning demands and metrical considerations to pragmatic purposes (Ferreira, 2007; Wagner & Watson, 2010; Zellner, 1994). Among pragmatic reasons, speakers pause longer before sentences containing new information (Gee & Grosjean, 1984), when initiating new topics (Swerts & Geluykens, 1994) or when highlighting words or phrases (Dahan & Bernard, 1996; Gu & Lee, 2007; Huang & Liao, 2002). Given that young children produce more between-word silent pauses than adults (e.g. Redford, 2013), we asked whether children systematically manipulate the location and duration of such pauses for focus-marking purposes. As pausing for focus has already been described in adults, we also wanted to know whether differences in pausing patterns could be observed between adults and children performing the same task.

The chapter is structured as follows. In the background section (7.2) previous work on pausing and information structure are presented (7.2.1), followed by a review of

[37] A previous version of this chapter has been published as Romøren & Chen (2015a).

previous work on focus in child and adult Dutch (Section 7.2.2). In Section 7.3 we present our research questions, and in Section 7.4 we present our methodology. In Section 7.5 we present our analyses and results, and in Section 7.6 we discuss our findings, and what they tell us concerning the relationship between focus and pausing in child and adult Dutch.

7.2 Background

7.2.1 Pausing and information structure

Theories of 'information structure' or 'information packaging' (e.g. Chafe, 1976; Halliday, 1967) are concerned with the various manners in which speakers *package* the information they wish to communicate according to the knowledge state of the listener, or more precisely, to the common ground shared between speaker and listener. As in the previous chapters, the information structural category investigated here is focus, and sentences with varying focus structure are elicited by means of wh-questions (see Chapter 2 for a more elaborate discussion of focus and information structure). To the best of our knowledge, the use of pausing for focus has not been systematically investigated in children's speech production, despite reports that information structure does affect pausing patterns in adult speech. Below we will start by reviewing work on the relationship between pausing and focus in adult speech, before returning to the topic of pausing in child speech.

Pausing in adult speech production has been a popular topic in the last 60 or so years (see Ferreira, 2007; Wagner & Watson, 2010; Zellner, 1994, for reviews). Particularly relevant for our study are reports that speakers tend to pause longer when adding new information to a narrative (Gee & Grosjean, 1984), when adding new information in instruction monologues (Swerts & Geluykens, 1994) and when highlighting certain information within sentences (Dahan & Bernard, 1996; Gu & Lee, 2007; Huang & Liao, 2002). These reports are made based on both spontaneous and more controlled speech styles. The finding that adults pause to single out new information in a larger discourse already suggests a potential link between information status and pausing. Nevertheless, the papers on pausing for within-sentence highlighting are particularly interesting in light of the current study, and will therefore be described in more detail below.

The first paper to be discussed comes from Dahan & Bernhard (1996), who used a reading task to investigate acoustic manifestations of emphasis in four adult speakers of French. Emphasis was implemented through asking the speakers to

'insist' on underlined target words in the emphatic condition; frequencies and durations of pauses preceding and following the target words were extracted and compared between 'emphatic' and 'not emphatic' conditions. The authors did not specify what they defined as pauses, but one might assume that they were silences of a certain dB, and that a certain durational threshold was applied. Although the pause frequencies (e.g. the number of pauses observed preceding the target) only increased in the 'emphatic' condition in one speaker, emphasis made the durations of pre-target pauses significantly longer in three out of the four speakers. Interestingly, in a follow-up perception study, the pre-target pauses were found to contribute significantly to perceived emphasis, suggesting that listeners also treat such pauses as meaningful cues.

Similar findings are reported by Gu & Lee (2007) for Cantonese. In their study, pre-target pauses were significantly longer before focal targets than before 'neutrally produced' targets. Focus was operationalized by using questions to elicit contrastive focus on target non-words within a fixed sentence frame. As found for one of the speakers in Dahan & Bernhard's study, Gu & Lee also reported on pauses being inserted before the focal constituent. Finally, Huang & Liao (2002) similarly postulated that pauses could be used for highlighting certain constituents in Mandarin Chinese.

In the studies by Dahan & Bernhard (1996) and Gu & Lee (2007), the pre-target pauses occasionally co-occurred with plosive word-onsets. The authors therefore suggested that the effect of emphasis or focus might be articulatorily based, in that focus led to lengthening of the silent part of a plosive, but only one speaker inserted pauses before focal words regardless of whether this word started with a plosive or not. While it is true that pausing was confounded with plosive closures in these investigations, other researchers have warned against using too strict thresholds when investigating pausing phenomena in speech. According to Hieke, Kowal & O'Connell (1983), stop-closures of consonants can vary between 80 and 250 ms (as shown by Dalton & Hardcastle, 1977), making it hard to establish an unambiguous cut-off point where pauses can no longer be attributed to articulatory processes. Resorting to perceptual arguments to justify duration thresholds is equally vulnerable, as the percievability of a pause varies substantially depending on the speech context in which it appears (Rochester, 1975). Investigating the origins of shorter pauses in read-out poems and political speeches, Hieke and colleagues (1983) found that most pauses ranging between 130 ms and 250 ms were attributable to effects such as emphasis, segmentation, or punctuation, rather than articulatory processes. Following these findings, the authors concluded that

dismissing pauses within this time range on articulatory grounds might lead to interesting patterns being ignored. [38] In a more recent cross-linguistic study, Campione & Veronis (2002) reached a similar conclusion. They extracted pause durations from a corpus of read and spontaneous speech in five languages, showing how a simple comparison between spontaneous and read speech could lead to completely different conclusions depending on the threshold applied (Campione & Veronis, 2002). It is interesting to note that most of the previous studies reporting relationships between focus and pausing are conducted using fairly controlled speech material, predominantly read speech. Conversely, our data was elicited as part of a game, allowing for investigating whether pause patterns observed in controlled speech samples are also attested in more spontaneous-like speech.

Whereas pausing has received considerable attention in research on adult speech, pausing in child language has been much less studied (see Sabin, Clemmer, O'Connell, & Kowal, 1979, for a review of early studies). This can partly be explained by the prevalence of traditional competence-based approaches to acquisition, in which pausing and disfluencies are assumed to be irrelevant for describing children's linguistic knowledge (e.g. Wijnen, 1990). To the best of our knowledge there are no previous systematic investigations of pausing and information structure in children. However, in a recent study, Redford (2013) speculates on a possible link between newness of information and pausing. Using a narrative-task, she compared pausing patterns of five-year-olds to those of adults. In addition to the finding that pauses were generally longer and more frequent in the children's speech, she also found a comparatively larger number of 'ungrammatical pauses' in the children's utterances (defined as pausing after a determiner, conjunction, or copula, or between an auxiliary and a verb, between a transitive verb and its direct object, or between a preposition and a following noun phrase). Redford suggested that the children's 'ungrammatical' pauses preceding focal elements might be wrongly categorized as such, as the pauses could in fact be there for 'prosodic purposes' (e.g. *and then he fell into…the lake!*). Tentatively, we interpret these prosodic purposes along the line of pausing to emphasize upcoming information. The fact that 7 % of the pauses produced by the adults were also found

[38] Hieke et al. did not examine pauses shorter than 130 ms. The 130 ms minimum applied in their study was justified by making reference to Butcher (1981), who claims most pauses of this kind to be caused by "(...) prolonged articulatory closures", and that they " (...) create measurement problems in both manual and automatic methods of analysis" (Butcher, 1981:48).

in 'ungrammatical locations' might suggest that adults also pause to emphasize in English, as reported for French and Chinese.[39]

In a related study, Maloney, Payne and Redford (2012) addressed the question of whether pause durations are correlated with the strength of syntactic boundaries. They hypothesised that pauses would increase in length from weaker boundaries (e.g. between a determiner and the head noun) through stronger ones (e.g. between the head verb and the NP which it dominates) to the strongest ones (between the subject NP and the VP). Narratives were elicited from five-year-olds, seven-year-olds, and adults, and pauses were measured following the same procedure as in Redford (2013). The three groups were similar in pausing the least at the weakest boundary (i.e. determiner-head) and the most at the strongest boundary (i.e. subject-VP), but they behaved differently at the medium-strength boundary between the head verb and its argument NP, where the children paused much more often than the adults. Maloney and colleagues also considered information structure as a possible explanation for the children's pausing patterns. Following Chafe's (1987) suggestion that speakers tend to plan and produce phrases that contain maximally one piece of new information, Maloney and colleagues suggested that the pauses occurring between the verb and its argument might be triggered by both constituents containing new information (e.g. not previously mentally activated), causing the children to divide them into two phrases through pausing.

As we have seen, several studies on pausing in adult and child speech point to a potential relationship between pausing and information structure. However, in the studies of adult speech, pauses are mostly included as one out of several dependent variables investigated, and the finding that pausing might play a role in marking focus is granted relatively little attention. In the two studies on pauses in child speech, information structure is presented as a possible interpretation of the pausing patterns observed, but without this being empirically investigated. In addition, in the latter two studies, a relatively high threshold for pausing was applied where the pauses co-occurred with plosive closures, despite the fact that this approach runs the risk of dismissing psychologically relevant pauses on somewhat arbitrary grounds (Campione & Véronis, 2002; Hieke et al., 1983).

---

[39] Careful measures were taken to avoid including plosive closures in the pause measurements. For details, we refer the reader to the original paper (Redford, 2013).

7.2.2 Prosodic focus marking in adult and child Dutch

In West-Germanic languages, focus is predominantly marked using prosody, whereby focal information is accented and post-focal information is de-accented (e.g. Gussenhoven, 2004; Ladd, 2008; Féry, 2013, see also Chapter 2.3). Previous work on Dutch, together with our own work presented in Chapter 6, shows that Dutch-speaking four- to five-year-olds accent focal information and de-accent post-focal information at the age of four or five, in line with what is described for adults (see also Chen, 2011a, 2011b). At the same time, we have seen that Dutch-speaking children at this age do not yet show the adult preference for falls for marking focus sentence-finally (Chen, 2011a), that they do not yet use phonetic manipulations for distinguishing between focal and non-focal accents sentence-initially (Chen, 2009), and that they do not yet de-accent post-focal information sentence-medially (see Chapter 6). Finally, children are less consistent in marking focus in SVOA (subject-verb-object-adverbial) sentences than they are in SVO (subject-verb-object) sentences (Romøren & Chen, 2014). In other words, studying the use of pausing in children between four and five allows for determining whether these children, who are not yet adult-like in their use of the canonical cues to focus used by adults, make use of pausing as a compensatory strategy.

7.3 Research questions

As we have seen, Dutch-speaking children can make some prosodic adjustments to mark focus at the age of four or five, but their proficiency is still developing beyond this age, especially regarding accent type preferences, phonetic cues when accent placement does not suffice to mark focus, and the use of prosodic manipulations sentence-medially (Chen, 2009, 2011a, 2011b, Romøren & Chen, 2014). Against this background, we asked whether pausing might be an additional parameter available for children to use for focus marking. Our research questions were thus:

1. Do Dutch-speaking four- to five-year-olds insert or lengthen silences before narrowly focal constituents as compared to non-focal ones?

2. Do Dutch-speaking adults insert or lengthen silences before narrowly focal constituents as compared to non-focal ones?

3. Do Dutch-speaking four- to five-year-olds differ from Dutch-speaking adults in the way they use pausing to mark narrowly focal versus non-focal target words?

7.4 Method

7.4.1 Participants

Ten Dutch-speaking children (6 boys, 4 girls, range: 4;4-4;11, mean 5;2) and 9 female Dutch-speaking adults (mean 23;10) participated in the study.[40] All the participants were native speakers of Dutch without any history of language disorders, hearing problems or other known developmental disorders. The children were recruited from primary schools around the city of Utrecht, and their parents gave written consent for them to be tested and for their speech to be recorded. The adult participants were recruited from the participant pool of the Linguistics Lab at Utrecht Institute of Linguistics. They were all university students, but none of them were studying linguistics at the time of testing.

7.4.2 Procedure and materials

All the participants were recorded individually in a quiet room; the children in a designated test room at their school and the adults in a sound attenuated booth at the Linguistics Lab at Utrecht University. Two female experimenters were trained to carry out the testing according to a detailed instruction, and all sessions were video recorded to control for consistency across sessions. The audio recordings were made using a portable ZOOM H1 handy recorder, with a 44.1 kHz sampling rate and 16-bit accuracy. The procedure for eliciting sentence-productions with varying information structure was the same as described in Chapter 3, but different from the datasets included in Chapters 4, 5 and 6, as the dataset included here included both subject-verb-object (hereafter SVO) and subject-verb-object-adverbial (hereafter SVOA) sentences.

The choice to include both SVO and SVOA sentences in this study was made in order to investigate whether pausing patterns we might find in SVO sentences would also be generalizable to a more complex sentence structure. Further, in a previous study, we found Dutch-speaking children to be less consistent in accenting focal

---

[40] Eight of the children and all of the adults included in this study were also included in the analyses presented in Chapter 6.

constituents in SVOA sentences than in SVO sentences (Romøren & Chen, 2014). If children exploit pausing for focus more in cases where they are less proficient in their use of canonical cues to focus, one would predict more use of pausing for this purpose in SVOA than in SVO sentences. Additional reasons for choosing to elicit SVOA sentences were that they lie well within the syntactic complexity four- to five-year-olds can handle, that they were easy to construct and illustrate using child-friendly words, and that such constructions could be easily integrated into the picture-game, following the same structure that was used for the SVO sentences.

The picture-matching game elicited of 30 test trials and 10 practice trials, divided into an SVO part and an SVOA part, where trials pertaining to each part were kept together. In each part, the test trials were spread over five sentence conditions, namely narrow focus on the initial constituent (narrow initial), narrow focus on the medial constituent (narrow medial), narrow focus on the final constituent (narrow final) and contrastive focus on the medial constituent (contrastive medial) and broad focus on the whole sentence (broad focus) (see Tables 48 and 49). The SVO and SVOA parts were each preceded by five practice trials, one from each sentence condition. In the analysis presented here, the broad focus condition was excluded from the analysis.

Sentence condition	**Example context/ question**
Narrow initial	Look, the flower! It looks like someone is hiding the flower. Who is hiding the flower?
Narrow medial	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?
Narrow final	Look, the frog. It looks like the frog is hiding something. What is the frog hiding?
Contrastive medial	Look, the flower! And there is also a frog. It looks like the frog is doing something with the flower. I'll guess: The frog is painting the flower. (What do you say?)
Broad focus	Look! A ghost picture, I can't see anything at all. What is happening in the picture?

Table 47: Examples of context and questions implementing the five sentence conditions, SVO (repeated)

Sentence condition	Example context/ question
Narrow initial	Look, the flower! And there is also a basket. It looks like someone is throwing the flower into the basket. Who is throwing the flower into the basket?
Narrow medial	Look, the baker! And there is also a basket. It looks like the baker is throwing something into the basket. What is the baker throwing into the basket?
Narrow final	Look, the flower! And there is also a baker. It looks like the baker is throwing the flower into something. Where is the baker throwing the flower?
Contrastive medial	Look, the baker! And there is also a basket. It looks like the baker is throwing something into the basket. I'll guess: The girl is throwing the cake into the basket. (What do you say?)
Broad focus	Look! A ghost picture, I can't see anything at all. What is happening in the picture?

Table 48: Examples of context and questions implementing the five sentence conditions, SVOA

Within the experimental trials, six medial and six final target constituents were carefully distributed over the four conditions so that each medial and final target occurred once in every condition. We also spread five initial constituents over the four conditions. When creating and ordering the stimuli we ensured that each combination of initial, medial and final constituent only occurred once in the whole set. Further, two consecutive trials never realized the same condition and always differed by minimally two constituents. Following these constraints, the experimental trials were arranged into two different stimulus orders. Since we also randomized the order of the SVO and SVOA sets, this left us with a total of four trial orders, to which the participants were randomly assigned.

7.4.3 Data selection and coding

Each test session resulted in a 20-40 minute long recording, which was segmented into trials using Praat (Boersma & Weenink, 2010). The responses to the experimenter's questions were then evaluated, and only responses following the scripted speech context were included in the analysis. Responses were also excluded

if they contained deviant word orders, deviant word choices or elided constituents, as well as self-repairs, stuttering, filled pauses, or background noise. This selection was made in order to ensure that the prosodic comparisons were made across the same words or phrases, and that the experimental conditions were properly controlled for. Further, since we needed the word boundaries (at which the pauses were measured) to be the same for all responses, we did not include non-target sentences (e.g. sentences that did not contain the words presented in the naming task and introduction in SVO or SVOA order). Among the excluded responses from the children, 33 were excluded because the speech context could not be completely controlled, 35 were excluded because they contained filled pauses, stuttering or repairs, and 30 were excluded because they contained the wrong words, lacked certain constituents or had non-target constituents added to them. Finally, 8 responses were excluded because of laughter, background noise or other disturbances making the recordings unfit for analysis. The final dataset from both groups consisted in 188 SVO sentences and 176 SVOA sentences. The average response inclusion rate was 65% (range 40% - 87%) in the children, and 92,2% (range 83% - 100%) in the adults.

The included responses were orthographically transcribed and segmented into words using Praat. When segmenting, we relied on changes in the waveform in addition to the formant transitions shown in the spectrogram (Turk, Nakai, & Sugahara, 2006). Conventions were established for how to segment the words at particularly challenging boundaries (e.g. onset plosives were segmented right before the burst).

For our measure of pauses, a pause was defined as a between-word interval of any duration with no or insignificant amplitude.[41] Pauses were coded by combining the automatic silence detection function from Praat (minimum silence threshold 25db, minimal silence duration 20 ms) with manual visual inspection. In the manual checking of the automatically detected silences, between-word silences shorter than 20 ms were also included when observed, but such short pauses were very rare. Since this definition of pausing meant that closures of unvoiced plosives (where the beginning of the closure has no acoustic trace) were counted as pauses, we decided to also include the pre-burst part of voiced plosives as pauses. As discussed in the introduction, the use of arbitrary thresholds for pausing runs the risk of leaving out potentially relevant data. As this was an exploratory study, we decided not to separate plosive-induced between-word silences from silences that did not co-occur with plosives.

---

[41] Sometimes there was background noise or breathing noises, giving rise to some minor energy distributions in the spectrogram.

We investigated pause durations related to sentence-medial and sentence-final target constituents. The between-word boundaries where pauses were measured are illustrated in Figures 34 and 35 below. In the SVO sentences the medial targets were verbs, and the final targets were object NPs. In the SVOA sentences the medial targets were object NPs, and the final targets were adverbial PPs. In the illustration, squares illustrate medial and final target constituents. Further, large brackets mark larger constituents (e.g. NP or PP), whereas smaller brackets mark words.

In our analysis of the SVO-sentences, our medial analysis involved comparing pauses preceding the verb, whereas our final analysis involved comparing pauses at the boundary between the verb and the object NP, as well as at the boundary between the determiner and the noun within the object NP. In the SVOA-sentences, the medial analysis involved comparing pauses at the boundary between the verb and the object NP, as well as between the determiner and the noun within the object NP. The final analysis involved comparing pauses at the boundary between the object NP and the adverbial PP, as well as between the preposition and the NP, and between the determiner and the noun, within the PP.

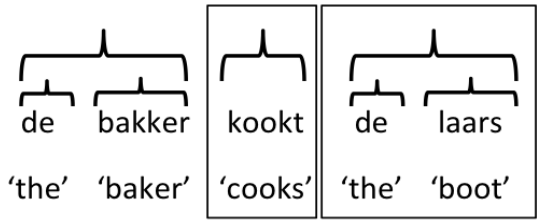

Figure 34: Target constituents and potential pause locations, SVO.

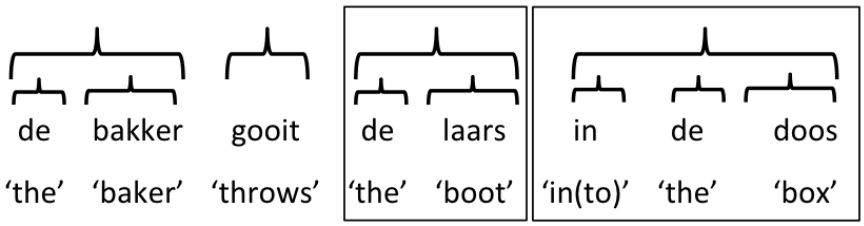

Figure 35: Target constituents and potential pause locations, SVOA

In the analysis, each word boundary was given a designated number, and pause coding was based on these numbers, so that each potential between-word pause location carried a unique label. Pause durations were extracted using a Praat script,

and samples were taken from the output file to check for tracking and measuring errors.

7.5 Analysis and results

7.5.1 Introduction

Previous investigations of prosodic focus marking in adult speech have revealed only subtle differences between contrastive and narrow focus in adult Dutch (Hanssen et al., 2008). Similarly, we found no significant differences in pause durations either before or within target phrases, when comparing the 'contrastive medial' and the 'narrow medial' condition. Based on these results we decided to collapse these conditions in the rest of the analysis, in order to include as many data points as possible. Our 'no focus' condition contained all of our sentence conditions that did *not* render a specific target constituent focal, e.g. 'narrow initial' and 'narrow final' for medial comparisons and 'narrow initial', 'contrastive medial', and 'narrow medial' for final comparisons. We also ran separate analyses to check for differences between the conditions collapsed in the no-focus condition, and there were no significant differences in pause durations either before or within target phrases when comparing across these.

Linear Mixed Effect Modelling (LMM) in R (R Core Team, 2014) were used to assess the effect of focus on pause durations before and within medial and final target constituents, with the factors 'focus' ('focus' vs. 'no focus') and 'group' ('four- to five-year-olds' vs. 'adults') as fixed factors and 'participant' and 'item number' as crossed random factors. The analysis procedure was the same as the one described in Chapter 4.5.4, whereby factors were added in a step-wise fashion in order to establish the best fitting model.

In cases where the interaction between 'focus' and 'group' significantly improved the model fit, new models were built for each group separately, to explore whether the interaction was caused by a difference in the degree to which focus influenced pause duration between the groups, or in the absence of any effect of focus in one of the groups. All analyses were carried out separately for SVO and SVOA sentences, as the boundaries at which pausing could take place preceding medial and final targets differed between the two sentence types, and corresponded to different kinds of syntactic junctures (see Figures 34 and 35). Below we first report the results from the analysis of the SVO sentences, and then present the results from the SVOA sentences.

7.5.2 Subject-verb-object- sentences (SVO)

For the SVO sentences, models were built for pauses at three different locations; preceding the verb, preceding the object NP, and within the object NP (Figure 34). With respect to the boundary preceding the verbs, both 'focus' (p=.034), 'group' (p=.005) and the interaction between 'focus' and 'group' (p=.049) improved the model fit. Re-running the models on the data split by group showed that narrow focus on the medial target significantly increased the pause duration preceding this target in both adults and children (children: p=.054, adults: p=.000), but the increase was larger in the children than in the adults. Mean pause durations split by group and focus conditions are presented in Figure 36.

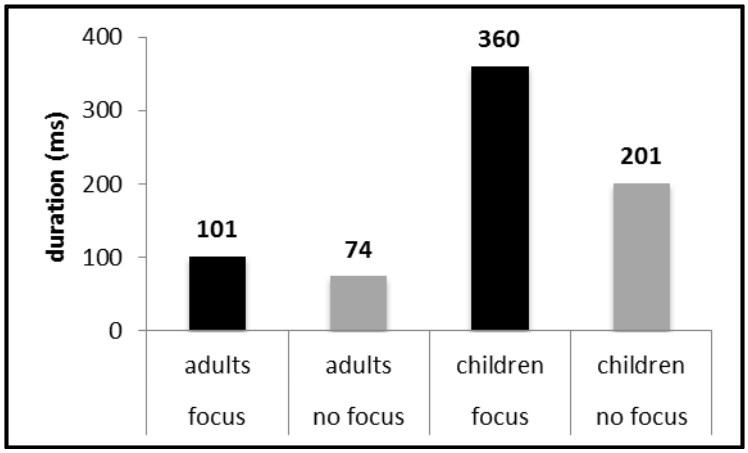

Figure 36: Duration of pre-medial pauses (before verb) by focus and group, SVO

Pauses in sentence-final position were analysed both on the boundary *before* the final object NP and on the boundary between the determiner and the noun *within* this NP. Neither 'focus', 'group', or the interaction between 'focus' and 'group' turned out to be significant predictors for pause durations in these locations, thus neither the children nor the adults varied pause durations according to focus at these boundaries.

7.5.3 Subject-verb-object-adverbial sentences (SVOA)

The medial analyses of the SVOA sentences concerned both pause durations preceding the medial object NPs and pause durations preceding the final noun *within* these NPs. The final analysis involved comparisons within three different pause locations, the one preceding the whole PP, the one preceding the NP within the PP, and the one preceding the final noun of the PP (see Figure 35).

With respect to the pause durations preceding the medial object NPs (Figure 37), adding 'focus' to the baseline model did not significantly improve it (p=.443). However, both 'group' and the interaction between 'focus' and 'group' significantly improved the model (group: p=.010, group × focus: p=.042). Re-running the models split by group revealed a significant effect of focus in the children's data (p=.011), but no such effect in the adult data (p=.352).

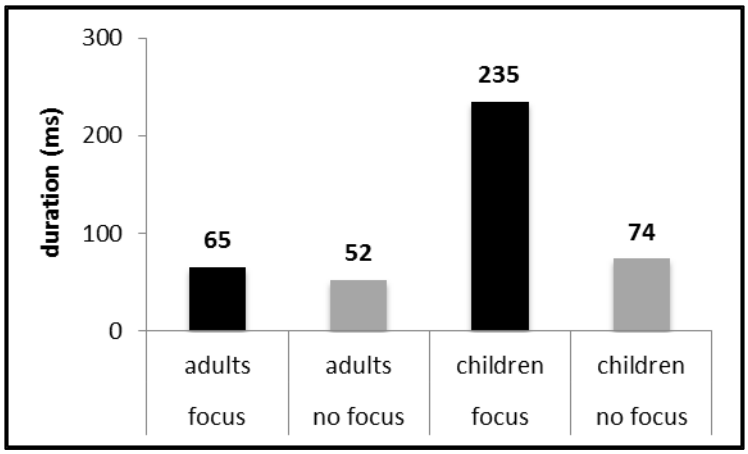

Figure 37. Duration of pre-medial pauses (before object NP) by focus and group, SVOA

The analysis of pause durations before the final target PP revealed a main effect of 'focus' (p=.051), a main effect of 'group' (.000) and an interaction effect between 'focus' and 'group' (p=.005) (see Figure 38). Our follow-up analysis split by group showed a significant effect of focus on the pre-PP pauses in both groups, but that the effect was stronger in the children than in the adults.

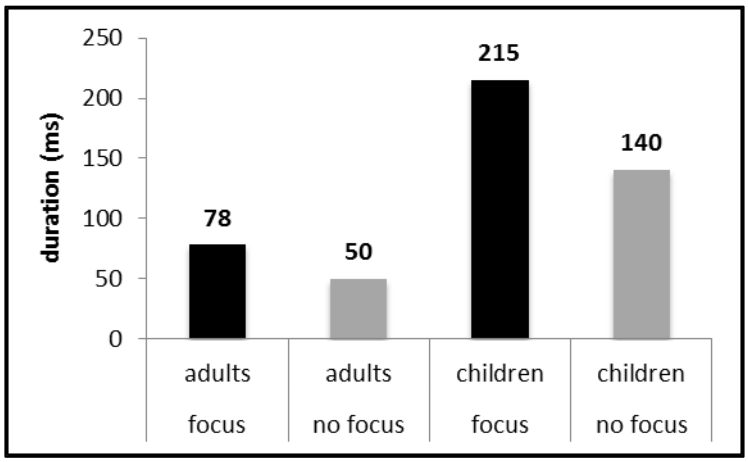

Figure 38. Duration of pre-final pauses (before adverbial PP), SVOA

We also ran models examining the effect of focus on the pause preceding the NP within the PP, but there were no significant effects either of 'focus' (p=.257), 'group' (p=.345) or the interaction between 'focus' and 'group' (p= .201) for this position. However, preceding the final noun *within* the PP, main effects of 'focus' (p= .030) and 'group' (p= .019), were found, but the interaction between 'focus' and 'group' did not reach significance (p=.139) (see Figure 39).

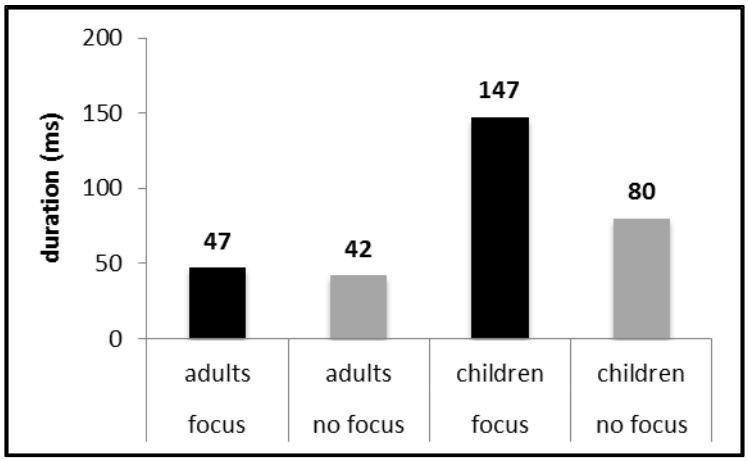

Figure 39: Duration of within-final pauses (before the final noun of the PP), SVOA

Both groups paused much less frequently in this location than they did in the other ones, but when a pause *was* observed it was systematically longer when the PP was focal then when it was not. Even if we see in Figure 39 that the effect of focus is hardly present in the adult data, the interaction between focus and group did not reach significance for this location.

7.6 Discussion

A general observation from our data is that the children paused longer and in more locations than the adults, similar to the findings from Redford (2013) and Maloney et al. (2012). In terms of pausing mediated by focus, both groups paused systematically longer before focal verbs in the SVO sentences, and before focal PPs in the SVOA sentences, as compared to their non-focal counterparts. This pattern is in line with what was reported by Dahan & Bernhard (1996), Gu & Lee (2007) and Huang & Liao (2002). In contrast to the children, the adults tended to avoid pausing at weaker syntactic junctures, (e.g. at the boundary between the verb and its internal argument), which is similar to the findings reported by Maloney et al. (2012). Crucially, our results provide empirical evidence for a consistent relationship between pre-target pause duration and focus in both child and adult Dutch, suggesting that pausing may be an available parameter for children to make use of at a stage where their access to pitch and duration cues to focus is still not completely adult-like (e.g. Chen, 2011a, Chen & Romøren, 2014). In terms of our research questions, the answer to our first question, about whether children manipulate pause durations for focus, is positive. Similarly, our question as to whether adults also manipulated pauses for focus is also answered positively, but our adults were more restricted in their use of pausing for focus than the children. This difference also serves to answer our third question about group differences: Differences were observed between children and adults in the extent to which pausing was used for marking focus, where adults were more constrained by syntactic restrictions than the children were.

The focus-mediated pauses between the subject and the VP in the SVO sentences, as well as between the object NP and the adjunct PP in the SVOA sentences, both took place at strong syntactic boundaries, thus pausing in these locations may be seen as more natural than in other locations (e.g. Maloney et al., 2012). The finding that these pauses were systematically lengthened for focus in the adult data suggests that pre-target pauses can be used by adults as an additional phonetic cue to focus, at least in locations where pausing is syntactically appropriate. The location where only the children lengthened pauses for focus (i.e. before the medial object NPs in

the SVOA sentences) was at weaker syntactic boundary, indicating that children are less constrained by syntax than adults when pausing for focus.

The focus-mediated pauses before the final nouns in SVOA stand out from the other pauses observed in our data, as they occurred *within* the target constituent rather than before it (e.g. before the PP). A closer look at the data from this location shows that the children only paused there in about half of the responses, and the adults in about a quarter. Still, in the cases where pauses were observed, they were consistently longer when the PP was focal. Given that the questions eliciting final focus actually contained the relevant preposition in the Dutch version (e.g. *waarin* ('in what'), *waaronder* ('under what'), one might ask whether it is really the case that the whole PP is focal, in the sense that it contains new information, as the preposition is mentioned in the scripted context. Nevertheless, the fact that focus lengthened the pauses before the PP suggests that the participants *did* treat this phrase as a focal constituent. Even if only the NP rather than the whole PP were focal, one might not expect the speakers to pause between the determiner and the noun, but rather before the NP (e.g. Maloney et al., 2012). In the whole dataset we observed remarkably few pauses between the preposition and the determiner, suggesting that there is a general tendency for speakers to keep these items together. One might speculate whether this tendency is caused by a more prosodic type of constraint than the syntactic ones we have discussed so far. The fact that there are languages that merge prepositions and determiners before nouns (e.g. *em* 'in' + *a* 'the (fem.)' is lexicalized as *na* in Portuguese) might suggest that there is some prosodic pressure to keep prepositions and determiners phrased together, which might explain the patterns we observe.

In addition to our hypothesis that the participants make use of the pre-target pauses as an additional cue to focus, two alternative interpretations also merit mentioning here. One is that the pauses measured are primarily segmental, originating from plosive word onsets found in our elicited sentences. All the target verbs and most of the target NPs in our dataset (due to the article *de*) had plosive onsets. However, as the participants often used the indefinite article *een* in their NPs, and as all of the final PPs (which was where both groups lengthened pauses for focus in the SVOA sentences) began with non-plosives, the patterns found in our data cannot be explained by plosive closures alone. Importantly, lengthening a silence already present or inserting a pause where the segmental content of a word does not require one might both result in a silent stretch that, in addition to canonical cues like accentuation, could contribute to the signalling of focus (e.g. Dahan & Bernhard, 1996).

A second alternative interpretation of our findings relates the observed pausing patterns to processing, or more specifically, to lexical access. The speed of lexical access is affected by previous mention (e.g. Ferreira & Hudson, 2011) and this effect could also come into play in our experimental design, as presenting the non-focal items in the trial context could make focal items less primed than non-focal ones. In this way, the longer pauses observed before focal targets could be explained by the focal targets being harder for the participants to retrieve. Still, the rather limited set of target words included in our game should make the lexical accessibility of these generally high, also because they were all introduced in the naming task and the introduction to the game, as well as repeated randomly across trials. Further, the participants had no constraints in terms of the time used between looking at their picture and answering the question, and were thus allowed ample time for planning the response, in contrast to what tends to be the case in priming studies. We therefore conclude that the pre-target pausing found in this study is used as a means to prosodically highlight focal information. More work is needed in order to determine how systematically such patterns occur in spontaneous speech, as well as the extent to which listeners also exploit this use of pausing as a cue to focus in their speech comprehension.

The current study has useful methodological implications for research on pausing. Our choice of avoiding a minimum threshold for pause durations led us to the discovery that pause durations co-vary with information structure. Since the average pause durations found were sometimes relatively short, choosing a cut-off point such as the 250 ms threshold suggested by Goldman-Eisler (1968) would most likely have caused us to miss the patterns we observe. In order to prevent relevant data from being excluded a-priori, we suggest that future research attempts to separate articulatory from linguistically relevant pauses not by applying pre-determined thresholds, but rather by strictly controlling the segmental makeup of the target words included.

Part III:
The closing

8 Conclusion

8.1 Introduction

We start this chapter by summarizing our main findings, sketching the developmental path to prosodic focus marking in Swedish (Section 8.2) and Dutch (Section 8.3), as it emerges from the studies presented in Chapters 4, 5, 6 and 7. In Section 8.4 we will return to the hypotheses presented in Chapter 1.4, considering which of these are supported by our findings, and which ones are not. In Section 8.5 we present a general discussion of our results. In this section, we discuss cross-linguistic factors that seem to shape the development of prosodic focus marking in Swedish and Dutch, also commenting on general developmental tendencies that appear to be shared between children learning these two languages. In the last section (Section 8.6) we suggest some directions for future studies of the development of prosodic focus marking in children.

8.2 The acquisition of prosodic focus marking in Swedish

Starting with the adult model, the analyses presented in Chapter 4 confirm previous descriptions of prosodic focus marking in CS-speaking adults: they predominantly add prominence H to narrowly focal constituents, and they largely avoid this tone on pre- and post-focal constituents (e.g. Bruce, 1977, 1998; Heldner, 2001; Ambrazaitis, 2009; Myrberg, 2009, 2013). Comparing the pitch range between focal words carrying prominence H and post-focal words lacking it showed that whereas adding prominence H to accent 1 words increased the over-all pitch range on that word, adding this tone to accent 2 words did not have such an effect. Additionally, the duration analyses presented in Chapter 5 showed a systematic effect of narrow focus on word duration; focal words were consistently longer than post-focal ones, similar to previous reports from Heldner & Strangert (2001). Lastly, we have also seen that adult speakers of CS tend to use non-modal voice quality on post-focal constituents, a pattern that has not previously been systematically described for Swedish, but that resembles patterns attested in Finnish (Arnhold, 2011; Arnhold & Féry, 2014). With regards to the prosodic differentiation of different focus types, we have seen that adult speakers assign prominence H in parallel ways on contrastively and non-contrastively focal constituents. We have also seen that the probability of

prominence H on both medial and final targets was lower when these were part of broad focus sentences than when they were narrowly focal (c.f. Myrberg, 2013).

The results from our five-year-olds suggest that they are already reasonably proficient in prosodic focus marking at this age; they use prominence H much more on narrowly focal targets than on post or pre-focal targets, adding prominence H increases the pitch range on accent 1 words but not on accent 2 words, and they produce narrowly focal targets with longer duration than post-focal ones. We have also seen that already at this age they do not differ from adults in the way they use prominence H across broad, narrow and contrastive focus. Nevertheless, significant differences were observed between our adults and our five-year-olds in two of our analyses. Firstly, the five-year-olds showed a weaker effect of narrow versus post focus on the use of prominence H sentence-medially than the adult controls did. Secondly, there was no association between post-focal status and voice quality in this group, even if a trend was found sentence-finally, suggesting that the adult pattern of post-focal voice alternations is already developing at this age.

Moving on to the eight-year-olds, we find no differences between this group and our adult controls on any of our measures, except that the effect of post-focal status on the use of non-modal voice was only trending in this group sentence-medially. In other words, by the age of eight, CS-speaking children assign prominence H in similar ways as adults both medially and finally, with similar phonetic consequences for focal as compared to post-focal contours. In addition, CS-speaking eight-year-olds systematically lengthen focal as compared to post-focal target words, and they use non-modal voice quality on post-focal targets in sentence-final position.

The CS-speaking eleven-year-olds also did not differ from the adults in our measures related to prominence H, and they used non-modal voice quality in line with the adults post-focally. However, in Chapter 5 we saw that the eleven-year-olds differed from the adults in their use of duration for focus, showing no effect of focus on duration in accent 2 words sentence-medially, and a reversed effect of focus on word duration sentence-finally. We will return to these unexpected results in the general discussion (Section 8.5).

Summarizing the developmental path to prosodic focus marking in CS it appears that the use of both prominence H and word duration is close to adult-like already at four to five years, and completely adult-like by seven to eight years. Further, it seems that CS-speaking children reach adult proficiency in the use of duration manipulations earlier than they reach adult proficiency in the use of pitch-based manipulations (i.e. prominence H). Finally, a significant relationship between non-modal voice quality and post-focal status was found in our adults and eleven-year-olds. This use of non-modal voice develops rather gradually between five and eight

years, with adult patterns attested earlier in the sentence-final position than in the sentence-medial position.

8.3 The acquisition of prosodic focus marking in Dutch

Starting again with the adult model, we have seen that our Dutch-speaking adults were less consistent in avoiding accentuation post-focally than has typically been assumed in the literature, and that this was particularly true in the sentence-medial position. Even if the adults accented narrowly focal targets more than post-focal ones, post-focal targets were accented 64% of the time sentence-medially and 37% of the time sentence finally. Our analysis of accent type choices showed that even if there was a general preference for falls ('H*L') and downstepped falls (!H*L') for focus, the adults often used the same accent type on both focal and post-focal targets, observed most prominently the use of downstepped falls ('!H*L'). Finally, our analysis of the effect of focus on pitch minimum, maximum, and range, showed our Dutch-speaking adults to increase the pitch maximum and pitch range on focal versus post-focal target words medially, but that they only raised the pitch maximum sentence finally. Notably, our analysis showed no effect of narrow focus on word duration in the adult group. Our results regarding post-focal de-accentuation, as well as our observation of downstepped falls being common both focally and post-focally, resemble previous findings from Chen (2007, submitted). However, whereas Chen's results showed adults to use non-downstepped falls ('H*L') 45% of the time, followed by downstepped falls ('!H*L') at around 20%, for marking sentence-final focus, our adults preferred downstepped falls (52%) followed by non-downstepped ones (29%). In terms of focus type effects, we found no differences between narrow and contrastive focus in their effect on accent placement, and similarly, no differences between broad and narrow focus in their effect on accent placement, in either sentence position.

Our analysis of accent placement by narrow focus in the five-year-olds confirms the patterns described by Chen (e.g. 2011a), showing that in sentence-final position, Dutch-speaking five-year-olds accent focus and de-accent post focus in the same way as adults. However, our analysis of accent placement sentence-medially showed that where the adults de-accented post-focal targets more than focal ones, the five-year-olds did not show this effect. A similar position effect was observed in the accent type analysis, where our youngest group resembled the adults sentence-finally, preferring '!H*L' followed by 'H*L' for marking focus, and de-accentuation followed by '!H*L' for post-focus, whereas non-adult patterns such as using falls

('H*L') and high/rising accents ('!H*') post-focally, were observed sentence-medially. Our pitch measures partly reflected the results from our analysis of accent placement and type, showing more adult-like effects of focus on pitch sentence-finally (raised pitch maximum), whereas none of our pitch measures were affected by focus sentence-medially. Interestingly, in the sentence-medial position, where our analyses of pitch suggest that the five-year-olds are not yet making the adult adjustments, they did show significant effects of focus on word duration. Finally, the analyses presented in Chapter 7 showed that four- to five-year-olds also make use of pre-target pausing to mark focus at this stage.

Our group of eight-year-olds resembled the adults and five-year-olds in their use of accent placement sentence-finally, accenting narrowly focal targets more than post-focal ones, in line with findings from Chen (2011a). Nevertheless, sentence-medially they did not differentiate between focal and post-focal targets, thus the use of accent placement for focus is still not mastered in this position by eight-year-olds. Development takes place between five and eight in terms of accent type preferences sentence-medially, where the eight-year-olds were less variable than the five-year-olds. Interestingly, where our adult participants used downstepped falls ('!H*L') both focally and post-focally, the eight-year-olds preferred non-downstepped falls for focus and downstepped falls for post-focus, thereby creating a clearer labour division between certain accent types and focus. Interestingly, Chen (2011a) also describes the use of particular accent types to differentiate focus from non-focus in this age group, but her analysis concerned the sentence-initial position. Lastly, our phonetic analysis showed no systematic effects of focus on any of our phonetic measures in the eight-year-olds. While this makes sense sentence-medially, where the eight-year-olds rarely de-accented post-focal targets, the fact that their adult-like de-accentuation patterns sentence-finally do not show up as effects on pitch is more surprising. It is possible that the patterns from the eight-year-olds were acoustically too variable to cause systematic effects on pitch, even if the accent type analyses showed systematic prosodic differentiation between focal and post-focal target words.

Moving from our eight-year-olds to our eleven-year-olds, this group differentiated focal from post-focal targets using accent placement in both positions. In fact, our accent placement analysis showed the eleven-year-olds to de-accent post-focal targets more than the adults did, in addition to using accent types more consistently than the adults when differentiating between focal and post-focal targets. In contrast to our findings from the eight-year-olds, the accentuation patterns from the eleven-year-olds were also reflected in our phonetic analyses; focal targets were produced with a higher pitch maximum, a wider pitch range and increased duration sentence

medially, and with a higher pitch maximum and a wider pitch range sentence-finally, as compared to post-focal targets.

To summarize, the following developmental path toward adult prosodic focus marking emerges from our Dutch results: Accent placement and adult accent-type preferences are largely in place already at four to five, in sentence-final position. At this stage, children may use temporal adjustments such as pausing or lengthening of focal words as a strategy to prosodically highlight focus, particularly in the sentence-medial position, were their mastery of pitch-based manipulations is still developing. At seven to eight, Dutch children are still not adult-like in their use of accent placement for focus sentence-medially. At the same time, their use of accent types for marking focus is becoming more systematic, thus they seem to go through a phase where they rely more on accent *type* contrasts than on accent *placement,* for marking sentence-medial focus. Finally, by ten to eleven the use of post-focal de-accentuation is mastered in both sentence-positions.

8.4 Hypotheses revisited

In Chapter 1 we presented six hypotheses for how cross-linguistic effects between Swedish and Dutch could affect the development toward adult prosodic focus marking in these two languages. The first four hypotheses were related to pitch manipulations, whereas the last two were related to the use of duration.

Our first two hypotheses concerned two different ways the presence versus absence of a lexical pitch accent contrast could be assumed to affect the acquisition of pitch-based cues to focus in Dutch and CS. The first hypothesis proposed that early sensitivity to pitch would give the CS-speaking children a head start in the acquisition of pitch cues to focus, as compared to Dutch-speaking children. The second hypothesis suggested that that the complexity associated with the combination of word and sentence-level tones in CS would make CS-speaking children slower than their Dutch-speaking peers in acquiring the relevant pitch manipulations used for focus. Our third hypothesis was related to the repertoires of pitch accent-like contours assumed in the two languages. This hypothesis proposed that the relatively larger repertoire of pitch accents used in Dutch would make it harder to detect what the relevant accentual categories are, and we predicted this would slow down the acquisition of pitch-based cues to focus in Dutch as compared to CS. Finally, our fourth hypothesis was related to predictability in the mapping between prosody and focus. Whereas the association between prominence H and focus in CS is described as fairly consistent, work on the association between accentuation and focus in Dutch, as well as in descriptions of related languages like

German and English, suggests that this mapping may be more variable in Dutch than it is in CS. Based on such differences in prosody-focus mapping we hypothesized that the more predictable CS system would be easier to acquire than the Dutch system, leading to an earlier acquisition in CS than in Dutch. These four hypotheses are summarized in Table 49 below.

	Domain	**Order**
1	Lexical accents cause early tonal sensitivity	CS > DU
2	Prominence H + lexical accent is a complex category to acquire	CS < DU
3	A large repertoire of pitch accents makes it harder to detect the relevant categories	CS > DU
4	A less consistent prosody-to-focus mapping makes a system harder to acquire	CS > DU

Table 49: Hypotheses 1-4, summarized

As can be observed in Table 49, all but one of our hypotheses predicted an earlier mastery of pitch-based cues to focus in CS than in Dutch, and our results also point in this direction: CS-speaking children between four and five years are already close to adult-like in their use of prominence H for marking narrow focus. Further, our pitch analysis also showed that CS-speaking five-year-olds already match CS-speaking adults in the pitch range effect of adding or avoiding prominence H. Conversely, even if the youngest Dutch-speaking group resembled the Dutch-speaking adults in their use of accent placement and accent type sentence-finally, they did not de-accent post-focal targets sentence-medially, and neither did the eight-year-olds. Additional evidence for non-adult patterns was also found in the phonetic results from Dutch, where the results were generally mixed in our two youngest groups. Since our CS-speaking children largely performed in line with adults in both sentence-positions already from the age of four to five, we conclude that the Dutch-speaking children lag behind their CS-speaking peers in the acquisition of pitch-based cues to focus. We thus we reject hypothesis 2, which predicted an opposite pattern, but consider hypotheses 1, 3 and 4 to be confirmed by our data.

In addition to our four hypotheses concerning pitch-based differences between Dutch and CS, two additional hypotheses were also formulated in Chapter 1, related to the relationship between focus and word duration in the two languages. Generally, previous work on duration and focus do not differentiate durational effects of focus from durational effects of accentuation, and accentuation and focus

typically coincide in the stimuli (see Section 5.2.1). Assuming that the acquisition of word duration for focus may shed light on this issue, two competing hypotheses were proposed regarding the development of focal/accentual lengthening in the two languages. Our first hypothesis suggested that duration is linked to accentuation or prominence H, and thus that the lengthening associated with focus is related to the phonological category used for marking focus in the respective language, not with focus itself. If this were the case, one would expect the acquisition of duration for focus to follow that of accent placement/ prominence H. Conversely, if the lengthening is associated with focus more directly, without being mediated by accentuation/prominence H, the use of focal lengthening may develop differently than the use of prominence H/accentuation. Our two hypotheses on duration are summarized in Table 50 below.

	Domain	**Order**
5	Increased duration is related to prominence H/ accentuation	Duration = accent/prominence H
6	Duration for focus is manipulated independently of focus	Duration ≠ accent/prominence H

Table 50: Hypotheses 5 and 6, summarized

Generally, our results show a more robust relationship between focus and word duration in CS than in Dutch: Focus systematically lengthened the word duration in our CS-speaking adults, five-year-olds and eight-year-olds. In the Dutch data, duration effects of focus, if any, were rather spurious, and no effects of focus on word duration were found in our adult data, counter to previous reports (Sluijter & van Heuven, 1995; Cambier-Langeveld & Turk, 1999). In terms of association/dissociation, Hypothesis 5 seems to be at least partly supported by our adult CS data, where robust effects of focus on word duration were found in parallel with systematic effects of focus on prominence H. A similar relationship was found in the Dutch eleven-year-olds, who were more consistent in de-accenting post-focal information than the adults, and for whom a significant effect of focal lengthening was also found. At the same time, we also have some evidence for dissociation between duration and accentuation/prominence H, in that the Dutch-speaking five-year-olds, who did not de-accent sentence-medially at all, showed significant effects of focus on word duration. The same duration effect was also found in the CS-speaking five-year-olds, who were not completely adult in their use of prominence H sentence-medially. It seems that our data does not allow for concluding on the issue of association/dissociation, and it may well be that the use of duration and accentuation/prominence H is associated in some groups but independent in others.

A larger dataset is required for isolating the effect of accentuation from the effect of focus in different age groups and languages. Further, isolating these effects also requires systematic elicitation of non-focal accents/prominence H, something our methodology was not designed to do.

8.5 General discussion

As hypotheses 1, 3 and 4 are all supported by our data, we are left with three possible explanations for why our CS-speaking children acquire the use of prominence H earlier than their Dutch-speaking peers acquire the use of accent placement and type: pitch sensitivity, size of contour repertoire, or predictability in prosody-focus mapping. Starting with tonal sensitivity, it should be noted that the notion of higher tonal sensitivity among children learning tone languages is primarily based on infant studies. A general pattern emerging from this literature is all infants are sensitive to tonal contrasts from birth, but that this sensitivity drops toward the end of the first year in children learning non-tone languages. Still, even children learning non-tone languages re-gain their tonal sensitivity by around 18 months, thus the period where infants learning tone languages show an advantage corresponds to a fairly small time window in children's language development, and the children included in the current study are far outside that age range. This said, it is not clear how early tonal sensitivity relates to later production, and it could still be the case that lagging effects of early sensitivity affects production at later stages. To the best of our knowledge the relationship between early tonal sensitivity in perception and later tonal sensitivity in production has not been systematically studied. It is possible that tonal sensitivity remains a relevant factor for later production of pitch categories (see e.g. Liu & Kager, 2014; Frota; Mattock, Molnar, Polka & Burnham, 2008 and Butler & Vigário, 2014, for reviews on infant tone perception).

Whereas the CS-speaking adults predominantly used prominence H for focus and avoided this tone on non-focal constituents, the accentuation/de-accentuation pattern was much less consistent in our Dutch adults, where substantial overlap was found in the use of certain accentual categories across focal and non-focal target words. This finding confirms our assumption of a less reliable mapping between focus and prosodic categories in Dutch than in CS. Interestingly, where our Dutch adults seemed to combine the use of accent placement and the use of accent type for marking focus, the Dutch children between seven and eight seemed to go through a phase where they relied more on accent types than on accent placement for marking focus, before using both strategies in parallel at the age of ten to eleven. This developmental pattern may be taken to reflect that both accent placement and accent

type are used to mark focus in Dutch, but that children, as a developmental strategy, focus on one at a time, at least in certain sentence-positions. Greater tonal variability may well delay the children in terms of determining what the accentual categories are and how they are used, but when they do discover that certain accent types are more common for focus than others, they make use of this pattern in adult-like ways, even if the pattern is over-generalized at certain stages of development.

A few words should be said about position effects, which surfaced in different ways in our dataset, but were observed in both languages. Firstly, the data from our Dutch-speaking adults suggest that focus is marked in different ways sentence-medially and sentence-finally, with de-accentuation of post-focal targets being much more common sentence-finally than sentence-medially. Secondly, it seems that both Dutch and CS-speaking children master prosodic focus marking earlier in the sentence-final position than the sentence-medial position, and this pattern is strikingly robust across a range of different measures. As was discussed in Chapters 4 and 6, this position effect may be related to the fact that default alignment of prosodic prominence is the rightmost position in the IP for both Dutch and Swedish, thus what we observe is that prosodic emphasis is mastered earlier in the unmarked final position than in the marked non-final ones. Earlier acquisition of default prominence is also evidenced in our results from the broad-focus comparisons, showing that both Dutch and Swedish-speaking children behave in line with adults in assigning an accent/ prominence H to sentence-final targets under broad focus.

Other possible explanations why the sentence-final position is mastered earlier than the sentence-medial one is that the sentence-medial position may require more prosodic integration (e.g. the scaling of tonal targets according to adjacent tonal events on either side) than the sentence-final position, making prosodic focus-marking in this position more demanding from a production point of view. It is also possible that the sentence-final position is perceptually more salient, and that children therefore detect the pattern for focus marking earlier in this position than for sentence-internal ones. The salience of the sentence-final position may also be helped by the fact that important prosodic functions like turn-taking or interrogativity are typically marked here (e.g. House, 2003, on Swedish). Finally, our position effect may also be interpreted as an effect of grammatical category, as our sentence-medial targets were always verbs and our sentence-final targets were always nouns. As suggested by Röhr, Baumann & Grice (2015), it may be less common to focus on actions than on objects, and such a frequency effect may give children less experience with focus marking on verbs than on nouns, causing them to master prosodic focus marking earlier on nouns than on verbs.

The data from our oldest children, the eleven-year-olds, included some unexpected patterns, and this was the case for both the Dutch and the Swedish-speaking group. Whereas the Dutch-speaking eleven-year-olds frequently seemed to 'hyper articulate' the prosodic patterns we observed in our adults, the CS-speaking eleven-year-olds showed non-adult patterns in the use of word duration for focus. [42] In Chapter 4 we suggested that the striking duration-results from the CS-speaking eleven-year-olds could be a reflection of unengaged speech (e.g. Burkhardt & Sendlmeier, 2000), whereby the children occasionally lengthened post-focal target words, cancelling out the effect of focus found in the other groups. The hyper-articulated accentuation patterns produced by the Dutch-speaking eleven-year-olds may similarly have to do with our task being more suitable for children between four and eight than for children as old as eleven. Our goal of distracting the children from their own speech (Chapter 3) was typically reached in the five and eight-year-olds, who played along and did not detect that the pictures were pre-ordered. Conversely, the eleven-year-olds understood that what we wanted was their speech, and that the experimenter did not actually need their help in sorting the pictures. Because we predicted that the adults would similarly not be fooled by the game setup, we included extra information when introducing the game to the adult controls, telling them that the game was designed for children, and that they should just play along, speaking like they normally would. This made them talk casually and indeed play along, and they did not seem overly conscious of how they spoke.

In retrospect we believe that giving the eleven-year-olds a similar explanation would have made them play along in a more natural way, possibly avoiding the odd patterns we observed in some of the data from these groups. This said, we do believe that the patterns produced by the eleven-year-olds are representative of the natural repertoire they use for marking focus, or in showing disinterest for that matter, but the fact that the task was both less engaging and more transparent to them may have moderated the way they spoke in ways that were not attested in the other groups. Interestingly, the adult participants in Chen (2007, 2011a, submitted) resemble our Dutch eleven-year-olds in preferring non-downstepped accents for focus. It is possible that the setup we used, excluding the robot 'helper', made the task more conversational than was the case when the robot was present, making the adults produce less 'neat' patterns in our study than what was found in Chen's study.

As we have seen, existing work on the prosodic marking of information structure is somewhat heterogeneous in terms of the elicitation method used, the kind of

---

[42] The fact that we do not find over-shooting of prosodic patterns in our CS-speaking eleven-year-olds may be that both adults and eleven-year-olds were already at ceiling in the use of prominence H in our CS dataset.

information structural contrast looked at, and the type of speech that is elicited. In terms of acquisition it is interesting to note that whereas studies of adult speech tend to use read or more controlled speech samples, studies involving children typically require the use of more interactional methodologies (see Den Ouden, Noordman & Terken, on prosodic characteristics of read speech). This may explain why different results are sometimes reported from the same population, when studies aimed at adults are compared to studies designed for children (e.g. Baumann, Grice & Steindamm, 2006, versus Müller, Höhle, Schmitz & Weissenborn, 2006, on German, see also DeRuiter, 2010). Studies like the ones from Breen and colleagues (2010), showing that speakers prosodically differentiate contrastive from narrow focus if the task of prosodically differentiating the two is made explicit, suggest that that the way in which the data is elicited is important, and may shape the prosodic patterns found in a certain study. Eliciting read or otherwise strictly controlled speech where the same words or the same sentences are repeated across different information structural contexts is expected to make the task fairly transparent, which may cause adult participants strive to produce the relevant contrasts more clearly, possibly resulting in a more consistent marking of information structure than what may be found in more spontaneous types of speech. Our finding that the Dutch adults did not systematically lengthen focal versus post-focal words is an example of a case where we failed to replicate patterns described in more controlled speech samples (e.g. Cambier-Langeveld, 2000; Eefting, 1991; Sluijter, 1995).

In the studies presented in this thesis we combined a number of different prosodic measures, with the benefit of gaining more information about the kinds of manipulations the children make, also in cases where the children were found to use non-adult strategies for marking focus. An important methodological insight from our analyses is that the choice of output measure has important consequences for the conclusions that are drawn. Whereas our Dutch accent placement analysis showed that four- to five-year-olds did not distinguish focal from post-focal targets at all in sentence-medial position, our analysis of accent types, together with our analysis of pitch and duration, showed that this group resemble adults in using downstepped falls post-focally, and that they in addition increased the word duration for focus sentence-medially. Similarly, comparing the development of prominence H for focus with the development of word duration for the same purpose, revealed that CS-speaking children reach adult proficiency earlier in the use of duration than they do in the use of prominence H, suggesting a possible dissociation between these different manipulations in the children's development. Our study shows that performing a wider set of analyses than those typically performed has the benefit of capturing more information about the children's prosodic development than what you get by relying on one measure only, making it possible to see children's development more as a gradual process than as a question of mastery or not.

The goal of this thesis was to establish the developmental path toward adult prosodic focus marking in CS and Dutch. This work was carried out in order to determine how cross-linguistic differences between prosodic systems shape the developmental path toward adult prosodic focus marking, as well as to determine whether common developmental patterns can also be observed, despite such cross-linguistic differences. We have seen that the ability to prosodically mark focus develops pretty gradually in both Dutch and CS, and that the prosodic system a child is acquiring shapes the developmental path in various ways. We have also seen that CS-speaking children reach adult proficiency in prosodic focus marking earlier than their Dutch-speaking peers. It is impossible to determine whether early sensitivity to pitch, a more constrained repertoire of contours, or a more reliable mapping between prominence H and focus is what causes the the CS system to be acquired earlier than the Dutch one, and it is possible that all three factors contribute. Indications of common developmental patterns were also observed between our CS and Dutch-speaking children, primarily in that prosodic focus marking is mastered earlier in sentence-final position than in sentence-medial position. In addition, we have also seen that both Dutch and CS-speaking five-year-olds tended to lengthen focal as compared to post-focal words, even at stages when their pitch-manipulations for focus are not yet completely in place.

8.6 Directions for future work

The picture-matching game constitutes a useful methodology for gathering naturalistic speech data from children between four and eleven years, particularly if the suggested adaptations are made when gathering data from older children. Our general impression was that the children mostly enjoyed the game, and that even the youngest children understood the task. Our experience of adapting the game for Korean, Mandarin Chinese, Bai, Dutch, and Swedish, has shown the methodology to be fit for cross-linguistic work. Future versions of the game can easily be constructed based on the database of pictures we have built, which is publicly available for other researchers to use (www.prosodicdevelopment.com). The results of the current study have provided interesting insights into the acquisition of prosodic focus marking that occurs between four and eleven years. This said, our results show several manipulations are already in place by this age, thus future investigations should include children younger than four. The challenge in using the picture-matching game with children below four is primarily related to the requirement that they produce complete SVO-sentences. If the game could be simplified, for example by eliciting adjective-noun pairings or other more simple structures, younger children could be studied using a similar procedure. Keeping the methodology fairly parallel would allow for comparing data from younger children

with the findings presented in this thesis, as well as comparable data from Korean, Mandarin Chinese and Bai (Yang & Chen, 2014; Yang, Cho, Kim & Chen, 2015; Yang, forthchoming; Liu, Chen & van de Velde, 2014, 2015).

Finally, all our analyses approach prosodic focus marking from a paradigmatic perspective: The same words are compared across different utterances, representing different focus conditions. This approach is typically chosen in studies of prosodic manipulations for focus, as it allows for keeping target words and sentence positions consistent across trials. However, such an analysis partly ignores the relative aspects of prominence, despite the intuition that the extent to which a constituent is perceived as prominent depends on its salience relative to neighbouring constituents. Studies of the development of prosodic focus marking should also take this perspective into account, using perception experiments for determining what kind of prosodic manipulations listeners attend to when interpreting various focus structures. Perception experiments should also be constructed in order to determine the actual communicative efficiency of children's developmental strategies (see Ito, 2014, for examples of such experiments). As we find our children to use non-adult manipulations (pausing, over-using certain accent types, relying on certain parameters only) at stages when their proficiency at using more canonical cues is still developing, studies on how such strategies are received by listeners is an important next step for broadening our understanding of the acquisition of prosodic focus marking in children.

Bibliography

Ambrazaitis, G. (2009). Nuclear intonation in Swedish: Evidence from experimental-phonetic studies and a comparison with German. PhD thesis, Lund University.

Arnhold, A. (2011). Multiple prosodic parameters signaling information structure: Parallel focus marking in Finnish. Proceedings of the 17th International Congress of Phonetic Sciences (ICPHS), pp. 248-251.

Arnhold, A. (2013). *Finnish prosody: Studies in intonation and phrasing.* PhD thesis, Johann Wolfgang Goethe University.

Arnhold, A. & Féry, C. (2014). The role of syntactic flexibility and prosody in marking given/new distinctions in Finnish. *Finno-Ugric Languages and Linguistics, 2,* pp. 2-32.

Arnold, J. E. (1998). *Reference form and discourse patterns.* PhD thesis, Stanford University.

Avesani, C. & Vayra, M. (2003). Broad, narrow and contrastive focus in Florentine Italian. *Proceedings of the 15th International Congress of Phonetic Sciences,* pp. 1803-1806.

Balog, H. L., & Snow, D. (2007). The adaptation and application of relational and independent analyses for intonation production in young children. *Journal of Phonetics, 35*, pp. 118-133.

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

Bard, E. G., & Aylett, M. (1999). The dissociation of deaccenting, givenness, and syntactic role in spontaneous speech. In J. J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, & A. C. Bailey (Eds.). *Proceedings of the 14th International Congress of Phonetic Sciences (ICPHS),* pp. 1753-1756.

Basbøll, H. (2005). *The phonology of Danish.* Oxford University Press.

Baumann, S. (2006). *The intonation of givenness: Evidence from German.* Tübingen: Niemeyer.

Baumann, S., & Grice, M. (2006). The intonation of accessibility. *Journal of Pragmatics, 38*, pp. 1636-1657.

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

Baumann, S., Becker, J., Grice, M., & Mücke, D. (2007). Tonal and articulatory marking of focus in German. *Proceedings of the 16th International Congress of Phonetic Sciences (ICPHS),* pp. 1029-1032.

Beckman, M. E. (1986). *Stress and non-stress accent.* Dordrecht: Foris.

Beckman, M. E., & Pierrehumbert, J. B. (1986). Intonational structure in Japanese and English. *Phonology, 3*, pp. 255-309.

Beckman, M. E., Hirschberg, J., & Shattuck-Hufnagel, S. (2005). The original ToBI system and the evolution of the ToBI framework. *Prosodic models and transcription: Towards prosodic typology,* pp. 9-54.

Berglund, E., & Eriksson, M. (2000). Communicative development in Swedish children 16-28 months old. The Swedish early communicative development inventory - words and sentences. *Scandinavian Journal of Psychology, 41*, pp. 133-144.

Birch, S., & Clifton, C. (1995). Focus, accent, and argument structure: Effects on language comprehension. *Language and Speech, 38*, pp. 365-391.

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

Bolinger, D. L. (1961). Contrastive accent and contrastive stress. *Language, 37*, pp. 83-96.

Bortfeld, H., & Morgan, J. (2010). Is early word-form processing stress-full? How natural variability supports recognition. *Cognitive Psychology, 60*, pp. 241–266.

Braun, B. (2006). Phonetics and phonology of thematic contrast in German. *Language and Speech, 49*, pp. 451-493.

Breen, M., Fedorenko, E., Wagner, M., & Gibson, E. (2010). Acoustic correlates of information structure. *Language and Cognitive Processes, 25*, pp. 1044-1098.

Bruce, G. (1977). *Swedish word accents in sentence perspective.* Lund University: Gleerup.

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

Bruce, G. (1983). Accentuation and timing in Swedish. *Folia Linguistica, 17*, pp. 221-238.

Bruce, G. (1987). How floating is focal accent? In K. Gregersen, & H. Basbøll (Eds.). *Nordic Prosody, 4,* pp. 41–49. *Odense University Press*,

Bruce, G. (1998). *Allmän och svensk prosodi.* Lunds Universitet: Institutionen för Lingvistik.

Bruce, G. (1999). Word tone in Scandinavian languages. In H. van der Hulst (Ed.). *Word prosodic systems in the languages of Europe*, *20,* pp. 605-633. Berlin, New York: Mouton de Gruyter.

Bruce, G. (2005). Intonational prominence in varieties of Swedish revisited. In S. A. Jun (Ed.). *Prosodic typology: The phonology of intonation and phrasing*, pp. 410-430. Oxford: Oxford University Press.

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.

Burkhardt, F., & Sendlmeier, W. F. (2000). Verification of acoustical correlates of emotional speech using formant-synthesis. *Proceedings of the ISCA Workshop on Speech and Emotion,* pp. 151-156.

Butcher, A. (1981). Aspects of the speech pause: Phonetic correlates and communication functions. *Arbeitsberichte Kiel, 15,* pp. 1-233.

Büring, D. (2003). On D-trees, beans, and B-accents. *Linguistics and Philosophy 26,* pp. 511-545.

Büring, D. (2006). Focus projection and default prominence. In V. Molnár & S. Winkler (Eds.) *The Architecture of Focus,* pp. 321–346. Berlin, New York: Mouton de Gruyter,

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

Cambier-Langeveld, T. (1999). The interaction between final lengthening and accentual lengthening: Dutch versus English. *Proceedings of the 14th International Congress of Phonetic Sciences (ICPHS),* pp. 467-470.

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

Campione, E., & Véronis, J. (2002). A large-scale multilingual study of silent pause duration. *Proceedings of Speech Prosody*, pp. 199-202.

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.

Chafe, W. (1987). Cognitive constraints on information flow. In Tomlin, R. (Ed.). *Coherence and grounding in discourse.* Amsterdam: John Benjamins Publishing.

Chafe, W.L. (1994). *Discourse, consciousness, and time.* University of Chicago Press, Chicago.

Chen, A. (2007*).* Intonational realisation of topic and focus by Dutch-acquiring 4-to 5-year-olds. *Proceedings of the 16th International Congress of Phonetic Sciences (ICPHS)*, pp. 1553-1556.

Chen, A. (2009). The phonetics of sentence-initial topic and focus in adult and child Dutch. In M. C. Vigario, S. Frota, & M. J. Freitas (Eds.). *Phonetics and phonology: Interactions and interrelations,* pp. 91-106. Amsterdam: John Benjamins Publishing.

Chen, A. (2010). Is there really an asymmetry in the acquisition of the focus-to-accentuation mapping? *Lingua, 120,* pp. 1926-1939.

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

Chen, A. (2011b). The developmental path to phonological focus-marking in Dutch. In S. Frota, (Ed.). *Prosodic categories: Production, perception and comprehension,* pp. 93-110. Dordrecht: Springer Publishing.

Chen, A. (2012). The prosodic investigation of information structure. In M. Krifka, & R. Musan (Eds.). *The expression of information structure,* pp. 249-286. Berlin: Mouton de Gruyter.

Chen, A. (2015). Children's use of intonation in reference and the role of input. In L. Serratrice, & S. Allen (Eds.). *The acquisition of reference,* pp. 83-104. Amsterdam: John Benjamins Publishing.

Chen, A. (submitted). A phonetics-phonology interface-based approach to prosodic realization of topic and focus in Dutch.

Chen, A., & Fikkert, P. (2007). Intonation of early two-word utterances in Dutch. *Proceedings of the 16th International Congress of Phonetic Sciences (ICPHS)*, pp. 1553-1556.

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

Cruttenden, A. (1985). Intonation comprehension in ten-year-olds. *Journal of Child Language 12,* pp. 643–661.

Cruttenden, A. (2006). The de-accenting of given information: A cognitive universal? *Empirical Approaches To Language Typology, 20*, pp. 311-356.

Dahan, D., & Bernard, J. (1996). Interspeaker variability in emphatic accent production in French. *Language and Speech, 39*, pp. 341-374.

Dalton, P., & Hardcastle, W. J. (1977). *Disorders of fluency and their effects on communication.* Elsevier, Amsterdam.

De Ruiter, L. E. (2009). The prosodic marking of topical referents in the German 'Vorfeld' by children and adults. *Linguistic Review, 26*, pp. 329-354.

De Ruiter, L. E. (2010). *Studies on intonation and information structure in child and adult German.* PhD Thesis, Max Planck Institute for Psycholinguistics.

De Ruiter, L. E. (2014). How German children use intonation to signal information status in narrative discourse. *Journal of Child Language, 41,* pp. 1015-1061.

De Ruiter, L. E. (2015). Information status marking in spontaneous vs. read speech in story-telling tasks–Evidence from intonation analysis using GToBI. *Journal of Phonetics, 48,* pp. 29-44.

Den Ouden, H., Noordman, L., & Terken, J. (2009). Prosodic realizations of global and local structure and rhetorical relations in read aloud news reports. *Speech Communication, 51*, pp. 116-129.

Dilley, L., Shattuck-Hufnagel, S., & Ostendorf, M. (1996). Glottalization of word-initial vowels as a function of prosodic structure. *Journal of Phonetics, 24*, pp. 423-444.

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

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

Engstrand, O. (1995). Phonetic interpretation of the word accent contrast in Swedish. *Phonetica, 52*, pp. 171-179.

Engstrand, O. (1997). Phonetic interpretation of the word accent contrast in Swedish: Evidence from spontaneous speech. *Phonetica, 54,* pp. 61-75.

Engstrand, O., Williams, K., & Strömqvist, S. (1991). Acquisition of the tonal word accent contrast. *Actes du XIIème Congres International Des Science Phonétiques,* pp. 324-327.

Fant, G. & Kruckenberg, A. (1989). Preliminaries to the study of Swedish prose reading and reading style. *STL-QPSR, 2,* pp. 1-83.

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

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

Ferreira, F. (2007). Prosody and performance in language production. *Language and Cognitive Processes, 22*, pp. 1151-1177.

Ferreira, V. S., & Hudson, M. (2011). Saying "that" in dialogue: The influence of accessibility and social factors on syntactic production. *Language and Cognitive Processes, 26*, pp. 1736-1762.

Féry, C. (1993). *German Intonation Patterns.* Tübingen: Niemeyer.

Féry, C. (2008). Information structural notions and the fallacy of invariant correlates. *Acta Linguistica Hungarica, 55*, pp. 361-379.

Féry, C. (2013). Focus as prosodic alignment. *Natural Language & Linguistic Theory, 31*, pp. 683-734.

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

Frid, J. (2003). *Lexical and acoustic modelling of Swedish prosody.* PhD Thesis, Lund University.

Frota, S. & Vigário, M. (2008). Early intonation in European Portuguese. Talk given at *The Third Conference on Tone and Intonation (TIE3),* Lisbon.

Frota, S., Butler, J., & Vigário, M. (2014). Infants' perception of intonation: Is it a statement or a question? *Infancy, 19*, pp. 194-213.

Fuchs, A. (1984). Deaccenting and default accent. In D. Gibbon, & H. Richter (Eds), *Intonation, Accent and Rhythm*, pp. 134–164. Berlin: Mouton de Gruyter.

Gee, J. P., & Grosjean, F. (1984). Empirical evidence for narrative structure. *Cognitive Science, 8,* pp. 59-84.

Gobl, C., & Ní Chasaide, A. (2010). Voice Source Variation and Its Communicative Functions. In W. J. Hardcastle, J. Laver, & F. E. Gibbon (Eds.). *The Handbook of Phonetic Sciences,* pp. 378-423. Hoboken: John Wiley & Sons.

Goldman-Eisler, F. (1968). *Psycholinguistics: Experiments in spontaneous speech.* Academic Press.

Goldsmith, J. (1976). An overview of autosegmental phonology. *Linguistic analysis, 2,* pp. 23-68.

Gordon, M., & Ladefoged, P. (2001). Phonation types: a cross-linguistic overview. *Journal of Phonetics, 29*, pp. 383-406.

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.

Gu, W., & Lee, T. (2007). Effects of focus on prosody of Cantonese speech – A comparison of surface feature analysis and model-based analysis. *Proceedings of the international workshop Paralinguistic Speech.* pp. 59-64.

Gundel, J. K., & Fretheim, T. (2004). Topic and focus. In L. R. Horn, & G. Ward (Eds.). *The Handbook of Pragmatics,* pp. 175-196. Hoboken: Wiley-Blackwell.

Gundel, J. K., Hedberg, N., & Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. *Language, 69*, pp. 274-307.

Gussenhoven, C. (1983). Testing the reality of focus domains. *Language and speech, 26*, pp. 61-80.

Gussenhoven, C. (1984). *On the grammar and semantics of sentence accents*. Berlin: Mouton de Gruyter.

Gussenhoven, C. (1999). On the limits of focus projection in English. In P. Bosch, & R. van der Sandt (Eds.), *Focus: Linguistic, Cognitive, and Computational Perspectives,* pp. 43–55. Cambridge: Cambridge University Press

Gussenhoven, C. (2002). Phonology of intonation. *Glot International, 6*, pp. 271-284.

Gussenhoven, C. (2004). *The phonology of tone and intonation.* Cambridge: Cambridge University Press.

Gussenhoven, C. (2005). Transcription of Dutch intonation. In S. A. Jun (Ed.), *Prosodic typology: The phonology of intonation and phrasing*, pp. 118 – 145. Oxford: Oxford 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.

Gussenhoven, C. (2011). Sentential prominence in English. In M. van Oostendorp, C. J. Ewen, E. V. Hume, & K. Rice (Eds.). *The Blackwell Companion to Phonology,* pp. 2778-2806. Hoboken: John Wiley & Sons.

Gussenhoven, C., & Bruce, G. (1999). Word prosody and intonation. In H. van der Hulst (Ed.). *Word prosody systems in the languages of Europe.* Berlin: Mouton de Gruyter,

Gussenhoven, C., Rietveld, T., & Terken, J. (1999). Transcription of Dutch intonation (ToDI). Courseware, available at http://todi.let.kun.nl/ToDI /home.htm

Gussenhoven, C., & Van der Vliet, P. (1999). The phonology of tone and intonation in the Dutch dialect of Venlo. *Journal of Linguistics, 35*, pp. 99-135.

Hallé, P. A., De Boysson-Bardies, B., & Vihman, M. M. (1991). Beginnings of prosodic organization: Intonation and duration patterns of disyllables produced by Japanese and French infants. *Language and Speech, 34*, pp. 299-318.

Halliday, M. A. K. (1967). *Intonation and grammar in British English*. Berlin: Mouton de Gruyter.

Hanssen, J., Peters, J., & Gussenhoven, C. (2008). Prosodic effects of focus in Dutch declaratives. *Proceedings of Speech Prosody,* pp. 609-612.

Hansson, P. (2003). *Prosodic phrasing in spontaneous Swedish.* PhD Thesis, Lund University.

Hedberg, N., & Sosa, J. M. (2008). The prosody of topic and focus in spontaneous English dialogue. In In C. M. Lee, M. Gordon, & D. Büring. (Eds.). *Topic and Focus,* pp. 101-120. Dordrecht: Springer Publishing.

Heldner, M. (2001). Focal accent-f0 movements and beyond. Phd Thesis, Umeå University.

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

Heldner, M., & Strangert, E. (2001). Temporal effects of focus in Swedish. *Journal of Phonetics, 29*, pp. 329-361.

Henton, C. G., & Bladon, A. (1987). Creak as a sociophonetic marker. In L. Hyman, & C. Li (Eds). *Language, speech and mind: studies in honour of Victoria A. Fromkin.*

Hieke, A. E., Kowal, S., & O'Connell, D. C. (1983). The trouble with "articulatory" pauses. *Language and Speech, 26*, pp. 203-214.

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.

Horne, M. (1991). Why do speakers accent ‘‘given’’ information?. *Proceedings of Eurospeech 1991*, pp. 1279-1282.

Horne, M., Strangert, E., & Heldner, M. (1995). Prosodic boundary strength in Swedish: Final lengthening and silent interval duration. *Proceedings of the 13th International Congress of Phonetic Sciences (ICPHS)*, pp. 170-173.

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.

House, D. (2005). Phrase-final rises as a prosodic feature in wh-questions in Swedish human–machine dialogue. *Speech Communication, 46*, pp. 268-283.

Huang, B., & Liao, X. (2002). *Modern Chinese.* China Higher Education Press, Beijing.

Hyman, L. M. (2009). How (not) to do phonological typology: the case of pitch-accent. *Language Sciences, 31*, pp. 213-238.

Ito, K. (2014). Children's pragmatic use of prosodic prominence. *Pragmatic Development in First Language Acquisition, 10,* pp. 199-218.

Ito, K., & Speer, S. R. (2008). Anticipatory effects of intonation: Eye movements during instructed visual search. *Journal of Memory and Language, 58*, pp. 541-573.

Ito, K., Jincho, N., Minai, U., Yamane, N., & Mazuka, R. (2012). Intonation facilitates contrast resolution: Evidence from Japanese adults and 6-year olds. *Journal of Memory and Language, 66*, pp. 265-284.

Jun, S. A. (2005). *Prosodic typology: the phonology of intonation and phrasing*. Oxford University Press.

Kadin, G., & Engstrand, O. (2005). Tonal word accents produced by Swedish 18- and 24-month-olds. *Proceedings of Fonetik 2005*, pp. 67-70.

Karttunen, L. (1974). Presupposition and linguistic context. *Theoretical linguistics, 1*, pp. 181-194.

Kiss, K. É. (1998). Identificational focus versus information focus. *Language, 74,* pp. 245-273.

Klatt, D. H., & Klatt, L. C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *Journal of the Acoustical Society of America, 87*, pp. 820-857.

Kochanski, G., & Shih, C. (2003). Prosody modelling with soft templates. *Speech Communication, 39*, pp. 311-352.

Krahmer, E., & Swerts, M. (2001). On the alleged existence of contrastive accents. *Speech communication, 34*, pp. 391-405.

Krifka, M. (2007). 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.

Kristoffersen, G. (2006). Markedness in urban east Norwegian tonal accent. *Nordic Journal of Linguistics, 29,* pp. 95-135.

Kristoffersen, G. (2007). Jamvektseffekten. En fonetisk analyse av jamvektseffekten i nordgudbrandsdalsdialekten. *Norsk Lingvistisk Tidsskrift, 25,* pp. 187-232.

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

Laan, G. P. (1997). The contribution of intonation, segmental durations, and spectral features to the perception of a spontaneous and a read speaking style. *Speech Communication, 22*, pp. 43-65.

Ladd, D. R. (1983). Phonological features of intonational peaks. *Language, 59*, pp. 721-759.

Ladd, D. R. (1996). *Intonational Phonology, First Edition.* Cambridge University Press.

Ladd, D. R. (2008). *Intonational phonology, Second Edition.* Cambridge 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.

Lee, S., Potamianos, A., & Narayanan, S. (1999). Acoustics of children's speech: Developmental changes of temporal and spectral parameters. Journal of the Acoustical Society of America, 105, pp. 1455-1468.

Lewis, D. (1979). Scorekeeping in a language game. *Journal of Philosophical Logic, 8*, pp. 339-359.

Li, C. N., & Thompson, S. A. (1977). The acquisition of tone in Mandarin-speaking children. *Journal of Child Language, 4*, pp. 185-199.

Liberman, M. Y. (1975). *The intonational system of English.* PhD thesis, Massachusetts Institute of Technology.

Liberman, M. Y., & Pierrehumbert, J. (1984). Intonational invariance under changes in pitch range and length. In M. Aronoff, & R. T. Oehrle (Eds.). *Language Sound Structure,* pp. 157-233. Cambridge: MIT 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.

Maloney, E. M., Payne, L., & Redford, M. A. (2012). What children's pause patterns indicate about their constituent structure. *Proceedings of BUCLD 36,* pp. 1-12.

Mattock, K., Molnar, M., Polka, L., & Burnham, D. (2008). The developmental course of lexical tone perception in the first year of life. *Cognition, 106*, pp. 1367-1381.

Molnár, V. (2002). Contrast - from a contrastive perspective. *Language and Computers, 39*, pp. 147-161.

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

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.

Myrberg, S., & Riad, T. (2015). The prosodic hierarchy of Swedish. *Nordic Journal of Linguistics, 38*, pp. 115-147.

Ní Chasaide, A., & Gobl, C. (1997). *Voice source variation. The handbook of phonetic sciences,* pp. 427-461. Blackwell Science, Oxford.

Nooteboom, S. G., & Kruyt, J. G. (1987). Accents, focus distribution, and the perceived distribution of given and new information: An experiment. *Journal of the Acoustical Society of America, 82*, pp. 1512-1524.

Ota, M. (2006). Children's production of word accents in Swedish revisited. *Phonetica, 63*, pp. 230-246.

Pierrehumbert, J. B. (1980). *The phonology and phonetics of English intonation.* Phd Thesis, Massachusetts Institute of Technology.

Pierrehumbert, J. B. (1995). Prosodic effects on glottal allophones. In O. Fujimura, & M. Hirano (Eds.). *Vocal Fold Physiology: voice quality control*, pp. 39-60. San Diego: Singular Publishing Group

Pierrehumbert, J., & Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. In P. Cohen, J. Morgan, & M. Pollack (Eds.). *Intentions in Communication*, pp. 271-311. Cambridge: MIT Press.

Pierrehumbert, J., & Talkin, D. (1992). Lenition of/h/and glottal stop. In D. R. Docherty, & D. Ladd. (Eds.). *Papers in laboratory phonology II,* pp. 90-117. Cambridge: Cambridge University Press

Pitrelli, J. F., Beckman, M. E., & Hirschberg, J. (1994). Evaluation of Prosodic Transcription Labelling Reliability in The Tobi Framework. *Proceedings of the International Conference on Spoken Language Processing,* pp. 123-126.

Prieto, P., & Vanrell, M. D. M. (2007). Early intonational development in Catalan. *Proceedings of the 16th International Congress of Phonetic Sciences (ICPHS)*, pp. 309-314.

Prieto, P., Estrella, A., Thorson, J., & Vanrell, M. D. M. (2012). Is prosodic development correlated with grammatical and lexical development? Evidence from emerging intonation in Catalan and Spanish. *Journal of Child Language, 39*, pp. 221-257.

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, 59*, pp. 413-425.

Redford, M. A. (2013). A comparative analysis of pausing in child and adult storytelling. *Applied Psycholinguistics, 34*, pp. 569-589.

Riad, T. (1998). Towards a Scandinavian accent typology. In W. Kehrein, & R. Wiese (Eds.). *Phonology and morphology of the Germanic languages,* pp. 77-109. Tübingen: Niemeyer.

Riad, T. (2003). Diachrony of the Scandinavian accent typology. In P. Fikkert, & H. Jacobs (Eds.). *Development in prosodic systems,* pp. 91-144. Berlin: Mouton de Gruyter.

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.

Roberts, C. (1996). Information structure in discourse: Towards an integrated formal theory of pragmatics. In J. H. Yoon, & A. Kathol (Eds.). OSU *Working Papers in Linguistics,* pp. 91-136. Columbus: Ohio State University.

Rochester, S. R. (1975). Defining the silent pause in speech. *Journal of the Ontario Speech and Hearing Association, 8*, pp. 1-14.

Röhr, C. T., & Baumann, S. (2010). Prosodic marking of information status in German. *Proceedings of Speech Prosody,* pp. 1-4.

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)*.

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

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

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

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

Rooth, M. (1985). Association with focus. PhD Thesis, University of Massachussets at Amherst.

Rooth, M. (1992). A theory of focus interpretation. *Natural Language Semantics, 1*, pp. 75-116.

Rump, H. H., & Collier, R. (1996). Focus conditions and the prominence of pitch-accented syllables. *Language and Speech, 39*, pp. 1-17.

Sabin, E., Clemmer, E., O'Connell, D., & Kowal, S. (1979). A pausological approach to speech development. In A. Siegman, & S. Feldstein (Eds.). *Of speech and time: Temporal speech patterns and interpersonal contexts*, pp. 35-55. Mahwah: Lawrence Erlbaum Associates.

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

Selkirk, E. (1984). Phonology and syntax: The relation between sound and structure. PhD Thesis, Massachusetts Institute of Technology.

Selkirk, E. (1995). Sentence prosody: Intonation, stress, and phrasing. In J. A. Goldsmith (Ed.). *The Handbook of Phonological Theory*, pp. 550-569

Silverman, K. E., Beckman, M. E., Pitrelli, J. F., Ostendorf, M., Wightman, C. W., Price, P., & Hirschberg, J. (1992). TOBI: a standard for labeling English prosody. *Proceedings of the 2th International Conference on Spoken Language Processing (ICSLP)*, pp. 13-16.

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

Sluijter, A. (1995). Phonetic correlates of stress and accent. PhD thesis, University of Leiden.

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

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

Smith, B. L. (1992). Relationships between duration and temporal variability in children's speech. *Journal of the Acoustical Society of America, 91*, pp. 2165-2174.

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

Snow, D., & Balog, H. L. (2002). Do children produce the melody before the words? A review of developmental intonation research. *Lingua, 112*, pp. 1025-1058.

Stalnaker, R. C. (1974). Pragmatic presuppositions. In M. K. Munitz, & P. K. Unger (Eds.). *Semantics and Philosophy*, pp. 197-214. New York: New York University Press.

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.

Strangert, E., & Heldner, M. (1998). On the amount and domain of focal lengthening in Swedish. *Proceedings of the 5th International Conference on Spoken Language Processing (ICSLP)*, pp. 3305-3308.

Swerts, M., & Geluykens, R. (1994). Prosody as a marker of information flow in spoken discourse. *Language and Speech, 37*, pp. 21-43.

Swerts, M., Krahmer, E., & Avesani, C. (2002). Prosodic marking of information status in Dutch and Italian: A comparative analysis. *Journal of Phonetics, 30*, pp. 629-654.

t' Hart, R. C., & Cohen, A. (1990). *A perceptual study of intonation.* Cambridge: Cambridge University Press.

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

Terken, J. M. (1984). The distribution of pitch accents in instructions as a function of discourse structure. *Language and Speech, 27*, pp. 269-289.

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

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

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.

Wagner, M., & Watson, D. G. (2010). Experimental and theoretical advances in prosody: A review. *Language and Cognitive Processes, 25*, pp. 905-945.

Welby, P., (2003). Effects of pitch accent position, type, and status on focus projection. *Language and Speech, 46*, pp. 53-81.

Wells, B., Peppé, S., & Goulandris, N. (2004). Intonation development from five to thirteen. *Journal of Child Language, 31*, pp. 749-778.

Wetterlin, A. (2007). *The lexical specification of Norwegian tonal word accents.* PhD thesis, University of Konstanz.

Wetterlin, A. (2010). *Tonal accents in Norwegian: Phonology, morphology and lexical specification.* Berlin: Mouton de Gruyter.

Wieman, L. (1976). Stress patterns in early child language. *Journal of Child Language, 3,* pp. 283-286.

Wijnen, F. (1990). The development of sentence planning. *Journal of Child Language, 17*, 651-675.

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

Xu, Y., & Wang, Q. E. (2001). Pitch targets and their realization: Evidence from Mandarin Chinese. *Speech communication, 33*, pp. 319-337.

Xu, Y., & Xu, C. X. (2005). Phonetic realization of focus in English declarative intonation. *Journal of Phonetics, 33*, pp. 159-197.

Yang, A., & Chen, A. (2014). Prosodic focus marking in Chinese four-and eight-year-olds. *Proceedings of Speech Prosody,* pp. 713-717.

Yang, A., Cho, T., Kim, S., & Chen, A. (2015). Phonetic focus-marking in Korean-speaking 7- to 8-year-olds and adults. *Proceedings of the 18th International Congress of Phonetic Sciences (ICPHS).*

Yang, A. (to appear). Prosodic focus-marking in Mandarin Chinese-speaking and Seoul Korean-speaking children. PhD Thesis, Universiteit Utrecht

Yip, M. (2002). *Tone.* Cambridge: Cambridge University Press.

Zellner, B. (1994). Pauses and the temporal structure of speech. In E. Keller (Ed.). *Fundamentals of speech synthesis and speech recognition*, pp. 41-62. Hoboken: John Wiley.

Appendices

A.1. Swedish sentence list

	Swedish	English translation
1	Lejonet tvättar bollen	The lion washes the ball
2	Hunden målar bilen	The dog paints the ball
3	Björnen gömmer tåget	The bear hides the train
4	Tomten kastar blomman	The Santa throws the flower
5	Grodan köper stöveln	The frog buys the boot
6	Bagaren kokar tårtan	The baker boils the cake
7	Bagaren gömmer bollen	The baker hides the ball
8	Lejonet kastar bilen	The lion throws the car
9	Tomten målar tåget	The Santa paints the train
10	Björnen tvättar blomman	The bear washes the flower
11	Grodan kokar stöveln	The frog boils the boot
12	Hunden köper tårtan	The dog buys the cake
13	Grodan kokar bollen	The frog boils the ball
14	Björnen tvättar bilen	The bear washes the car
15	Hunden gömmer tåget	The dog hides the train
16	Tomten målar blomman	The Santa paints the flower
17	Bagaren köper stöveln	The baker buys the boot
18	Lejonet kastar tårtan	The lion throws the cake
19	Grodan målar bollen	The frog paints the ball
20	Björnen gömmer bilen	The bear hides the car
21	Hunden kokar tåget	The dog boils the train
22	Lejonet köper blomman	The lion buys the flower
23	Tomten kastar stöveln	The Santa throws the ball
24	Bagaren tvättar tårtan	The baler washes the cake
25	Tomten kastar bollen	The Santa throws the ball
26	Bagaren tvättar bilen	The baker washes the car
27	Hunden köper tåget	The dog buys the train

28	Lejonet gömmer blomman	The lion hides the flower
29	Grodan kokar stöveln	The frog boils the boot
30	Björnen målar tårtan	The bear paints the cake

A.2. Dutch sentence list

	Dutch	**English translation**
1	De hond tovert de wortel	The dog 'enchants' the carrot
2	De leeuw kookt de laars	The lion cooks the boot
3	De hond tekent de lepel	The dog draws the spoon
4	De poes tovert de lepel	The cat 'enchants' the spoon
5	De bakker kookt de wortel	The baker boils the carrot
6	De hond tekent de laars	The dog draws the boot
7	Het meisje tovert de wortel	The girl 'enchants' the carrot
8	De poes kookt de laars	The cat boils the boot
9	Het meisje tekent de lepel	The girl draws the spoon
10	De hond tovert de laars	The dog 'enchants' the boot
11	De poes kookt de lepel	The cat boils the spoon
12	De leeuw tekent de wortel	The lion draws the carrot
13	De bakker tovert de lepel	The baker 'enchants' the spoon
14	De leeuw kookt de wortel	The lion boils the carrot
15	De bakker tekent de laars	The baker draws the boot
16	De leeuw vindt de bloem onder de hoed	The lion finds the flower under the hat
17	De hond verstopt het boek in de mand	The dog hides the book in the basket
18	De poes verstopt de auto in de doos	The cat hides the car in the box
19	Het meisje gooit de bloem in de doos	The girl throws the flower in(to) the box
20	De bakker verstopt het boek onder de hoed	The baker hides the book under the hat
21	Het meisje verstopt de auto in de mand	The girl hides the car in the basket

22	De leeuw vindt de bloem in de mand	The lion finds the flower in the basket
23	De poes vindt het boek onder de hoed	The cat finds the book under the hat
24	De leeuw gooit de auto in de doos	The lion throws the car in(to) the box
25	De bakker vindt de bloem onder de hoed	The baker finds the flower under the hat
26	Het meisje vindt het boek in de mand	The girl finds the book in the basket
27	Het meisje vindt de auto in de doos	The girl finds the car in the box
28	De poes verstopt de bloem in de doos	The cat hides the flower in the box
29	De bakker vindt het boek in de mand	The baker finds the book in the basket
30	De hond vindt de auto onder de hoed	The dog finds the car under the hat

Nederlandse samenvatting

Om efficient to kunnen communiceren, moeten kinderen hun uitingen aanpassen aan de gemeenschappelijke kennis die ze delen met hun conversatiepartners. Eén manier om dit te doen is door gefocuste informatie prosodisch te benadrukken. Dit proefschrift is een cross-linguïstisch onderzoek naar hoe Nederlands- en Zweedstalige kinderen het vermogen ontwikkelen om gefocuste informatie prosodisch te markeren in hun spraakproductie.

Het begrip 'focus' is misschien wel één van de meest bestudeerde aspecten van informatieordening. Over het algemeen verwijst dit naar 'nieuwe' of 'belangrijke' informatie in een zin, die in contrast staat tot 'gegeven' of 'vaststaande' informatie (bv. Krifka & Musan, 2012). Wereldwijd hebben talen verschillende strategieën tot hun beschikking om focus te markeren, zoals syntactische alternanties, morfologische markeerders, en prosodie. De onderzoeken in dit proefschrift gaan allemaal over prosodische manieren om focus te markeren. In navolging van vele anderen gebruiken we de term 'prosodie' voor akoestische variatie in toonhoogte, duur en intensiteit, maar ook voor taalkundige begrippen zoals klemtoon, accenten, en tonen.

De keuze om de prosodische ontwikkeling van kinderen in het Standaard Nederlands en in het Centraal Zweeds te vergelijken is niet toevallig. Het Centraal Zweeds (CS) is een taal met lexikale toonhoogteaccenten, waarin toonhoogte contrastief werkt op het woordniveau. Het Standaard Nederlands (Nederlands) is een zogenaamde 'alleen intonatie'-taal, waarin toonhoogte niet onderscheidend werkt (Yip, 2002; Hyman, 2009). Prosodie, en toonhoogte in het bijzonder, wordt in beide talen gebruikt om focus te markeren, maar de preciese prosodische mechanismen die daarbij een rol spelen verschillen tussen de talen. Het bestuderen van de ontwikkeling van het volwassen gebruik van prosodische focusmarkering in het CS en het Nederlands is dan ook een manier om vast te stellen of en hoe systematische verschillen tussen de prosodische systemen het ontwikkelingspad naar volwassen competentie beïnvloeden. Tegelijkertijd stelt de vergelijking van de prosodische ontwikkeling in deze twee talen ons ook in staat om te bepalen of er algemene ontwikkelingspatronen te zien zijn in kinderen die verschillende prosodische systemen leren.

Dit proefschrift heeft een experimentele benadering van de studie van prosodische verwerving. Een plaatjescombineerspel werd ontwikkeld, waarin de taak van het kind was om de experimentator te helpen de juiste combinaties van plaatjes te

vinden. In het spel waren vraagwoorden ingebed, zodat de informatiestructuur van de subject-werkwoord-objectzinnen die de kinderen produceerden gecontroleerd was. Het spel werd gespeeld met vier leeftijdsgroepen in elke taal: vier- tot vijfjarigen, zeven- tot achtjarigen, tien- tot elfjarigen en volwassenen. Voor het gemak noemen we deze groepen de vijfjarigen, achtjarigen en elfjarigen. In dit proefschrift worden de resultaten van vier onderzoeken besproken: twee over de verwerving van focusprosodie in het CS en twee over de verwerving van focusprosodie in het Nederlands. De leeftijdsgroepen en methodologie waren hetzelfde voor alle onderzoeken, maar de talen en metingen die gedaan werden, waren verschillend. De onderzoeken worden hieronder kort samengevat.

In Hoofdstuk 4 bespreken we een onderzoek naar de prosodie en focus in kinder-CS. Het doel van dit onderzoek was om te beschrijven hoe CS-sprekende kinderen tussen de vier en elf jaar oud het vermogen ontwikkelen om prominentie-H te gebruiken om focus te markeren. Prominentie-H wordt gezien als de belangrijkste indicator voor focus in het CS, maar er is nog geen eerder onderzoek gedaan naar hoe kinderen het gebruik verwerven van deze toon om gefocuste constituenten te markeren. Onze voornaamste bevinding in dit onderzoek was dat CS-sprekende kinderen al opvallend volwassen zijn in de manier waarop ze prominentie-H gebruiken op vier- tot vijfjarige leeftijd, en dat ze volwassen vaardigheid bereiken op de leeftijd van zeven tot acht jaar. Om precies te zijn verschilden geen van de kindergroepen significant van de volwassenen in de manier waarop ze prominentie-H gebruikten om nauwe focus in zinsfinale positie te markeren, maar in zinsmediale positie waren de vijfjarigen minder consistent dan de volwassenen, zelfs terwijl ze op dezelfde manier onderscheid maakten tussen gefocuste en niet-gefocuste doelelementen (woorden of frases). Ten tweede bevestigen de data van onze volwassen groep eerdere onderzoeken in de conclusie dat prominentie-H systematisch geassocieerd wordt met nauwe focus, en dat prominentie-H consistent vermeden wordt in post-focus positie. Onze vergelijking van het gebruik van prominentie-H in nauwe en wijde focus in zinsmediale positie wees uit dat al onze groepen prominentie-H significant meer gebruikten op doelelementen in nauwe focus dan op doelelementen die deel uitmaakten van een uiting met wijde focus. In zinsfinale positie maakten geen van onze groepen onderscheid tussen doelelementen in nauwe focus en doelelementen in wijde focus: allebei werden bijna maximaal geaccentueerd. In de vergelijking van doelelementen in nauwe focus en doelelementen in contrastieve focus in zinsmediale positie vonden we ook geen verschil tussen onze groepen. Ten slotte wees het bestuderen van het effect van het toevoegen of weglaten van prominentie-H op het toonhoogtebereik uit dat het gebruik van prominentie-H hetzelfde gevolg had op het toonhoogtebereik in alle vier de groepen: in zinsmediale positie vergrootte het gebruik van deze toon het toonhoogtebereik op woorden met accent 1, maar niet op woorden met accent 2. In

zinsfinale positie was er een soortgelijk patroon te zien, maar op deze positie was de interactie tussen focus en het lexikale accent niet significant.

In Hoofdstuk 5 bespreken we ons tweede onderzoek over prosodie en focus in het CS. Het doel van dit onderzoek was om vast te stellen hoe CS-sprekende kinderen tussen de vier en elf jaar oud duur en niet-modale stemkwaliteit gebruiken om focus te markeren. Eerder werk over prosodische focusmarkering in het CS heeft uitgewezen dat in de spraak van volwassenen focus een vrij robuust effect heeft op woordlengte (Heldner & Strangert, 1997; Heldner, 2001), maar het gebruik van woordlengte als focusmarkering door kinderen is nog niet eerder onderzocht. Onze oorspronkelijke reden voor het kijken naar stemkwaliteit was dat we regelmatig datapunten misten in onze toonhoogteanalyses, doordat post-focus constituenten werden geproduceerd met onregelmatige stemhebbendheidpatronen (voornamelijk met zogenaamde 'breathy voice' of 'creaky voice'). Voor dit onderzoek werd dezelfde dataset gebruikt als voor het onderzoek in Hoofdstuk 4. Onze belangrijkste bevindingen waren dat focus de duur van onze doelwoorden systematisch verlengde voor de volwassenen, de vijfjarigen en de zevenjarigen. Verrassend genoeg was dit effect anders voor de elfjarigen. Voor hen had focus een omgekeerd effect: de duur van woorden met accent 2 in zinsmediale positie werd verkort, evenals de duur van woorden met accent 1 of 2 in zinsfinale positie. Het resultaat van de oudste kindergroep interpreteren we als een taakeffect, het gevolg van het feit dat het plaatjescombineerspel minder interessant was voor kinderen van tien of elf. Een langzamer spraaktempo wordt vaak geassocieerd met verveling of gebrek aan interesse (bv. Burkhardt & Sendlmeier, 2000), en we geloven dat de enigszins verveelde elfjarigen hun spraaktempo verlaagden bij de productie van post-focus doelelementen die informatie bevatten die de experimentator al had. De resultaten van onze analyse van de stemkwaliteit wees uit dat post-focus status het gebruik van niet-modale stemkwaliteit deed toenemen bij de volwassenen en de elfjarigen, maar dit effect was zwakker voor de achtjarigen en bijna afwezig voor de vijfjarigen. Wat betreft het gebruik van niet-modale stemkwaliteit concluderen we dan ook dat dit voor CS-sprekende volwassenen een betrouwbaar correlaat is van focus, maar dat kinderen de stemkwaliteit in post-focus positie pas systematisch wijzigen vanaf acht jaar.

In Hoofdstuk 6 bespreken we de eerste van onze twee studies over prosodische focusmarkering in kinder-Nederlands. Het doel van dit onderzoek was om vast te stellen hoe Nederlandssprekende kinderen tussen de vier en elf jaar oud a) accentplaatsing, b) accenttype, en c) algemene manipulaties van toonhoogte en duur gebruiken om focus te markeren. Wat betreft accentplaatsing deden alle kindergroepen hetzelfde als de volwassenen in zinsfinale positie, waar gefocuste constituenten vaker geaccentueerd werden dan post-focus constituenten. In

zinsmediale positie maakten de vijf- en achtjarigen echter geen onderscheid tussen gefocuste en post-focus constituenten in hun accentplaatsing, in tegenstelling tot de volwassenen en de elfjarigen. Een vergelijking tussen nauwe, contrastieve en wijde focus liet zien dat alle groepen maximaal accentueerden in elk van deze drie condities. Er werden geen verschillen gevonden tussen de leeftijdsgroepen. Wat betreft het gebruik van verschillende accenttypes vonden we dat onze vijfjarigen soms zogenaamde 'non-downstepped falls' gebruikten in post-focus positie, terwijl de oudere kinderen en de volwassenen dit vermeden. Over het algemeen lieten zowel de vijf- als achtjarigen meer variatie zien in de distributie van hun accenttypes dan de volwassenen en de elfjarigen, maar net als bij accentplaatsing lieten de kinderen een volwassen gebruik van accenttypes eerder zien in zinsfinale positie dan in zinsmediale positie. Onze fonetische analyses moeten worden bevestigd in een grotere dataset, maar onze voorlopige bevinding duidt erop dat er verschillen zijn tussen de groepen wat betreft de manier waarop nauwe focus verschillende aspecten van de toonhoogte en duur beïnvloedde. Focus had zelden een systematisch effect op duur in onze groepen, terwijl het de maximum toonhoogte vaak beïnvloedde, soms zelfs in groepen die accentplaatsing of accenttypes niet op een systematische manier gebruikten.

In Hoofdstuk 7 bespreken we onze tweede studie naar prosodie en focus in het Nederlands, waarin we het gebruik van pauzes voor focus door volwassenen en vijfjarigen bestudeerden. Om precies te zijn was onze onderzoeksvraag of Nederlandse kinderen van deze leeftijd, wiens beheersing van de gebruikelijke indicatoren (d.w.z. de plaatsing en het type van toonhoogteaccenten) nog onder ontwikkeling is, pauzes inlassen om gefocuste constituenten te benadrukken. In de onderzoeken die werden besproken in Hoofdstuk 4-6 bestonden de data uit subject-werkwoord-objectzinnen. In dit onderzoek werd de procedure uitgebreid met iets ingewikkeldere zinsstructuren, namelijk subject-werkwoord-object-bijwoordzinnen (SWOB), zodat het effect van syntactische complexiteit op de pauzepatronen van kinderen kon worden bestudeerd. De belangrijkste bevinding van dit onderzoek was dat Nederlandse vijfjarigen pauzes inlassen of verlengen voorafgaand aan gefocuste constituenten. Ook al was dit effect het meest systematisch op sterkere syntactische grenzen (bv. tussen een subject en een werkwoordsfrase), het werd ook gevonden op zwakkere grenzen (tussen een determinator en een naamwoord). Volwassenen daarentegen bleken pauzes alleen te gebruiken voor focus op de sterkste syntactische grenzen, en in mindere mate dan de kinderen.

Op basis van deze vier onderzoeken concluderen we dat CS-sprekende kinderen iets voorlopen op hun Nederlandssprekende leeftijdsgenoten bij de verwerving van de prosodische manipulaties die gebruikt worden om focus te markeren in hun taal. Hoewel de verhoogde toongevoeligheid van de toonlerende CS-kinderen een

mogelijke verklaring hiervoor is, wijzen de data van de Nederlands- en CS-sprekende volwassenen er ook op dat de consistentie waarmee accentplaatsing en accenttypes gebruikt worden om focus te markeren in het Nederlands verschilt van de consistentie waarmee prominentie-H gebruikt wordt in het CS. Met andere woorden, prosodische verschillen tussen het Nederlands en het Zweeds die verder reiken dan de aan- of afwezigheid van lexikale tonen kunnen ook van invloed zijn op de weg naar een volwassen gebruik van prosodische focusmarkering in deze twee talen.

Curriculum vitae

Anna Sara Hexeberg Romøren was born in 1985 in Kristiansand, Norway. After finishing high school in Kristiansand, she enrolled in a BA programme in Spanish at the University of Oslo. Even if she enjoyed taking courses on Latin-American history, Spanish theatre traditions and the grammar of Brazilian Portuguese, her interest in general linguistics made her take additional courses in linguistics, and after this to enroll in the MA programme in linguistics offered at UiO. A year of working as an assistant teacher sparked her interest in language disorders in children, shaping the topic of her MA thesis *Steady steps on tonal feet: Tonal accent production in children with and without high functioning autismon the acquisition of the Norwegian lexical accent contrast in children with and without high-functioning autism*. While writing her thesis, she also worked as a research assistant at the Norwegian Institute of Public Health, on a population-based study on the development of autism in children (the ABC-study). After submitting her MA thesis in 2011 she was offered a PhD position at Utrecht University. This dissertation is the result of work that Anna Sara carried out between 2011 and 2015 as a PhD-candidate at UiL OTS, Utrecht University. During her PhD she presented her work at several international workshops and conferences. She also spent several periods in Sweden as a visiting researcher, both at KTH in Stockholm and at Lund University. Since spring 2016 Anna Sara holds a position as Lecturer/ Assistant Professor at Oslo and Akershus University of Applied Sciences (HIOA), continuing her research on language acquisition, combined with teaching various courses in the Department of Early Childhood Education.