

# Quiet is the New Loud: Pausing and Focus in Child and Adult Dutch

Language and Speech

2015, Vol. 58(1) 8–23

© The Author(s) 2015

Reprints and permissions:

sagepub.co.uk/journalsPermissions.nav

DOI: 10.1177/0023830914563589

las.sagepub.com

**Anna Sara H Romøren**

Utrecht University, The Netherlands

**Aoju Chen**

Utrecht University, The Netherlands;

Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

## Abstract

In a number of languages, prosody is used to highlight new information (or focus). In Dutch, focus is marked by accentuation, whereby focal constituents are accented and post-focal constituents are de-accented. 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. Considering that Dutch-speaking 4 to 5 year olds are not yet completely proficient in using accentuation for focus and that children generally pause more than adults, we asked whether pausing might be an available parameter for children to manipulate for focus. Sentences with varying focus structure were elicited from 10 Dutch-speaking 4 to 5 year olds and 9 Dutch-speaking adults by means of a picture-matching game. Comparing pause durations before focal and non-focal targets showed pre-target pauses to be significantly longer when the targets were focal than when they were not. Notably, the use of pausing was more robust in the children than in the adults, suggesting that children exploit pausing to mark focus more generally than adults do, at a stage where their mastery of the canonical cues to focus is still developing.

## Keywords

Dutch, focus, language acquisition, pause, prosody

## Introduction

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

### Corresponding author:

Anna Sara H Romøren, Utrecht Institute of Linguistics OTS, Utrecht University, Trans 10, Utrecht, The Netherlands.

Email: a.s.h.romoren@uu.nl

(Gee & Grosejan, 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 (Redford, 2013), we asked whether pausing may be an available parameter for young children to use in focus marking. 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 rest of the introduction consists of three subsections. In the first subsection we discuss some basic notions of information structure, and in the second subsection we review earlier work on pausing in the speech of adults and children. In the third subsection we briefly describe prosodic focus marking in adult Dutch, before summarizing past work on prosodic focus marking in English, German and Dutch-speaking children. In the Methodology section, we describe the picture-matching game that we used to elicit sentence production with varying focus structure, along with the procedures for extracting and analysing the speech data gathered. Finally, we present the results of our analyses, discussing how they provide new insight into the developmental path to prosodic focus marking in Dutch.

### 1.1 Information structure and focus

Theories of ‘information structure’ or ‘information packaging’ (Chafe, 1976; Halliday, 1967) treat 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. What is assumed as part of the common ground is continuously updated through the course of a conversation, and this has consequences for the packaging speakers decide to use, for example in their choice of referring expressions, syntactic structures or prosodic patterns. The current study concerns whether pauses tend to be longer before constituents referring to information that is added to the common ground (e.g., new) as opposed to information that is already present in the common ground (e.g., given).

Gundel and Fretheim (2004) distinguish two different dimensions of givenness–newness relations, namely ‘referential’ versus ‘relational’ newness–givenness. Whereas the ‘referential’ level describes a relation between a referent and a non-linguistic entity in the speaker’s or hearer’s mind (as in the case of referring expressions), ‘relational’ givenness–newness describes a relation that applies *within* a sentence (as in theme–rheme, topic–comment or focus–given dichotomies; Krifka & Musan, 2012). At the relational level, the conceptual representation of a sentence is divided into two complimentary parts, *X* and *Y*, where *X* is what the sentence is about and *Y* is what is said about *X*.

In the following, we will use the term ‘information status’ to refer to referential givenness–newness, and the term ‘information structure’ to refer to relational givenness–newness, following Vallduví and Engdahl (1996). Furthermore, we will refer to the *Y* of Gundel and Fretheim (2004) as the ‘focus’ of a sentence. In our experiment, we manipulate the information structure of elicited target sentences through the use of *wh*-questions, rendering initial, medial or final constituents focal and the rest of the constituents non-focal. This kind of question–answer paradigm is frequently applied in studies of prosodic focus marking, as it is seen as a relatively straightforward way to control the information structure of elicited responses (Roberts, 1996).

Another notion frequently appearing in the discussion of information packaging is ‘contrast’, which can apply to both focal and topical referents (Molnár, 2002). In the following, we will use the term ‘contrastive focus’ when referring to cases where alternative candidates are explicitly mentioned in the preceding context, as illustrated in Table 1.

**Table 1.** Example of context rendering the final constituent contrastively focal.

	Dutch (original)	English
<b>Experimenter:</b>	Kijk, de hond! Het lijkt net alsof de hond iets kookt. Ik doe een gok: de hond kookt de laars.	Look, the dog! It looks like the dog is cooking something. I'll make a guess: the dog is cooking the boot.
<b>Child:</b>	De hond kookt. [ <b>DE WORTEL</b> ] <small>contrastive focus</small>	The dog is cooking. [ <b>THE CARROT</b> ] <small>contrastive focus</small>

## 1.2 Pausing and information structure

Pausing in adult speech production has been a popular topic in the last 60 or so years (Ferreira, 2007; Wagner & Watson, 2010; Zellner, 1994). Particularly relevant for our study are reports that speakers tend to pause longer when adding new information to a narrative (Gee & Grosejan, 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). The finding that adults pause to single out new information in 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 later.

The first paper to be discussed comes from Dahan and Bernard (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, and frequencies and durations of pauses preceding and following the target words were extracted and compared between ‘emphatic’ and ‘not emphatic’ conditions.<sup>1</sup> 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 and Lee (2007) for Cantonese. In this 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 in one of the speakers in Dahan and Bernard’s (1996) study, Gu and Lee (2007) also reported on pauses occasionally being inserted before the focal constituent. Finally, Huang and Liao (2002) similarly postulated that pauses could be used for highlighting certain constituents in Mandarin Chinese.

In the studies by Dahan and Bernard (1996) and Gu and Lee (2007), the pre-target pauses occasionally occurred simultaneously 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 in one speaker to pauses being inserted independently of plosives. 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, and O’Connell (1983), stop-closures of consonants can vary between 80 and 250 ms (as shown by Dalton and 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 perceivability of a pause varies substantially depending on the speech context in which it appears (Rochester, 1975). Investigating the origins of shorter pauses in read-aloud poems and

political speeches, Hieke et al. (1983) found that most pauses ranging between 130 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.<sup>2</sup> In a more recent cross-linguistic study, Campione and Véronis (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 & Véronis, 2002).

Whereas pausing has received quite a lot of attention in research on adult speech, pausing in the language of children is studied less often (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 irrelevant for describing children's linguistic knowledge (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 5 year olds to those of adults. In addition to 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 its noun phrase). Redford (2013) 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!'*). 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 in ungrammatical locations might suggest that adults also pause to emphasize in English, as reported for French and Chinese.<sup>3</sup>

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 hypothesized 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 noun phrase which it dominates) to the strongest ones (between the subject noun phrase and the verb phrase). Narratives were elicited from 5 year olds, 7 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 in pausing the most at the strongest boundary (i.e., subject-verb phrase), but they behaved differently at the medium-strength boundary between the head verb and its argument noun phrase, where the children paused much more often than the adults. Maloney et al. (2012) 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 et al. (2012) 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 the speech of adults and children 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 occurred simultaneously 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).

### 1.3 Prosodic focus marking in adult and child language

In West Germanic languages, focus is predominantly marked using prosody. In Dutch, this is done by accenting focal information, often leading to expanded pitch range and increased duration on the accented word (Chen, 2009, 2011a, 2011b; Gussenhoven, 1984; Hanssen, Peters, & Gussenhoven, 2008). Speakers can use a range of different pitch accent types to mark focus (e.g., fall '*H\*L*', rise '*L\*H*', sustained high pitch '*H\**' or sustained low pitch '*L\**'), but the most frequent pattern is the falling pitch accent '*H\*L*', regardless of sentence position (Chen, 2007). Non-focal constituents are predominantly de-accented post-focally, but in sentence-initial position they are nearly always accented, mostly with the same fall ('*H\*L*') that is also used for focus. In this case, focal falls are phonetically distinguished from non-focal ones by being produced with a larger pitch range (mainly due to a lowering of the low tonal target) and longer duration (Chen, 2009).

Dutch children have been shown to accent focal information and de-accent post-focal information at the age of 4 or 5, in line with what is described for adults (Chen, 2007, 2009, 2011a, 2011b). However, a closer look at the children's accentual patterns reveals differences between the two groups. First, the adults showed a preference for falls (*H\*L*) or downstepped falls (!*H\*L*) for marking final focus, whereas the children's accent choices were more variable, with a large proportion of rising (*L\*H*) accents. Second, in sentence-initial position, adults distinguished focal from non-focal falls by means of pitch range and duration, but the children did not do this (Chen, 2009; see also Romøren & Chen, 2014).

A few words can be added about prosodic focus marking in English- and German-speaking children. A series of studies have shown English 3 to 4 year olds to use accentuation, pitch and intensity to distinguish contrastive from given information (Hornby & Hass, 1970; MacWhinney & Bates, 1978; Wieman, 1976; Wonnacott & Watson, 2008), but there are also reports of further development towards the age of 6 (MacWhinney & Bates, 1978) and even 13 (Wells, Peppé, & Goulandris, 2004). A paper on prosodic focus marking in German-speaking children showed 4 to 5 year olds to produce new and contrastive referents with a higher mean pitch than previously mentioned ones (Müller, Höhle, Schmitz, & Weissenborn, 2006), but another investigation described non-adult-like accent choices in 5 to 7 year olds (De Ruiter, 2009). This final finding is similar to what was reported for Dutch-speaking children (Chen, 2011b).

As can be seen from this brief review, children can make prosodic adjustments to mark contrast, as well as differences between relational givenness–newness when newness simultaneously occurs with contrastivity, at the age of 4 or 5 (Chen, 2014). Further, children's ability to mark focus is still developing beyond this age, especially regarding choice of accent type and the use of phonetic cues when accent placement and choice of accent type do not suffice for this purpose (Chen, 2009, 2011a, 2011b; De Ruiter, 2009). Against this background, we asked whether pausing might be an additional parameter available for children to use for focus marking.

## 2 Methodology

In our experiment, we used a game setting to simulate natural mini-conversations about a restricted set of referents. The information structure of target sentences was manipulated by explicitly presenting relationally given referents and asking *wh*-questions about relationally new ones. Given that all referents were introduced in a picture-naming task preceding the experiment proper, the referents can be considered referentially accessible, following Chafe (1987) and Lambrecht (1994).

## 2.1 Participants

Ten Dutch-speaking children (six boys, four girls, range: 4;4–4;11, mean 5;2) and nine female Dutch-speaking adults (mean 23;10) participated in the study. All participants were native speakers of standard 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 the Utrecht Institute of Linguistics. They were all university students, but none of them was studying linguistics at the time of testing.

## 2.2 Procedure and materials

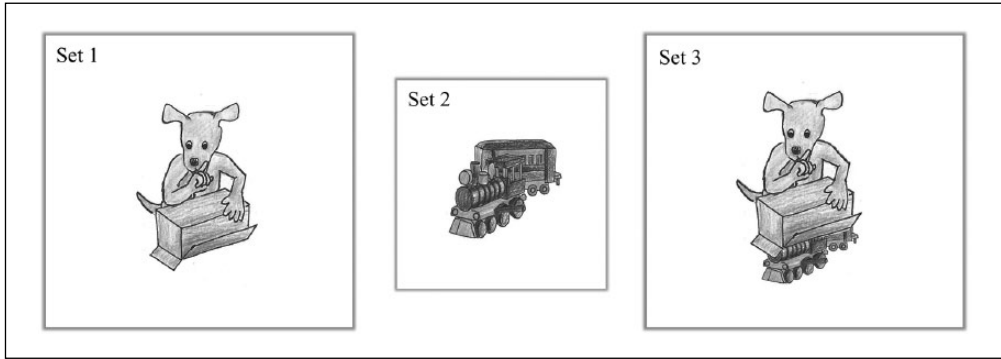
All participants were tested individually in a quiet room; the children in a designated test room at their school and the adults in a sound attenuated booth at the Linguistics Lab at Utrecht University. Two female experimenters were trained to do the testing according to detailed instructions, and all 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. Subject–verb–object (SVO) and subject–verb–object–adverbial (SVOA) sentences were elicited through an interactive picture-matching game, adopted from Chen (2011a).

The choice to include both SVO and SVOA sentences 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. In a recent study, we found Dutch-speaking children to be less consistent in accenting focal 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 was that they lie well within the syntactic complexity 4 to 5 year olds can handle, that they were easy to construct and illustrate using child-friendly words, and that they could be integrated into the game following the same structure that was used for the SVO sentences.

The picture-matching game was preceded by a picture-naming task. Detailed instructions were created for both tasks, 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 picture-matching game. We also made conventions for the intonation pattern to be used by the experimenter, making sure that each trial and each session was conducted in the same manner.

**2.2.1 The picture-naming task.** The picture-naming task was constructed to familiarize the participants with the nouns appearing in the picture-matching game, in order for them to use the intended words when playing the game. In the picture-naming task, the participants were instructed to name figures and objects illustrated in 17 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 cat a dog), the experimenter explained what the relevant figure/object 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 dog; it’s a cat. Do you see the whiskers?*’). The target verbs were not a part of the picture-naming game, but were presented, illustrated and explained in the introduction to the game (e.g., ‘*Look, this is “finding”, and when someone finds something they always look happy.*’).

**2.2.2 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



**Figure 1.** One picture from each set, representing the sentence 'The dog is hiding *THE TRAIN*'.

questions about her pictures. Scripted contexts were created for all experimental trials to make the focal elements relationally new and the non-focal elements relationally given, following the terminology of Gundel and Fretheim (2004). In terms of referential newness–givenness, the baseline was that all target referents were made accessible (Lambrecht, 1994), both through the picture naming and through repeated mention during the course of game.

The materials consisted of three separate sets of pictures, two for the experimenter and one for the participant (see Figure 1). The experimenter's first set (set 1) was piled face down in front of her. These pictures always lacked one constituent, for example, the subject, the verb, the object or the adverbial. 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 always pre-ordered before each session, so that corresponding pictures always appeared in the same trial.

Each trial was conducted as follows: the experimenter first picked up a picture from set 1, drawing the participant's attention to it, uttering the context sentences as illustrated in Table 2. After the target question was asked, the participant could look at his/her complete picture in order to answer the question. Once the answer was provided, the experimenter could look for the 'missing piece' of 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 in the context and target questions, consisting in a falling accent ( $H^*L$ ) on 'look' as well as on the nouns and verbs, when these were introduced for the first time. In the questions, the experimenter used the same falling accent ( $H^*L$ ) on the *wh*-word, and no accent on the following words.

The game consisted of 24 test trials and 8 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 four sentence conditions, namely narrow focus on the initial constituent (NF-i), narrow focus on the medial constituent (NF-m), narrow focus on the final constituent (NF-f) and contrastive focus on the medial constituent (CF-m) (see Tables 3 and 4). The SVO and SVOA parts were each preceded by four practice trials, one from each sentence condition.

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.

**Table 2.** Example of conversational exchange for the trial represented in Figure 1.

	Dutch (original)	English
<b>Experimenter:</b>	Kijk! Een hond. Het lijkt net of de hond iets verstopt. Wat verstopt de hond?	Look, a dog! It looks like the dog is hiding something. What is the dog hiding?
<b>Child:</b>	De hond verstopt. [DE TREIN] <small>focus</small>	The dog is hiding. [THE TRAIN] <small>focus</small>

**Table 3.** Example of trial context for the four sentence conditions, subject–verb–object (SVO).

Sentence condition	Example context/question
<b>Narrow focus on initial constituent (NF-i)</b>	Look, the carrot! It looks like someone is drawing the carrot. <b>Who is drawing the carrot?</b>
<b>Narrow focus on medial constituent (NF-m)</b>	Look, the carrot! And there is also a girl. It looks like the girl is doing something with the carrot. <b>What is the girl doing with the carrot?</b>
<b>Narrow focus on final constituent (NF-f)</b>	Look, the girl. It looks like the girl is drawing something. <b>What is the girl drawing?</b>
<b>Contrastive focus on medial constituent (CF-m)</b>	Look, the carrot! And there is also a girl. It looks like the girl is doing something with the carrot. I'll guess: <b>the girl IS COOKING the carrot.</b> (What do you say?)

**Table 4.** Example of trial context for the four sentence conditions, subject–verb–object–adverbial (SVOA).

Sentence condition	Example context/question
<b>Narrow focus on initial constituent (NF-i)</b>	Look, the flower! And there is also the basket. It looks like someone is throwing the flower into the basket. <b>Who is throwing the flower into the basket?</b>
<b>Narrow focus on medial constituent (NF-m)</b>	Look, the baker! And there is also the basket. It looks like the baker is throwing something into the basket. <b>What is the baker throwing into the basket?</b>
<b>Narrow focus on final constituent (NF-f)</b>	Look, the flower! And there is also the baker. It looks like the baker is throwing the flower into something. <b>Where is the baker throwing the flower?</b>
<b>Contrastive focus on medial constituent (CF-m)</b>	Look, the baker! And there is also the basket. It looks like the baker is throwing something into the basket. I'll guess: <b>the girl is throwing THE CAKE into the basket.</b> (What do you say?)

We also spread five initial constituents over the four conditions. When creating and ordering the stimuli, we made sure that each combination of initial, medial and final constituent only occurred once in the whole set. Furthermore, two consecutive trials never realized the same condition and always differed by a minimum of two constituents. Following these constraints, the experimental trials were arranged into two different stimulus orders. Because 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.



### 2.3 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. The choice of being rather strict in the inclusion of responses was made in order to make sure that the prosodic comparisons were made across the same words or phrases, and that the experimental conditions were properly controlled for. Furthermore, 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). As a consequence of our strict inclusion criteria, the average response inclusion rate was 65% (range 40.0–86.7) in the children, and 92.2% (range 83.3–100) in the adults. Among the excluded responses from the children, 33 were excluded because 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). Thirty-five 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, eight responses were excluded because of laughter, background noise or other disturbances making the recordings unfit for analysis. The final dataset from both groups consisted of 188 SVO sentences and 176 SVOA sentences.

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, the boundary between *de* and *hoed* was segmented at the onset of friction).

A pause was defined as a between-word interval of any duration with no or insignificant amplitude.<sup>4</sup> Pauses were coded by combining the automatic silence detection function from Praat (minimum silence threshold 25 dB, 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. 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 also to 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 occur simultaneously with plosives.

We investigated pause durations related to medial and final target constituents. The between-word boundaries where pauses were measured are illustrated in Figure 2. In the SVO sentences, the medial targets were verbs and the final targets were object noun phrases (hereafter NPs). In the SVOA sentences, the medial targets were object NPs and the final targets were adverbial prepositional phrases (hereafter PPs). Large square brackets mark between-constituent boundaries while small horizontal brackets mark within-constituent boundaries. Next, we will refer to comparisons of pause durations preceding medial targets as the medial analysis, and comparisons of pause durations preceding final targets as the final 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.

### 3 Analysis and results

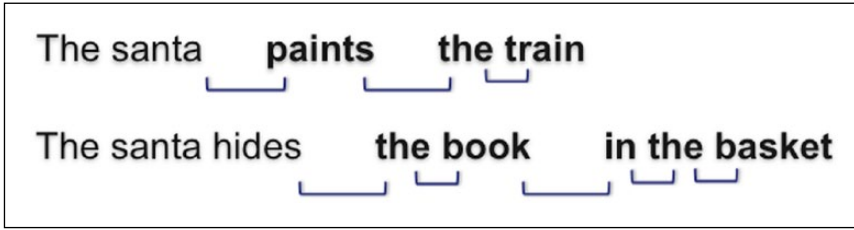
Previous investigations of prosodic focus marking in adult speech have revealed only subtle differences between contrastive and narrow focus in adult Dutch (Hanssen et al., 2008). Similarly, we found no significant differences in pause durations either before or within target phrases when comparing between the CF-m and the NF-m condition. Based on these results we decided to collapse the NF-m and CF-m conditions in the rest of the analysis, in order to include as many data points as possible. The no focus condition contained all the sentence conditions that did *not* render a specific target constituent focal, for example, NF-i and NF-f for medial comparisons and NF-i, CF-m and NF-m 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 was used to assess the effect of focus on pause durations before and within medial and final target constituents, with the factors ‘focus’ (two levels: focus vs. no focus) and ‘group’ (child vs. adult) as fixed factors and ‘participant’ and ‘item number’ as random factors.

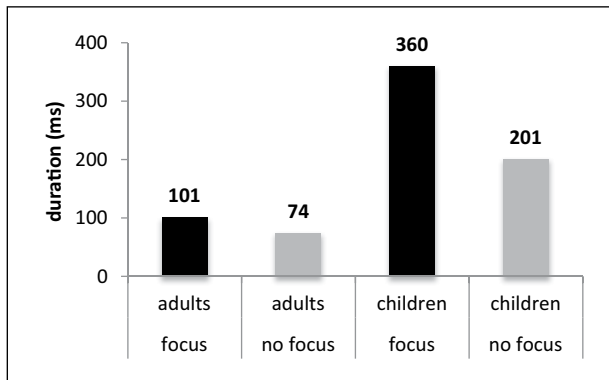
Each analysis was run using the *lmer4* package in R. We started out with a baseline model (hereafter model 0) in which only the random factors were included. From this starting point, we extended the model in a stepwise fashion by first adding the factor ‘focus’ in model 1, then adding the factor ‘group’ in model 2 and finally adding the interaction between ‘focus’ and ‘group’ in model 3. Only factors that significantly improved the previous model were included in subsequent models. In order to assess the improvement of the model fit from models 0 through 3, we used R’s ‘ANOVA’ function to compare pairs of models. A *p*-value below 0.05 in the model comparison was taken to indicate that the model with the added parameter (main effect or interaction) fit the data significantly better than a model without this parameter. This was then taken as evidence that the parameter had a significant effect on the outcome variable, that is, the pause duration at a certain location. 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 done 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 Figure 2). We will first report the results from the analysis of the SVO sentences, and then present the results from the SVOA sentences.

#### 3.1 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 (see Figure 2). With respect to the boundary preceding the verbs, adding the factor ‘focus’ significantly improved the 0 model ( $p = 0.034$ ), as



**Figure 2.** Sentences, target constituents and potential pause locations in the two sentence types.



**Figure 3.** Pre-medial pause durations by focus and group, subject–verb–object (SVO).

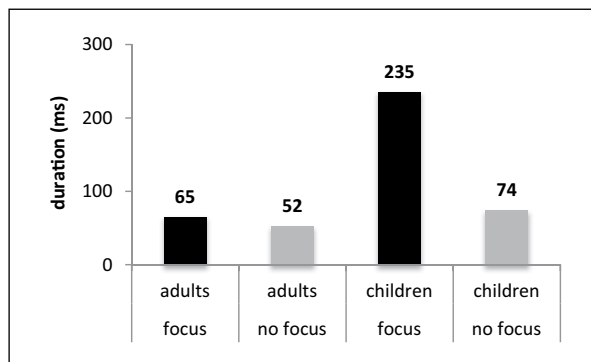
did adding ‘group’ to the model with ‘focus’ ( $p = 0.005$ ) and adding the interaction ‘focus  $\times$  group’ to the model with main effects only ( $p = 0.049$ ). Re-running the models on the data split by group showed that focus on the verb significantly increased the pause duration in both adults and children (children:  $p = 0.054$ , adults:  $p = 0.000$ ), but the increase was larger in the children than in the adults. Mean pause durations split by group and focus condition are presented in Figure 3.

The analysis of pauses in the sentence final position was done 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’ nor the interaction ‘focus  $\times$  group’ came out as significant predictors for pause durations in these locations; thus, neither children nor adults varied pause duration according to focus in final position in the SVO sentences.

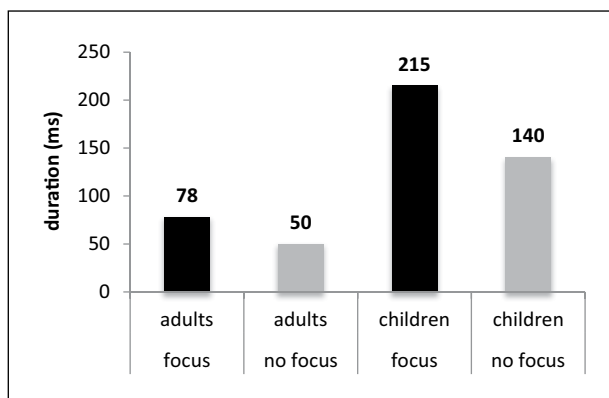
### 3.2 SVOA

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 2).

With respect to the pause durations preceding the medial object NPs (Figure 4), adding ‘focus’ to the baseline model did not significantly improve it ( $p = 0.443$ ). However, both ‘group’ and the interaction ‘focus  $\times$  group’ significantly improved the previous models (group:



**Figure 4.** Pre-medial pause durations by focus and group, subject-verb-object-adverbial (SVOA).



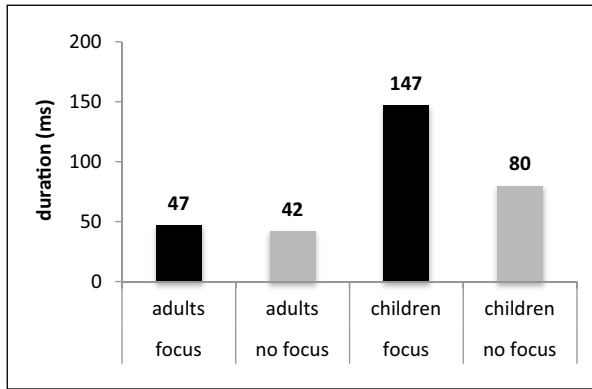
**Figure 5.** Pre-final pause durations (before final PP), subject-verb-object-adverbial (SVOA).

$p = 0.010$ , group  $\times$  focus:  $p = 0.042$ ). Re-running the models split by group revealed a significant effect of focus in the children's data ( $p = 0.011$ ), but no effect in the adult data ( $p = 0.352$ ) (see Figure 4).

The analysis of pause durations before the final target PP revealed a main effect of 'focus' ( $p = 0.051$ ), a main effect of 'group' (0.000) and an interaction effect between 'focus' and 'group' ( $p = 0.005$ ) for the pause durations preceding the PP (see Figure 5). The follow-up analysis split by group showed that there was a significant effect of focus on the pre-noun pauses in both groups, but that the effect was stronger in the children.

We also ran models examining the effect of focus on the pause preceding the NP within the PP, but there were no significant effects of 'focus' ( $p = 0.257$ ), 'group' ( $p = 0.345$ ) or the interaction 'focus  $\times$  group' ( $p = 0.201$ ). However, preceding the final noun *within* the PP, main effects of 'focus' ( $p = 0.030$ ) and 'group' ( $p = 0.019$ ) were found, but no effect of the interaction between 'focus' and 'group' was found ( $p = 0.139$ ) (see Figure 6).

Both groups paused much less frequently in this location than they did in the other ones, but where a pause was observed it was systematically longer when the PP was focal than when it was not. Even if we see in Figure 6 that the effect of focus is hardly present in the adult data, this interaction did not reach significance.



**Figure 6.** Within-final pauses (before final noun of the PP), subject–verb–object–adverbial (SVOA).

## 4 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 and Bernard (1996), Gu and Lee (2007) and Huang and Liao (2002). Different from the children, the adults tended to avoid pausing at weaker syntactic junctures (e.g., at the boundary between the verb and its internal argument), similar to the findings from 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 (Chen, 2011a; Romøren & Chen, 2014).

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 (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 a weaker syntactic boundary, indicating that children are less constrained by syntax than are 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, as the preposition is mentioned in the scripted context. However, the effect of focus on the pauses before the PP suggests that the participants did treat this phrase as a focal constituent.

Furthermore, 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 (Maloney et al., 2012). In the entire 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 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 (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 (Dahan & Bernard, 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 (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. However, the lexical accessibility of the limited set of targets included in the game should generally be high, as they are all introduced in the naming task in the introduction to the game, as well as repeated randomly across trials. Furthermore, 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, different from what tends to be the case in priming studies.

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. Because the average pause durations found were sometimes relatively short, choosing a cut-off point like the 250 ms threshold suggested by Goldman–Eisler (1968) would most likely have caused us to miss the patterns we observed. 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.

## Acknowledgements

We are grateful to research assistants Paula Cox, Frank Bijlsma, Saskia Verstegen and Martine Veenendaal for invaluable help with experiment design, data collection and coding. We also thank the five anonymous reviewers for their useful comments and suggestions.

## Funding

This work was supported by The Netherlands Organization for Scientific Research (NWO), grant number 276-89-001, awarded to the second author.

## Notes

1. 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.
2. Hieke et al. (1983) 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).
3. 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).
4. Sometimes there was background noise or breathing noises in the recordings, giving rise to some minor energy distributions in the spectrogram.

## References

- Boersma, P., & Weenink, D. (2010). PRAAT: Doing phonetics by computer (version 5.1.25) [computer program]. Retrieved from <http://www.praat.org/>
- Butcher, A. (1981). Aspects of the speech pause: Phonetic correlates and communicative functions (Arbeitsberichte Nr. 15). Institut für Phonetik, Universität Kiel, Germany.
- Campione, E., & Véronis, J. (2002). A large-scale multilingual study of silent pause duration. *Proceedings from Speech Prosody 2002*, Aix-en-Provence, France, 2002.
- Chafe, W. (1976). Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In C. Li (Ed.), *Subject and topic*. New York, NY: Academic Press.
- Chafe, W. (1987). Cognitive constraints on information flow. In R. Tomlin (Ed.), *Coherence and grounding in discourse*. Amsterdam, The Netherlands: John Benjamins.
- Chen, A. (2007). Intonational realisation of topic and focus by Dutch-acquiring 4-to-5-year-olds. *Proceedings of the International Congress of Phonetic Sciences XCI*, Saarbrücken, Germany, 2007.
- Chen, A. (2009). The phonetics of sentence-initial topic and focus in adult and child Dutch. In M. C. Vigario, S. Frota & M. João Freitas (Eds.), *Phonetics and phonology: Interactions and interrelations* (pp. 91–106). Amsterdam, The Netherlands: John Benjamins.
- Chen, A. (2011a). The developmental path to phonological focus-marking in Dutch. In S. Frota (Ed.), *Prosodic categories: Production, perception and comprehension*. Studies in natural language and linguistic theory 82 (pp. 93–110). London, UK: Springer Science and Business Media B.V.
- Chen, A. (2011b). Tuning information packaging: Intonational realization of topic and focus in child Dutch. *Journal of Child Language*, 38, 1055–1083.
- Chen, A. (2014). Children's use of intonation in reference and the role of input. In L. Serratrice & S. Allen (Eds.), *The acquisition of reference*. Amsterdam, The Netherlands: John Benjamins.
- Dahan, D., & Bernard, J. (1996). Interspeaker variability in emphatic accent production in French. *Language and Speech*, 39(4), 341–374.
- Dalton, P., & Hardcastle, W. J. (1977). *Disorders of fluency and their effects on communication*. London, UK: Elsevier.
- De Ruiter, L. E. (2009). The prosodic marking of topical referents in the German 'Vorfeld' by children and adults. *The Linguistic Review*, 26(2/3), 329–354.
- Ferreira, F. (2007). Prosody and performance in language production. *Language and Cognitive Processes*, 22(8), 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(10), 1736–1762.
- Gee, J. P., & Grosejan, F. (1984). Empirical evidence for narrative structure. *Cognitive Science*, 8, 59–84.
- Goldman-Eisler, F. (1968). *Psycholinguistics: Experiments in spontaneous speech*. New York, NY: Academic Press.
- 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 2007*, Saarbrücken, Germany, 2007.

- Gundel, J. K., & Fretheim, T. (2004). Topic and focus. In L. R. Horn & G. Ward (Eds.), *The handbook of pragmatics*, (pp. 175–196). Cornwall, UK: Blackwell.
- Halliday, M. A. K. (1967). Notes on transitivity and theme in English (II). *Journal of Linguistics*, 3, 199–244.
- Hanssen, J., Peters, J., & Gussenhoven, C. (2008). Prosodic effects of focus in Dutch declaratives. *Proceedings of the 4th International Conference on Speech Prosody*, Campinas, Brazil, 2008.
- Hieie, A. E., Kowal, S., & O'Connell, D. C. (1983). The trouble with “articulatory” pauses. *Language and Speech*, 26(3), 203–214.
- Hornby, P. A., & Hass, W. A. (1970). Use of contrastive stress by preschool children. *Journal of Speech and Hearing Research*, 13, 359–399.
- Huang, B., & Liao, X. (2002). *Modern Chinese*. Beijing, China: Higher Education Press.
- Krifka, M., & Musan, R. (2012). *The expression of information structure*. Berlin, Germany: De Gruyter Mouton.
- Lambrech, K., 1994. *Information structure and sentence form: Topics, focus, and the representations of discourse referents*. Cambridge, UK: Cambridge University Press.
- MacWhinney, B., & Bates, E. (1978). Sentential devices for conveying givenness and newness: A cross-cultural developmental study. *Journal of Verbal Learning and Verbal Behaviour*, 17, 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*, Boston, MA, 2011.
- Molnár, V. (2002). Contrast – from a contrastive perspective. *Language and Computers*, 39(1), 147–161.
- 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*, Cambridge, UK, 2005.
- Redford, M. A. (2013). A comparative analysis of pausing in child and adult storytelling. *Applied Psycholinguistics*, 34(1), 569–589.
- Roberts, C. (1996). Information structure: Towards an integrated theory of formal pragmatics. In J. H. Yoon & A. Kathol (Eds.), *OSU Working Papers in Linguistics*, Vol. 49. Columbus: The Ohio State University Department of Linguistics.
- Rochester, S. R. (1975). Defining the silent pause in speech. *Journal of the Ontario Speech and Hearing Association*, 8, 1–14.
- Romøren, A. S. H., & Chen, A. (2014). Accentuation, pitch and duration as cues to focus in Dutch 4- to 5-year-olds. *Proceedings from BUCLD 2013*, Boston, MA, 2013.
- 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). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Swerts, M., & Geluykens, R. (1994). Prosody as a marker of information flow in spoken discourse. *Language and Speech*, 37(1), 21–43.
- 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, Germany and New York, NY: Walter De Gruyter.
- Vallduví, E., & Engdahl, E. (1996). The linguistic realization of information packaging. *Linguistics*, 34, 459–519.
- Wagner, M., & Watson, D. G. (2010). Experimental and theoretical advances in prosody: A review. *Language and Cognitive Processes*, 25(7–9), 905–945.
- Wells, B., Peppé, S., & Goulrandris, N. (2004). Intonation development from five to thirteen. *Journal of Child Language*, 31, 749–778.
- Wieman, L. (1976). Stress patterns in early child language. *Journal of Child Language*, 3, 283–286.
- Wijnen, F. (1990). The development of sentence planning. *Journal of Child Language*, 17(3), 651–675.
- Wonnacott, E., & Watson, D. G. (2008). Acoustic emphasis in four year olds. *Cognition*, 107(3), 1093–1101.
- Zellner, B. (1994). Pauses and the temporal structure of speech. In E. Keller (Ed.), *Fundamentals of speech synthesis and speech recognition* (pp. 41–62). Chichester, UK: John Wiley.