# Focus marking in Dutch children with Autism Spectrum Disorder*

**Saskia Mooijman**
**Jasmijn Bosch**
**Veronica Miatto**
*ACLC, University of Amsterdam*

*Information structure has been suggested to be impaired in autistic populations, as they have consistently been found to experience difficulties with pragmatics and prosody (Eigsti et al. 2011; DePape et al. 2012). Therefore, the present study investigated the prosodic realization of focus by 37 Dutch children with autism spectrum disorder (ASD), in comparison with 29 typically developing (TD) children. The data, which we retrieved from the CHILDES database, were elicited in a structured storytelling task administered by Kuijper et al. (2015). Three acoustic measures were analyzed – accent type, pitch range and duration –, as these all have been associated with focus marking in Dutch (Gussenhoven 1983; Chen 2007, 2011; DePape et al. 2012). In addition, we looked at the position of accent placement within the sentence. The results show that even though in both groups children are more likely to produce a falling accent type to mark focus in sentence final position than in initial or medial position, this preference is significantly weaker for children diagnosed with ASD. However, the ASD group was not significantly different with regard to pitch range, duration or accent placement.*

## 1 Introduction

The study of information structure is concerned with the ways information is packaged in utterances (Halliday 1967; Chafe 1976). Dutch realizes information structure not only through word order, but also in a prosodic manner. Focus, a notion central to information structure, is marked through accent placement, a specific accent type, and manipulations in pitch range, intensity, and duration (Gussenhoven 1983; van Donzel 1999; Chen 2009, 2011). For example, the focus in *'Veronica ate tiramisu'* is on *tiramisu* when it answers the question '*What did Veronica eat?'*, since it contains the new and informative part of the

* We would like to thank Enoch Aboh, Silke Hamann, Vadim Kimmelman, and Jeannette Schaeffer for their valuable feedback and suggestions.

sentence (Vallduví & Engdahl 1996). In Dutch, focus is marked by intonation: a falling pitch tone, larger pitch range, and longer duration.

Adults and children diagnosed with autism spectrum disorder (ASD) are often found to show deviant prosodic characteristics in their speech (e.g., Shriberg et al. 2001). This finding was extended to the domain of information structure by DePape, Chen, Hall, and Trainor (2012), who demonstrated that the use of prosody to mark topic and focus was impaired in a subpopulation of high functioning autistic adults. However, research on the acquisition of focus marking is mostly limited to typically developing children. The present study therefore aims to investigate the realization of focus in a population of children diagnosed with ASD, and compare their marking strategies to age-matched typically developing (TD) children. This results in the following research question: *Which prosodic cues encoding focus in Dutch are employed by children with autism spectrum disorder, compared to typically developing children?* In answering this question, the three acoustic measures (accent type, pitch range and duration) used to mark focus in Dutch are systematically analyzed.

This study intends to fill a gap in the existing research literature by studying an autistic population by means of a detailed acoustic analysis. This will add to our understanding of how typical and atypical children express information structure. It is important to study the populations that may experience problems with realization of information structure, as inappropriate marking of information can create confusion between interlocutors and in achieving communicative intents (Fine, Bartolucci, Ginsberg & Szatmari 1991).

In the following section, we address the relevant theoretical background. Section 2.1 introduces the important notions of information structure, and Section 2.2 describes the prosodic realization of focus in Dutch. Section 2.3 then highlights research on the acquisition of focus, after which Section 2.4 concludes with a characterization of the linguistic difficulties that populations with ASD experience. The hypotheses and predictions of the present study are presented in Section 3, followed by an overview of the methods in Section 4. Finally, we present our findings in Section 5 and discuss them in relation to the theoretical background in Section 6.

## 2 Theoretical background

### *2.1 Information structure*

Information structure refers to the way sentences are structured in different blocks of information (Féry & Ishihara 2016). It is concerned with the various ways in which speakers "package" the information based on the assumed knowledge of the hearer (Halliday 1967; Chafe 1976). Stalnaker (1974) refers to this information mutually known to be shared between interlocutors as the *common ground*. As Krifka (2008) explains, the common ground is continuously updated as a conversation progresses, which in turn affects the information packaging of utterances.

Central notions in the theory of information structure are *focus, topic,* and *givenness.* Krifka (2008) defines *focus* as an indication of the presence of 'alternatives that are relevant for the interpretation of linguistic expressions' (p. 247). It is the informative, dominant, or unexpected part of an utterance (Vallduví & Engdahl 1996). Krifka describes *topic* as the entity about which the information is given. Finally, an expression is *given* if its denotation is present in the immediate common ground. Example (1) shows that *pasta* is in focus (underlined) because it contains the new and informative part of the utterance, whereas *Veronica's favorite food* is the element that is given and constitutes the topic of the sentence.

(1) What is Veronica's favorite food?

Veronica's favorite food is <u>pasta</u>.

This is an example of non-contrastive, presentational focus, which should be distinguished from contrastive focus (Krifka 2008). As can be seen in example (1), a non-contrastively focused element is the part of a proposition that answers an implicit or explicit WH-question. Contrastive focus, on the other hand, directly contrasts a proposition with an alternative that is available in the common ground (Hertel 2003). This is illustrated by example (2), which negates a presupposition based on the statement made in example (1). The present study is concerned with non-contrastive focus only, as previous research has shown that languages may adopt different prosodic marking strategies to express contrastive and non-contrastive focus (Gussenhoven 2004).

(2) Veronica's favorite food is <u>pizza</u> (not pasta).

Information structure can be realized in various ways. Some languages adopt prosodic means, others morphological means or syntactic reordering. A language may also combine strategies to realize information-structural categories. Dutch marks focus prosodically as well as syntactically (e.g., by means of direct object scrambling). The prosodic strategies observed in Dutch are discussed in the next section.

## *2.2 The prosodic realization of focus in Dutch*

West Germanic languages predominantly mark information structure through prosodic means. However, phonology and syntax are closely intertwined. According to Cinque's (1993) Nuclear Stress Rule, the most deeply embedded constituent in the sentence is the most prominent constituent and is marked with phrasal stress. Syntactic operations such as topicalization or object scrambling can alter this by placing new-information focus in sentence initial position (in the case of topicalization). Phrasal stress can be realized phonetically through the manipulation of duration, $F_0$, and intensity, as described by Breen, Fedorenko, Wagner and Gibson (2010), who studied the acoustic parameters of focus realization in English. Breen and colleagues also claim that their findings presumably extend to all West Germanic languages.

Dutch is no different: focus is realized prosodically by duration and manipulation of the $F_0$, which results in a particular pitch accent and a broader pitch range (see Gussenhoven 1983; Chen 2009, 2011). Chen (2007) describes how speakers can use different accent types to realize focus. The preferred accent type is a fall from high to low pitch, independent of the sentence position in which focus occurs. This accent type is notated as *H*L* in the ToDI notation system (Transcription of Dutch Intonation, Gussenhoven 2005). Other accent types that are associated with focus marking in Dutch are rises (L*H), a sustained high (H*), or a sustained low (L*).

These studies on Dutch do not take sound acoustic *intensity* into consideration in their analyses of focus marking, unlike Breen et al. (2010). For example, van Donzel (1999) briefly states that intensity is indeed one of the acoustic cues that mark information structure, but does not analyze this parameter in her study. The reason for this is that the studies that analyze how focus is realized in Dutch are particularly concerned with its phonological realization (e.g., pitch accent) rather than its acoustical correlates.

## *2.3 Acquisition of focus marking*

Various studies have shown that children as young as three to four years old can use prosody to distinguish focused elements. Manipulations of accent, pitch, and intensity have been described for English-speaking children (Hornby & Hass 1970; Wieman 1976; MacWhinney & Bates 1978; Wonnacot & Watson 2008), and German children (Müller, Höhle, and Schmitz & Weissenborn 2006). However, other studies report that children are older when they acquire the ability to mark information structure. For English, non-adult-like focus realizations have been reported at the ages of six, or even thirteen (MacWhinney & Bates 1978; Wells, Peppé & Goulandris 2004). Chen (2011) argues that the differences in these findings are due to contrastivity. Children are able to mark contrastive elements at a younger age, but are still developing their ability to mark non-contrastive elements beyond the age of four to five years old.

For Dutch, Chen (2009) showed that four- to five-year-olds can accent focused elements, and de-accent post-focal information. The children in her study did, however, differ from adults in that they did not show a preference for falls (H*L) or downstepped falls (!H*L) in sentence-final focus. Furthermore, adults differentiated foci from non-foci by means of alternations in pitch range and durations, whereas children did not use such acoustic manipulations. In addition, Chen (2011) investigated the intonational realization of topic and focus in children acquiring Dutch, by looking at accent placement and accent type. She found a discrepancy between topic and focus, as the four- to five-year-olds were able to mark topic correctly, whereas they did not show an adult-like preference for the accent type H*L to mark focus. In contrast, the seven- to eight-year-olds in her study did show a preference for H*L for focused elements in sentence-final position. They were also able to distinguish between topic and focus in initial position, showing an adult-like pattern.

### *2.4 Information structure in autism spectrum disorder*

Autism spectrum disorder (ASD) is characterized by persistent deficits in social communication, and restricted, repetitive patterns of behavior or interests. The symptoms are present in the early developmental period and can cause impairments in social, occupational, and other important areas of functioning (DSM-5, American Psychiatric Association 2013). Eigsti, de Marchena, Schuh, and Kelley (2011) reviewed the extensive literature on associated language difficulties in ASD. Due to the heterogeneity of ASD populations and the complexity of the disorder, it is difficult to draw firm conclusions. However, two linguistic domains in which ASD populations have consistently been shown to experience difficulties are pragmatics and prosody. Several studies report pragmatic impairment in storytelling tasks, such as problems with constructing a story with clear links between the events (Diehl, Bennetto & Young 2006), or difficulties communicating the causal structures of narratives (Kelley, Paul, Fein & Naigles 2006). Eigsti et al. (2011) further describe that prosodic impairments have been found in every study of children with ASD in this domain, though these studies are few in number. These impairments concern prosodic “unnaturalness”, like misplaced stress, or slowed speech and deviant resonance qualities (Shriberg et al. 2001). Diehl, Bennetto, Watson, Gunlogson, and McDonough (2008) found that the ASD group in their study also had issues regarding comprehension, as they were less able to use prosody to resolve syntactic ambiguities. These issues seem to persist through late development (Shriberg et al. 2001; Diehl, Watson, Bennetto, McDonough & Gunlogson 2009).

Difficulties with prosody and pragmatics, two linguistic domains central to information structure, may lead to problems with the realization of information structural concepts. McCaleb and Prizant (1985) analyzed contrastive stress and lexicalization strategies to mark new information in utterances of four- to fourteen-year-old children with ASD and found that the participants did not differentiate between new and old information. However, the authors only considered contrastive stress and base their conclusions on hearer judgments rather than an acoustic analysis. Another study shows that populations with ASD tend to accentuate the beginning of a sentence, regardless of informational value (Shriberg et al. 2001).

DePape et al. (2012) were the first to conduct a well-controlled study with a detailed acoustic analysis to investigate the use of prosody and information structure in high functioning autistic adults. Surprisingly, they found that ASD participants with high language abilities did not mark information structure prosodically, even though overall they used a larger pitch range than control participants. The ASD participants with moderate language abilities, on the other hand, produced smaller pitch ranges than controls but did not show difficulties marking information structure. The authors explain these counterintuitive results by suggesting that the ASD group with high language abilities have problems with the more basic automatic communication strategies, like how to vary pitch and duration, while the poorer language group's communicative issues may have a different origin.

## *2.5 Present study*

Several studies have shown that in Dutch, focus is realized using prosodic cues, such as accent type, pitch range, and duration.[1] These three acoustic means to mark focus have never been investigated in a systematic manner in a population of children with ASD. Previous research thus leaves the question of how children with autism employ these prosodic cues to mark focus open for investigation. Therefore, the present study aims to investigate focus realization in a population of children diagnosed with ASD, and compare their focus marking strategies to age-matched TD children. Since TD children are still developing an adult-like realization of non-contrastive focus until the age of seven, this study focuses on children who are older than seven years old.

---

[1] As mentioned before, some authors, such as Breen et al. (2010) and van Donzel (1999), state that focus is also realized through manipulation of intensity. In the present study intensity is not taken into consideration due to (i) the quality of the recordings, whose reliability is compromised due to background noise; and (ii) the conditions in which the data were collected: we cannot be sure whether the children were recorded always the same distance, or whether they moved too much during the experiment session. This makes it impossible for us to analyze intensity.

# 3 Hypotheses and predictions

The general hypothesis is that children with ASD diverge from TD children with regard to their use of prosodic cues to mark focus in Dutch. More specifically, we hypothesize that they will show deviant strategies for accent type, pitch range, and duration as compared to the TD children. This hypothesis leads to the following predictions:

1. *Accent type:* The ASD-group is expected not to show a preference for H*L for foci in any position in the sentence, in contrast to the TD-group.
2. *Pitch range:* The ASD-group is predicted to make atypical use of manipulations of pitch range to mark focus, resulting in either a broader or narrower pitch range as compared to the TD-group.
3. *Duration:* The ASD-group is expected to produce foci with a shorter duration as compared to the TD-group, because children with ASD are expected to make less use of duration as a focus marker.

In addition, we will investigate whether our study confirms previous research with regard to accent placement. Based on Shriberg et al.'s (2001) findings for adults, we expect that the ASD group will show a tendency to accentuate the initial constituent of an utterance, regardless of information structure.

# 4 Methods

## *4.1 Participants*

The data analyzed in this study are available on CHILDES (MacWhinney 2000) and were collected for a research project carried out at the University of Groningen (Kuijper, Hartman & Hendriks 2015). In total, 84 children participated in this research project, of whom 46 children with ASD. For the purpose of our study we excluded 13 participants, because they were younger than seven years old. We decided on this threshold because this is the age at which Dutch TD children are expected to have mastered the ability to mark information structure. In addition, we had to exclude five more participants due to the quality of the audio recordings. These decisions lead to a total number of 66 Dutch monolingual participants: 37 children diagnosed with ASD (5 girls and 32 boys[2]) aged 7;4-12;10 (*M* = 9;9), and 29 TD children (11 girls and 18 boys) aged 7;1-12;0 (*M* = 9;4). Children in the ASD group were diagnosed with either Autistic Disorder, PDD-NOS, or Asperger's Syndrome by qualified clinicians.

[2] ASD is 4.5 times more prevalent among boys than girls (1 in 42 versus 1 in 189, respectively), according to Christensen et al. (2014).

### *4.2 Materials*

The materials of the present study consist of transcripts and audio files for each participant. The participants had to perform a structured storytelling task, consisting of four stories with six pictures each (Appendix A). Every story included an introductory page with storybook figures, depicted in stereotypical color drawings. In addition, there were two practice stories consisting of two pictures each. The child had to describe the pictures, and the utterances were recorded with an Olympus voice recorder. The WMA files are downloadable from CHILDES.

### *4.3 Procedure*

The structured storytelling task was part of a larger session in which the participants were tested for other linguistic and cognitive abilities. All children were tested individually at a lab at the University of Groningen, between May 2009 and August 2011. In the present study, only the data from the storytelling task were analyzed.

### *4.4 Analysis*

First, the focused elements were selected in the transcripts while listening to the audio files by the first and second author, who are both native speakers of Dutch. Decisions were based on auditory perception and crosschecked by both listeners to ensure reliability of the judgments. In order to make the data comparable across groups, only simple declarative sentences were taken into consideration. Moreover, instances of contrastive focus were disregarded. The focused elements were segmented in the speech-analysis software *Praat* (Boersma & Weenink 2017), and the position in the sentence and the story in which they appeared were annotated, together with accent type and the phrase.

The accent types were coded using an adaptation of the ToDI notation system (Gussenhoven 2005). In this notation system, pitch is used to describe the contour of a sentence as high and low tones. We distinguished between the following accent types: falls (H*L), rises (L*H), sustained high and low (H and L, respectively), and a delayed fall and rise (L*HL and H*LH respectively). We created a category for instances in which the participant used a pitch contour that did not fit the predefined categories (such as H*LHL and HLH*L), which constitutes the *accentric* category.[3]

In addition to accent type, the measures duration (in milliseconds) and pitch range (measured by subtracting the minimum $F_0$ in Hertz from the maximum) were extracted from the segmentations through a Praat script. With

[3] "Accentric" is not an existing word; it is a portmanteau word consisting of "accent" and "eccentric".

this script, we created a table consisting of the speaker information, phrases, and acoustic measures. Figure 1 is an example of an annotated utterance from the corpus.

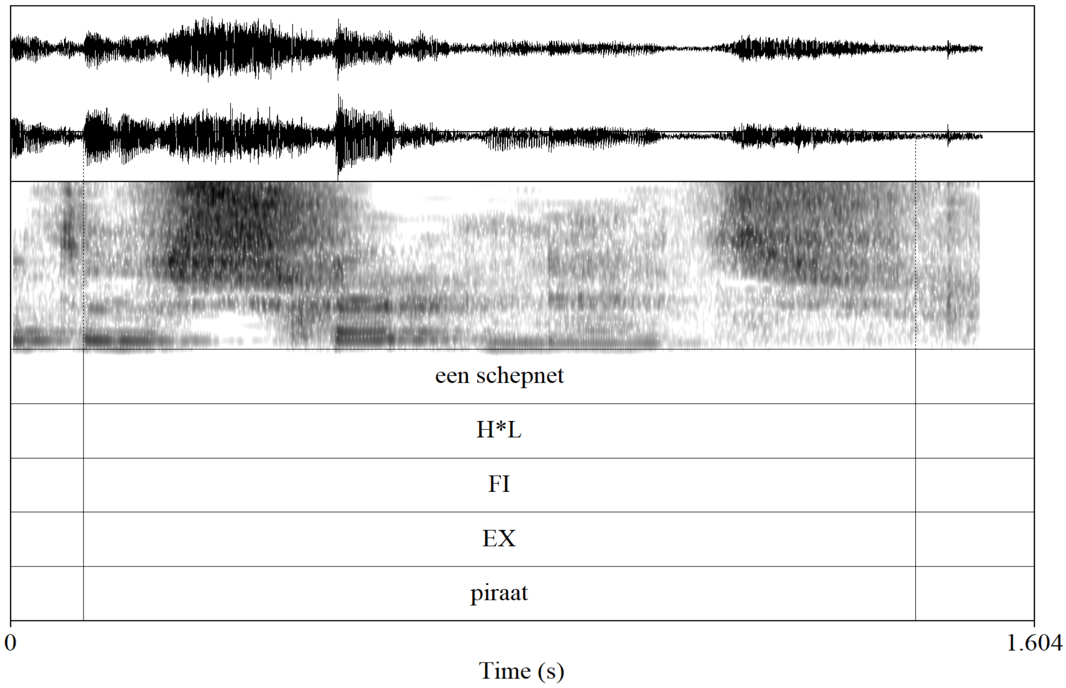



**Figure 1**: Example of annotated utterance

Unfortunately, the quality of the recordings was in some cases very poor, due to background noise. Tokens were excluded for three reasons: (i) presence of a pause within a prosodic word; (ii) poor visibility in the spectrogram which prevented the token to be segmented reliably; (iii) lack of visible pitch contour during a voiced consonant or a vowel. Figure 2 is an example of a token that was excluded due to a lack of pitch in the first vowel. Participants were excluded whenever at least 50% of the total number of the tokens they produced was judged to be unreliable, for the reasons explained above. Following this criterion, five participants were excluded.

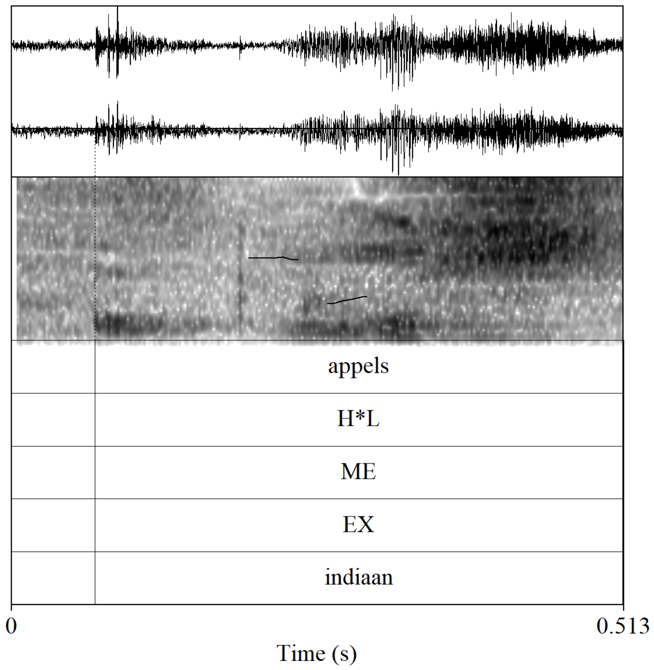



**Figure 2:** Example of excluded token

Furthermore, in particular circumstances, we decided to treat delayed falls (L*HL) as simple falls. Figure 3 shows the pattern L*HL in a noun phrase. As can be seen, the first low is located on the article *een* ‘an’, while the noun *appel* ‘apple’ carries the fall. In these instances, we considered the noun to be properly focused, and the article to be deaccented. Therefore, a fall was deemed the proper accent type for delayed falls on noun phrases starting with an article.

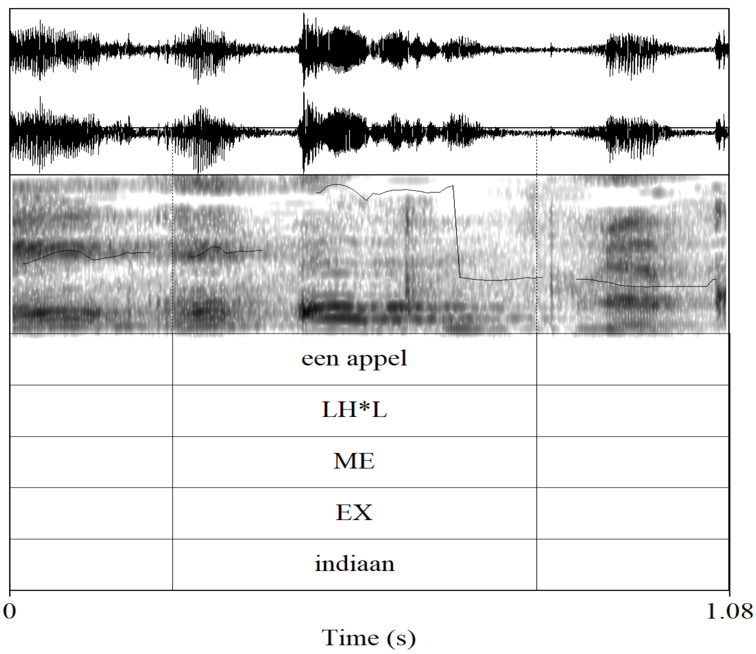



**Figure 3:** Example of delayed falls successively treated as a simple fall

### *4.5 Statistical analysis*

The first step in the statistical analysis was to establish the distribution of accent types per participant group (ASD versus TD). We used Chi-square Tests of Independence in SPSS (IBM Corp. 2013) to investigate whether the overall distribution of accent types and the distributions per position in the sentence were significantly different for ASD children than for TD children. After these preliminary analyses, the rest of the analysis was conducted in R, using the lme4 package (Bates, Maechler, Bolker & Walker 2016; R Core Team 2016). A generalized mixed effects regression model was used to account for the fact that multiple observations were derived from one participant, and to take counterbalancing predictors and random factors into account.[4] The dependent variable was a binary outcome (fall or no fall), the research question-answering predictors were ‘group’ (ASD or TD) and ‘position’ within the sentence (initial or medial or final), while ‘gender’ and ‘age’ were included as counterbalancing factors. All predictors were coded appropriately (i.e., contrast coded or centered around the mean). We used the most complete model that would still converge, which included a random intercept for ‘subject’ with a random slope for ‘position’.

For the analyses of pitch range and duration, we used linear mixed effects regression models, with pitch range in Hertz and duration in milliseconds as numeric outcome variables. The main predictors were ‘group’ and ‘position’, the counterbalancing factors were ‘age’ and ‘gender’, and random intercepts for ‘speaker’ and ‘phrase’ were included, with a random slope for ‘position’. In order to make a more direct comparison we also conducted separate analyses for duration for individual words, for which we selected the four most frequently annotated phrases in the data.

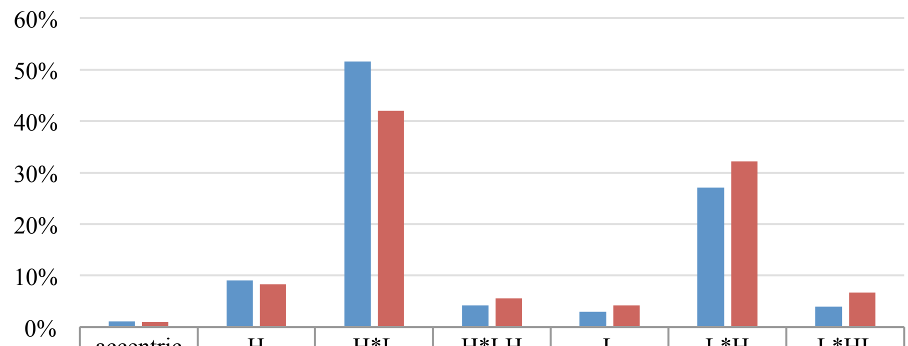
## 5 Results

### *5.1 Accent type*

The overall proportions of accent types per group are illustrated in Figure 4. Both groups show a preference for falls (H*L) to mark focus, which seems to be

[4] We realize that the statistical analysis for accent type is not optimal, and a multinomial logistic regression model would have been more appropriate. However, due to practical limitations and time constraints, we decided to use a generalized linear mixed effects model with a binary outcome and a Chi-square Test instead. The combination of these two tests is expected to give us roughly the same insight in the data as a multinomial logistic regression. The Chi-Square Test and the descriptive statistics provided us with a general picture of the distribution of all accent types, which justifies our decision to focus on odds ratios of falls versus no falls in the main model of accent type.

greater for TD children (52% falls) than for children with ASD (42% falls). However, a considerable number of rises (L*H) was used; TD children used a rise in 27% of cases, while children with ASD did so 32% of the time. The Chi-Square Test of Independence shows that the distribution of accent types is significantly different for children with ASD than for TD children $\chi^2$ (6) = 21.566, *p* =.001. A post-hoc analysis of the contingency table indicates that the odds of a fall are 1.47 times higher for foci uttered by TD children than for foci uttered by children with ASD.




| | accentric | H | H*L | H*LH | L | L*H | L*HL |
|---|---|---|---|---|---|---|---|
| TD | 1% | 9% | 52% | 4% | 3% | 27% | 4% |
| ASD | 1% | 8% | 42% | 6% | 4% | 32% | 7% |

**Figure 4:** Proportions of accent type used in foci by TD and ASD group

However, the separate Chi-Square Tests per position revealed that the distribution of accent types of the ASD group deviated from that of the TD group for focused elements in sentence-final position only. There was no significant association between group and accent type in sentence-initial position $\chi^2$ (6) = 11.97, p =.063, nor in medial position $\chi^2$ (6) = 8.689, p =.192. There was, however, a significant association between group and accent type in final position $\chi 2$ (6) = 27.26, p < .001. This represents the finding that, based on the odds ratio of the contingency table, the odds of a fall in final position were 2.32 times higher for foci uttered by TD children than for foci uttered by children with ASD. In the TD group foci in final position were accented with a fall 79% of the time, while this happened only in 62% of cases in the ASD group. The distributions of accent types per sentence position are illustrated in Figures 5, 6 and 7.

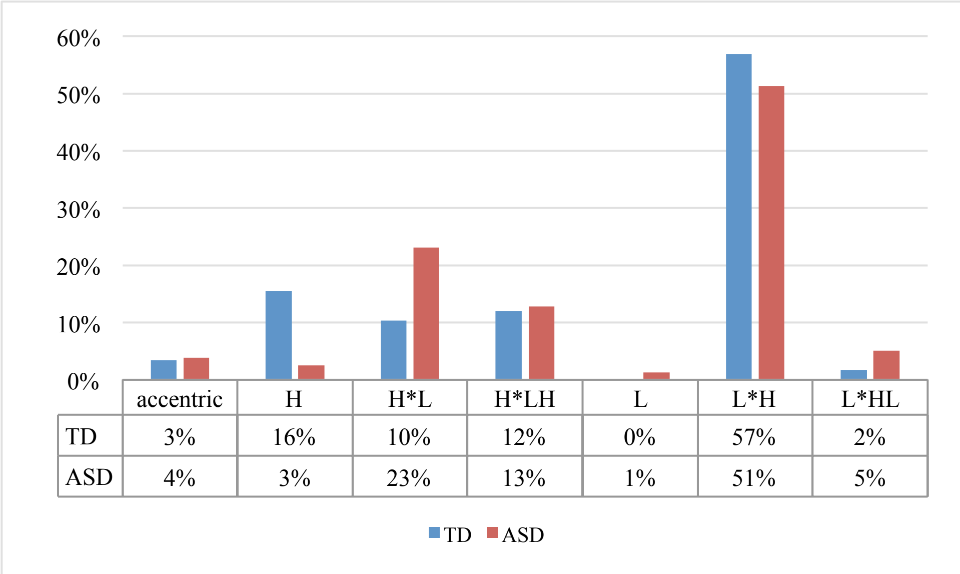



**Figure 5**: Accent types used in sentence-initial foci by TD and ASD group

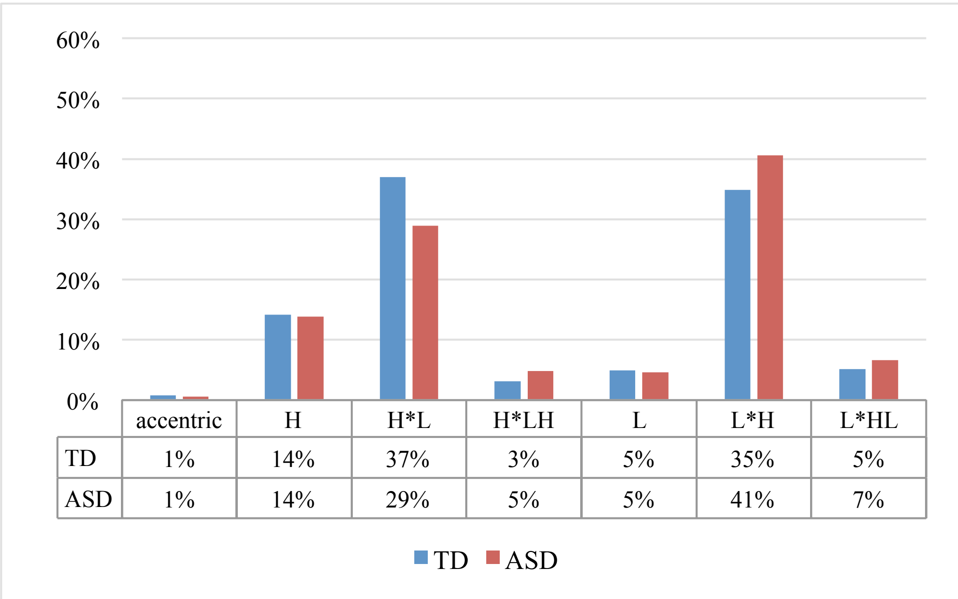
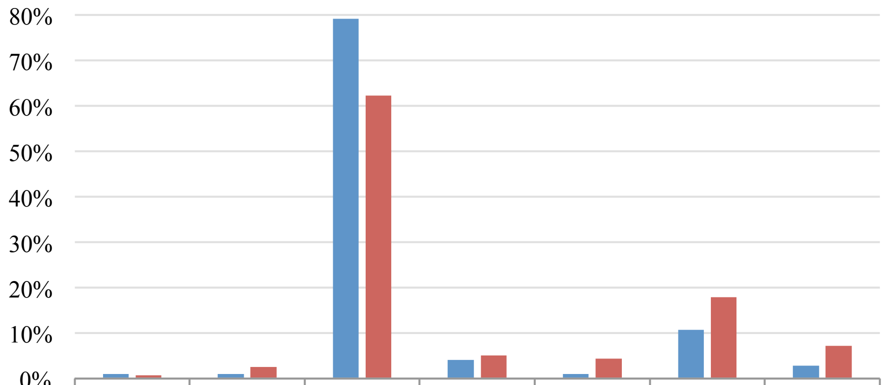



**Figure 6**: Accent types used in sentence-medial foci by TD and ASD group




| | accentric | H | H*L | H*LH | L | L*H | L*HL |
|---|---|---|---|---|---|---|---|
| TD | 1% | 1% | 79% | 4% | 1% | 11% | 3% |
| ASD | 1% | 3% | 62% | 5% | 4% | 18% | 7% |



**Figure 7:** Accent types used in sentence-final foci by TD and ASD

The findings discussed above are confirmed by the results of the generalized linear mixed effects model, which treated accent type as a binary outcome variable (fall or no fall). For this analysis we grouped simple falls (H*L) and delayed falls (L*HL) together to form a larger group of falls. The results showed that the odds that focus was expressed by a fall are estimated to be 11.5 times greater in sentence-final position than in the other positions (95% CI = 4.83...27.33, $p$ <.001). A significant interaction between ‘group’ and ‘position’ showed that this preference for a fall in final position was larger for TD children than for children with ASD (estimated odds ratio = 12.33, 95% CI = 4.83...27.33, $p$ = .003). There was, however, no significant main effect of group (95% CI = 0.21...1.74). This means that despite the finding that children with ASD are less likely than TD children to produce a fall to mark focus in final position, in the present study they were not significantly different with respect to the odds of producing a fall in focused elements overall.

### *5.2 Pitch range*

We did not find a significant difference between the TD group and the ASD group in terms of pitch range (95% CI = -54.10…22.82 Hz). The standardized residuals of the children with ASD were plotted to check whether there are two sub-groups within the ASD group, with one group using more variation in pitch than the other. Since there was no bimodal distribution, it can be concluded that this is not the case in the present study. As is shown by Figure 8 and Figure 9, the ASD group as well as the TD group shows a pitch range distribution with a strong positive skew. This means that in both groups the majority of the children used a relatively narrow pitch range up to 100 Hz, while in several extreme cases, children used wider pitch ranges up to 500 Hz. The mean pitch range was 98.37 ms for TD children (SD = 102.2) and 111.5 ms for children with ASD (SD = 113.3).

In addition, there is a significant negative effect of the counterbalancing factor 'age'; as children get older, their pitch range becomes smaller and thus more adult-like (estimate = -15.92 Hz, $t$ = -2.19, 95% CI = -29.99...-0.83 Hz).

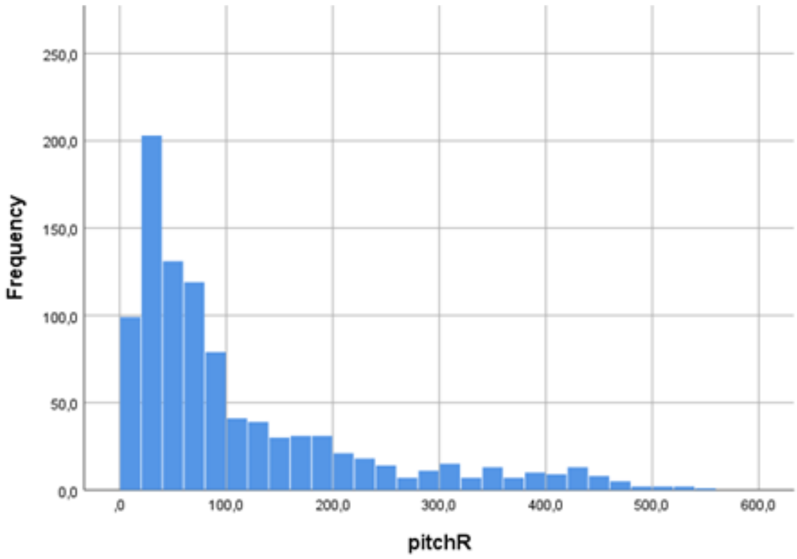



**Figure 8:** Distribution of pitch range for ASD

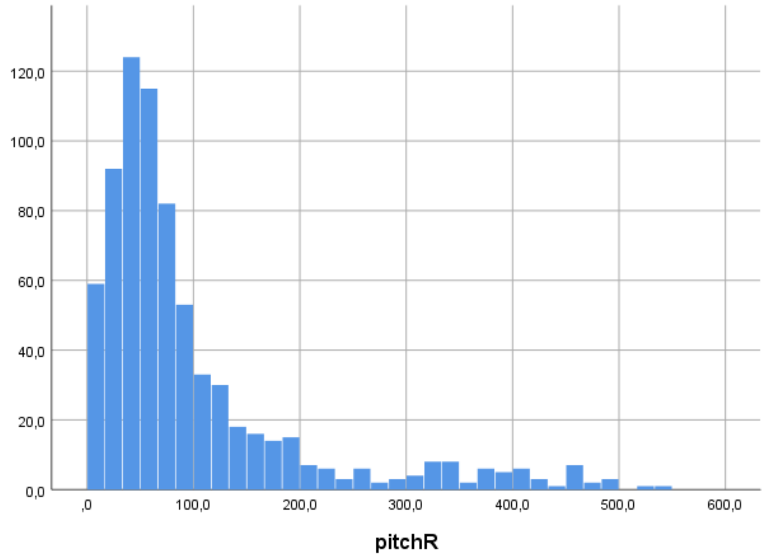



**Figure 9:** Distribution of pitch range for TD

### *5.3 Duration*

The mean duration of focused elements overall was 583.5 ms for TD children (SD = 244.8) and 560 ms for children with ASD (SD = 256.5). The difference between groups was not significant (95% CI = -35.50...107.31 ms), and there was no significant effect of age (95% CI = -47.03...6.14 ms). This may at least partly have been due to the fact that duration depends on speech rate, which differs greatly between individual speakers. Moreover, the tokens analyzed differ in their number of syllables, which made the comparison between the two groups less reliable.

In order to make a more direct comparison, we decided to examine individual noun phrases in more detail. The four noun phrases that were annotated most frequently in the data were *de bal* 'the ball' (N = 47), *de indiaan* 'the indian' (N = 44), *de heks* 'the witch' (N = 35), and *een appel* 'an apple' (N = 33). Three out of four of the separate analyses that used the duration of these phrases as an outcome variable did not reveal any significant differences between groups. Only the model for the duration of *de bal* 'the ball' showed a significant effect; on average this phrase was produced 138 ms longer by children with ASD than by TD children (95% CI: 47.8…228 ms). It should be noted, however, that this finding may not be reliable, since the sample sizes for this analysis were relatively small. *De bal* 'the ball', which was the most frequent phrase in the data, was annotated 22 times for TD children and 25 times for children with ASD.

### *5.4 Accent placement*

The absolute numbers and proportions of foci per position within the sentence are provided in Table 1. In this study the children with ASD did not show a tendency to accent the beginning of the sentence, as they only marked focus in sentence-initial position 8% of the time. As can be seen from the table, children with ASD are remarkably similar to TD children in terms of accent placement; in both groups focus was realized most often in medial or final position.

**Table 1:** Distributions of accent placement on focused elements.

| | **TD** | **ASD** |
|---|---|---|
| *Initial* | 58 (**7.9%**) | 77 (**8%**) |
| *Medial* | 388 (**52.8%**) | 498 (**51.5%**) |
| *Final* | 289 (**39.3%**) | 392 (**40.5%**) |

## 6 Discussion

Our results for accent type show that the ASD group deviates significantly from the TD group in their distribution of accent types used to mark focus. It has become clear that this difference lies in their preference for accent types in final position, as the likelihood of a sentence-final fall is smaller for children with

ASD than for TD children. This is an important finding, as Gussenhoven (1983) and Chen (2007, 2009, 2011) characterize the fall as the main accent type for foci. Chen (2007) found that this preference for a fall is independent of the position in which the focus occurs (i.e. initial or final). However, our data reflect a different pattern, showing a strong preference for a fall in sentence-final position, and a more variable picture for initial and medial position in both groups. The differences between these findings are perhaps caused by the nature of the administered tasks; Chen designed a relatively constrained task, which consisted of question-answer pairs. The participants were asked to repeat the answer that was first uttered by a robot, without any accentuation. Naturally, this is very different from a structured storytelling task eliciting semi-spontaneous speech.

Concerning accent placement, our results indicate that the ASD group does not differ from the TD group in this respect, as their distributions of accent placement across the three positions in the sentence are strikingly similar. This is in contrast to Shriberg et al. (2001), who describe a tendency for high-functioning autistic adults and adolescents to accentuate the beginning of a sentence, irrespective of informational value. However, it should be noted that Shriberg and colleagues did not place their study in the theory of information structure. Instead, they judged the appropriateness of stress placement in spontaneous speech based on listener intuitions.

The fact that the distribution of accent placement is similar between the two groups, shows that the syntactic realization of focus, i.e., the most embedded position in the sentence (Nuclear Stress Rule; Cinque 1993), appears to be intact in children with ASD. This is consistent with previous studies showing that syntax is not specifically impaired in autistic populations (see Eigsti et al. (2011) for a review). This gives rise to the idea that it thus becomes less crucial to mark focus by prosody, as it is already in the default focus position. If children with ASD rely more heavily on syntax to mark focus to compensate for their prosodic deficit, this could partially explain the ASD group's weaker preference for a fall in sentence-final position.

Duration is often described as one of the acoustic correlates of foci, with focused elements being longer than non-focused elements. Chen (2009, 2011) investigated the use of manipulations of duration in child speech, and found that these can be employed by seven- to eight-year-olds to mark focus. Consequently, we hypothesized that the TD children in our study would show longer durations than the children with ASD. This prediction was not borne out by our data, as we did not find significant differences between the groups. A possible explanation for the discrepancy between Chen's findings and the results of our study is the fact that we had to limit the present study to foci only. Because of this, we were unable to investigate the relative difference in duration between focused and non-focused elements as produced by TD and ASD children.

Therefore, any follow-up studies should include a direct comparison between focused and non-focused elements. This would contribute to the informativeness of the results, not only for duration, but also for the other measures. It is in fact possible that ASD populations employ similar accent types, duration, and pitch range for both focused and non-focused elements, and this would imply that they are not able to convey information structure through prosody.

Our findings for pitch range did not show a significant difference between the ASD group and the TD group. The majority of all children used a relatively small pitch range (up to 100 Hz), but there were several children in both groups who used more extreme pitch ranges (up to 500 Hz). DePape et al. (2012) found two patterns in their group of high-functioning autistic adults; one sub-group produced larger-than-controls pitch ranges while the other subgroup showed smaller-than-controls pitch ranges. However, this was not confirmed by our results, as we did not find a bimodal distribution for the ASD group.

Currently we have two suggestions to account for these diverging findings. The first possible explanation is a difference in patient characteristics. As people with autism are on a spectrum, with great variability between individuals, such individual differences can influence the results. Importantly, DePape et al. (2012) found that the differences in pitch range correlated with language ability; participants with high language ability used larger pitch ranges than participants with lower language ability. Thus, it is possible that our sample consists of a less heterogeneous group, resulting in a unimodal distribution for pitch range. Unfortunately, due to the fact that this was a corpus-based study, we were unable to link children's performance to the nature and severity of the disorder in the individual ASD participants. We only had access to the overall distribution of ASD types in all participants. Therefore, we are not able to check whether it is indeed the case that language ability influenced the results. An alternative explanation for the discrepancy between our results and the findings reported by DePape and colleagues is that they differentiated between different accent types. Whereas they found an effect of pitch range for falls only, we did not take accent type into account in the analysis of pitch range.

Our results further show that the average pitch range decreases significantly with age. We argue that this is due to a more general development of speech characteristics, rather than to the acquisition of focus marking. Lee, Potamianos, and Narayanan (1999) found that children until the age of 14 show more variability in temporal and spectral speech characteristics than adults, and it is not unlikely that age-related differences in pitch range variability also exist.

In future research, the acoustic measure *intensity* should also be taken into consideration, as this has been found to be an acoustic correlate for information structure in English (Breen et al. 2010) as well as in Dutch (van Donzel 1999). However, measuring intensity relies heavily on high quality recordings, preferably made in soundproof studios, which was not the case for the recordings we worked with. An additional reason for not including intensity in this study is that we cannot be sure whether the participants maintained the same distance from the microphone throughout the task.

Finally, it should be emphasized that the differences between children diagnosed with autism and typically developing children in terms of prosodic marking of information structure may be limited. Although our results suggest that children with ASD show a weaker preference for a falling accent type in final position than TD children, they did exhibit the same general pattern. In addition, we were able to identify focus for each child, indicating that they have mastered focus marking at least to the extent at which it is recognizable to a listener. This suggests that children with ASD have a notion of information structure, despite subtle differences regarding its prosodic realization. Although research in the field of atypical language acquisition tends to focus on finding asymmetries between clinical and typical populations, it is crucial that the similarities are acknowledged as well.

## 7 Conclusion

The present study investigated the prosodic cues to encode focus as produced by Dutch children diagnosed with ASD. In order to establish any deviant pattern, their utterances elicited in a structured storytelling task were compared to those of TD children. We analyzed three acoustic measures that correspond with focus realization in Dutch: accent type, pitch range, and duration. It can be concluded that in a storytelling task, the overall distribution of accent types is different for children with autism, compared to TD children. Both TD children and children with ASD are more likely to produce a fall to mark focus in sentence-final position than in initial or medial position, but this preference for a fall in final position is significantly weaker for children diagnosed with autism. In contrast to Shriberg et al. (2001), children with ASD did not show a general tendency to accent the beginning of the sentence. In fact, children with ASD and TD children exhibited an identical pattern in terms of accent placement, with the majority of foci occurring in medial or final position. Furthermore, we failed to find significant differences between the two groups for pitch range and duration; the evidence for these measurements thus remains inconclusive. All in all, despite the difference in accent type preference in final position, the similarities in focus marking strategies between children with ASD and TD children in this study appeared to be more striking than the differences.

## 8 References

American Psychiatric Association. 2013. *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.). Washington: Author.

Bates, D., M. Maechler, B. Bolker & S. Walker. 2015. Fitting linear mixed effects models using lme4. *Journal of Statistical Software* 67(1): 1–48.

Boersma, P. & D. Weenink. 2016. Praat: doing phonetics by computer [Computer program]. Version 6.0.36. Retrieved from http://www.praat.org/.

Breen, M., E. Fedorenko, M. Wagner & E. Gibson. 2010. Acoustic correlates of information structure. *Language and Cognitive Processes* 25(7-9): 1044–1098. DOI: 10.1080/01690965.2010.504378.

Chafe, W.L. 1976. Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In C. Li (ed.), *Subject and Topic*. New York: Academic Press, 25–56.

Chen, A. 2007. Intonational realisation of topic and focus by Dutch-acquiring 4- to 5-year-olds. In J. Trouvain & J.B. William (eds.), *Proceedings of the 16th International Congress of Phonetic Sciences*. Dudweiler: Pirrot GmbH, 1553–1556.

Chen, A. 2009. The phonetics of sentence-initial topic and focus in adult and child Dutch. In M. Vigário, S. Frota & M.J. Freitas (eds.), *Phonetics and Phonology: Interactions and Interrelations.* Amsterdam: Benjamins, 91–106.

Chen, A. 2011. Tuning information packaging: Intonational realization of topic and focus in child Dutch. *Journal of Child Language* 38(5): 1055–1083.

Christensen, D. L., J. Baio, K.V. Braun, D. Bilder, J. Charles, J.N. Constantino, J. Daniels, M.S. Durkin, R.T. Fitzgerald, M. Kurzius-Spencer, L.C. Lee, S. Pettygrove, C. Robinson, E. Schulz, C. Wells, M.S. Wingate, W. Zahorodny & M. Yeargin-Allsopp. 2014. Prevalence of autism spectrum disorder among children aged 8 years-autism and developmental disabilities monitoring network, 11 sites, United States, 2010. *Morbidity and Mortality Weekly Report. Surveillance Summaries (Washington, DC: 2002)* 63(2): 1–21.

Cinque, G. 1993. A null theory of phrase and compound stress. *Linguistic Inquiry* 24: 239–298.

DePape, A.M.R., A. Chen, G.B. Hall & L.J. Trainor. 2012. Use of prosody and information structure in high functioning adults with autism in relation to language ability. *Frontiers in Psychology* 3(72): 1–13.

Diehl, J.J., L. Bennetto & E.C. Young. 2006. Story recall and narrative coherence in high-functioning autism spectrum disorders. *Journal of Abnormal Child Psychology* 34(1): 87–102.

Diehl, J.J., L. Bennetto, D. Watson, C. Gunlogson & J. McDonough. 2008. Resolving ambiguity: A psycholinguistic approach to understanding prosody processing in high-functioning autism. *Brain and Language* 106(2): 144–152.

Diehl, J., D. Watson, L. Bennetto, J. McDonough & C. Gunlogson. 2009. An acoustic analysis of prosody in high-functioning autism. *Applied Psycholinguistics* 30(3): 385–404.

van Donzel, M. 1999. *Prosodic Aspects of Information Structure in Discourse*. The Hague: LOT.

Eigsti, I.M., A.B. de Marchena, J.M. Schuh & E. Kelley. 2011. Language acquisition in autism spectrum disorders: A developmental review. *Research in Autism Spectrum Disorders* 5(2): 681–691.

Féry, C. & S. Ishihara. 2016. Introduction. In C. Féry & S. Ishihara (eds.), *The Oxford Handbook of Information Structure.* Oxford: Oxford University Press, 1–15.

Fine, J., G. Bartolucci, G. Ginsberg & P. Szatmari. 1991. The use of intonation to communicate in pervasive developmental disorders. *Journal of Child Psychology and Psychiatry* 32(5): 771–782.

Gussenhoven, C. 1983. Testing the reality of focus domains. *Language and Speech* 26(1): 61–80.

Gussenhoven, C. 2004. *The Phonology of Tone and Intonation*. Cambridge: Cambridge University Press.

Gussenhoven, C. 2005. Transcription of Dutch intonation. In S. Jun (ed.), *Prosodic Typology and Transcription: A Unified Approach*. Oxford: Oxford University Press, 118–145.

Halliday, M.A.K. 1967. *Intonation and Grammar in British English.* The Hague: De Gruyter.

Kuijper, S.J.M., C.A. Hartman & P. Hendriks. 2015. Who is he? Children with ASD and ADHD take the listener into account in their production of ambiguous pronouns. *PLoS ONE 10(7):* e0132408.

Hertel, T.J. 2003. Lexical and discourse factors in the second language acquisition of Spanish word order. *Second Language Research* 19(4): 273–304.

Hornby, P.A. & W.A. Hass. 1970. Use of contrastive stress by preschool children. *Journal of Speech and Hearing Research* 13(2): 359–399.

IBM Corp. 2013. IBM SPSS Statistics for Windows, Version 22.0. Armonk, USA: IBM Corp.

Kelley, E., J.J. Paul, D. Fein & L.R. Naigles. 2006. Residual language deficits in optimal outcome children with a history of autism. *Journal of Autism and Developmental Disorders* 36(6): 807–828.

Krifka, M. 2008. Basic notions of information structure. *Acta Linguistica Hungarica* 55(3-4): 243–276.

Lee, S., A. Potamianos & S. Narayanan. 1999. Acoustics of children's speech: Developmental changes of temporal and spectral parameters. *The Journal of the Acoustical Society of America* 105(3): 1455–1468.

MacWhinney, B. 2000. *The CHILDES Project: Tools for Analyzing Talk* (3rd ed.). Mahwah: Lawrence Erlbaum Associates.

MacWhinney, B. & E. Bates. 1978. Sentential devices for conveying givenness and newness: A cross cultural developmental study. *Journal of Verbal Learning and Verbal Behaviour* 17(5): 539–555.

McCaleb, P. & B.M. Prizant. 1985. Encoding of new versus old information by autistic children. *Journal of Speech and Hearing Disorders* 50(3): 230–240.

Müller, A., B. Höhle, M. Schmitz & J. Weissenborn. 2006. Focus-to-stress alignment in 4- to 5-year-old German-learning children. *Proceedings of Generative Approaches to Language Acquisition* 7: 393–407.

R Core Team. 2016. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna. URL https://www.R-project.org/.

Shriberg, L.D., R. Paul, J.L. McSweeny, A. Klin, D.J. Cohen & F.R. Volkmar. 2001. Speech and prosody characteristics of adolescents and adults with high-functioning autism and Asperger syndrome. *Journal of Speech, Language, and Hearing Research* 44(5): 1097–1115.

Stalnaker, R. 1974. Pragmatic presuppositions. In M. Munitz & P. Unger (eds.), *Semantics and Philosophy*. New York: New York University Press, 197–214.

Swerts, M., E. Krahmer & C. Avesani. 2002. Prosodic marking of information status in Dutch and Italian: a comparative analysis. *Journal of Phonetics* 30(4): 629–654. DOI:10.1006/jpho.2002.0178.

Terken, J.M.B. 1984. The distribution of pitch accent in instructions as a function of discourse structure. *Language and Speech* 27(3): 269–289.

Vallduví, E. & E. Engdahl. 1996. The linguistic realization of information packaging. *Linguistics* 34(3): 459–520.

Wells, B., S. Peppé & N. Goulandris. 2004. Intonation development from five to thirteen. *Journal of Child Language* 31(4): 749–778.

Wieman, L. 1976. Stress patterns in early child language. *Journal of Child Language* 3(2): 283–286.

Wonnacott, E. & D.G. Watson. 2008. Acoustic emphasis in four year olds. *Cognition* 107(3): 1093–1101.

s.mooijman@let.ru.nl

jasmijn.bosch@unimib.it

veronica.miatto@stonybrook.edu

## Appendix

Pictures used in story-telling task