



Acquiring Complex Focus-Marking: Finnish 4- to 5-Year-Olds Use Prosody and Word Order in Interaction

Anja Arnhold^{1,2*}, Aoju Chen³ and Juhani Järvikivi²

¹ Department of Linguistics, University of Konstanz, Konstanz, Germany, ² Department of Linguistics, University of Alberta, Edmonton, AB, Canada, ³ Department of Languages, Literature and Communication, Utrecht Institute of Linguistics OTS, Utrecht University, Utrecht, Netherlands

OPEN ACCESS

Edited by:

Judit Gervain,
Université Paris Descartes, France

Reviewed by:

Aritz Irurtzun,
French National Centre for Scientific
Research (CNRS), France
Kriszta Szendroi,
University College London, UK
Balazs Suranyi,
Research Institute for Linguistics of
the Hungarian Academy of Sciences,
Hungary

*Correspondence:

Anja Arnhold
anja.arnhold@gmail.com

Specialty section:

This article was submitted to
Language Sciences,
a section of the journal
Frontiers in Psychology

Received: 09 June 2016

Accepted: 15 November 2016

Published: 01 December 2016

Citation:

Arnhold A, Chen A and Järvikivi J
(2016) Acquiring Complex
Focus-Marking: Finnish 4- to
5-Year-Olds Use Prosody and Word
Order in Interaction.
Front. Psychol. 7:1886.
doi: 10.3389/fpsyg.2016.01886

Using a language game to elicit short sentences in various information structural conditions, we found that Finnish 4- to 5-year-olds already exhibit a characteristic interaction between prosody and word order in marking information structure. Providing insights into the acquisition of this complex system of interactions, the production data showed interesting parallels to adult speakers of Finnish on the one hand and to children acquiring other languages on the other hand. Analyzing a total of 571 sentences produced by 16 children, we found that children rarely adjusted input word order, but did systematically avoid marked OVS order in contrastive object focus condition. Focus condition also significantly affected four prosodic parameters, f_0 , duration, pauses and voice quality. Differing slightly from effects displayed in adult Finnish speech, the children produced larger f_0 ranges for words in contrastive focus and smaller ones for unfocused words, varied only the duration of object constituents to be longer in focus and shorter in unfocused condition, inserted more pauses before and after focused constituents and systematically modified their use of non-modal voice quality only in utterances with narrow focus. Crucially, these effects were modulated by word order. In contrast to comparable data from children acquiring Germanic languages, the present findings reflect the more central role of word order and of interactions between word order and prosody in marking information structure in Finnish. Thus, the study highlights the role of the target language in determining linguistic development.

Keywords: information structure, prosody, word order, child language, focus, contrast, givenness

1. INTRODUCTION

To become successful communicators, children need to learn to transmit information in a way that is appropriate for the given context and knowledge state of the interlocutors. For example, answering *Who wants a banana?* with *I WANT a banana* (where capitals indicate a prominent accent) is pragmatically inappropriate and may lead to confusion, even though the sentence itself is morphologically, syntactically and phonologically well-formed and would be perfectly natural in another context. That is, children need to learn appropriate information packaging or information structure marking as part of successful language acquisition. Although languages use different

linguistic devices to mark information structure, prosodic marking is often central, as in the use of accentuation in the English example above.

The present study investigates information structure marking in 4- to 5-year-olds acquiring Finnish, a language characterized by employing word order alongside prosody and showing interactions between the two in encoding information structure. We used a language game to gather semi-spontaneous data with different word orders and information structures. Our prosodic analyses revealed effects of information structure on f_0 range, duration and the use of pauses. The study further found significant effects of information structure on voice quality, a dimension that has, to our knowledge, not previously been investigated for information structure marking in child language (but on focus effects in adult speech, Epstein, 2002; Ní Chasaide et al., 2011, for English, Vainio et al., 2010; Arnhold, 2016, for Finnish).

Before turning to the experimental methods in Section 2 and results in Section 3, we will give an overview of basic concepts of information structure that are essential to the current study, existing research on information structure marking in child language and information structure marking in adult Finnish, and state the research questions in the rest of Section 1. We will discuss how the findings relate to the research questions in Section 4 and conclude with Section 5.

The term information structure refers to the way interlocutors organize their utterances to match the common ground, i.e., the information that is shared and known to be shared among them (see, e.g., Krifka, 2008, for an introduction to basic notions information structure). Information structure is often discussed in terms of binary partitions of utterances. Although there is some variation in the literature, scholars generally distinguish between focus and background, between given and new, and between topic and comment. Focus, commonly marked by square brackets and a subscript letter F, is the part of the utterance which indicates that alternatives are relevant for its interpretation, while the rest of the sentence is often referred to as the background (Rooth, 1985, 1992). For example, for the utterance *My [sister]_F hates broccoli*, with the word *sister* in focus, relevant alternatives include *My brother hates broccoli*, *My mother hates broccoli*, etc., but not *My sister hates strawberries* or *My sister loves broccoli*. This information structure might appear as an answer to the question *Which of your family members hates broccoli?*, while a different information structure, with the object noun phrase *broccoli* in focus would be induced by a question like *Which vegetable does your sister hate?* These examples contain focus on a single word or noun phrase, but larger constituents and even whole sentences can be focused, e.g., *What's that noise?—[Our neighbors are renovating]_F*. This distinction is commonly called narrow vs. broad focus following Ladd (1980) (but see Katz and Selkirk, 2011, for an argument against the use of these terms). In this article, we will reserve the use of the term “broad focus” for cases where the whole sentence is in focus and the term “narrow focus” for focus constituents consisting of single words.

A second fundamental distinction is that between given and new, i.e., between denotations that are present in the common

ground and those that are not. In fact, this distinction is often characterized as a scale involving for example inferable information in addition (e.g., Gundel et al., 1993, but see for example Schwarzschild 1999, for an account using a binary distinction). In the present materials, all constituents were either mentioned in the immediate context, i.e., clearly given, or were not present in the preceding context at all, i.e., new. Additionally, the given/new distinction and focus/background division were correlated, such that all focused constituents were new and all background constituents were given. While this is frequently the case in naturally occurring discourse as well, the two dimensions are independent in principle, as illustrated by cases like second-occurrence focus (Beaver et al., 2007; Féry and Ishihara, 2009; but see e.g., Lambrecht 1994, for a newness-based definition of focus).

A third important partition distinguishes topic, i.e., what an utterance is about, and comment, i.e., the information given about it (Reinhart, 1981). In English, topics can be marked syntactically for example through fronting, as in the present sentence.

Finally, regarding the notion contrast, focus can be non-contrastive, for example when providing requested information, e.g., *What's the time?—It's [quarter to seven]_F* or contrastive (Dik et al., 1981; Gussenhoven, 2008; Krifka, 2008, for more on focus types). A prototypical case of contrastive focus, and the only one to appear in the present materials, is correction, e.g., *Is that your coat?—No, it's [my mother's]_F coat*. Contrastive topics are possible as well, but did not feature in the present study.

Children's speech shows an influence of information structure even at very early developmental stages, but adult-like ability to mark information structure is attained quite late¹. Wieman (1976) observed that 2-year-old English-speaking children deviated from their default realization of two-word utterances by accenting words in non-contrastive narrow focus. However, this observation was based on only seven utterances in Wieman's study and did not emerge in a systematic investigation of Dutch-speaking children, who predominantly accented both words (Chen and Fikkert, 2007). Likewise, Behrens and Gut's (2005) case study of a 2-year-old boy acquiring German found that both words were stressed in most two-word utterances. Recently, Grünloh et al. (2015) have suggested that differences between 2- and 3-year-old children and adults in accentuation, particularly young children's failure to de-accentuate given material, is at least partially due to specific characteristics of caregiver speech.

Analyzing descriptions of picture pairs that differed in one feature corresponding to either subject, object or verb in the description, Hornby and Hass (1970) found that English-acquiring 3- to 4-year-olds frequently produced the contrastive constituent with falling accents with wide f_0 ranges, especially for subject constituents. In a similar study, MacWhinney and Bates (1978) reported that the placement of prominent accents on new and focused constituents was already acquired by age three, but its use significantly increased in frequency between

¹For comparability with the present study, this overview concentrates on language production; on children's sensitivity to information structure and its marking in perception, see e.g. Ito and Speer (2008); Fernald et al. (2010); Ito et al. (2014); Järvikivi et al. (2014); see Chen (2010) for evidence that 4- to 5-year-olds possess similar skills in production and perception of prosodic focus marking.

age three and age six. However, this finding held most clearly for their English-acquiring participants and to a lesser degree for the children acquiring Italian. By contrast, children learning Hungarian (and Hungarian adults) did not systematically use accents to mark focus, but showed most variation in word order. Note also that focused referents in Hornby and Hass's (1970) and MacWhinney and Bates's (1978) materials were contrastive as well as new. Similarly, Müller et al. (2006) found that German-acquiring 4- to 5-year-olds consistently placed accents on focused constituents with materials in which all focused constituents were contrastive.

Even when patterns of accent placement are overall similar to those of adults, children's speech may still differ in crucial ways. Chen (2011) found that Dutch 4- to 5-year-olds accented foci more frequently than given topics and used a similar set of accents as adults. However, in children's speech focus was less clearly associated with falling accents (see de Ruiter, 2014, for somewhat different results regarding givenness). Chen (2011) only found a completely adult-like use of accent type in 7- to 8-year-olds, but even at this age, focus marking in terms of duration and alignment of f_0 turning points was not yet completely adult-like (Chen, 2009). Wells et al. (2004) have reported that children's ability to mark focus condition prosodically and especially their ability to correctly identify information structure in language input continues to improve between age five and age thirteen. While children underuse f_0 range and duration as markers of focus compared to adults, they employ pause durations more extensively, using longer pauses before focal target words than before non-focal ones (Romøren and Chen, 2015).

Regarding word order, several studies suggest that in contrast to a cross-linguistic tendency to place given before new information in adult language, children aged three to six generally place new before given constituents, although contrasting findings have also been reported (Narasimhan and Dimroth, 2008, and references therein). In a study using the same materials and method as the present study, Sauermann et al. (2011) found no general tendency for either given-before-new or new-before-given order in the productions of German 4-year-olds. Children's productions in this task generally did not reflect information structure effects on word order described for adult language, but it is worth noting that the adult control group exclusively produced unmarked SVO order in response to the same task. Prosodic variation, by contrast, was more extensive and more unified in both groups. Children as well as adults showed significant differences in f_0 and duration between broad focus, narrow non-contrastive and narrow contrastive focus conditions, although not all effects were identical for the two groups. An analogous study on Dutch children yielded very similar results (Chen and Höhle, submitted). Interestingly, prosodic effects were only significant for subject nouns in both languages.

In sum, previous research has indicated crucial development in children's ability to mark information structure between the ages of three and six. It has also provided evidence of cross-linguistic differences in the devices acquired, although most studies so far have focused on children acquiring West Germanic languages.

TABLE 1 | An example for the connection between word order and information structure in Finnish.

	Contrast	Topic	Rest (focus final)	English equivalent
(a) SVO		<i>Kissa</i>	<i>söi hiiren</i> cat.NOM ate mouse.ACC	"The cat ate the mouse."
(b) OVS		<i>Hiiren</i>	<i>söi kissa</i>	"The mouse was eaten by the cat."
(c) OSV	<i>Hiiren</i>	<i>kissa</i>	<i>söi</i>	"It is the mouse that the cat ate."
(d) SOV	<i>Kissa</i>	<i>hiiren</i>	<i>söi</i>	"It is the cat that ate the mouse."
(e) VSO	<i>Söi</i>	<i>kissa</i>	<i>hiiren</i>	"The cat did indeed eat the mouse."
(f) VOS	<i>Söi</i>	<i>hiiren</i>	<i>kissa</i>	"The mouse was indeed eaten by the cat."

Finnish marks information structure through syntax, prosody, and, to some degree, morphology. Some clitics like *-kin* "also" mark their host as focused or contrastive, but the role of morphology is least well researched and will not be discussed further here (see Nevis, 1986, for more details).

Syntactically, Finnish is a discourse configurational language, i.e., word order is largely determined by information structure, while grammatical roles are coded through case marking (Vilkuna, 1989, 1995; Vallduví and Vilkuna, 1998). Thus, all constituent permutations are grammatical, but indicate differences in information structure. **Table 1** illustrates this, adopting Vilkuna's division of Finnish sentences into the contrast position, the topic position and the rest of the sentence (called K-position, T-position and V-field and identified as Spec(CP), Spec(IP) and I', respectively, by Vilkuna, 1995)². By default, the finite verb is the beginning of the "Rest" and a constituent directly preceding it will be interpreted as topical and/or given, occupying the topic position. Constituents preceding the topic position are generally contrastive and can be topics or foci, whereas non-contrastive foci and new information appear in absolutely final position, at the end of the "Rest"³.

Unmarked SVO word order is possible with all information structures. Thus sentence (a) in **Table 1** is a good answer to different questions like "What happened?" (broad focus), "What did the cat eat?" (narrow focus on the object), and even "Who ate the mouse?" (narrow focus on the subject), even though by default the pre-verbal subject is interpreted as the topic and the final object as being in narrow focus. The OVS sentence in (b), however, is a felicitous answer to "Who ate the mouse?" but not to the other questions. It marks the object as a topic and the subject as non-contrastively focused, a constellation that is sometimes

²These terms somewhat simplify Vilkuna's description of the correspondence between information structural roles and positions. For generative accounts of Finnish discourse configurationality, also see Holmberg et al. (1993); Kaiser (2000, 2006); Holmberg and Nikanne (2002); Molnár and Järventausta (2003).

³Note that Vilkuna (1989, 1995) and Vallduví and Vilkuna (1998) do not employ the alternatives-based definition of focus adopted here, but a newness-based definition of focus (and topic). This difference is not relevant here and will be ignored in the following, since these dimensions were correlated in the present data.

expressed through passivization in English. Finally, the elements in the contrast position are interpreted as contrastive in (c)-(f), so that (d), for example would be an appropriate answer to “Did the dog eat the mouse?” while (e) and (f) can be corrections of claims that the event did not take place (the use of VSO word order (e) is less restricted than that of VOS order (f), see Välimaa-Blum 1988, p. 71; Vilkuna 1995; Jokinen 2005; also note that the use of the contrast position is generally not obligatory and that it is in fact relatively rare in written corpora, as discussed in Section 4.1).

Finnish prosody is affected by information structure in several ways. Finnish f_0 contours normally consists of a series of rise-falls in broad focus, with peaks on all constituents except for finite verbs (Välimaa-Blum, 1993; Iivonen, 1998; Suomi et al., 2008). For constituents in narrow focus, the f_0 range of these rise-falls expands while the f_0 range of the other constituents is compressed (Mixdorff et al., 2002; Vainio and Järvikivi, 2007). The rising-falling shape is not altered by information structural variation, and Finnish is frequently described as having just a single accent, with the same tonal targets and the same alignment realized in virtually all contexts (Välimaa-Blum, 1993; Suomi et al., 2008, but see Arnhold, 2014; Arnhold, submitted, for a different analysis employing phrase tones). Crucially, no account of Finnish prosody has suggested contrasting accents for constituents with different information structural roles. This constitutes a major difference from Germanic languages, where the choice of accent type frequently marks pragmatic distinctions, including information structure. For example, constituents in narrow focus frequently carry falling accents, while contrastive topics are often realized with a rise (e.g., Hedberg and Sosa, 2008, on English and Féry, 1993; Braun, 2006, on German).

In addition to f_0 range adjustments, Finnish prosody is affected by information structure in several ways. Specifically, constituents in narrow focus have longer duration (Mixdorff et al., 2002; Suomi, 2007), higher intensity (Vainio and Järvikivi, 2007; Arnhold, 2016) and are followed by pauses more often than constituents in broad focus (Arnhold, 2016). The prosody of given constituents shows the opposite characteristics, i.e., shorter durations, reduced intensity and absence of following pauses. Further, Vainio et al. (2010) reported a less tense voice quality in narrow focus, while Arnhold (2016) found increased use of different kinds of non-modal voice quality (e.g., creaky voice and whisper) on the second syllables of constituents in narrow focus and all following words in the same sentence.

Finally, prosodic and syntactic marking of information structure interact. For example, Vainio and Järvikivi (2006) found that compared to final words in sentences with unmarked word order, listeners perceived a word that appeared in the sentence-final focus position due to the use of a marked word order as prosodically more prominent, even though both cases were manipulated to have the same prosodic characteristics. Conversely, in production speakers compensated for a mismatch between syntax and information structure by reducing the prosodic prominence of a constituent located in the focus position when a context question implied that another constituent was in narrow focus (Vainio and Järvikivi, 2007). Furthermore, Arnhold and Féry (2013) found that speakers produced more consistent prosodic focus marking in

scripted speech with fixed unmarked word order than in semi-spontaneous productions where they were free to use both word order and prosody to mark information structure.

This study investigates children’s acquisition of the complex Finnish system of prosodic information structure marking, syntactic information structure marking and interactions between both components. Specifically, it addresses the following questions:

- How do Finnish-acquiring 4- to 5-year-olds use prosody and word order in various focus conditions?
- How do they differ from their peers acquiring West Germanic languages?
- How do they differ from adult Finnish speakers?

Given that children learning Germanic languages attain adult-like information-structure marking quite late, we hypothesize that also children acquiring Finnish differ from adults at ages four to five, and do not yet make full use of all available prosodic and syntactic tools. We further expect that Finnish-acquiring children differ from children learning Germanic languages by showing at least some interactions between word order and prosody in information structure marking, since these interactions are well-attested in adult Finnish speech.

The comparison with adult speakers of the same language on the one hand and children acquiring other languages on the other hand will thereby allow some insight into a further research question: What is the relative importance of universal tendencies and language-specific characteristics for the trajectory of language acquisition?

2. MATERIALS AND METHODS

To elicit short sentences with systematically varied information structures in a controlled way, participants were asked to teach a robot (depicted in **Figure 1**) to speak Finnish like a human. To this end, the experimenter asked the robot questions about visual stimuli and the children repeated and—if they chose to—adjusted the robot’s answers. Crucially, to encourage children to produce sentences with natural information structure marking, the robot’s utterances included both contextually inappropriate and appropriate word orders and were produced with unnatural flat f_0 (see below for details). The same method and materials, which allow for controlling information structural conditions in a naturalistic game-type setting, were used in experiments on children acquiring German and Dutch (Sauermann et al., 2011; Chen and Höhle, submitted).

2.1. Procedure

Participants were recorded individually in a quiet room at their respective day care facility in Helsinki or Espoo by a female research assistant.

Before the start of the experiment, each participant saw a slide with the picture of the robot, who introduced herself and solicited the participant’s help in improving her human language skills. After the procedure was explained to the participant, the introduction concluded with three practice trials.

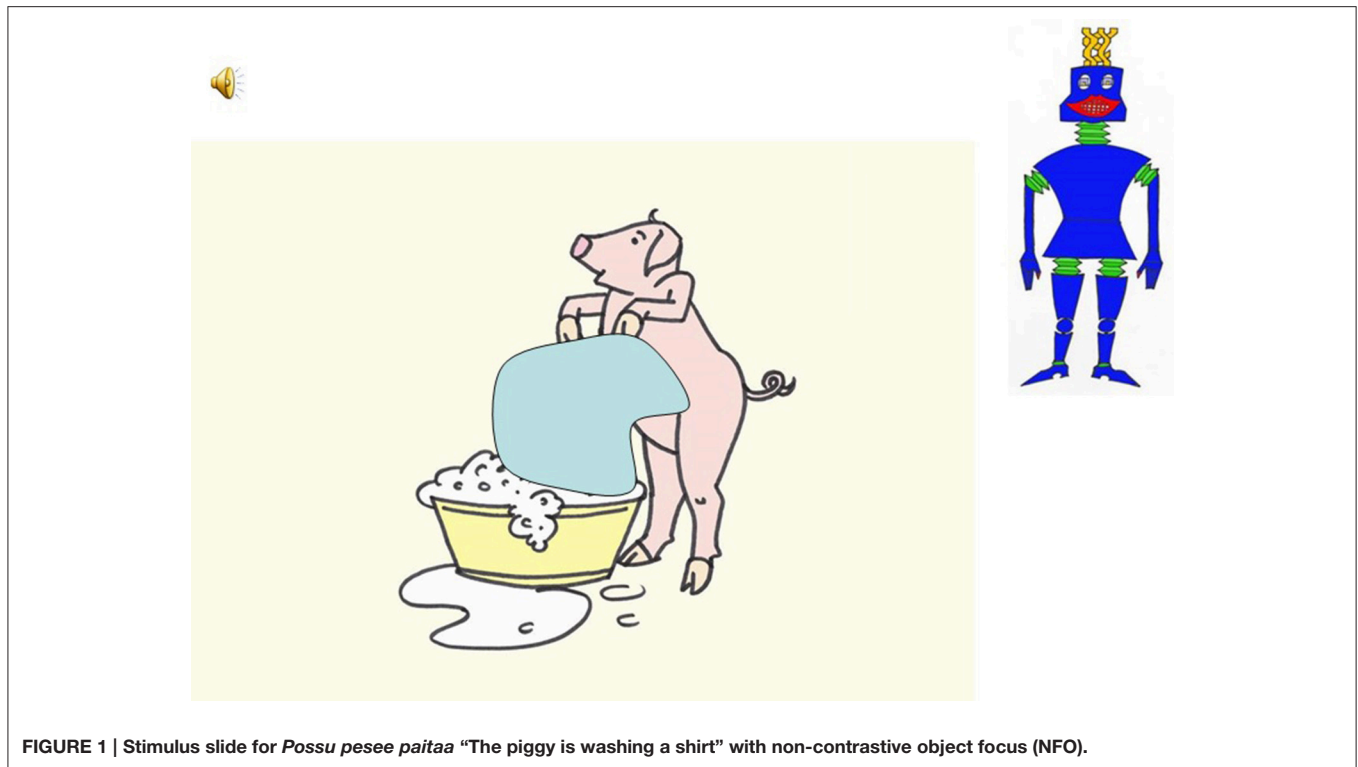


FIGURE 1 | Stimulus slide for *Possu pesee paitaa* “The piggy is washing a shirt” with non-contrastive object focus (NFO).

At the beginning of each trial, the participant saw a slide as illustrated in **Figure 1**, containing a picture of the robot and visual display of a scene. A part of this scene was covered by a blue shape, but the participant was told that the robot could see the complete scene. The experimenter described the visible part of the scene and then asked the robot about the covered part. For example for the item in **Figure 1**, the introductory description was *Possu pesee jotain ammeessa* “A piggy is washing something in a tub,” followed by the question *Mitä possu pesee?* “What is the piggy washing?” The robot answered the question, in this case with *Possu pesee paitaa* “The piggy is washing a shirt.” After this, the experimenter posed the same question to the participant, who answered the question based on the robot’s answer. The participant’s answer thus had the same information structural context as the robot’s utterance, i.e., inducing non-contrastive narrow focus on the object *paitaa* “shirt” for the present example. Finally, the experimenter removed the blue cover from the picture to reveal the complete scene. For broad focus items, the central part of the picture was completely covered and the experimenter introduced these items with “Now we cannot see what is happening” or “Now the whole picture is covered again” before asking a broad focus question like shown in **Table 2** below.

Every participant was presented with 48 experimental items in a pseudo-randomized order. For half of the participants, the presentation order was reversed. Participants’ speech was recorded directly onto the hard drive of a laptop computer with a sampling frequency of 44,100 Hz, using a high-quality head-mounted microphone.

2.2. Materials

Five different information structural conditions appeared in the robot’s as well as the participants’ answers: broad focus (BF), contrastive narrow focus on the subject (CFS), contrastive narrow focus on the object (CFO), non-contrastive narrow focus on the subject (NFS), and non-contrastive narrow focus on the object (NFO). For the sake of simplicity, we will refer to contrastive narrow focus as “contrastive focus” and to non-contrastive narrow focus as “narrow focus” in the following. Sixteen trials elicited broad focus, a further 16 trials elicited focus on the subject (8 NFS and 8 CFS), and another 16 trials elicited focus on the object (8 NFO and 8 CFO).

Table 2 shows examples of the experimenter’s questions and the robot’s answers for all conditions. Note that the information structure of the (robot’s and participants’) answers was not only set up in the experimenter’s question, but also in her preceding description of the the picture, as well as in the picture itself. Thus, for broad focus, the whole scene was covered in the picture and the experimenter introduced the trial by stating something like *The whole picture is covered* or *The picture is again covered so that we cannot see what is happening*. By contrast, in all contrastive and non-contrastive narrow focus conditions, only the character who was supposed to be focused in the answer was covered in the picture (e.g., the shirt in **Figure 1**). The experimenter’s introductions of these trials accordingly replaced the hidden entity with *someone* or *something*, while naming the other entity and the action (see above for the introductory description for **Figure 1**, which mentions that the piggy is washing something). Thus, the named entities can be treated as given and the focused

TABLE 2 | Example question-answer pairs for all experimental conditions.

	SVO	OVS
BF	Q: Mitä tässä tapahtuu? "What is happening here?"	Q: Mitä tässä tapahtuu? "What is happening here?"
	A: [Tyttö lakaisee katua] _F . girl.NOM sweeps street.PRT "A girl is sweeping a street."	A: [Paitaa silittää poika] _F . shirt.PRT irons boy.NOM "A boy is ironing a shirt."
CFS	Q: Poimiiko leijona mansikkaa? "Is a lion picking the strawberry?"	Q: Vetääkö kala autoa? "Is a fish pulling the car?"
	A: [Varsa] _F poimii mansikkaa foal.NOM picks strawberry.PRT "A foal is picking the strawberry."	A: Autoa vetää [kameli] _F . car.PRT pulls camel.NOM "A camel is pulling the car."
CFO	Q: Peseekö prinsessa autoa? "Is the princess washing a car?"	Q: Ompeleeko noita nallea? "Is the witch sewing a teddy bear?"
	A: Prinsessa pesee [maljakkoa] _F . princess.NOM washes vase.PRT "The princess is washing a vase."	A: [Paitaa] _F ompelee noita. shirt.PRT sews witch.NOM "The witch is sewing a shirt."
NFS	Q: Kuka raaputtaa ikkunaa? "Who is scratching the window?"	Q: Kuka pitää haarukkaa? "Who is holding the fork?"
	A: [Kameli] _F raaputtaa ikkunaa camel.NOM scratches window.PRT "A camel is scratching the window."	A: Haarukkaa pitää [lapsi] _F . fork.PRT holds child.NOM "A child is holding the fork."
NFO	Q: Mitä isoäiti antaa? "What is grandmother giving?"	Q: Mitä lapsi ostaa? "What is the child buying?"
	A: Isoäiti antaa [pallon] _F . grandmother.NOM gives ball.ACC "Grandmother is giving a ball."	A: [Paitaa] _F ostaa lapsi. shirt.PRT buys child.NOM "The child is buying a shirt."

BF, broad focus; CFS, contrastive focus on the subject; CFO, contrastive focus on the object; NFS, narrow focus on the subject; NFO, narrow focus on the object.

entity as focused in the answer not only because of the question, but also based on the previous conversational turn. For the contrastive focus condition, the context first set up the narrow focus the same way before introducing the contrast between the incorrect replacement for *someone/something* suggested in the experimenter’s question and the correct replacement given in the robot’s answer (see **Table 2**) and revealed when the blue cover was removed from the picture.

The second condition varied in the experiment was the word order of the robot’s utterance; she produced 24 sentences in SVO order and another 24 in OVS order. Recall that both word orders are grammatical in Finnish, but while SVO is the unmarked order, OVS is not always information-structurally appropriate. We restricted our attention to SVO and OVS, since these word orders are more frequent than word orders involving the contrast position (Hakulinen and Karlsson, 1995, see discussion in Section 4.1) and it can thus be assumed that children acquire them earlier. Furthermore, this restriction reduced the complexity of the experiment and allowed for better comparability to parallel studies on Dutch and German (Sauermaun et al., 2011; Chen and Höhle, submitted).

Both factors were crossed as illustrated in **Table 2**. The item list thus consisted of 48 sentences. All sentences were unique, but most of the subject and object words appeared in more than one sentence, although always in a different information structural condition and frequently combined with a different verb and a different object or subject, respectively (see for example the subject *kameli* “camel” in conditions NFS-SVO and CFS-OVS and the object *paitaa* “shirt” in the conditions BF-OVS, CFO-OVS, and NFO-OVS in **Table 2**). Subjects were always in nominative case, while objects carried partitive or accusative case, depending on the verb. The complete list of question prompts and robot answers is available as Supplementary Material.

All robot utterances were created from words spoken by a 24-year-old female native speaker of Finnish from Helsinki. To create stimuli devoid of natural sentence-level prosody, the speaker first recorded all words separately in random order. Second, flat f_0 of around 200 Hz was imposed for all words, using the program Praat (Boersma and Weenink, 1992–2014). The temporal structure of the words was preserved, since Finnish has lexical quantity distinctions, so that manipulating segment durations could impact perceptibility. Finally, the words were concatenated into sentences, with 200 ms pauses between all words, as well as at the beginning and end of each sentence. As mentioned above, the robot’s utterances included pragmatically inappropriate word orders in addition to this unnatural prosody. To further encourage the participants in their role as language teachers, the robot’s introduction before the beginning of the experiment contained morphological agreement errors. The experimental prompts, i.e., the robot’s answers to the experimenter’s questions, however, did not contain this type of error to minimize task load.

2.3. Participants

Twenty four- to five-year-old Finnish children from the Helsinki area participated in the study (10 male and 10 female, mean age: 5;1, range: 4;6–5;6). One participant was excluded from the analysis because the daycare teacher expressed concerns regarding language development, three other participants were excluded because they had trouble following the experimental protocol. Thus, we analyzed data from 16 participants.

Dates of birth were not recorded for three of the participants in the final data set. Their data were retained as they were confirmed to be within the target age range. Alternative subset analyses excluding these three participants were conducted for all dependent measures. They confirmed the same result patterns as the analyses reported below, although some effects were weakened as is to be expected when noticeably reducing the number of data points. Since one of the three participants without exact age information also reported to speak English, we performed further subset analyses excluding only his data, which only showed different significance levels for one of the measures, i.e., the use of pauses after sentence-medial verbs, but otherwise returned the same results as the models of the complete data set. We therefore report the results from the complete data set.

The experiments of the present study were non-invasive and were carried out in accordance with Finnish law and adhered to the guidelines of the Declaration of Helsinki, the

American Psychological Association, and the ethical policies of the University of Helsinki. The university abides by the guidelines of the Finnish Advisory Board on Research Integrity on the responsible conduct of research and procedures for handling allegations of misconduct as well as on ethical review in human sciences. As we obtained parental consent for the participants' research participation, this study was exempt from ethics review and approval by the University of Helsinki Ethical Review Board in the Humanities and Social and Behavioral Sciences.

3. RESULTS

We removed responses from trials that were unsuitable for analysis, for example because the experimenter had asked the wrong question, inducing a different information structure than intended, because the participant imitated the robot and spoke with flat f_0 or because the participant gave an elliptical one-word answer. Altogether, we excluded 26% of trials, retaining 571 sentences for analysis⁴.

We analyzed the data with respect to f_0 range, duration, pauses and voice quality, because effects of information structure on those measurements have been found in Finnish adults. We additionally evaluated choice of word order in the participants' utterances. Intensity was not analyzed since recording quality varied too widely to allow reliable measurements, despite the use of a head-mounted microphone.

The analyses of all the above-mentioned dependent variables were performed using linear mixed-effect models in R (Baayen, 2008; R Core Team, 2015), as implemented in the package lme4 (Bates et al., 2015). These models test the significance of predictor variables, as well as including random effects to model for example participant variation in the data (Baayen et al., 2008). Here, for the analysis of word order produced by our participants we tested information structure induced by the experimenter's question (levels: BF, NFO, NFS, CFO, CFS) and word order (levels: SVO, OVS) as predictor variables (fixed effects). For the prosodic analyses, since measures were obtained for individual words, we additionally included the predictor constituent, which coded the grammatical role of the word (levels: subject, object). As random effects, we tested participant and lexical item. We determined the model with the best fit to the data by comparing the log likelihood of models including different variables with the anova function. Only variables significantly contributing to an improved model fit were retained. P - and χ^2 values for the significance of predictors were obtained from these model comparisons. For models of categorical dependent variables (here: word order, voice quality, and occurrence of pauses), which were binomial (also see Jaeger, 2008), which were binomial, the model output included p -values indicating the significance of differences between factor levels. For models of continuous variables (here: f_0 range, duration, and pause duration), p -values

were obtained with the package lmerTest (Kuznetsova et al., 2015).

3.1. Word order

Participants used SVO word order in 68% of their utterances and OVS order in 32% of their utterances (388 and 180 cases, respectively), while they used OSV and VOS word order in less than one percent of the entire data set. Since the number of sentences produced with other word orders was negligible, we conducted statistical modeling for SVO vs. OVS responses with binomial models. Model comparison suggested that word order in the participants' utterances ("output word order") was significantly affected both by the word order used by the robot ("input word order") and by the information structure induced by the experimenter's question [$p < 0.001$, $\chi^2 = 327.4$ and $p = 0.006$, $\chi^2 = 14.5$, respectively]. A more complex model including an interaction between the two factors did not converge for this data set, i.e., there were not enough data points to compute a reliable interaction model. The best linear mixed-effects model of output word order indicated that participants produced SVO order significantly more frequently in response to SVO input word order than in response to OVS input word order [$estimate = 5.7509$, $SE = 0.6346$, $z = 9.0617$, $p < 0.001$; positive estimates indicate more, negative ones less SVO productions]. When the input word order was unmarked SVO, participants also produced SVO utterances in 99% of the cases, with almost no difference between the information structures (see the right panel of Figure 2). Only one OVS utterance appeared in broad focus, contrastive object focus and narrow object focus each. When the robot's utterance used OVS, participants also retained the input word order in the majority of cases (60% overall).

However, the model also revealed an effect of information structural conditions: Compared to the broad focus condition, participants uttered significantly more SVO sentences only in the contrastive object focus condition [$estimate = 1.5129$, $SE = 0.4316$, $z = 93.5054$, $p < 0.001$]. As can be seen in the left panel of Figure 2, participants produced about a third of the input OVS sentences with an output SVO order in most

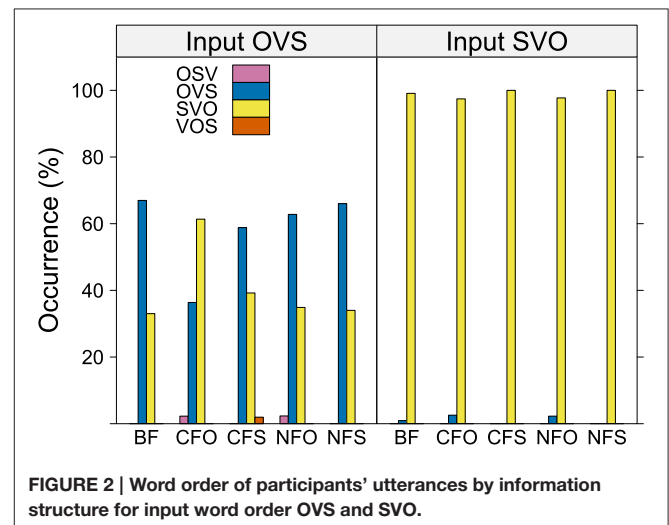


FIGURE 2 | Word order of participants' utterances by information structure for input word order OVS and SVO.

⁴Conditions were not distributed exactly equally in the removed items, but there were no large differences either. Of removed sentences with SVO input word order, 18 had BF as the intended information structure, 25 CFO, 26 CFS, 20 NFO, and 18 NFS. Among removed sentences with input OVS order, 22 were intended BF, 20 CFO, 13 CFS, 21 NFO, and 14 NFS.

information structural conditions, but realized 61% sentences of input OVS sentences with output SVO order in contrastive object focus. Even though the interaction between word order and focus condition could not be tested for the data set as a whole, the figure thus strongly indicates that the significant difference between the conditions BF and CFO stems from sentences with input OVS word order. To assess this, we created a subset of the data and performed a linear mixed-effects analysis only for the two critical information structural conditions. Modeling results were inconclusive, but provided some limited evidence for an interaction⁵. Note that in the condition combining input OVS order with CFO focus condition, the robot's production constituted a mismatch between the information structure induced by the experimenter's question (contrastive object focus) and the information structure implied by the word order (narrow subject focus). Participants almost never corrected this mismatch with an explicit syntactic marking of contrastive object focus, i.e., by placing the object in the sentence-initial contrast position (Table 1 above). A possible interpretation is that at age five, Finnish children have not yet acquired the use of the sentence-initial position for marking contrast. Instead, they frequently realized SVO word order and placed the contrastive object sentence-finally, in the default position for narrow focus. Recall, however, that as the unmarked word order, SVO is felicitous in most information structural contexts. Therefore, substituting it for the more marked OVS order achieved a better alignment between syntax and information structure for contrastive object focus. However, this happened only for contrastive object focus condition, whereas no significant difference from broad focus was observed for the other information structural conditions, including narrow object focus.

To sum up, participants were strongly influenced by the word order of the input and overwhelmingly used SVO and OVS sentences. Less than 1% of their utterances had a filled contrast position. Notably, only one significant difference appeared between information structural conditions, and there was some limited evidence that it stemmed from sentences with input OVS word order: Participants used the unmarked SVO order more often in contrastive object focus than in broad focus, placing the focused constituent in sentence-final position.

3.2. F₀

We analyzed prosody for the 568 sentences that the participants produced with either SVO or OVS order. Out of the 1136 subject and object nouns, 42 subjects and 53 objects could not be analyzed with respect to their *f*₀ due to bad sound quality or the presence of non-modal voice quality (see Section 3.5). Thus, we analyzed the *f*₀ of 1041 words. Figure 3 plots the average values of three pitch measurement points for subject and object nouns in different information structural conditions for both word

⁵Significant main effects of input word order [*estimate* = 6.4843, *SE* = 1.1024, *z* = 5.882, *p* < 0.001] and information structure [*estimate* = 1.7947, *SE* = 0.4841, *z* = 3.707, *p* < 0.001] appeared, while the interaction between them was marginal [*p* < 0.080, χ^2 = 3.1; *estimate* = -2.9450, *SE* = 1.5258, *z* = -1.930, *p* = 0.054]. The interaction appeared significant when a by-participant random effect of information structure was included in the model [*estimate* = -3.3543, *SE* = 1.6483, *z* = -2.035, *p* = 0.042], but this did not significantly improve model fit [*p* = 0.755, χ^2 = 0.6].

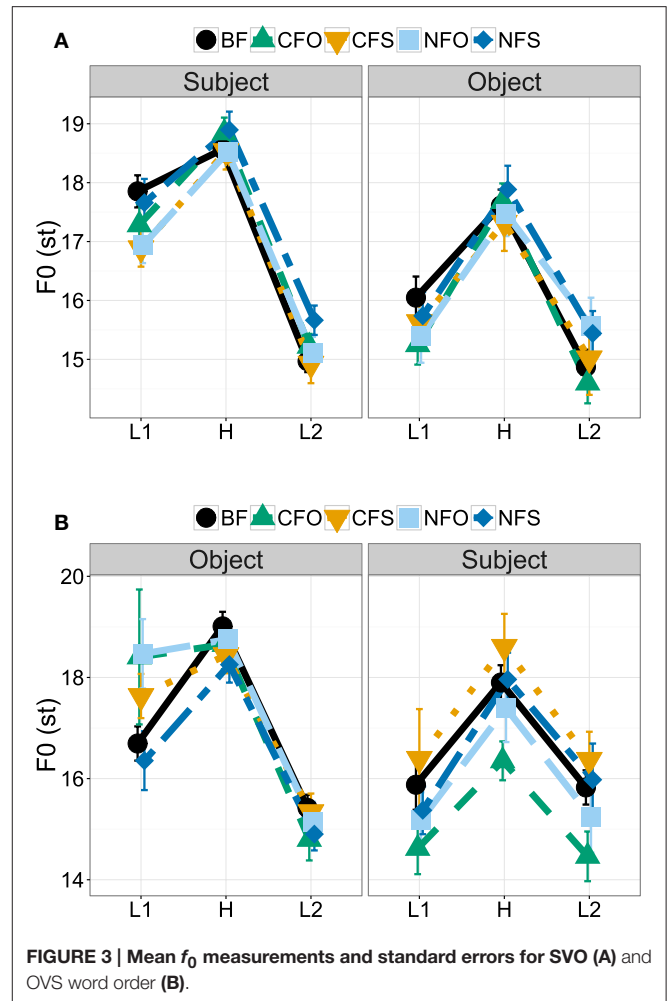
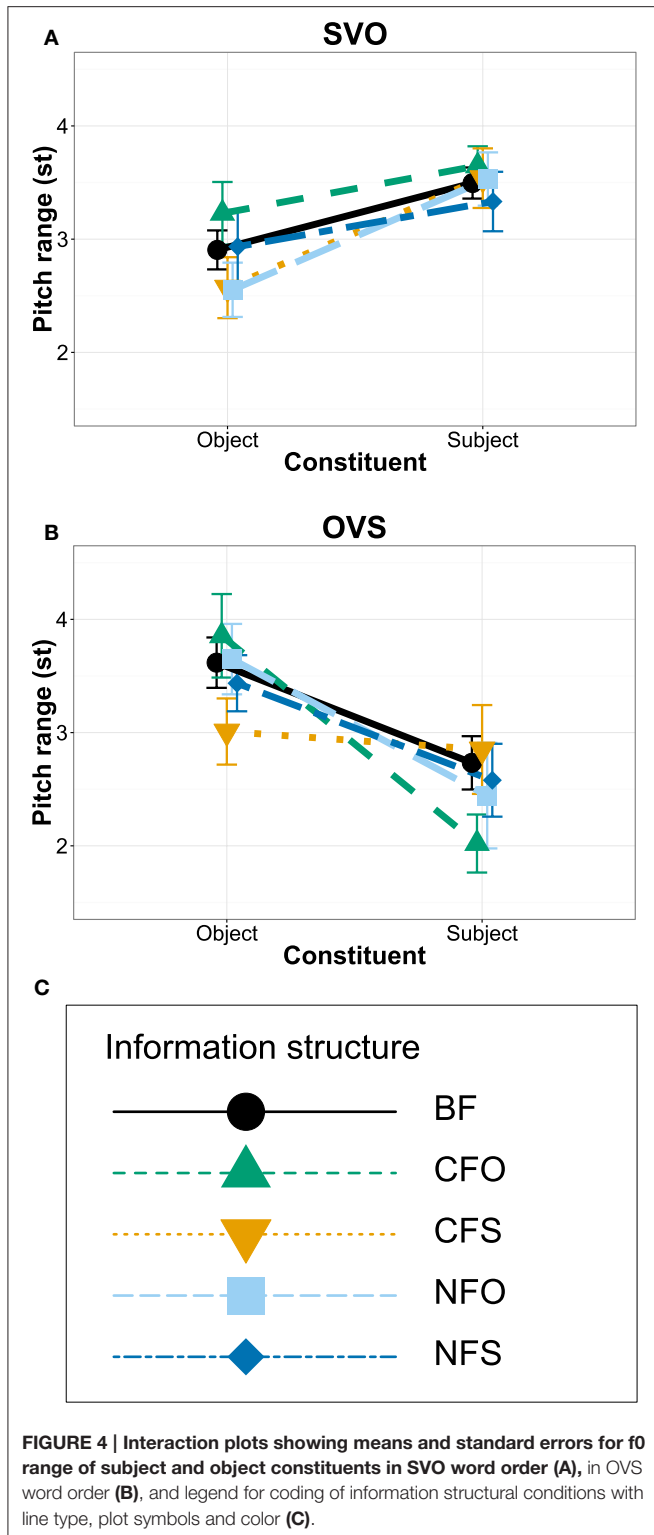


FIGURE 3 | Mean *f*₀ measurements and standard errors for SVO (A) and OVS word order (B).

orders: the *f*₀ maximum (H) and the minimum before (L1) and after it (L2) within the same word. Figure 4 directly compares *f*₀ range in different conditions (calculated as the distance between the maximum H and the lower one of the two minima). All measurements were converted to semitones (st) relative to a reference value of 100 Hz and statistical analyses evaluated *f*₀ range.

The best linear mixed-effects model included a significant interaction between word order and constituent [*p* < 0.001, χ^2 = 44.9]. It suggested that *f*₀ ranges decreased over the course of the utterance, i.e., *f*₀ ranges were larger for objects than for subjects in OVS sentences and larger for subjects than for objects in SVO sentences, as illustrated in Figures 3, 4. Accordingly, the model contained significant main effects indicating overall smaller *f*₀ ranges for subject constituents than for objects [*estimate* = -0.9148, *SE* = 0.1927, *t* = -4.747, *p* < 0.001] and smaller *f*₀ ranges in SVO output order than for the OVS intercept [*estimate* = -0.7711, *SE* = 0.1629, *t* = -4.735, *p* < 0.001], together with an interaction suggesting that subjects in SVO sentences had larger *f*₀ ranges [*estimate* = 1.5768, *SE* = 0.2326, *t* = 6.778, *p* < 0.001]. Figures 3, 4 show a clear peak downstep and, as a result, a reduced *f*₀ range on the sentence-final constituent in both word orders.



There was little indication of a significant effect of information structure for the data set as a whole, as adding this predictor did not significantly improve model fit [$p = 0.348, \chi^2 = 4.5$]. A model including an interaction between information structure and constituent likewise did not provide a significantly better

fit [$p = 0.315, \chi^2 = 9.3$], even though it included significant interactions and effects of information structure indicating that f_0 range was marginally larger in broad focus than for contrastive subject focus [$estimate = 0.4606, SE = 0.2396, t = 1.922, p = 0.056$] and significantly larger in contrastive object focus than in contrastive subject focus [$estimate = 0.7713, SE = 0.2865, t = 2.692, p = 0.008$], while the difference between contrastive subject focus and contrastive object focus was significantly smaller for subjects than for objects [$estimate = -0.8893, SE = 0.4074, t = -2.183, p = 0.030$] and the difference between contrastive and narrow subject focus was marginally smaller for subjects than for objects [$estimate = -0.6646, SE = 0.3840, t = -1.731, p = 0.084$].

Figures 3, 4 illustrate that the largest differences between the conditions CFO and CFS appeared in OVS word order, while lines representing different focus conditions in SVO order overlap in Figure 3A. Indeed, a subset model of SVO sentences suggested no significant effect of information structure [$p = 0.486, \chi^2 = 3.4$], only effects of an f_0 downtrend over the course of the sentence, i.e., a significant difference between subjects and objects [$p < 0.001, \chi^2 = 24.9$], with larger f_0 ranges for the initial subjects [$estimate = 0.6634, SE = 0.1317, t = 5.038, p < 0.001$]. For OVS order, by contrast, Figures 3B, 4B display lower f_0 maxima (H) and smaller ranges of the sentence-final subjects in contrastive object focus condition, but raised maxima and larger ranges when the subjects themselves were contrastively focused. To test these differences in the face of contradictory evidence from modeling the whole data set, we conducted subset analyses for the two contrastive focus conditions in OVS order. Here, the best model contained an effect suggesting overall marginally smaller f_0 ranges for subject focus (CFS) than for object focus (CFO) [$estimate = -0.8707, SE = 0.4462, t = -1.951, p = 0.055$] and significantly smaller f_0 ranges for subjects than for objects [$estimate = -1.9552, SE = 0.5470, t = -3.574, p < 0.001$], in addition to significant interaction between constituent and information structure [$p = 0.01, \chi^2 = 6.7; estimate = 1.7852, SE = 0.6730, t = 2.653, p = 0.01$]. This suggests that the f_0 range of contrastively focused constituents was significantly extended in this word order, while the f_0 range of unfocused constituents was compressed at the same time.

In summary, there was some evidence that participants' pitch scaling differed between the two output word orders: In sentences that the participants produced with SVO word order, there was no significant effect of information structure. In OVS sentences, by contrast, a subset model suggested that f_0 range was expanded when the constituent itself was contrastively focused, but was compressed for contrastive focus on the other constituent.

3.3. Duration

We measured the duration of all subjects and objects in the data set and analyzed subjects and objects in the 568 sentences with SVO and OVS output order (1136 items). The length of target words varied in terms of number of syllables and segments, so that evaluating total word duration was not informative. Therefore, we divided total word duration by number of syllables and evaluated the resulting measure of mean syllable duration. Although Finnish is a quantity language with some

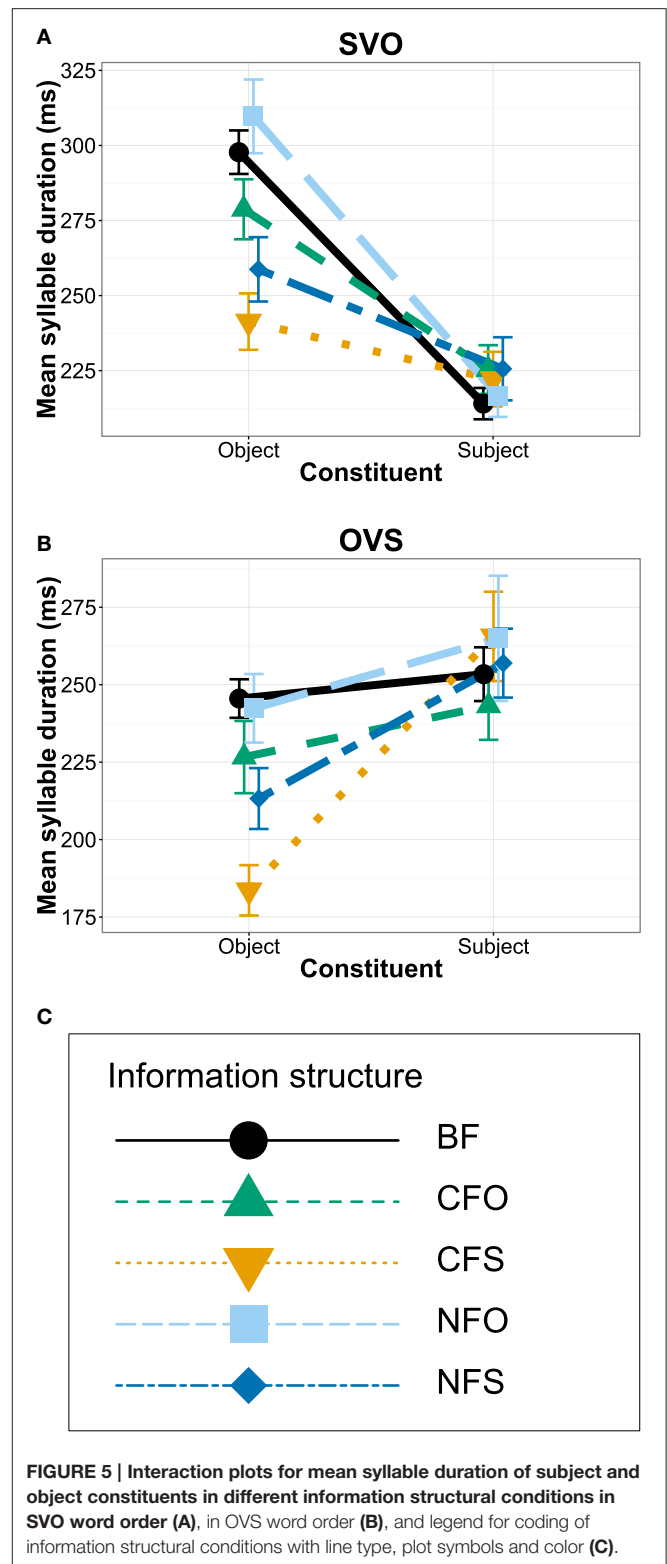
isochronous tendencies (Iivonen, 1974, found longer phoneme durations in shorter words, Suomi and Ylitalo, 2003, found no tendency for syllable isochrony, Suomi and Ylitalo, 2004, observed isochronous tendencies for disyllabic feet, but not for longer ones), this considerably reduced variability between items in size, diminishing overall standard deviation from 196 ms for words to 79 ms for syllables.

Statistical modeling of the data suggested significant final lengthening, i.e., longer durations for constituents in sentence-final position: The model indicated shorter syllable durations in output SVO order than in OVS order [*estimate* = -36.9634, *SE* = 5.9302, *t* = -6.2331, *p* < 0.001] and a positive interaction between the factors constituent and word order suggested that durations were longer in SVO word order for objects than subjects [*p* < 0.001, $\chi^2 = 128.8$; *estimate* = 97.1528, *SE* = 8.3127, *t* = 11.6873, *p* < 0.001]. Thus, only final objects had longer durations in SVO order, while initial subject durations were shorter. This is illustrated in **Figure 5A**, which shows mean syllable duration in the different information structural conditions. By contrast, **Figure 5B** displays generally longer durations for subjects in OVS word order. While the main effects of information structure were not significant, the best model contained a significant interaction indicating an effect of information structure on object durations [*p* = 0.012, $\chi^2 = 12.9$]: Durations of object nouns were significantly shorter in both conditions with narrow focus on the subjects (CFS and NFS) than in broad focus condition. The effect was larger for contrastive focus, but significant for narrow focus as well [*estimate* = -59.9199, *SE* = 17.9352, *t* = -3.3409, *p* = 0.001 and *estimate* = -40.3638, *SE* = 17.4382, *t* = -2.315, *p* = 0.023, respectively]. Adding an interaction between information structure and word order or a three-way interaction between information structure, word order and constituent (subject vs. object) did not improve the model significantly [*p* = 0.825, $\chi^2 = 1.5$ and *p* = 0.819, $\chi^2 = 4.4$, respectively], indicating that the effect of information structure did not differ between word orders. The analysis thus indicated a consistent strategy of shortening object durations in subject focus conditions for both SVO order, where it curtailed final lengthening, and OVS order, where the shortened object was in topic position (see the shorter durations for objects for conditions NFS and CFS in **Figure 5**).

Thus, this section presented information structure effects on the duration of objects in both word orders—SVO and OVS—and for both focus types—contrastive or narrow focus. Object nouns showed significantly reduced mean syllable durations when subject nouns were in focus.

3.4. Pauses

There are two relevant positions for evaluating the occurrence of pauses, defined as a perceivable silence irrespective of duration (see below), in the short SVO and OVS sentences analyzed here: first, before the verb, i.e., after the subject in SVO sentences and after the object in OVS sentences, and second, after the verb, i.e., before the object in SVO sentences and before the subject in OVS sentences. As is common in child data (Redford, 2013), pauses were frequent in our participants' productions. Altogether, 567



pauses appeared in the 568 sentences evaluated here, of which 45% preceded the verb and 55% followed the verb. **Figure 6** shows their distribution by information structural condition and output word order.

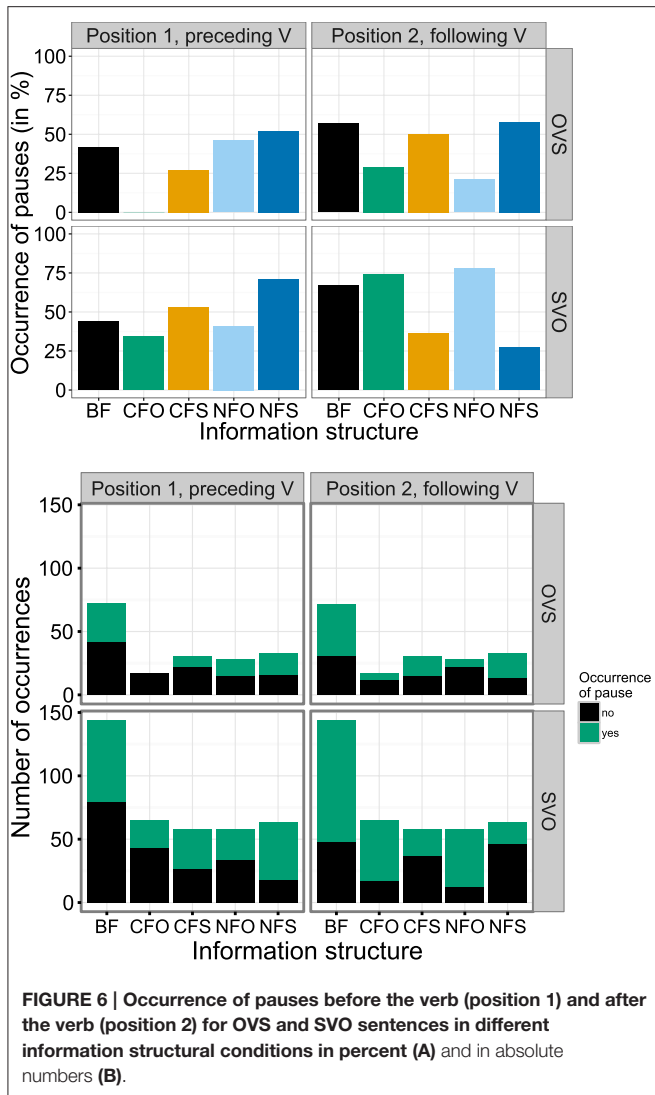


FIGURE 6 | Occurrence of pauses before the verb (position 1) and after the verb (position 2) for OVS and SVO sentences in different information structural conditions in percent (A) and in absolute numbers (B).

The best binomial model of pause occurrence in the first, pre-verbal position included significant effects of both output word order and information structure [$p = 0.006, \chi^2 = 7.5$ and $p < 0.001, \chi^2 = 29.7$, respectively]. It suggested that pauses preceding verbs were overall more frequent in SVO than in OVS word order [$estimate = 0.531, SE = 0.1957, z = 2.7133, p < 0.01$; positive estimates indicate more frequent pause occurrence]. Regarding information structure, the model indicated that these pauses were significantly less frequent in contrastive object focus (CFO) and significantly more frequent in narrow subject focus (NFS) compared to BF condition [$estimate = -0.8460, SE = 0.2916, z = -2.9008, p = 0.004$ and $estimate = 0.8912, SE = 0.2590, z = 3.4412, p < 0.001$, respectively]. The left panels of Figure 6 illustrate that this was true in both word orders.

With respect to pauses following the verb, the best model included interactions between information structure and word order [$p < 0.001, \chi^2 = 43.7$]. Main effects of information structure reflected the distribution in the intercept word order

OVS, i.e., marginally fewer pauses in contrastive object focus (CFO) [$estimate = -1.1328, SE = 0.5866, z = -1.9312, p = 0.053$] and significantly fewer pauses in narrow object focus (NFO) [$estimate = -1.5983, SE = 0.5214, z = -3.0653, p = 0.002$] than in broad focus (see top right panel of Figure 6). In SVO word order, by contrast, post-verbal pauses were significantly more frequent in both object focus conditions (CFO and NFO) [$estimate = 1.4826, SE = 0.6749, z = 2.1967, p = 0.028$ and $estimate = 2.1545, SE = 0.6362, z = 3.3866, p < 0.001$, respectively], while being significantly less frequent in narrow subject focus (NFS) and marginally less frequent in contrastive subject focus (CFS) [$estimate = -1.7060, SE = 0.5430, z = -3.1417, p = 0.002$ and $estimate = -0.9927, SE = 0.5470, z = -1.8150, p = 0.069$, respectively], see bottom right panel in Figure 6. Thus, pauses after the verb occurred more often before a constituent in narrow focus in SVO word order. When the other, sentence-initial constituent was focused, pauses after the verb occurred less often in both word orders.

The analyses of pause occurrence were based on a native speaker's annotation, who inspected the waveforms and the spectrograms, and labeled all perceivable pauses irrespective of their duration (see Romøren and Chen, 2015, on advantages of this method over setting a minimum pause duration). The shortest annotated pause was a little over 1 ms long, while the longest was 458 ms, with a median of 9 ms and a standard deviation of 44 ms. An analysis of pause duration showed no significant effect of information structure [$p = 0.548, \chi^2 = 3.1$]. The best model for the duration of all 567 pauses realized by the participants also included an effect of pause position [$p < 0.001, \chi^2 = 16.4$], which indicated that pauses preceding the verb were overall significantly longer than those following the verb [$estimate = -14.1695, SE = 3.4734, t = -4.0795, p < 0.001$]. An alternative model additionally containing a marginal effect of word order, which indicated longer durations for pauses in SVO sentences, only provided a marginally better fit to the data [$p = 0.076, \chi^2 = 3.1$]. Modeling the duration of pauses preceding the verb separately, there was no significant effect of information structure [$p = 0.626, \chi^2 = 2.6$] and a model including word order was only marginally better than a null model without any predictor variables [$p = 0.060, \chi^2 = 3.5$]. For the subset of pauses following the verb, neither word order nor information structure had a significant effect [$p = 0.592, \chi^2 = 0.3$ and $p = 0.954, \chi^2 = 0.7$, respectively].

Thus, we found no relevant variation in pause duration, but a significant increase in the occurrence of pauses both before and after constituents in narrow focus for several conditions. Conversely, pauses occurred significantly less often before or after an unfocused constituent, i.e., when another constituent in the sentence was focused. In other words, the data indicated a tendency toward prosodic separation of focused constituents into prosodic phrases of their own, whereas non-focused constituents were more frequently not separated and rather phrased with the verb. Contrastive and narrow focus conditions generally displayed the same pattern, although effects did not always reach significance. Interestingly, a significant increase in pauses only appeared for SVO sentences, either separating initial subjects from the rest of the sentence with following pauses or

disconnecting the final object and the preceding constituents. Recall that while sentences with unmarked SVO word order are by default divided into information structural fields as shown in **Table 1**, this word order is appropriate with other information structures as well. Participants might thus have used the insertion of pauses as part of a prosodic strategy for disambiguation.

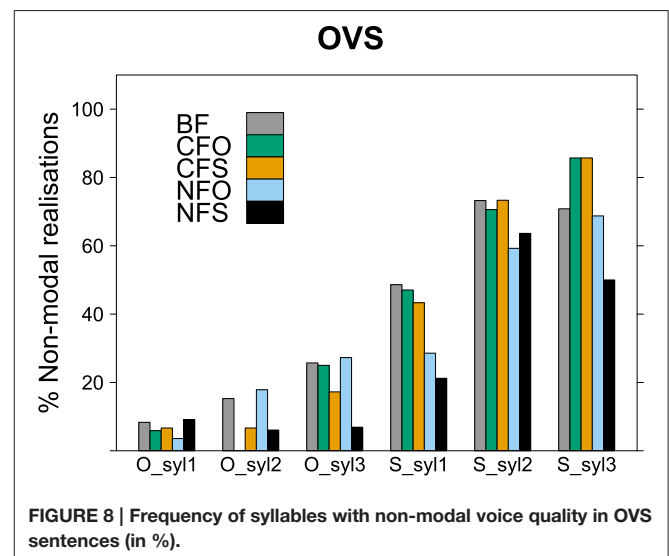
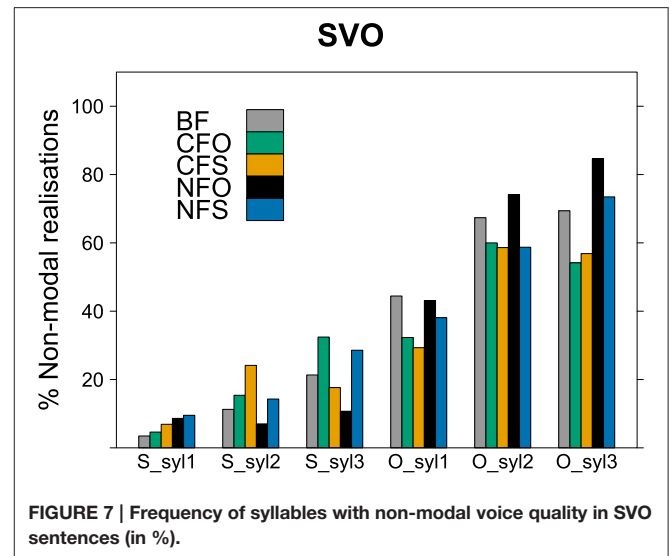
3.5. Voice Quality

We evaluated voice quality in a binary fashion here, determining for each syllable whether it was realized with modal voice throughout or whether it was realized (partially or completely) with non-modal voice, e.g., creaky, breathy or whispery voice. This annotation was based on waveform, spectrogram and auditory impression. Syllable-based evaluation was chosen as a compromise between a rather coarse word-by-word analysis and a detailed phoneme-level approach, which would make comparisons between different lexical items difficult. A binary syllable-level evaluation of voice quality has shown effects of information structure in adult Finnish, see Arnhold (2016) who also discusses differences between this method and a study on the effects of focus on voice quality in Finnish by Vainio et al. (2010) using inverse filtering (also see Epstein, 2002; Ní Chasaide et al., 2011, on English). Out of the 1136 words considered in the prosodic analyses, 1131 (nearly 100%) were disyllabic or longer and 536 (47%) had at least three syllables. We therefore restricted our analyses to voice quality of the first three syllables, discarding later syllables of longer words for the sake of comparability.

With the original coding of the data, some interactions could not be tested conclusively because more complex models failed to converge. The most complex converging model included an effect of syllable number indicating that second and third syllables were realized with non-modal voice quality more often than first syllables [$p < 0.001$, $\chi^2 = 113.0$; for second syllables: $estimate = 0.88107$, $SE = 0.09880$, $z = 8.918$, $p < 0.001$; for third syllables: $estimate = 1.06692$, $SE = 0.12089$, $z = 8.825$, $p < 0.001$]. It also contained significant interactions between word order and focus condition [$p = 0.018$, $\chi^2 = 11.9$] and between constituent and focus condition [$p = 0.006$, $\chi^2 = 14.6$]⁶. They indicated that there were less non-modal realizations in narrow subject focus overall compared to broad focus, but more in narrow subject focus for SVO sentences [$estimate = -0.85446$, $SE = 0.24766$, $z = -3.450$, $p < 0.001$ and $estimate = 0.90481$, $SE = 0.27325$, $z = 3.311$, $p < 0.001$, respectively] and that there were overall less non-modal realizations of subject than object nouns and less in contrastive subject focus than in broad focus, but significantly more non-modal realizations of subjects in

contrastive subject focus condition [$estimate = -0.82535$, $SE = 0.14079$, $z = -5.862$, $p < 0.001$, $estimate = -0.74142$, $SE = 0.24652$, $z = -3.008$, $p = 0.003$ and $estimate = 0.75097$, $SE = 0.25669$, $z = 2.926$, $p = 0.003$, respectively]. When SVO word order was taken as the reference level instead of OVS, the interaction between word order and information structure showed a significant difference between broad focus and narrow object focus, indicating less non-modal realizations for OVS word order in narrow object focus [$estimate = -0.90475$, $SE = 0.27321$, $z = -3.311$, $p < 0.001$].

As illustrated in **Figures 7, 8**, the use of non-modal voice in final narrow focus (NFO in SVO order and NFS in OVS; black bars) deviated from all other information structures in both word orders (lighter bars). The figures show the percentage of syllables with non-modal voice in the different information structural conditions for output SVO and OVS order, respectively. The first group of bars illustrates the percentages for the first syllable of



⁶Separate comparisons with a simple model also indicated a significant interaction between constituent and word order [$p < 0.001$, $\chi^2 = 655.6$], suggesting that non-modal voice quality was generally more frequent later in the sentence, i.e., more non-modal voice for subjects than for objects and more non-modal voice in SVO order than in OVS, but less non-modal voice for subjects in SVO sentences, where they were in initial position [$estimate = 2.8597$, $SE = 0.1978$, $z = 14.454$, $p < 0.001$, $estimate = 2.6861$, $SE = 0.1808$, $z = 14.858$, $p < 0.001$ and $estimate = -5.4059$, $SE = 0.2472$, $z = -21.871$, $p < 0.001$, respectively]. However, a model adding this interaction to the other two did not converge anymore, suggesting that model complexity was too high for the amount of data. To assess interactions between sentence position and focus condition, we therefore report the results of modeling with a simplified coding below.

the sentence-initial word (subject for SVO; object for OVS) and the second and third group give percentages for the second and third syllable of the initial word. The three groups of bars on the right of each panel indicate the occurrence of non-modal voice for the first three syllables of the sentence-final word (object for SVO; subject for OVS).

A model comparing sentence-final narrow focus with all other information structural conditions provided an equally good or better fit to the data as models with the original coding of information structural conditions. This model found a significant interaction of information structure with syllable position [$p = 0.024$, $\chi^2 = 13.0$]. It indicated that creaky, breathy or whispery realizations were significantly more frequent for all later syllables than for first syllables of sentence-initial constituents, i.e., subjects in SVO and objects in OVS, respectively [for second syllables in initial words: $estimate = 1.0294$, $SE = 0.2424$, $z = 4.2475$, $p < 0.001$; for third syllables in initial words: $estimate = 1.7515$, $SE = 0.2593$, $z = 6.7548$, $p < 0.001$; for first syllables in final words: $estimate = 2.5986$, $SE = 0.2258$, $z = 11.5097$, $p < 0.001$; for second syllables in final words: $estimate = 3.7753$, $SE = 0.2297$, $z = 16.4367$, $p < 0.001$; for third syllables in final words: $estimate = 3.8912$, $SE = 0.2551$, $z = 15.2514$, $p < 0.001$]. As the figures demonstrate, voice quality was mostly modal early in the sentence, whereas the percentage of non-modal realizations rose steadily throughout the sentence in both word orders and for most information structural conditions. The model further indicated that in sentences with narrow focus on the final constituent (NFO in SVO order and NFS in OVS order), non-modal realizations were significantly less frequent on the second and third syllable of the sentence-initial word than in the other information structural conditions [second syllables: $estimate = -1.3535$, $SE = 0.6274$, $z = -2.1573$, $p = 0.031$; third syllables: $estimate = -1.5827$, $SE = 0.6701$, $z = -2.3618$, $p = 0.018$]. This reflects the deferred rise in the percentage of non-modal realizations for the NFO condition in SVO word order and for the NFS condition in OVS visible for the black bars in **Figures 7, 8**, respectively. Additionally, the model contained an interaction between information structure and word order [$p = 0.014$, $\chi^2 = 6.0$], suggesting that sentences with narrow focus on the final constituent displayed overall more non-modal realizations when the word order was SVO than when it was OVS [$estimate = 0.7488$, $SE = 0.3092$, $z = 2.4218$, $p = 0.016$], compare the last three black bars in **Figures 7, 8**.

To sum up, our analyses showed an interaction between word order and information structure: While the use of non-modal voice quality increased steadily throughout the sentence in most information structural conditions, a different pattern consistently emerged in conditions with non-contrastive final focus. In these conditions (NFO in SVO word order and NFS in OVS), non-modal voice remained infrequent throughout the first constituent and then increased abruptly on the final focused constituent. With the narrow focus occupying the final position, these utterances showed a perfect agreement between word order and information structure (recall **Table 1**). The prosodic realization underscored this agreement: Since non-modal voice quality is associated with finality in Finnish (Iivonen, 1998; Myers and Hansen, 2007; Nakai et al., 2009), avoiding non-modal voice

quality earlier in the sentence but increasing its use on the focused constituent highlighted the fact that the focused constituent occupied the final position.

4. DISCUSSION

Using a language game elicitation task, we found several significant effects of information structure on prosody, as well as some adjustments of word order, in the speech of Finnish 4- to 5-year-olds. The data showed both differences from and similarities with Finnish adults' marking of information structure as described in the literature. The next sections discuss this for our findings regarding word order, prosody, and the relationship between them, before Section 4.3 compares the present findings to research on children acquiring other languages.

Overall, an interesting observation is that the results frequently showed differences between contrastive and narrow focus conditions, with only one or the other showing significant differences from broad focus for word order, f_0 range and voice quality. Different marking for contrastive and non-contrastive focus has been observed for several languages. For example, Northern Bizkaian Basque possesses prosodic means to mark contrastive focus that are not applied to mark non-contrastive narrow focus on the same word (Elordieta, 2008) and Catalan uses different pitch accents to mark contrastive and non-contrastive focus (Prieto, 2014). For adult Finnish, Arnhold (2016) observed the same prosodic strategies marking contrastive and non-contrastive focus, with larger effect sizes for contrastive focus. This contrasts with the present findings for child language. However, in terms of word order, the difference between contrastive and non-contrastive focus is firmly entrenched in adult Finnish, as detailed in the introduction.

4.1. Word Order

Our participants mostly retained the word order provided in the input and only showed a significant effect of information structure in one condition, changing input OVS to SVO word order when the experimenter's question induced contrastive focus on the object. Most noticeably, the participants did not employ word order to mark contrast. They almost exclusively produced SVO and OVS word orders, which leave the sentence-initial contrast position unfilled. Thus, participants never used SOV word order in contrastive subject focus (or any other information structure condition) and only one sentence with OSV order appeared in contrastive object focus (with another one appearing in narrow object focus). This might indicate that at age five, Finnish children have yet to acquire the correspondence between word order and information structure shown in **Table 1**. At any rate, participants in the present study exhibited no signs of competence in the usage of the sentence-initial contrast position. Instead, the only significant change in word order moved the focused constituent into sentence-final condition, which might suggest that participants have over-generalized the function of the sentence-final focus position to accommodate both contrastive and narrow focus or indeed that it is reserved for contrastive foci in their grammar. This change did however lead to improved congruence between word order and information

structure compared to the input, as discussed in Section 3.1. This suggests that children have acquired the relation between information structure and word order at least to a certain extent. Prosodic findings further strengthen this conclusion, as discussed below.

In fact, syntactic marking of contrast by using the contrast position is not obligatory in adult Finnish. Non-contrastive narrow focus can be marked by prosody alone in unmarked word order (Välilä-Blum 1988, p. 75; Vainio and Järvikivi 2007). Also, contrastively focused subjects and objects can appear in SVO sentences with appropriate prosody, although it is not clear whether prosodic and word order marking are completely equivalent, especially for contrastive subjects (Heinämäki, 1982; Vallduví and Vilkuna, 1998; Kaiser, 2000; Molnár and Järventaua, 2003; Kaiser, 2006; Karlsson, 2008). At least in written language, use of the contrast position seems relatively rare: A corpus study of about 10,000 sentences found that 49% of the sentences had SVX order (where X stands for any non-subject NP, including objects), while SXV and XSV order occurred for only 1 and 3%, respectively (Hakulinen and Karlsson, 1995, p. 311). Finally, a corpus study by Kaiser (2000) found three discourse functions of OSV order. While marking the object as contrastive was a sub-case of the most common function appearing for 55% of OSV sentences, 36% of OSV sentences placed salient, but non-contrastive given information in initial position, while about 9% fronted new information.

Altogether, the use of the contrast position is thus a complex and non-obligatory strategy, making the fact that children did not employ it in the present study less surprising. By contrast, children's prevalent use of SVO word order in the present study is in line with the fact that Finnish is undoubtedly an SVO language both in terms of frequency and in terms of the acceptability of SVO in most information structural conditions. Further, while our participants only significantly adjusted input word order in one condition, this adjustment did improve the congruence between information structure and word order.

4.2. Prosody

With respect to prosody, the results of the present study evidenced effects of information structure on all investigated phonetic measures affected in adult speech, i.e., duration, pausing, voice quality and, to a limited degree, f_0 range. These effects did not appear for all combinations of word order and information structural condition. However, where they did occur, they overwhelmingly reflected the prosodic focus marking strategies used by Finnish adults (e.g., Arnhold, 2016), suggesting that children possess some competence in the use of these strategies at age five. Where information structural effects appeared in our materials, f_0 range was larger for the noun in contrastive focus than for the other noun in the sentence. Durational effects lead to shorter durations for unfocused, i.e., previously mentioned or given nouns. Preceding and following pauses were more frequent in several narrow focus conditions. At the same time, final narrow focused constituents showed an increased use of non-modal voice quality following an infrequent use of non-modal voice on the other, unfocused constituents preceding them in the same sentence. These effects were in congruence with Finnish adults' prosodic focus marking with one

exception: Pauses preceding constituents in narrow focus have not been reported for adult Finnish, but are in line with other research on child language (see Section 4.3).

None of the prosodic parameters analyzed here displayed significant differences between all information structural conditions. However, all combinations of word order and information structure showed significant effects for at least one of the prosodic measures. For example, narrow subject focus in SVO sentences did not differ from broad focus in terms of f_0 range, but did differ in terms of voice quality, while contrastive object focus in OVS sentences showed an indication of f_0 range adjustments, but no effects on voice quality.

The distribution and nature of prosodic effects can be related to the interplay between word order and information structure by distinguishing three cases, discussed in detail below: matching word order and information structure, mismatch between word order and information structure, and unmarked word order with non-default information structure. Only duration displayed the same pattern across both word orders produced by the children: Durations of object nouns were significantly shorter in subject focus conditions—both contrastive and narrow—compared to broad focus. This can be interpreted as an effect of givenness, shortening the unfocused word, or possibly as an indirect way of making the focused subject constituent appear relatively longer. The other phonetic measures showed an influence of both word order and information structure. Interestingly, a division of labor appeared between the different measures; while some underscored a congruence between word order and information structure, some compensated for mismatches and some disambiguated unmarked cases.

4.2.1. Matching Word Order and Information Structure

The default interpretation of the SVO order predominantly used by the children is that the subject occupying the topic position is a topic and/or mentioned in the previous discourse (given) while the sentence-final object is in non-contrastive focus and/or new, see (a) in **Table 3**. For OVS, correspondingly, objects are usually given while subjects are focused, see (b). Thus, these word orders perfectly matched the information structure induced by the experimenter's questions in narrow object focus (NFO) and narrow subject focus condition (NFS), respectively. In precisely these two conditions, significant effects appeared for voice quality, with a delayed onset of sentence-final non-modal voice highlighting the finality of the focused constituent and thus the match between word order and information structure (see Section 3.5). Otherwise, sentences in these conditions did not differ much from broad focus realizations, in line with an information structure with final narrow focus being the default interpretation. Heinämäki (1982) in fact argues that both SVO and OVS are default word orders in Finnish, but it is noteworthy that SVO sentences additionally showed a significant increase in the use of pauses before final narrowly focused objects.

4.2.2. Unmarked Word Order and Information Structure

While a perfect match between word order and information structure appears in (a) in **Table 3**, unmarked SVO word order

TABLE 3 | Discourse configurational analysis for conditions with word order matching information structure.

	Contrast	Topic	Rest (focus final)
(a)		S	V[O] _{NF}
(b)		O	V[S] _{NF}

may appear felicitously with other information structures. In addition to broad focus, participants also produced SVO word order with contrastive and non-contrastive subject focus, as well as with the object in contrastive focus (conditions CFS, NFS, and CFO, respectively). Two of these conditions, NFS and CFO, were realized with significantly more pauses preceding or following the constituent in narrow focus, while the third showed an insignificant increase in the occurrence of pauses compared to the broad focus condition. In sentences with unmarked word order, the use of pauses thus frequently distinguished information structural conditions.

4.2.3. Mismatch between Word Order and Information Structure

One of the most striking findings of the present study was that f_0 contours of SVO sentences showed no significant differences between information structural conditions. Changes in f_0 range are a well-established part of prosodic focus marking in Finnish (e.g., Välimaa-Blum, 1993; Mixdorff et al., 2002; Suomi et al., 2003) and should be expected especially in the default word order, i.e., in the absence of syntactic marking of information structure (Arnhold and Féry, 2013). However, as mentioned in the introduction and discussed further in Section 4.3, children frequently underuse f_0 as a cue to information structure while overusing other prosodic cues like pauses compared to adults. In the present data, there was only limited evidence for significant adjustments of f_0 range in two conditions, OVS sentences with contrastive focus on the subject (CFS) and OVS sentences with contrastive focus on the object (CFO). As illustrated in (a) in **Table 4**, productions of OVS orders in CFO condition showed a mismatch between the information structure implied by the use of the more marked word order and the information structure induced by the experimenter’s question. Although participants frequently avoided this mismatching constellation by changing input word order, they were overall likely to retain the input word order. By expanding the f_0 range of the contrastively focused constituent and compressing the f_0 range of the other noun in the sentence, the prosody of children’s productions compensated for this mismatch when they did produce it.

The other constellation with an indication of significant f_0 range adjustments, CFS with OVS word order, appears in (b) in **Table 4**. In this condition, participants made the focused subject more prominent by expanding its f_0 range. Recall that children practically never placed a constituent in the sentence-initial contrast position and may not have had acquired its use yet (cf. Section 4.1). Therefore, it is somewhat unclear whether the constellation in (b) in **Table 4** constitutes a (slight) mismatch in their grammar—as the position for non-contrastive foci contains

TABLE 4 | Discourse configurational analysis for conditions with mismatch between word order and information structure.

	Contrast	Topic	Rest (focus final)
(a)		[O] _{CF}	VS
(b)		O	V[S] _{CF}
(c)		[O] _{NF}	VS

a contrastive focus—or whether it is instead a perfect match—as the focused constituent occupies the only focus position they have acquired at this point. Interestingly, while children may have adjusted f_0 range for this constellation, they did not do so for the mirror image SV[O]_{CF} (i.e., CFO with SVO order). This is in line with adult intuitions that while contrastive objects are acceptable in the sentence-final position, contrastive subjects are not (Kaiser, 2000).

Finally, in a further mismatching condition, OVS sentences with narrow object focus (NFO; see (c) in **Table 4**), another type of prosodic compensation appeared. Here, the frequency of pauses preceding the subject decreased significantly, prosodically grouping the subject with the preceding verb. Recall that by contrast, focused constituents were more frequently separated from the verb by a pause.

Overall, our data suggest that, like Finnish adults (Vainio and Järvikivi, 2007), 4- to 5-year-olds are able to apply prosodic compensation for a mismatch between word order and information structure. This suggests that although participants rarely changed the input word order, they demonstrated competence regarding the correspondence between word order and information structure. Further, the findings indicated that children are already able to bring together two areas of grammar, word order and prosody, setting them on the path to acquiring information structure marking as a complex system of interacting strategies.

4.3. Comparison with Children Acquiring Other Languages

The present data did not display the tendency toward a new-before-given word order frequently reported for child speech (Narasimhan and Dimroth, 2008). This can in part be explained by the task, since the input word order, which significantly influenced children’s production, was perfectly balanced between given-before-new and new-before-given. However, it is noteworthy that children’s only significant deviation from the input placed the narrowly focused constituent in final position, resulting in a given-before-new order. This finding was not only in accordance with adult Finnish grammar, as argued above, but also contradicts a universal tendency for children of the age range under investigation to prefer new-before-given (note that Narasimhan and Dimroth, 2008, review other studies contradicting this tendency).

Regarding task effects, the study by Sauer mann et al. (2011), investigating German-learning 4-year-olds, provides an interesting opportunity for comparison with the present one. Using the same design as the current study, they, too, found

an effect of input word order. As in the present experiment, significant differences between information structural conditions appeared only with the more marked OVS input, but in contrast to the present results, these effects did not mirror information structure effects reported in the literature on adult German (the adult control group consistently used unmarked SVO order in all conditions). Instead, children retained input OVS order most frequently in broad focus, i.e., in an information structural condition when it is highly inappropriate in adult German, which the authors explain with reference to memory constraints. Chen and Höhle (submitted) report the same finding for Dutch 4- to 5-year-olds. Thus, these results suggest that children acquiring West Germanic languages have limited competence regarding the connection between information structure and word order, but are more influenced by general cognitive factors. In comparison, the finding of a significant information structural effect on word order in the present study seems all the more remarkable.

With respect to prosody, the current results mirror existing studies in showing that 4- to 5-year-olds use prosody to mark information structure, but are not yet employing it in a completely adult-like manner (e.g., Wells et al., 2004; Chen, 2007, 2009, 2011). More specifically, our participants deviated from adult speech through incomplete application of prosodic focus marking strategies. The only exception concerns the use of pauses. Here, children displayed a focus marking strategy not reported for adult speakers of Finnish by showing an increased frequency of pauses not only after words in (non-contrastive or contrastive) narrow focus, but also preceding them. While children generally pause more frequently than adults (e.g., Redford, 2013), the present data showed significant effects of information structure, with the use of pauses increasing specifically in the context of narrow focus. Interestingly, this finding is reminiscent of Romøren and Chen's (2015) observation that Dutch 4- to 5-year-olds showed longer pause durations before narrow focus constituents, while the same strategy was less prevalent in adult speech. Their hypothesis that the extended use of pausing is related to the reduced use of other prosodic markers fits with the current data set as well: Just as Dutch children show incomplete mastery of accentuation, participants in the present study underused the prosodic focus markers employed by Finnish adults, especially f_0 range. In adult Finnish, overall significant effects appear for all the prosodic measures investigated here. By contrast, in the present study none of measures showed significant distinctions between all three basic information structural conditions, i.e., broad focus, narrow focus and givenness, for both constituents and in both word orders— although all measures did show some significant differences. Thus, children seemed to have acquired all prosodic focus markers, but did not use any of them as prevalently as adults.

Most notably, our participants did not employ f_0 range adjustments in unmarked SVO order, but reserved the use of f_0 for range adjustments in unmarked SVO order, but reserved the for contrastive focus in the more marked OVS word order. This finding is not only in contrast to research on adult Finnish, but also differs from results for Dutch and German children

completing the same task. In direct opposition to our Finnish findings, Dutch 4- to 5-year-olds employed f_0 adjustments only in unmarked word order (Chen and Höhle, submitted), whereas German children showed f_0 effects in both word orders (Sauermaun et al., 2011). Based on the relative prevalence of prosodic vs. word order effects of information structure in their data, Sauermaun et al. (2011) suggest that children may prefer prosodic strategies due to their comparative simplicity. The results of the current study indicate that this analysis is not suitable for Finnish, not only because a significant word order effect of information structure appeared, but also more crucially because prosodic effects were consistently modulated by the (lack of) congruence between information structure and word order.

In German-speaking children, significant prosodic effects only appeared for subject nouns (Sauermaun et al., 2011). A somewhat parallel finding of the present study was the restriction of durational effects to object nouns in both word orders. Otherwise, however, prosodic effects never emerged only for subjects or only for object nouns in principle. Instead, significant differences either appeared for subjects and objects equally (f_0 range in contrastive focus, see Section 3.2) or were restricted to either subject or object nouns in specific combinations of information structural conditions and word order. For example, an increased frequency of realizations with non-modal voice quality appeared for narrowly focused objects in SVO sentences, but not for the subjects, and for narrowly focused subjects, but not objects, in OVS sentences (see Section 3.5). As argued in the previous sections, the distribution of the presence or absence of prosodic effects suggests that Finnish 4- to 5-year-olds have already partially acquired the correspondence between word order and information structure in Finnish, as well as interactions between prosody and word order. This contrasts with children acquiring Germanic languages who showed clearer differences between grammatical functions, i.e., subjects differing from objects in the same way for SVO and OVS word order. This observation may be related to the principled differences between the languages. In German, word order variation is a possible tool for marking information structure, although prosodic marking is more prevalent. In Finnish, word order is largely determined by information structure (Vilkuna, 1989, 1995), and word order and prosody interact and can compensate each other in marking information structure (Vainio and Järviö, 2007; Arnhold and Féry, 2013). The comparison between our results and findings of previous studies on a Germanic language applying the same method suggests that this difference in grammatical organization is also reflected in acquisition: At age four to five, children acquiring a discourse configurational Uralic language showed a complex interplay between prosody and word order, suggesting a more central role of word order than for children acquiring a Germanic language. Thus, cross-linguistic differences in grammatical organization and functional weight are reflected in language development. Our results further underline the importance of representing a broad sample of languages when investigating the acquisition of prosody as well as other aspects of grammar.

5. CONCLUSIONS

Based on a semi-spontaneous production experiment, the present study indicated that Finnish 4- to 5-year-olds use both prosody and word order to mark information structure. Although the use of syntactic markers and prosodic markers was very limited compared to adult speech, significant effects appeared for nearly all prosodic parameters reported to mark focus in Finnish adults (the exception being intensity, which we did not investigate due to the nature of the data).

Crucially, prosodic marking interacted with word order: Differing from adult Finnish, significant adjustments of f_0 only appeared when compensating a mismatch between word order and information structure. By contrast, a distinct pattern in the use of non-modal voice quality, also found for adult speakers, highlighted perfect matches between word order and information structure. A significant increase in the use of pauses before and/or after constituents in narrow focus only appeared in sentences with unmarked SVO word order, while adjustments in word duration appeared for both word orders, but affected only object nouns.

The modulation of prosodic effects by word order reflects the central role of the relationships between information structure, word order and prosody in Finnish, contrasting with the development of children acquiring Germanic languages, who predominantly show purely prosodic information structure marking at the same age.

REFERENCES

- Arnhold, A. (2014). *Finnish Prosody: Studies in Intonation and Phrasing*. Ph.D thesis, Goethe University Frankfurt am Main.
- Arnhold, A. (2016). Complex prosodic focus marking in Finnish: expanding the data landscape. *J. Phonet.* 56, 85–109. doi: 10.1016/j.wocn.2016.02.002
- Arnhold, A., and Féry, C. (2013). The role of syntactic flexibility and prosody in marking given / new distinctions in Finnish. *Finno-Ugric Lang. Linguist.* 2, 2–32. Available online at: <http://full.btk.ppke.hu/index.php/FULL/article/view/19>
- Baayen, R. H. (2008). *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R*. Cambridge: Cambridge University Press.
- Baayen, R. H., Davidson, D., and Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *J. Mem. Lang.* 59, 390–412. doi: 10.1016/j.jml.2007.12.005
- Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01
- Beaver, D., Clark, B. Z., Flemming, E., Jaeger, T. F., and Wolters, M. (2007). When semantics meets phonetics: acoustical studies of second-occurrence focus. *Language* 83, 245–276. doi: 10.1353/lan.2007.0053
- Behrens, H., and Gut, U. (2005). The relationship between prosodic and syntactic organization in early multiword speech. *J. Child Lang.* 32, 1–34. doi: 10.1017/S0305000904006592
- Boersma, P., and Weenink, D. (1992–2014). *Praat. Doing Phonetics by Computer*. Available online at: www.praat.org
- Braun, B. (2006). Phonetics and phonology of thematic contrast in German. *Lang. Speech* 49, 451–493. doi: 10.1177/00238309060490040201
- Chen, A. (2007). “Intonational realisation of topic and focus by Dutch-acquiring 4- to 5-year olds,” in *Proceedings of the 16th International Congress of Phonetic Sciences*, eds J. Trouvain and W. J. Barry (Saarbrücken: Universität des Saarlandes), 1553–1556.

AUTHOR CONTRIBUTIONS

AC and JJ constructed the experiment and created materials. JJ oversaw running the experiment. AA analyzed the data. AA, AC, and JJ wrote the paper.

FUNDING

AC was supported by a VIDI grant from the Netherlands Organization for Scientific Research (NWO-276-89-001) during the writing of the manuscript.

ACKNOWLEDGMENTS

We would like to thank Martti Vainio for hosting the experiment at the University of Helsinki, as well as Heini Kallio for help in stimuli creation and data collection. We also acknowledge Barbara Höhle’s input in the preparation stage of this study. We are further grateful to participating children and their parents, as well as to the staff at the daycare facilities Leipuri, Myllytupa, Siilitie, and Tahvonlahti in Helsinki and at Nuottaniemi and Kalamiehen päiväkoti in Espoo.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpsyg.2016.01886/full#supplementary-material>

- Chen, A. (2009). “The phonetics of sentence-initial topic and focus in adult and child Dutch,” in *Phonetics and Phonology: Interactions and Interrelations*, eds M. C. Vigario, S. Frota, and M. João Freitas (Amsterdam: John Benjamins), 91–106. doi: 10.1075/cilt.306.05che
- Chen, A. (2010). Is there really an asymmetry in the acquisition of the focus-to-accentuation mapping? *Lingua* 120, 1926–1939. doi: 10.1016/j.lingua.2010.02.012
- Chen, A. (2011). Tuning information packaging: intonational realization of topic and focus in child Dutch. *J. Child Lang.* 38, 1055–1083. doi: 10.1017/S0305000910000541
- Chen, A., and Fikkert, P. (2007). “Intonation of early two-word utterances in Dutch,” in *Proceedings of the 16th International Congress of Phonetic Sciences*, eds J. Trouvain and W. J. Barry (Saarbrücken: Universität des Saarlandes), 315–320.
- de Ruiter, L. E. (2014). How German children use intonation to signal information structure in narrative discourse. *J. Child Lang.* 41, 1015–1061. doi: 10.1017/S0305000913000251
- Dik, S. C., Hoffman, M. E., de Long, J. R., Sie Ing, D., Stroomer, H., and de Vries, L. (1981). “On the typology of focus phenomena,” in *Perspectives on Functional Grammar*, eds T. Hoekstra, H. van der Hulst, and M. Moortgat (Dordrecht: Foris Publications), 41–74.
- Elordieta, G. (2008). “Constraints on intonational prominence of focalized constituents,” in *Topic and Focus: Cross-Linguistic Perspectives on Meaning and Intonation*, *Studies in Linguistics and Philosophy* 82, eds C. Lee, M. Gordon, and D. Büring (Dordrecht: Springer), 1–22.
- Epstein, M. A. (2002). *Voice Quality and Prosody in English*. Ph.D thesis, University of California, Los Angeles.
- Fernald, A., Thorpe, K., and Marchman, V. A. (2010). Blue car, red car: developing efficiency in online interpretation of adjective noun phrases. *Cogn. Psychol.* 60, 190–217. doi: 10.1016/j.cogpsych.2009.12.002

- Féry, C. (1993). *German Intonational Patterns*. Tübingen: Linguistische Arbeiten 285. Max Niemeyer Verlag.
- Féry, C., and Ishihara, S. (2009). The phonology of second occurrence focus. *J. Linguist.* 45, 285–313. doi: 10.1017/S002226709005702
- Grünloh, T., Lieven, E., and Tomasello, M. (2015). Young children's intonational marking of new, given and contrastive referents. *Lang. Learn. Dev.* 11, 95–127. doi: 10.1080/15475441.2014.889530
- Gundel, J. K., Hedberg, N., and Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. *Language* 69, 274–307. doi: 10.2307/416535
- Gussenhoven, C. (2008). "Types of focus in English," in *Topic and Focus. Cross-Linguistic Perspectives on Meaning and Intonation, Studies in Linguistics and Philosophy* 82, eds C. Lee, M. Gordon, and D. Büring (Dordrecht: Springer), 83–100.
- Hakulinen, A., and Karlsson, F. (1995). *Nyky-suomen lauseoppi (Contemporary Finnish Grammar)*. Gummerus: Jyväskylä.
- Hedberg, N., and Sosa, J. M. (2008). "The prosody of topic and focus in spontaneous English dialogue," in *Topic and Focus: Cross-Linguistic Perspectives on Meaning and Intonation, Studies in Linguistics and Philosophy* 82, eds C. Lee, M. Gordon, and D. Büring (Dordrecht: Springer), 101–120.
- Heinäsmäki, O. (1982). "Problems of basic word order," in *Reports on text Linguistics. Approaches to Word Order*, eds N. E. Enkvist, and V. Kohonen (Turku: Åbo Akademi), 95–106.
- Holmberg, A., and Nikanne, U. (2002). "Expletives, subjects, and topics in Finnish" in *Subjects, Expletives, and the EPP*, ed P. Svenonius (Oxford: Oxford University Press), 71–106.
- Holmberg, A., Nikanne, U., Oraviita, I., Reime, H., and Trosterud, T. (1993). "The structure of INFL and the finite clause in Finnish," in *Case and Other Functional Categories in Finnish Syntax, Studies in Generative Grammar* 39, eds A. Holmberg and U. Nikanne (Berlin; New York: Mouton de Gruyter), 177–206.
- Hornby, P. A., and Hass, W. A. (1970). Use of contrastive stress by preschool children. *J. Speech Hear. Res.* 13, 359–399. doi: 10.1044/jshr.1302.395
- Iivonen, A. (1974). Äännekeston riippuvuus sanan pituudesta irrallaan äänneytyssä sanoissa (On the dependence of sound duration on word length in isolated words). *Virittäjä* 78, 134–151.
- Iivonen, A. (1998). "Intonation in Finnish," in *Intonation Systems: A Survey of Twenty Languages*, eds D. Hirst, and A. Di Cristo (Cambridge: Cambridge University Press), 311–327.
- Ito, K., Bibyk, S. A., Wagner, L., and Speer, S. R. (2014). Interpretation of contrastive pitch accent in six- to eleven-year-old English-speaking children (and adults). *J. Child Lang.* 41, 84–110. doi: 10.1017/S0305000912000554
- Ito, K., and Speer, S. R. (2008). Anticipatory effects of intonation: eye movements during instructed visual search. *J. Mem. Lang.* 58, 541–573. doi: 10.1016/j.jml.2007.06.013
- Jaeger, T. F. (2008). Categorical data analysis: away from ANOVAs (transformation or not) and towards logit mixed model. *J. Mem. Lang.* 59, 434–446. doi: 10.1016/j.jml.2007.11.007
- Järvikivi, J., Pykkönen-Klauck, P., Schimke, S., Colonna, S., and Hemforth, B. (2014). Information structure cues for 4-year-olds and adults: tracking eye movements to visually presented anaphoric referents. *Lang. Cogn. Process.* 29, 877–892. doi: 10.1080/01690965.2013.804941
- Jokinen, K. (2005). "A grammar for Finnish discourse patterns," in *Inquiries into Words, Constraints and Contexts. Festschrift for Kimmo Koskeniemi on his 60th Birthday, CSLI Studies in Computational Linguistics ONLINE*, eds A. Arppe, L. Carlson, K. Lindén, J. Piitulainen, M. Suominen, M. Vainio et al. (Stanford, CA: CSLI Publications), 227–240.
- Kaiser, E. (2000). "The discourse functions and syntax of OSV word order in Finnish," in *The Proceedings from the Main Session of the Chicago Linguistic Society's Thirty-Sixth Meeting, Vol. 36.1*, eds A. Okrent, and J. P. Boyle (Chicago, IL: Chicago Linguistic Society), 179–194.
- Kaiser, E. (2006). Negation and the left periphery in Finnish. *Lingua* 116, 314–350. doi: 10.1016/j.lingua.2004.08.008
- Karlsson, F. (2008). *Finnish. An Essential Grammar, Routledge Essential Grammars*. 2nd Edn., London; New York, NY: Routledge.
- Katz, J., and Selkirk, E. (2011). Contrastive focus vs. discourse-new: evidence from phonetic prominence in English. *Language* 87, 771–816. doi: 10.1353/lan.2011.0076
- Krifka, M. (2008). Basic notions of information structure. *Acta Linguist. Hung.* 55, 243–276. doi: 10.1556/ALing.55.2008.3-4.2
- Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2015). *lmerTest: Tests in Linear Mixed Effects Models*. R Package Version 2.0-29. Available online at: <http://CRAN.R-project.org/package=lmerTest>
- Ladd, D. R. (1980). *The Structure of Intonational Meaning: Evidence from English*. Bloomington: Indiana University Press.
- Lambrecht, K. (1994). *Information Structure and Sentence Form. Topic, Focus and the Mental Representations of Discourse Referents*. Cambridge: Cambridge University Press.
- MacWhinney, B., and Bates, E. (1978). Sentential devices for conveying givenness and newness: a cross-cultural developmental study. *J. Verb. Learn. Verb. Behav.* 17, 539–558. doi: 10.1016/S0022-5371(78)90326-2
- Mixdorff, H., Vainio, M., Werner, S., and Järvikivi, J. (2002). "The manifestation of linguistic information in prosodic features of Finnish," in *Proceedings of Speech Prosody 2002*, eds B. Bel and I. Marlien (Aix-en-Provence), 511–514.
- Molnár, V., and Järventausta, M. (2003). "Discourse configurability in Finnish and Hungarian," in *Structures of Focus and Grammatical Relations, Linguistische Arbeiten* 477, eds J. Hetland, and V. Molnár (Tübingen: Max Niemeyer Verlag), 111–148.
- Müller, A., Höhle, B., Schmitz, M., and Weissenborn, J. (2006). "Focus-to-stress alignment in 4- to 5-year-old German-learning children," in *Language Acquisition and Development. Proceedings of GALA 2005*, eds A. Belletti, E. Bennati, C. Chesì, E. D. Domenico, and I. Ferrari (Cambridge: Cambridge Scholars Press), 393–407.
- Myers, S., and Hansen, B. B. (2007). The origin of vowel length neutralization in final position: evidence from Finnish speakers. *Nat. Lang. Linguist. Theory* 25, 157–193. doi: 10.1007/s11049-006-0001-7
- Nakai, S., Kunnari, S., Turk, A., Suomi, K., and Ylitalo, R. (2009). Utterance-final lengthening and quantity in Northern Finnish. *J. Phonet.* 37, 29–45. doi: 10.1016/j.wocn.2008.08.002
- Narasimhan, B., and Dimroth, C. (2008). Word order and information status in child language. *Cognition* 107, 317–329. doi: 10.1016/j.cognition.2007.07.010
- Nevis, J. A. (1986). *Finnish Particle Clitics and General Clitic Theory*. Columbus, OH: The Ohio State University, Department of Linguistics.
- Ní Chasaide, A., Yanushevskaya, I., and Gobl, C. (2011). "Voice source dynamics in intonation," in *Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS 2011)*, eds W.-S. Lee, and E. Zee (Hong Kong), 1470–1473. Available online at: <https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2011/OnlineProceedings/RegularSession/Ni%20Chasaide/Ni%20Chasaide.pdf>
- Prieto, P. (2014). "The intonational phonology of Catalan," in *Prosodic Typology II. The Phonology of Intonation and Phrasing*, ed S.-A. Jun (Oxford: Oxford University Press), 43–80.
- R Core Team (2015). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Redford, M. A. (2013). A comparative analysis of pausing in child and adult storytelling. *Appl. Psycholinguist.* 34, 569–589. doi: 10.1017/S0142716411000877
- Reinhart, T. (1981). Pragmatics and linguistics: an analysis of sentence topics. *Philosophica* 27, 53–94.
- Romoren, A. S. H., and Chen, A. (2015). Quiet is the new loud: pausing and focus in child and adult Dutch. *Lang. Speech* 58, 8–23. doi: 10.1177/0023830914563589
- Rooth, M. (1985). *Association with Focus*. Ph.D thesis, University of Massachusetts, Amherst.
- Rooth, M. (1992). A theory of focus interpretation. *Nat. Lang. Semant.* 1, 75–116. doi: 10.1007/BF02342617
- Sauerermann, A., Höhle, B., Chen, A., and Järvikivi, J. (2011). "Intonational marking of focus in different word orders in German children," in *Proceedings of the 28th West Coast Conference on Formal Linguistics*, eds M. B. Washburn, K. McKinney-Bock, E. Varis, A. Sawyer, and B. Tomaszewicz, (McMerville, MA: Cascadilla Proceedings Project), 313–322.
- Schwarzschild, R. (1999). Givenness, AvoidF and other constraints on the placement of accent. *Nat. Lang. Semant.* 7, 141–177. doi: 10.1023/A:1008370902407
- Suomi, K. (2007). On the tonal and temporal domains of accent in Finnish. *J. Phonet.* 35, 40–55. doi: 10.1016/j.wocn.2005.12.001

- Suomi, K., Toivanen, J., and Ylitalo, R. (2003). Durational and tonal correlates of accent in Finnish. *J. Phonet.* 31, 113–138. doi: 10.1016/S0095-4470(02)00074-8
- Suomi, K., Toivanen, J., and Ylitalo, R. (2008). *Finnish Sound Structure. Phonetics, Phonology, Phonotactics and Prosody*. Oulu: University of Oulu, Studia Humaniora Ouluensia 9.
- Suomi, K., and Ylitalo, R. (2003). Syllable weight and segmental durations in Finnish. *Phonum Rep. Phonet.* 9, 37–40.
- Suomi, K., and Ylitalo, R. (2004). On durational correlates of word stress in Finnish. *J. Phonet.* 32, 35–63. doi: 10.1016/S0095-4470(03)00005-6
- Vainio, M., Airas, M., Järvikivi, J., and Alku, P. (2010). “Laryngeal voice quality in the expression of focus,” in *Proceedings of Interspeech, 11th Annual Conference of the International Speech Communication Association* (Makuhari, Chiba).
- Vainio, M., and Järvikivi, J. (2006). Tonal features, intensity, and word order in the perception of prominence. *J. Phonet.* 34, 319–342. doi: 10.1016/j.wocn.2005.06.004
- Vainio, M., and Järvikivi, J. (2007). Focus in production: tonal shape, intensity and word order. *J. Acoust. Soc. Am.* 121, EL55–EL61. doi: 10.1121/1.2424264
- Välímáa-Blum, R. (1988). *Finnish Existential Clauses - Their Syntax, Pragmatics and Intonation*. Ph.D thesis, The Ohio State University.
- Välímáa-Blum, R. (1993). A pitch accent analysis of intonation in Finnish. *Ural-Altäische Jahrbücher* 12, 82–94.
- Vallduví, E., and Vilkkuna, M. (1998). On rheme and kontrast. *Syntax Semant.* 29, 79–108.
- Vilkkuna, M. (1989). *Free Word Order in Finnish: Its Syntax and Discourse Functions*. Helsinki: Suomalaisen Kirjallisuuden Seuran Toimituksia.
- Vilkkuna, M. (1995). “Discourse configurationality in Finnish,” in *Discourse Configurational Languages*, ed K. É. Kiss (New York, NY; Oxford: Oxford University Press), 244–268.
- Wells, B., Peppé, S., and Goulandris, N. (2004). Intonation development from five to thirteen. *J. Child Lang.* 31, 749–778. doi: 10.1017/S030500090400652X
- Wieman, L. (1976). Stress patterns of early child language. *J. Child Lang.* 3, 283–286. doi: 10.1017/S0305000900001501

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Arnhold, Chen and Järvikivi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.