

## **to appear in Cognition**

### **“The tiger is hitting! The duck too!” 3-year-olds can use prosodic information to constrain their interpretation of ellipsis**

Letícia Kolberg <sup>a,\*</sup>, Alex de Carvalho <sup>b</sup>, Mireille Babineau <sup>c,e</sup>, Naomi Havron <sup>c,d</sup>,  
Anne-Caroline Fiévet <sup>c,e</sup>, Bernadete Abaurre <sup>a</sup> and Anne Christophe <sup>c,e</sup>

<sup>a</sup> *Departamento de Linguística, Universidade Estadual de Campinas, Campinas, Brasil;*

<sup>b</sup> *Laboratoire de Psychologie du Développement et de l'Éducation de l'Enfant (LaPsyDÉ), Department of Psychology, Université de Paris, CNRS, F-75005 Paris, France*

<sup>c</sup> *Laboratoire de Sciences Cognitives et Psycholinguistique, DEC-ENS / EHESS / CNRS, Ecole normale supérieure – PSL University, Paris, France*

<sup>d</sup> *University of Haifa*

<sup>e</sup> *Maternité Port-Royal, AP-HP, Université Paris Descartes, France*

\* *Corresponding author: correspondence concerning this article should be addressed to Letícia Kolberg, email: [leticia.kolberg@gmail.com](mailto:leticia.kolberg@gmail.com). Address: École Normale Supérieure, 29 Rue d'Ulm, 75005, Paris, France*

Word count: 14568 words

**Declarations of interest:** none.

## **Highlights**

- 3-year-olds use prosodic information to constrain their interpretation of ellipsis.
- 28-month-olds show only a trend in the expected direction.
- Children's use of prosody for parsing seems affected by the type of structure tested.
- Past failures might be related to methodological rather than crosslinguistic factors.

**“The tiger is hitting! The duck too!” 3-year-olds can use prosodic information to constrain their interpretation of ellipsis**

This work aims to investigate French children's ability to use phrasal boundaries for disambiguation of a type of ambiguity not yet studied, namely stripping sentences versus simple transitive sentences. We used stripping sentences such as "*Le tigre tape!* [*Le canard aussi!*]" ("[The tiger is hitting]! [The duck too]!", in which both the tiger and the duck are hitting), which, without the prosodic information, would be ambiguous with a transitive sentence such as "*Le tigre* [*tape le canard aussi!*]" ("[The tiger] [is hitting the duck too]!", in which the tiger is hitting the duck). We presented 3-to-4-year-olds and 28-month-olds with one of the two types of sentence above, while they watched two videos side-by-side on a screen: one depicting the transitive interpretation of the sentences, and another depicting the stripping interpretation. The stripping interpretation video showed the two characters as agents of the named action (e.g. a duck and a tiger hitting a bunny), and the transitive interpretation video showed only the first character as an agent, and the second character as a patient of the action (e.g. the tiger hitting the duck and the bunny). The results showed that 3-to-4-year-olds use prosodic information to correctly distinguish stripping sentences from transitive sentences, as they looked significantly more at the appropriate video, while 28-month-olds show only a trend in the same direction. While recent studies demonstrated that from 18 months of age, infants are able to use phrasal prosody to guide the syntactic analysis of ambiguous sentences, our results show that only 3-to-4-year-olds were able to reliably use phrasal prosody to constrain the parsing of stripping sentences. We discuss several factors that can explain this delay, such as differences in the frequency of these structures in child-directed speech, as well as in the complexity of the sentences and of the experimental task. Our findings add to the growing body of evidence on the role of prosody in constraining parsing in young children.

Keywords: Phrasal Prosody; Ellipsis; Syntactic Acquisition; Syntactic Ambiguity Resolution; Preferential Looking; Sentence Processing

## **1 Introduction**

One of the most interesting problems in language acquisition is how children discover the syntactic structure of their native language. As Pinker (1984, p.124) wrote, the abstract categories that compose the hierarchical structure of a language are "colorless, odorless and tasteless", that is, there is no obvious information in the speech stream that links words to them. How, then, do children access these categories and the grammatical relations between words by simply listening to the sentences addressed to them? In order to investigate this, linguists and psycholinguists look for cues in children's speech input that could possibly flag these relations and guide children's lexical and syntactic acquisition.

Several studies propose that the suprasegmental cues of speech (e.g., pitch, duration, energy, etc.) constitute a very important set of cues for the beginning of lexical and syntactic acquisition, as they indicate, among other things, prosodic phrase boundaries (Christophe, Dautriche, de Carvalho, & Brusini, 2016; Christophe, Millotte, Bernal, & Lidz, 2008; de Carvalho, Dautriche, Millotte, & Christophe, 2018; Hawthorne & Gerken, 2014; Hawthorne, Mazuka, & Gerken, 2015; Massicotte-Laforge & Shi, 2015, 2020; Morgan & Demuth, 1996). Young children have been shown to know a great deal about their native language's prosodic structure, and to use this knowledge for many different aspects of sentence parsing: for sentence disambiguation (e.g., Dautriche et al., 2014; de Carvalho et al., 2016a, 2016b, 2017; Snedeker & Yuan, 2008; Zhou, Su, Crain, Gao, & Zhan, 2012), for word and sentence segmentation (e.g., Bailey & Plunkett, 1998; Gervain & Werker, 2013; Graf Estes & Bowen, 2013; Hirsh-Pasek et al., 1987; Johnson, 2008; Johnson & Seidl, 2008; Jusczyk et al., 1992; Mandel, Jusczyk, & Nelson, 1994; Ramachers, Brouwer, & Fikkert, 2017; Shukla, White & Aslin, 2011; Soderstrom et al.,

2003); to decide whether a utterance is declarative or interrogative (e.g., Zhou, Crain, & Zhan, 2012); and for grammatical categorization of novel words (e.g., de Carvalho et al., 2019; Hawthorne & Gerken, 2014; Massicotte-Laforge & Shi, 2015).

Although there is no one-to-one correspondence between prosodic and syntactic boundaries, prosodic phrasing aligns with syntactic phrasing (Nespor & Vogel, 1986), such that whenever a prosodic phrase boundary occurs, it signals a syntactic phrase boundary (while the reverse is not true). According to some of the above-mentioned studies, if children know this relationship, they could use phrasal prosody to delimit syntactic phrases and to identify possible syntactic constituents. This would, in turn, help them to bootstrap syntactic acquisition.

Prosodic cues seem to be readily available to children even before they are born (Abboub, Nazzi & Gervain, 2016). Studies have shown that newborns can discriminate languages on the basis of their prosodic properties (e.g., Mehler et al., 1988). A few months later, around the age of 6 to 10 months, infants can rely on prosodic cues to constrain lexical segmentation of the speech stream (Gout, Christophe, & Morgan, 2004; Johnson, 2008; Johnson & Jusczyk, 2001; Jusczyk, Cutler, & Redanz, 1993; Morgan & Saffran, 1995; Shukla, et al., 2011). Infants of this age are also sensitive to the coalition of cues marking prosodic boundaries between groups of words (Hirsh-Pasek et al., 1987; Jusczyk et al., 1992; Soderstrom, Seidl, Kemler Nelson, & Jusczyk, 2003; van Ommen et al., 2020; Wellmann, Holzgrefe, Truckenbrodt, Wartenburger, & Höhle, 2012). The work of Jusczyk et al., (1992) and Hirsh-Pasek et al., (1987), for instance, shows that by 7 months of age, infants can identify prosodic phrasal boundaries, and pay more attention to recorded stories in which pauses are inserted correctly at those boundaries, as opposed to within the prosodic phrases. Infants have also been shown to understand, from 6 months onwards, that words cannot span prosodic boundaries, i.e., that there cannot be a

word which starts in a prosodic phrase but ends in a different phrase (e.g., Gout et al., 2004; Shukla et al., 2011; Soderstrom et al., 2003).

Studies such as the ones mentioned above corroborate the *Prosodic Bootstrapping* hypothesis (e.g. Christophe et al., 2008, 2016; Gutman, Dautriche, Crabbé, & Christophe, 2015; Höhle, Weissenborn, Schmitz, & Ischebeck, 2001; Morgan & Demuth, 1996), which states that children can use prosodic boundary information to construct a rudimentary syntactic structure that could help bootstrap lexical and syntactic acquisition (e.g. Christophe et al., 2016, 2008, 2003; de Carvalho, 2017; de Carvalho et al., 2018; Hirsh-Pasek et al., 1987; Morgan & Demuth, 1996). Because the prosodic structure of an utterance depends in part on its syntactic structure, these studies suggest that infants should be able to use their extensive experience with prosody to bootstrap their way into syntax. In other words, according to these studies, infants are precociously aware of the correlation between prosodic and syntactic boundaries and can use this information to make hypotheses about the boundaries delimitating syntactic constituents.

Direct empirical evidence in favor of the prosodic bootstrapping hypothesis comes from work on syntactic disambiguation through prosodic boundary cues, which investigates children's ability to differentiate sentences which are linearly similar (i.e., contain the same string of words) but have different syntactic relations between constituents, which is reflected in differences in the prosodic phrasing (e.g., de Carvalho, Lidz, Tieu, Bleam, & Christophe, 2016; de Carvalho, Dautriche, & Christophe, 2016; de Carvalho, Dautriche, Lin, & Christophe, 2017; Dautriche et al., 2014; Hirose & Mazuka, 2017; Snedeker & Yuan, 2008; Lehiste, Olive & Streeter, 1976). For instance, de Carvalho et al. (2017) found that French-learning 20-month-olds can exploit prosodic phrasing information to constrain their syntactic analysis and correctly interpret a homophonous word as either a noun or a verb. The authors exposed children to globally

ambiguous sentences in French such as “*Tu vois le bébé /sʊbi/*”, where the homophone /sʊbi/ could be interpreted as the verb “smile” or the noun “mouse” depending on the prosodic phrasing of the sentences. When there is a prosodic phrase boundary between *le bébé* and the homophone, the homophone has a verb interpretation (“[*Tu vois*]? [*Le bébé*] [*/sʊbi/*]!” – “[Do you see]? [The baby] [smiles]!”), with the brackets representing prosodic phrasing), but when all the words in the sentence are pronounced inside a single prosodic unit, it has a noun interpretation (“[*Tu vois le bébé /sʊbi/*]?” – “[Do you see the baby mouse]?”). When presented with two images depicting the two possible interpretations of the string of words (e.g., a baby smiling (verb interpretation) and a little mouse (noun interpretation)) side-by-side on a screen, children who listened to the sentences with the noun prosody looked longer towards the noun image, while children who listened to the sentences with the verb prosody looked longer towards the verb image. These results suggest that toddlers can successfully use prosodic information to constrain their parsing, and thus correctly interpret globally ambiguous sentences, associating the correct meaning to the ambiguous words.

Dautriche et al. (2014) showed that 28-month-old French-learning children can rely on prosodic boundary cues to solve an ambiguity involving sentences with right dislocation such as “[*Il<sub>i</sub> a mangé<sub>i</sub>*], [*le canard<sub>i</sub>*]” – “[He<sub>i</sub> ate], [the duck<sub>i</sub>]”, in which the duck is the agent of the eating action, and thus is co-referring with the pronoun “he” in the sentence, as opposed to transitive sentences without dislocation such as “[*Il<sub>i</sub> a mangé le canard<sub>k</sub>*]” – “[He<sub>i</sub> ate the duck<sub>k</sub>]”, in which the duck is the patient of the action (i.e., someone is eating the duck). In the first sentence, the boundary between “ate” and “the duck” indicates that they do not form a Verb Phrase unit, i.e., “the duck” is not the object of “ate”, whereas it clearly is the object in the second sentence, where all the words appear in a single prosodic unit. In a preferential looking task, children were presented with two

simultaneous videos: one with the right-dislocated interpretation of the sentences (e.g., a duck puppet eating bread, for the example above) and another one with the transitive interpretation (e.g., a tiger puppet eating the duck, for the example above). Children exposed to sentences with right-dislocated prosody looked longer towards the video where the mentioned character (e.g., the duck) was the agent, compared to children exposed to transitive sentences. These results show that 28-month-olds can use prosodic information to constrain their interpretation of the argument structure of sentences.

Recent findings such as the ones from Dautriche et al. (2014) and de Carvalho et al. (2017) show that by 20 months of age, French-learning children can already use prosodic boundary information to identify syntactic boundaries and correctly parse sentences. However, previous studies in languages other than French have shown mixed results. Some of the studies failed to observe an ability to use prosody for syntactic disambiguation in children between 3 and 6 years of age (e.g., Choi & Mazuka, 2003, with a pointing task in Korean; and Snedeker & Trueswell, 2001, with a visual world paradigm and act-out task in English), while others succeeded in showing this ability in children from 4 years of age (Snedeker & Yuan, 2008, with a visual world paradigm and act-out task, and de Carvalho et al., 2016b, with an oral completion task in English). Choi & Mazuka (2003), for instance, presented 5-6-year-old Korean-learning children with sentences such as "[kirin] [kwaja mægəyo]" ("[giraffe] [cookie eat]" - "[(A) giraffe] [eats cookies]"), which, without the prosodic boundary information, is not distinguishable from the sentence Ø "[kirin kwaja mægəyo]" ("[giraffe cookie eat]" - "[(somebody) eats giraffe-shaped cookies]"). Children were presented with one of the sentences above and were asked to choose the image corresponding to what they heard (i.e., an image of a giraffe eating cookies vs. another one of a boy eating giraffe-shaped cookies). Korean children

were not able to use the prosodic boundary information to correctly interpret the sentences, as they performed at chance level.

Work by Snedeker & Trueswell (2001) also failed to show the use of prosodic cues for sentence disambiguation with 5-years-old English-learning children. In this study, children and their mothers were invited to play a game; they stayed in opposite sides of an opaque screen, and an experimenter would lay some toys in front of the child, while another experimenter performed an action with a similar set of toys in front of the mother. The mother was then instructed to tell the child to perform the action she had just witnessed, using a sentence written on a card which was given to her. Children were either exposed to unambiguous sentences, such as “tap the frog who is carrying a flower”, or to ambiguous sentences, such as “tap the frog with the flower”, which had more than one possible interpretation given the available set of toys (e.g., a frog, a flower, and a frog carrying a small flower). Although the mothers’ prosodic phrasing correctly cued the right interpretation of sentences in the ambiguous contexts<sup>1</sup>, children's interpretation of these sentences was at chance, whereas for the unambiguous sentences children nearly always performed the correct actions. In a follow-up study from Snedeker & Yuan (2008), which controlled for lexical and perseveration biases in the task and improved the naturalness of the prosodic phrasing by slightly changing the sentence structure (“can you touch the frog with the flower?”), children did succeed in interpreting these sentences, although only when presented with only one kind of prosody (“modifier interpretation” vs. “instrument interpretation”).

These mixed results raise the question of whether the usefulness of prosodic information to constrain parsing in young children might vary depending on factors such

---

<sup>1</sup> A similar experiment was conducted with adults, who succeeded in disambiguating the sentences and performing the correct actions (Snedeker & Trueswell, 2003).

as crosslinguistic differences, the type of syntactic structure tested, and/or the experimental paradigms used. Regarding the crosslinguistic factor, some differences between English and French prosody, for instance, could explain why English-learning children might have more trouble using prosody for sentence parsing than their French-learning peers. As noticed by de Carvalho et al. (2016b), while in English prosodic cues are consistently used to mark word stress and emphasize information both at the word and sentence level (i.e., focus, as in “JOHN ate the apple”), French does not have lexical stress and prosodic cues to mark focus in French are rather optional and rarely used by speakers (see e.g., Féry, 2001). Instead, French speakers often use other syntactic strategies to mark focus, such as fronting (as in e.g., *Jean, il a mangé la pomme!* – “John, he ate the apple!”) or clefting (e.g., *C’est Jean qui a mangé la pomme*, “It is John who ate the apple”). Prosodic cues could therefore be more ambiguous for young children in English than in French, where they are used mainly to cue phrasal prosodic structure. In this case, learning how English uses prosodic cues for different aspects of language (i.e. to signal prominence vs to signal the boundaries between prosodic units) could take extra developmental time compared to French. If this were the case, children learning English could take longer to identify which suprasegmental cues indicate phrase boundaries in their language, whereas this learning process could be easier and/or faster in other languages such as French.

As observed in our last two examples, the differences observed between studies might also be due to the different methodologies adopted. Snedeker & Yuan's follow-up study, for instance, suggests that children's failure in Snedeker & Trueswell's (2001) study was most likely due to methodological issues, such as the choice of lexical items that favor one interpretation of the sentences over the other, rather than the language studied.

Finally, regarding the possible influence of different syntactic structures, in de Carvalho et al.'s studies (2017, 2016a) in French, and de Carvalho et al.'s study in English (2016b), the default prosodic structure directly reflected the syntactic structure of sentences: when the prosodic boundary falls before the critical word, the latter can only be interpreted as a verb; when the prosodic boundary falls after it, the word is interpreted as a noun. In contrast, in Snedeker & Trueswell's (2001) study, the default prosodic structure of the test sentences was the same for the two possible interpretations (e.g., "[touch] [the frog] [with the feather]", with three prosodic units), but adult speakers who are aware of the ambiguity can intentionally disambiguate the sentences by exaggerating one of the two prosodic breaks. Hence, children's failure in this task might be at least partly due to the fact that the disambiguating prosodic breaks are not part of the normal prosodic structure of these sentences.

Another evidence that syntactic and prosodic structure can affect sentence disambiguation comes from Lehiste et al.'s (1976) study with English-speaking adults. In this study, subjects listened to ambiguous sentences whose meaning could be represented by different syntactic bracketing, such as "the hostess greeted the girl with a smile", which can mean either that the hostess was smiling (the hostess [[greeted the girl]] [with a smile]) or that the girl was smiling (the hostess [greeted] [[the girl] [with a smile]]), or to sentences whose meaning was not distinguished through syntactic bracketing, such as "[[visiting] [relatives]] can be a nuisance", which can mean either "relatives who pay a visit can be a nuisance" or "visiting one's relatives can be a nuisance". The authors added artificial duration cues around the syntactic phrase boundaries of sentences, to favor one or another interpretation of each sentence. They found that subjects only correctly disambiguated the sentences that presented different surface phrase structure, but not sentences that presented the same surface structure. This shows that, even for mature

listeners, disambiguation through prosodic boundary information might depend on the type of syntactic structure or syntactic-prosodic relations tested.

In order to further clarify the role played by phrasal prosody in constraining syntactic analysis in children, and to better understand the reasons that have led to conflicting evidence in the literature, we chose to study another type of syntactic ambiguity in French, a language in which many studies have shown that young children successfully use prosody for parsing. If we can understand which type of syntactic information supports or hinders children's ability to disambiguate sentences through prosody, we may understand better the effectiveness of prosodic information to constrain parsing and to facilitate syntactic acquisition. With this in mind, we created an experiment with French-learning 3-to-4-year-olds and 28-month-olds (age groups that were attested to successfully use prosodic information to constrain parsing in previous studies in French) with a type of ambiguity not yet studied. Using the same type of preferential looking task as Dautriche et al.'s (2014), we aimed to investigate French-learning children's performance with a syntactic structure that is less frequent than those that have been tested before (i.e. right-dislocation, and the position of the prosodic boundary between noun and verb phrases), namely stripping sentences.

We used French sentences such as "*Le tigre tape le canard aussi*" ("The tiger is hitting the duck too"), which can have two interpretations depending on the prosodic phrasing. When produced with a major prosodic boundary (IP - Intonational Phrase boundary) after the verb (as in 1a), the string of words is interpreted as two main sentences, with both the tiger and the duck being the agents of the action "hit", and with the verb elided (i.e., deleted) in the second phrase via a syntactic process called *stripping* (Ross, 1969). However, when the sentence is produced with a minor prosodic boundary (PP - Phonological Phrase boundary) before the verb (i.e., between the subject noun

phrase and the verb phrase, as in 1b), all the words are part of the same main sentence and thus the tiger is the agent of the action, and the duck is the patient.

- (1) a) [*Le tigre tape*]! [*Le canard aussi*]! – [The tiger is hitting]! [The duck too]!  
b) [*Le tigre*] [*tape le canard aussi*]! – [The tiger] [is hitting the duck too]!

One reason why we believe children might have more trouble dealing with sentences such as (1a) than with the type of sentences used in Dautriche et al. and de Carvalho et al. is the fact that they convey an ellipsis, which means that there is an absent (deleted) element in the clause (i.e., the elided Tense Phrase) that needs to be recovered by the listener. Ellipsis sentences are very common amongst the world's languages (Barss, 2003), and stripping sentences are common in Romance languages such as French (Cyrino & Matos, 2002). Although some studies claim that children understand and produce ellipsis sentences from around 3 years of age (e.g., Foley, Núñez Del Prado, Barbier & Lust, 2003; Lindenbergh, Van Hout & Hollebrandse, 2015; Postman, Foley, Santelmann & Lust, 1997), a speaker that is still acquiring their native language may have more trouble solving the ambiguity of sentences such as (1a), since they require the recovery of an absent element.

Another reason why children may have more trouble interpreting our stripping sentences than Dautriche et al.'s right-dislocated sentences is that ellipsis sentences seem to be less frequent in French-learning children's input than sentences with dislocation. Dautriche et al., (2014) conducted a search for noun dislocation on the speech corpora of two French children from 1 to 4 years of age<sup>2</sup> and found that these structures composed 5% of all multiple-word utterances that were said to or around children<sup>3</sup>. We searched for

---

<sup>2</sup> Tim and Marie from the Lyon corpus in CHILDES (Rose & MacWhinney, 2014; Demuth & Tremblay, 2008). Total number of sentences: 33.491. The authors excluded sentences without verbs from the analysis.

<sup>3</sup> Perhaps because these are topicalized sentences, and topicalization is very frequent in spoken French.

ellipsis sentences with *aussi* on the same corpora<sup>4</sup> and found that these compose only 0.14% of all multiple-word utterances that were said to or around children, whereas sentences containing *aussi* in general composed 1.74% of these utterances. Although ellipses with *aussi* cannot account for all ellipsis sentences that can occur in French, these numbers suggest that ellipses are not as common in French-learning children's input as sentences with noun dislocation. Furthermore, if ellipses are not very frequent in children's input, *linearly ambiguous* ellipsis sentences seem to be even less frequent. This is because a speaker who is aware of the ambiguity between (1a) and (1b) could use another construction to avoid it, such as adding an overt coordination marker (i.e. "and"), as in "*le tigre tape et le canard aussi !*" ("the tiger is hitting **and** the duck too!"), or even repeating the verb in the second conjunct, as in "*le tigre tape ! le canard **tape** aussi !*".

In our study, while children listened either to stripping (e.g., (1a)) or transitive sentences (e.g., (1b)), they saw two videos side-by-side on a screen: one depicting the interpretation of the transitive test sentence (one-agent video: a tiger hitting a duck and a bunny with a stick alternately, for the example above), and another depicting the interpretation of the stripping sentence (two-agent video: a tiger and a duck both hitting a bunny with a stick at the same time, for the example above; see Figure 1 below). We measured 28-month-olds and 3-to-4-year-olds' looks towards the videos during test. In addition, for the 3-to-4-year-old children we also recorded their pointing responses to the request "show me which video the lady told you to look at" (*montre-moi quelle vidéo la dame t'a dit de regarder*). If participants exploit the relationship between prosodic and syntactic structures in our sentences to constrain their parsing, we expect that in the stripping condition children will look/point significantly more towards the two-agent

---

<sup>4</sup> Total number of sentences: 46.063. This number is higher than the one found in Dautriche et al. because we included the sentences without a pronounced verb in the analysis.

video than children in the transitive condition. If, however, children's looking behavior does not change between conditions, this could mean that children have more trouble using prosody to parse stripping sentences than to parse other syntactic structures that do not involve ellipsis, as attested in previous French studies with children of the same ages (Dautriche et al., 2014; de Carvalho et al., 2016a, 2017).



**Figure 1.** Example of test videos for the verb "hit".

To summarize, the reason we chose to test more complex, less frequent sentences than the ones from previous studies is because we wished to investigate whether children would be able to use prosody to constrain their parsing of these sentences. If, in this study, we observe similar results as previous studies in French, this would suggest that French-learning children can use prosodic boundary information to correctly parse stripping sentences, showing that this ability is not restricted to highly frequent structures which are more systematically disambiguated through prosodic information. This would also suggest that, by 28 months of age, children already understand complex stripping constructions, which involve an ellipsis. On the other hand, if French-learning children fail to correctly interpret stripping sentences differently from transitive sentences, then this could mean that they have more trouble parsing this type of sentence through prosodic

boundary information, since it is a more complex structure that is less frequent in children's input.

## **2 Materials and Methods**

The method, data analysis and criteria for exclusion of participants were pre-registered on the OSF (Open Science Framework) database before collecting data. The materials, collected data, data analysis and additional exploratory analyses are freely available to readers at [https://osf.io/spdgm/?view\\_only=26e20e2dd13c46ab80892b4fdc37cb70](https://osf.io/spdgm/?view_only=26e20e2dd13c46ab80892b4fdc37cb70).

### ***2.1 Participants***

For the preferential looking task (see section 2.3), fifty-one 3-to-4-year-olds and forty-eight 28-month-olds were included in the final analysis. In the 3-to-4-year-old group, 26 participants were assigned to the transitive condition (*Mage* = 40.2 months, range 36.3 to 46.0, 14 girls) and 25 to the stripping condition (*Mage* = 40.1 months, range 36.2 to 48.4, 10 girls). In the 28-month-old group, 24 participants were assigned to the transitive condition (*Mage* = 27.8, range 27 to 28.8, 12 girls) and 24 to the stripping condition (*Mage* = 27.8, range 27.1 to 28.7, 10 girls). All children were monolingual native French speakers and were not exposed to other languages at home more than 30% of the time. Their parents signed an informed consent form and filled out a vocabulary questionnaire with the nouns, verbs and the adverb used in the test sentences. 92% of the parents of the older group and 72% of the parents of the younger group reported that their children knew all the words from the test sentences. 3% of parents from both age groups reported that their children did not know the word “*aussi*” (too) in French, and the other nouns and

verbs were reported to be known by 96% of children or over, except for "dinosaur", which was reported by 24% of the parents not to be known by their children (all from the younger group).

An additional 31 children (twenty-one 3-to-4-year-olds and ten 28-month-olds) were tested but not included in the preferential looking analysis. Participants were excluded because of missing eye-tracking data<sup>5</sup> (five 28-month-olds and nine children from the 3-to-4-years-old group); fussiness (two younger and two older children); refusal to participate (two children from the older group); parental intervention (three children from the younger group); and technical problems (which affected only the first eight children tested, all from the older group)<sup>6</sup>. The children who were not included in the looking analysis because of missing eye-tracking data or technical problems did not have their pointing data discarded, as these factors did not influence their offline pointing responses, and the missing eye-tracking data was not judged to be caused by children's lack of attention to the videos (otherwise they would be classified as fussy children).

For the pointing task, performed only with the older group, 37 children were included in the final analysis<sup>7</sup>. From this group, 19 participants had been assigned to the transitive condition, (*Mage* = 40.7 months, range 36.1 to 48.5, 11 girls), and 18 to the stripping condition (*Mage* = 41.4 months, range 36.2 to 48.4, 10 girls). An additional eight children pointed but had their data discarded due to data loss<sup>8</sup>.

---

<sup>5</sup> Trials with more than 25% of data loss (which were counted as time looking away from the screen/missing tracks) were excluded from the analysis; participants with less than two exploitable trials were excluded from the analysis.

<sup>6</sup> The initial script for running the experiment contained the wrong onset times for the target words in the test sentences.

<sup>7</sup> We had more participants in the looking than in the pointing task because several children refused to point while still paying attention to the test videos.

<sup>8</sup> The video recordings of the children were lost, so the experimenter could not check their pointing answers after they completed the task.

## 2.2 Materials

Four pairs of test videos were created, using animal puppets. The four actions were the same used by Dautriche et. al. (2014): "eat", "push", "hit" and "carry". Each test trial started with an introduction video. These videos were intended to familiarize children with the verbs used in the test sentences and to provide a necessary context for the transitive sentences (one does not say "the tiger is hitting the duck too" without a preceding context where the tiger was hitting somebody else). The first character in each test sentence performed an action involving a third character, accompanied by a sentence such as "*Regarde ! C'est le tigre ! Et le lapin ! Oh, il le tape !*" ("Look! It's the tiger! And the rabbit! Oh, he is hitting him!"), for the "hit" action.

For each pair of test videos, one video corresponded to the interpretation of the simple transitive test sentences (i.e., the one-agent video – a tiger hitting a duck and a bunny with a stick alternately, for the "hit" action), and the other to the interpretation of the stripping sentence (i.e., the two-agent video – a tiger and a duck hitting a bunny with a stick at the same time, for the "hit" action; see Figure 1 for an example). For the verbs "eat", "push" and "hit", both test videos contained three animals which participated in the action, but the transitive videos had one agent and two patients, while the stripping videos had two agents and one patient. For the verb "carry", the transitive video contained two animals (i.e. a duck and a dinosaur) and a gift box: the duck was the agent, and the dinosaur and the gift box were the patients (i.e., the things being carried); and the stripping video contained the same animals but two gift boxes (which were the patients of the individual "carry" actions; see Appendix A for a detailed written description and illustrations for all the video stimuli used in this experiment, and the OSF for the full video stimuli). The quantity and quality of movements, as well as characters' position on

the screen was controlled for each video pair, so that both videos were as similar as possible in terms of complexity. Since the videos were inherently more complex than the ones in Dautriche et. al. (2014), which contained at most two characters and only one agent in each video, we increased the duration of our test trials from 9s to 12s, and the preview and contrast trial times from 7s to 9s, compared to the previous experiment, so that children would have more time to inspect the videos.

The sentences used in the experiment were recorded by a female native French speaker (the last author) in a sound-attenuated booth, using a recorder and a condenser microphone, and were included in the videos using the Filmora video editor. Before conducting the experiment with children, we assessed whether the sentences we created were indeed triggering the interpretations we expected by asking 38 French-speaking adults to listen and judge the recorded sentences via an online form, made with Google Forms. Subjects were asked to listen to the sentences through headphones, and to choose the best interpretation for them on a 5-point scale, 1 being the transitive interpretation of the sentences (one-agent interpretation: "*le tigre tape le canard*" ("the tiger is hitting the duck")), for instance), 5 being the stripping interpretation (two-agent interpretation: "*le tigre tape et le canard tape*" ("the tiger is hitting and the duck is hitting")) and 3 being completely ambiguous. The sentences were judged correctly in 89% of the answers (i.e., participants chose point 5 of the scale as the best interpretation for the stripping sentences and point 1 as the best interpretation for the transitive sentences)<sup>9</sup>.

---

<sup>9</sup> There were 1.86% of wrong judgements, all for transitive sentences (points 4 or 5), and 0.9% of completely ambiguous judgements (point 3, 0.55% for stripping sentences and 0.33% for transitive sentences). 7.67% of answers were correct but slightly ambiguous judgements (1.97% for stripping sentences (point 4); and 5.7% for transitive sentences (point 2)).

### 2.2.1 Acoustic analysis

To assess the prosodic differences between stripping and simple transitive sentences, we conducted acoustic measurements on the segments around the positions of the prosodic phrase boundaries in the sentences recorded for the experiment. We analyzed three acoustic parameters which usually cue phrase boundaries in natural languages (see e.g., Nespor & Vogel, 1986): presence and length of pauses; duration of the rhyme of the word preceding the prosodic boundary (e.g., [igrə] in "*tigre*"<sup>10</sup>; [ap] in "*tape*"); and pitch (F0) contour around the prosodic boundaries, calculated as the difference between the maximum pitch of the vowel preceding the boundary (e.g., [i] in "*tigre*") and the maximum pitch of the vowel following it (e.g., [a] in "*tape*"). Besides the boundary between the first noun and the verb for transitive sentences, and between the verb and the second noun for stripping sentences, we investigated the possibility that the transitive sentences also presented a boundary between the second noun and the adverb "*aussi*", due to the fact that this adverb is at the end of a long sentence (e.g., "*le tigre tape le canard aussi*"), whereas in the stripping sentences it is at the end of a short sentence (e.g. "*le canard aussi*"). So, we also analyzed the rhyme of the second noun (e.g. [ar] in "*canard*") and the pitch contour between the last vowel of the second noun and the first vowel of "*aussi*".

We found significant phrase-final lengthening effects, as expected from the literature (e.g., Delais-Roussarie, 1995; Jun & Fougeron, 2002; Millotte et al., 2008, 2007; Shattuck-Hufnagel & Turk, 1996): the rhyme of the first noun was significantly longer in the transitive sentences ( $M = 485\text{ms}$ ,  $SD = 59$ ) than in the stripping sentences ( $M = 262\text{ms}$ ,  $SD = 73$ ), and the rhyme of the verb was significantly longer in the stripping

---

<sup>10</sup> For the word [ˈtigrə], we considered [igrə] as the last rhyme and [i] as the last vowel instead of [ə], because [ə] is usually only pronounced in specific contexts where an epenthetic vowel is needed for pronunciation, as when followed by a consonant.

sentences ( $M = 335\text{ms}$ ,  $SD = 184$ ) than in the transitive sentences ( $M = 167\text{ms}$ ,  $SD = 95$ ). The rhyme of the second noun was also longer in the transitive ( $M = 298\text{ms}$ ,  $SD = 35$ ) than in the stripping sentences ( $M = 141\text{ms}$ ,  $SD = 28$ ), suggesting that there is indeed a second phrase boundary in this position in transitive sentences.

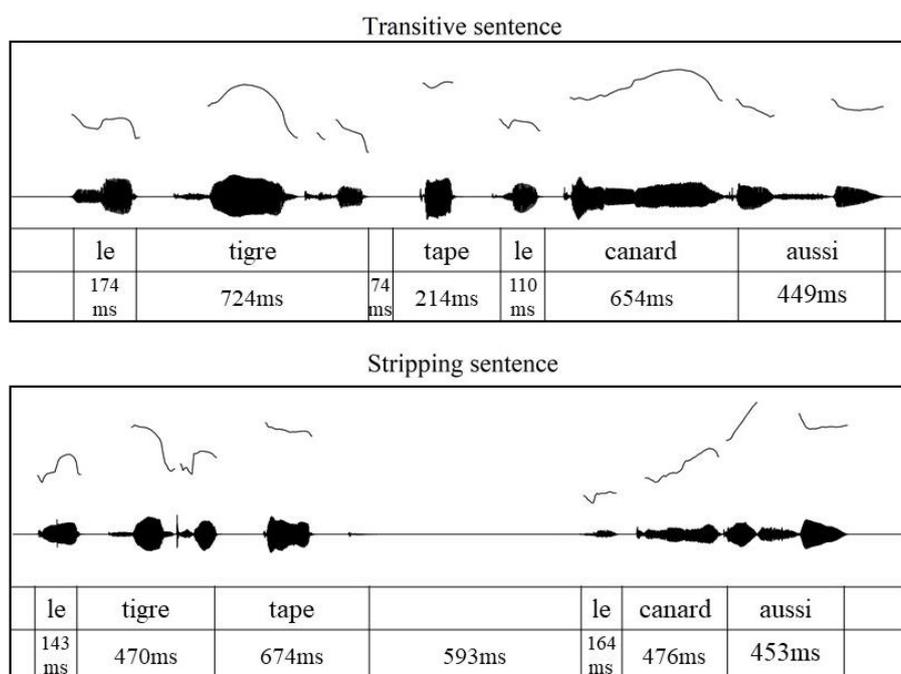
Regarding the pauses between boundaries, we found a consistent pause between the verb and the second noun in the stripping sentences, with mean length of  $593\text{ms}$  ( $SD = 190$ ), and a pause between the first noun and the verb in most of the transitive sentences, with mean length of  $74\text{ms}$  ( $SD = 67$ ).

We also found a significant difference in pitch (max.  $F_0$  of the vowel following the boundary *minus* max.  $F_0$  of the vowel preceding the boundary) around the boundary between the verb and the second noun, with a pitch discontinuity that was more important in the stripping sentences ( $M = -168\text{Hz}$ ,  $SD = 18$ ) than in the transitive sentences ( $M = -66\text{Hz}$ ,  $SD = 46$ ). A significant pitch difference between conditions was also found when comparing the maximum pitch value of the last vowel of the second noun and the first vowel of the adverb "aussi". In Transitive sentences we observed an overall pitch decrease between these vowels ( $M = -65\text{Hz}$ ,  $SD = 15$ ) while in stripping sentences we observed an overall increase in the pitch values ( $M = 154\text{Hz}$ ,  $SD = 54$ )<sup>11</sup>. However, we did not find a significant pitch difference around the boundary between the first noun and the verb across sentences. The average duration values can be found in Table 1, and the averaged values of  $F_0$  can be found in Table 2.

---

<sup>11</sup> These results are consistent with the fact that, in French, when there is a prosodic boundary we would expect the last vowel of the word preceding it to have a higher pitch (because of its position at the end of a phonological phrase) than the first vowel of the word following the boundary, which is located at the beginning of a new phonological phrase.

These analyses show significant duration and pitch differences between the transitive and stripping sentences around the intonational phrase boundary of stripping sentences (i.e., pause between the verb and the second noun; longer duration of the rhyme of the verb; and larger difference in F0 between the last vowel of the verb and the vowel of "le" from the last noun), and around the phonological phrase boundaries of transitive sentences (i.e., pause between the first noun and the verb; longer duration of the rhyme of the first and second nouns; and decrease in F0 between the last vowel of the second noun and the first vowel of "aussi").



**Figure 2.** Soundwave, pitch and duration for a pair of sentences used in this experiment. A stripping sentence ("Le tigre tape! Le canard aussi !") at the bottom and a transitive sentence ("Le tigre tape le canard aussi !") on top. The duration of each segment, averaged over all sentences from the four test trials, is indicated below the boxes containing the words.

**Table 1.** Mean duration (in ms) of the different segments around the prosodic boundaries for both conditions: stripping vs transitive sentences. The p-values from the analysis of the pauses come from a Wilcoxon rank sum test.

	Stripping sentences (e.g., test: [le tigre tape] [le canard aussi]; prompt: [le tigre va taper] [le canard aussi])		Transitive sentences (e.g., test: [le tigre] [tape le canard aussi]; prompt: [le tigre] [va taper le canard aussi])		t-value (p)
	Duration (ms)	SD	Duration (ms)	SD	
Rhyme of the 1 <sup>st</sup> noun (e.g. t[igre])	262	73	485	59	-6.74 (**p < 0.001)
Pause between the 1 <sup>st</sup> noun and the verb	0	0	74	67	**p = 0.0046
Rhyme of the verb (e.g. t[ape] or tap[er])	335	184	167	95	2.28 (*p = 0.038)
Pause between the verb and the 2 <sup>nd</sup> noun	593	190	0	0	**p = 0.0004
Rhyme of the 2 <sup>nd</sup> noun (e.g. can[ard])	141	28	298	35	-9.84 (**p < 0.001)

**Table 2.** Mean difference of F0 (in Hz) of the different segments around the prosodic boundaries for both conditions: stripping vs transitive sentences.

	Stripping sentences		Transitive sentences		t-value (p)
	F0 (in Hz)	SD	F0 (in Hz)	SD	
First vowel of the verb (e.g. t[a]pe or v[a]) <i>minus</i> last vowel of the 1 <sup>st</sup> noun (e.g., t[i]gre)	-10	70	-57	74	1.31 (p = 0.21)
Vowel of "le" from the 2 <sup>nd</sup> noun <i>minus</i> last vowel of the verb (e.g., t[a]pe or tap[e]r)	-168	18	-66	46	-5.83 (**p < 0.001)
First vowel of "aussi" <i>minus</i> last vowel of the 2 <sup>nd</sup>	154	54	-65	15	11.10 (**p < 0.001)

noun (e.g., can[a]rd)					
--------------------------	--	--	--	--	--

### ***2.3 Procedure***

Caregivers were contacted through a mailing list and invited to come to the babylab. Once at the babylab, they were asked to sign a consent form, and the experimenter explained the procedure to them. Caregivers were told they would sit in front of a screen inside a booth, with the child on their lap, and should remain as still as possible, not point at the screen nor talk to the child during the experiment, so as not to affect the child's behavior during the task.

Once the child was in the experimental booth, the experimenter put a sticker with a high-contrast pattern on their forehead, that served as a reference for the eye position for the eye-tracker. She then adjusted the eye-tracker camera, while the child watched a cartoon. After that, the caregiver was given headphones with masking music, so they could not listen to the test sentences and feel tempted to influence the child's interpretation. Then a five-point calibration of the eye-tracker began, followed by the experiment, which was run in Matlab (Mathworks, Natick, MA). For the older children (3-to-4-year-olds), the experimenter stayed inside the experimental booth with the caregiver and the child, sitting in a lower chair behind the caregiver and slightly to the child's right, to be able to request the pointing responses from the child at the end of each trial. The experimenter also wore headphones to ensure that they were blind to the exact stimuli the participant was watching. For the younger participants (28-month-olds), the experimenter stayed outside the booth and watched the child through a video camera during the experiment.

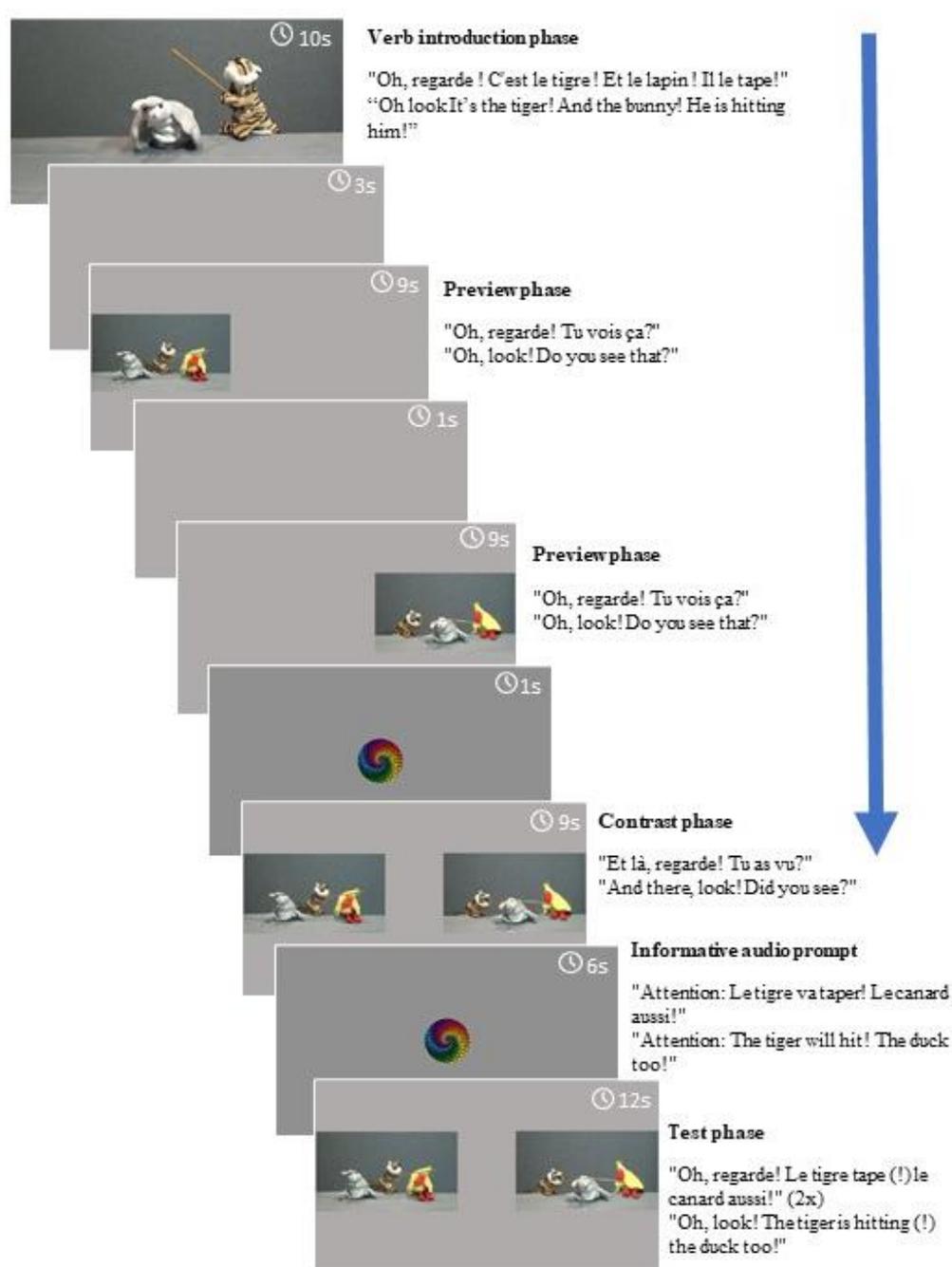
The videos were presented on a 27-inch screen, and the auditory stimuli were played through two speakers positioned on each side of the screen. The left-right position of the videos was pseudo-randomized, so children did not see more than two consecutive target videos on the same side. For half of the trials the target video appeared on the right side of the screen, and for the other half on the left. The order of item presentation was fully randomized, to avoid possible order biases in children's responses. The pointing responses were recorded by the experimenter through a keyboard (i.e., pressing the left or right arrow), and the eye data was collected via an Eyelink 1000 eye-tracker (operating in a remote mode and collecting data every 2ms).

The experiment started with an introduction video of each of the puppet characters: a bunny, a duck, a tiger, a dinosaur, and a monkey. Each puppet was presented in a five-second video in which the character appeared waving and dancing on the screen and was labeled once (e.g., "*Oh Regarde ! C'est le lapin !*" ("Oh Look! It's the bunny!")). After this familiarization preview, participants saw one training trial, with the same structure as the test trials (see Figure 3), involving the intransitive verb "*sauter*" ("to jump"): children saw two videos side-by-side, one displaying a bunny jumping and the other displaying a bunny and a monkey playing together, and heard the sentence "*le lapin saute !*" ("the bunny is jumping!"). This practice trial familiarized children with the procedure and showed them that the soundtrack matched only one of the videos on the screen.

In the test phase, the introduction videos (i.e., verb introduction phase) were followed by a preview phase (see Figure 3), where each action appeared individually on the screen for nine seconds, followed by a contrast phase, where both actions appeared at the same time also for nine seconds, with neutral prompt sentences such as "*Regarde ! Tu vois ça ?*" ("Look! Do you see that?"). At the end of the contrast phase, the screen turned

gray, and the first test sentence was played, in the future tense (e.g., “*Attention: le tigre va manger ! Le dinosaure aussi !*” - “Watch out: the tiger will eat! The dinosaur too!”). Then the test videos reappeared side-by-side on the screen for twelve seconds, while the test sentence was played two more times, but in the present tense (e.g., “*Le tigre mange ! Le dinosaure aussi !*” - “The tiger is eating! The dinosaur too!”). For the 3-to-4-year-old group, the videos froze on the screen at the end of each trial, and the experimenter asked the child to point at "the video the lady was talking about" (28-month-olds were not asked to point). If children refused to point at one of the videos, the experimenter encouraged them by saying that she could not hear the lady (since she was wearing headphones) and pointed at both videos herself, while asking "was it this one or this one?". If children still did not want to point, the experimenter proceeded with the experiment anyway, and pointing was recorded as missing. Pointing responses were systematically praised, whether they were correct or not.

Children saw a total of one training trial (see above) and four test trials (involving the actions "eat", "push", "carry" and "hit"). Once the experiment was over, caregivers were asked to fill out the specific vocabulary questionnaire. Children received a diploma of "honorary member of our babylab" as a gift for their participation.



**Figure 3.** Structure of a test trial. The clock icon followed by the number illustrates the time each video/grey screen appeared (it was not present in the actual videos).

## **2.4 Data analysis**

Eye-tracking data was analyzed through a cluster-based permutation analysis (Maris & Oostenveld, 2007) complemented by a t-test for overall looking-times (averaged over the entire duration of each test trial). Both analyses used the proportion of looks towards the two-agent action (i.e., stripping video) as the dependent variable. We used R (R core team, 2017) and the package `eyetrackingR` (Dink & Ferguson, 2016) to analyze the data. We first down-sampled the eyetracking data by averaging to one sample every 20ms. For the cluster-based permutation, the analysis searched the time-window(s) where a significant effect of condition was observed, indicating that children looked more at the two-agent action while listening to stripping sentences than while listening to transitive sentences. Since children listened to the test sentence once before the beginning of the test phase, they could possibly already show a preference for one video from the beginning of the trial; therefore, the cluster-analysis searched for time windows during the entire test trial (0 to 12000ms). Adjacent time points with a t-value greater than a predefined threshold ( $t = 1.5$ , for the comparison between groups) were grouped together into a cluster, and the probability of observing a cluster of the same size by chance was estimated by running the same analysis 1000 times on simulated data, in which groups were randomly assigned to participants. Our second analysis analyzed the difference in looking times averaged over the entire duration of the trials, with a t-test testing for the effect of condition.

For the pointing responses, we ran a generalized linear mixed-effects regression with the proportion of pointing towards the two-agent video as the dependent variable (coded as 0/1 in each trial), condition (sum coded) as a fixed effect with the stripping

condition as the base level, and participants as a random intercept<sup>12</sup>. If children can use the prosodic information of a sentence in order to figure out its syntactic structure, and therefore constrain their interpretation, then they should point more and look longer at the two-agent action when presented with stripping sentences than when presented with transitive sentences.

### 3 Results<sup>13</sup>

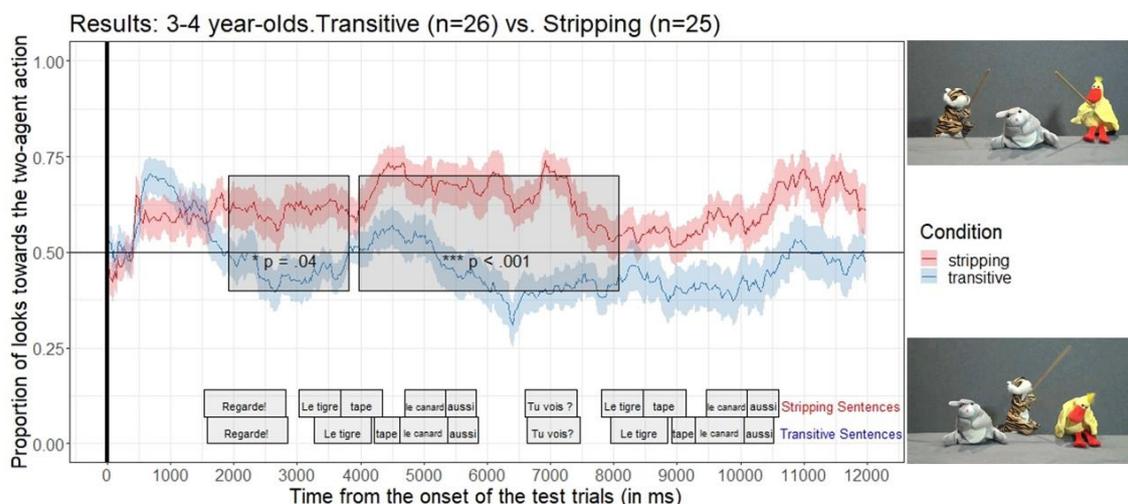
#### 3.1 Eye-tracking data: 3-to-4-year-olds

Figure 4 shows the time course of the proportion of looks to the two-agent video, during the test trials, for participants in the stripping condition (red line) and in the transitive condition (blue line). The cluster-based analysis found two clusters with a significant difference between conditions, which are indicated in Figure 4 by the gray rectangles with their respective  $p$ -values. The first cluster with a significant effect of condition started slightly before the onset of the test sentence (between 1920ms and 3820ms,  $p = .04$ ). This can be explained by the fact that children already heard the test sentence once, right before the beginning of the test trial. The second cluster started at 3980ms and lasted until 8080ms ( $p < .001$ ). This latter time-window coincides with the onset of the target verb during test phase.

---

<sup>12</sup> The maximal random effect structure that allowed the model to converge (Barr, Levy, Scheepers, & Tily, 2013).

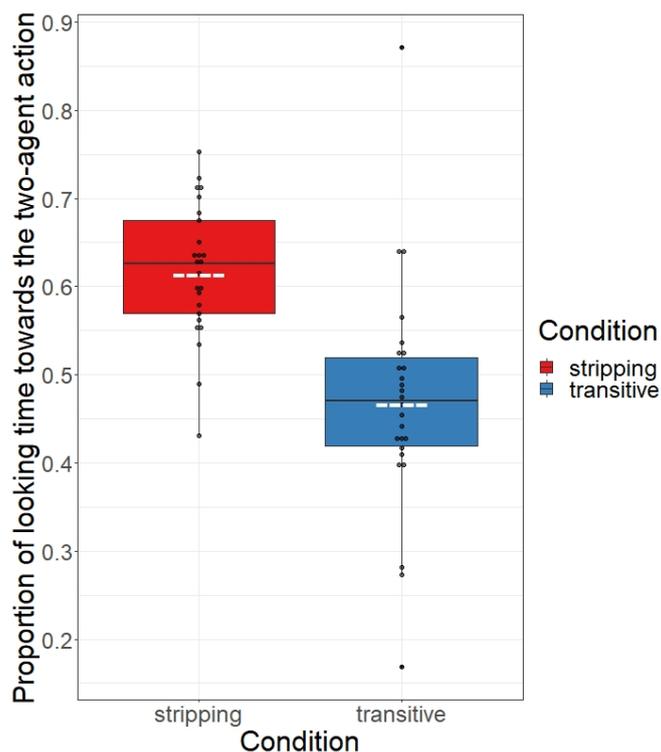
<sup>13</sup> Additional and exploratory analyses (not pre-registered) were conducted and are freely available to readers in a Supplementary Materials folder on our OSF page.



**Figure 4.** Proportion of looks towards the two-agent action throughout the entire test trial (12 seconds in each trial) in the stripping (red line) and transitive (blue line) condition for 3-to-4-year-olds. The significant time-clusters are delimited by the gray boxes with their respective  $p$ -values. The black vertical line marks the beginning of the test trial. The average times of the onsets of each word in the test sentences are indicated in the gray boxes at the bottom part of the graph (the upper boxes show the onsets for the stripping sentences, and the boxes below them show the onsets for the transitive sentences).

With regards to the overall looking behavior, Figure 5 shows the average proportion of looking time to the two-agent video averaged over the entire trial duration across all four items in the experiment, for the stripping condition (red box) and transitive condition (blue box). As expected, the results show that children who heard stripping sentences looked significantly longer towards the two-agent videos ( $M = .62s$ ,  $SD = .08s$ ) than children who heard the transitive sentences ( $M = .47s$ ,  $SD = .13s$ ,  $t(40.48) = 4.92^{14}$ ;  $p < .001$ ; Cohen's  $d = 1.36$ ).

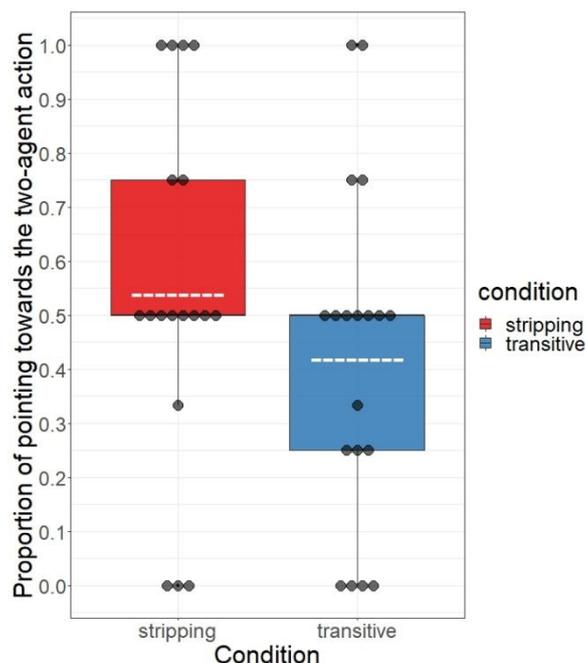
<sup>14</sup> The number of degrees of freedom is not whole because it compensates for the difference in number of participants between conditions ( $n = 25$  in the stripping condition, and  $n = 26$  in the transitive condition).



**Figure 5.** Average proportion of looking time towards the two-agent action in the stripping (red box) and transitive (blue box) condition for 3-to-4-year-olds. Each black dot represents the average looking time for one participant. The white dashed lines indicate the mean for each condition (.62 for the stripping condition and .47 for the transitive condition). The black horizontal lines within the squares represent the median (.63 for the stripping condition and .47 for the transitive condition). The lower and upper hinges correspond to the first and third quartiles. The top whiskers denote the maximum value, and the bottom whiskers the minimum value. The outliers (i.e., the black dots which are distant from the whiskers) were not removed from any of the reported analyses.

### 3.2 Pointing data: 3-to-4-year-olds

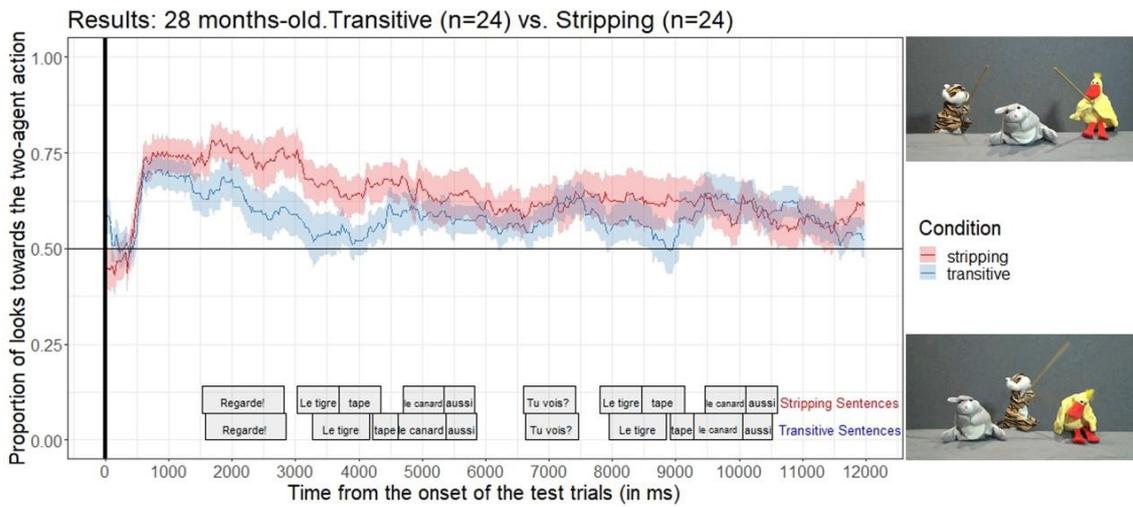
For the pointing analysis (Figure 6), the results show no significant difference between the stripping (red box) and transitive (blue box) conditions ( $SE = .251$ ,  $\beta = -.301$ ,  $z = -1.200$ ,  $p = .230$ ), although the results follow the expected direction, with children in the stripping condition pointing more towards the two-agent action than children in the transitive condition ( $M = 0.54$  in the stripping condition versus  $M = 0.42$  in the transitive condition).



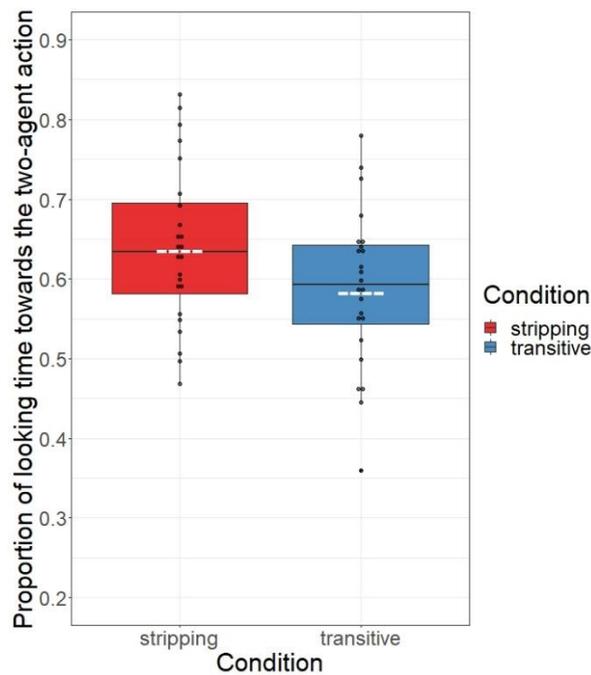
**Figure 6.** Average proportion of pointing towards the two-agent action video. The white dashed lines show the mean for each condition (.54 for the stripping condition (red bar) and .42 for the transitive condition (blue bar)). The median overlaps with the minimum value for the stripping condition and the maximum value for the transitive condition (e.g., .50). Each black dot represents the average proportion of pointing for one participant.

### 3.3 Eye-tracking data: 28-month-olds

For the 28-month-olds, no significant difference in proportion of looks between conditions was found in the cluster-based analysis, and only a marginally significant difference was found for the overall looking-times in the  $t$ -test analysis ( $t(46) = 1.828$ ;  $p = .074$ ; Cohen's  $d = .528$ ). Although toddlers in the stripping condition seem to have looked longer ( $M = .64s$ ,  $SD = .10s$ ) towards the stripping video than children in the transitive condition ( $M = .59s$ ,  $SD = .10s$ ), this difference was not significant. Visual inspection of the data suggests that the proportion of looks towards the two-agent action is above 50% during almost the entire trial for both conditions. However, the transitive condition is mostly below the stripping condition, suggesting that there is a tendency towards the same pattern of results observed for the 3-to-4-year-olds.



**Figure 7.** Proportion of looks towards the two-agent video throughout the whole test trial (12 seconds in each trial) in the stripping (red line) and transitive (blue line) condition for the 28-month-olds.



**Figure 8.** Average proportion of looking time towards the two-agent action in the stripping (red box) and transitive (blue box) condition for the 28-month-olds. Each black dot represents the average looking time for one participant. The white dashed lines show the mean for each condition (.64 for the stripping condition and .59 for the transitive condition). The black lines within the squares represent the median (.63 for the stripping and .59 for the transitive condition).

### 3.4 Age group comparison

In order to test the difference between the two age groups statistically, we performed an additional analysis comparing the overall proportion of looking times towards the two-agent action video averaged across the entire trials (12s in each trial), with Participants as the random factor, Condition (Stripping vs. Transitive), Age-group (28-month-olds vs. 3-to-4-year-olds), and an interaction between Condition and Age-group as between-participant factors, a random intercept for participant, and a random slope for condition by item<sup>15</sup> (this is a *post-hoc* comparison that was not pre-registered). We found a significant effect of Condition ( $\beta = -.061$ ,  $SE = .016$ ,  $t(30) = -3.677$ ;  $p < .001$ ) and Age-group ( $\beta = .054$ ,  $SE = .022$ ,  $t(283) = 2.416$ ;  $p = .016$ ) but no interaction between Condition and Age-Group ( $\beta = .034$ ,  $SE = .022$ ,  $t(283) = 1.518$ ;  $p = .13$ ). These results show that children's proportion of looks were influenced by condition (stripping vs. transitive), with children in the stripping condition looking longer towards the two-agent video ( $M = .59s$ ,  $SD = .20s$ ) than children in the transitive condition ( $M = .49s$ ,  $SD = .21s$ ). They also show that, overall, the younger children looked significantly less towards the two-agent video ( $M = .51$ ,  $SD = .20$ ) than the older children ( $M = .57$ ,  $SD = .21$ ). However, we found no evidence for a significant difference in the effect of condition between the two age groups.

## 4 Discussion

In the present work, we investigated French-learning children's ability to tell apart stripping sentences such as "[*Le tigre tape*]! [*Le canard aussi*]!" from simple transitive

---

<sup>15</sup> Maximal effect structure that allowed the model to converge. We used sum coding with the stripping condition in the older age group as the base level.

sentences such as "[*Le tigre*] [*tape le canard aussi*]!" by relying mainly on their different prosodic structures, which reflect their different syntactic structures. In order to do so, we ran a preferential looking experiment with a picture selection task where we presented children with one of the two types of sentence above, while displaying two simultaneous videos, one which corresponded to the interpretation of the transitive sentence (one-agent video), and another one which corresponded to the interpretation of the stripping sentence (two-agent video). We found that children in the stripping condition looked significantly longer towards the two-agent action than children in the transitive condition. These results show that 3-to-4-year-old children successfully exploit prosodic information to constrain their parsing and to interpret the two types of sentences differently. For the 28-month-olds, however, we only found a marginally significant difference between conditions, with a tendency in the expected direction. Since the comparison between age groups showed a significant effect of condition and age, but no interaction between these two factors, this might mean that the ability to tell apart transitive and stripping sentences is already emerging in 28-month-olds.

These results suggest that young children can use prosody as a cue for grouping words into syntactic constituents within a sentence. Moreover, our results suggest that 3-to-4-year-olds already understand complex stripping constructions, which involve an ellipsis with an identity relationship with the preceding sentence, conveyed by the adverb "*aussi*" ("too").

The current study dovetails nicely with other studies investigating children's interpretation of different types of ellipsis, which show that children understand and produce ellipsis constructions from around 3 years of age (Postman et al., 1997, and Foley et al., 2003, in English; Lindenbergh, Van Hout & Hollebrandse, 2015, in Dutch; among others). Our findings are also consistent with recent studies demonstrating that 2-to-3-

year-olds can use the information provided by adverbs to narrow down their interpretations of novel adjectives and verbs (Arunachalam & Syrett, 2014; Arunachalam, Syrett & Chen, 2016; Syrett, Arunachalam, & Waxman, 2014; Syrett & Lidz, 2010). For instance, in Arunachalam & Syrett (2014) and Arunachalam et al., (2016), 2-to-4-year-olds correctly interpreted novel verbs embedded in sentences such as “The boy and her sister *lorped* together” as referring to a synchronous action with two agents (i.e., two participants performing the same intransitive action at the same time, like the situation in our two-agent video), whereas they did not reach the same interpretation in the absence of the adverb ‘together’. In our study, similarly, we observed that children in the same age range understood the adverb “*aussi*” (“too”), which referred to the action being performed by the second agent in the stripping sentences (i.e., “too” = same action, e.g., the duck is *also* hitting), and to the patient of the action in the transitive sentences (i.e., “too” = additional patient, “the tiger is *also* hitting the duck (in addition to someone else)”).

One might wonder, however, whether children in our task might have overall looked longer at the two-agent videos when listening to the stripping sentences, even if they did not fully understand the ellipsis and/or the adverb “*aussi*”. For instance, if they encoded the sentences as “*Le tigre tape ! Le canard X !*”, X being an unknown word, they could still interpret the sentences as containing two agents, simply because *le tigre* and *le canard* are both at the beginning of intonational phrases, and this is the most common position for agents in French. Consequently, when confronted with the two test videos, they would choose the only one where both the tiger and the duck are agents, regardless of whether they are performing the same action or not. If this interpretation were correct, children would need either to consider that ‘*aussi*’ plays the role of a novel verb, or to ignore the fact that the second part of the stripping sentence has no verb, which would

make it ungrammatical. Neither of these options is particularly likely. Indeed, according to the first option, children would need to attribute a novel verbal meaning to the familiar adverb “*aussi*”, which according to parental reports, is a word that they already understand and produce. This would be made particularly difficult by the fact that there is no good candidate for that novel meaning: the duck is performing the same action as the tiger, and this action has been named already (with a familiar verb). In addition, children hear four test trials, with four different actions, and in all of them they would need to map “*aussi*” to each of these different actions. According to the second option, children would need to ignore the ungrammaticality of a sentence “without a verb” in order to stick to the two-agent interpretation.

While the looking data shows that children use prosodic boundary information to interpret stripping sentences differently from simple transitive sentences, the pointing results are less clear. As noted by an anonymous reviewer of this paper, a possible reason for this discrepancy could be that, while pointing tasks demand an overt judgement from children, looking tasks simply analyze children's spontaneous and online looking behavior. Tasks that include overt responses require not only a higher memory load, but also better inhibitory control and better capacity for overtly judging the meaning of a sentence, that is, in our case, the capacity to actively acknowledge that the sentence previously heard is correctly depicted by only one of the presented videos, and to identify which one is correct. However, since other studies show good performance on pointing tasks by children as young as two years of age (e.g., Bernal, Lidz, Millotte, & Christophe, 2007), one cannot say that the children tested here were too young to perform correctly in our pointing task. It is however still possible that the complexity of our task impacted our results; while previous experiments showing a success with pointing tasks in young children were usually word-mapping tasks where children were asked to look for referents

of novel verbs or nouns, our task required children to map an entire sentence structure to a dynamic scene (i.e., a video representing an event). The key difference is that in the word-mapping tasks, the experimenter asks children to point explicitly towards the referent of a novel word (e.g., “show me the *blick*/the one who is *blicking*”; or “find the apple/eating”), while in the present experiment we asked children to point at the video “the lady was talking about”, which is a more “abstract” instruction. This request seems to be less straightforward and to increase memory demand, since children need to recover a rather complex sentence they heard before in order to answer properly. Thus, it is possible that the request to retrieve complex sentences from the previous linguistic context, added to the task of actively matching the sentences with one of the presented videos, hindered children's performance in the pointing task. This could explain why several children refused to or were reluctant to point, as they might have been confused by the task. It is also important to notice that, due to this issue, we had a high amount of missing data, and because of that, the pointing analysis had less statistical power.

Although our study shows that young French-learning children can deal with prosodic information even when presented with a different syntactic ambiguity than those which have been studied before in French, the difficulty encountered by 28-month-olds in our task suggests that certain types of structures may be more difficult for children. Since 28-month-olds (Dautriche et al., 2014) and even 20-month-olds (de Carvalho et al., 2017) were able to use prosodic information to constrain their parsing in previous studies using different types of sentences, our findings with stripping versus transitive sentences indicate that young children's ability to use phrasal prosody to constrain parsing might depend on the type of syntactic structure being tested.

As mentioned before, stripping sentences might be harder for children to parse than the sentences used in previous studies in French, because they contain an ellipsis.

Although sentences with noun dislocation such as the ones in Dautriche et al. (2014) also involve a complex syntactic operation (i.e., a movement operation), ellipsis sentences require the interpretation of a silent element, and this task might be harder for young children. Furthermore, ellipsis sentences seem to be less frequent in children's input than sentences with noun dislocation in French, which means that children have less experience interpreting them. The possibly greater difficulty in parsing stripping sentences, combined with the smaller frequency of ellipsis in children's input, might explain why 28-month-olds had more trouble exploiting prosodic information to constrain their interpretation of stripping sentences in the current study than to use prosody to constrain their parsing of other types of sentences in previous studies.

Another possible explanation for the 28-month-olds' results comes from the observation that stripping structures might not be as often disambiguated through prosodic boundary information as the structures tested in previous studies. This is because ellipsis sentences are often produced as a continuation of a previously uttered sentence, such as when a speaker utters "me too" when another speaker makes a statement about something that also applies to them (e.g., "Mary wants ice cream. Me too!"). Additionally, coordinated sentences can also be produced with an overt coordination marker such as "and" (as in "the tiger eats *and* the duck too"): in this case, the sentence would no longer be ambiguous, so children would not need to rely on the prosodic break to realize that both the tiger *and* the duck are agents of the "eating" action.

In contrast, noun phrase dislocation is always marked by a prosodic phrase boundary between the dislocated elements and the elements that stay *in situ*, which signals the disruption of the sentence's canonical structure (Ashby, 1988; De Cat, 2007). As for the syntactic structures using noun-verb homophones in de Carvalho et al. (2017), children only need to interpret the prosodic boundary between the noun phrase and the

verb phrase as a syntactic boundary, something that applies to all sentences they hear daily, whether they contain ambiguous words or not (e.g., Fisher & Tokura, 1996; Gutman et al., 2015; Soderstrom, Blossom, Foygel, & Morgan, 2008). Therefore, the discrepancy between our 28-month-olds' results and previous results with French-learning toddlers might be due to the fact that the disambiguating prosodic breaks of stripping sentences are not as common in children's input as the prosodic structures that were used in previous studies (e.g., boundaries between noun and verb phrases in de Carvalho et al., and prosodic structure of the dislocated element to constrain the interpretation of right-dislocated sentences in Dautriche et al).

Finally, there are also methodological explanations for the 28-month-olds' results. While Dautriche et al.'s test videos involved one or two characters, and de Carvalho et al.'s stimuli consisted of simple still images to illustrate the noun or verb meaning of an ambiguous word, in our experiment each video presented three characters. Identifying the meaning of our test sentences while watching two complex videos side-by-side on the screen might have been more difficult for young children than inspecting simpler videos or still images in previous studies (an observation that is consistent with the recent findings of de Carvalho, Babineau, Trueswell, Waxman, & Christophe, 2019; and Valteau, Konishi, Golinkoff, Hirsh-Pasek, & Arunachalam, 2018). As we explained above, we had to use more complex videos in our experiment and to include three characters per video due to pragmatic reasons. To justify the use of "too" in the transitive sentences, we had to create a context in which the first-mentioned character performed the named action on another patient, as one does not utter a sentence such as "the tiger is hitting the duck too" without a previous context where the tiger was hitting someone else. Although the verb introduction phase already served as a previous context, we decided to also keep the second patient in the test videos, to reinforce the pragmatic antecedent. We

also chose to keep the same number of characters in the stripping videos to balance out the amount of information in each video. This methodological choice might have hindered younger children's performance in the task, as they might have been overwhelmed by the higher complexity of information in the videos, which might have taxed their processing abilities.

In the introduction we noted that some studies on languages other than French had shown that older children failed to use prosodic information to interpret ambiguous sentences (e.g., Choi & Mazuka, 2003 in Korean; Snedeker & Trueswell, 2001, in English). Although it may be tempting to attribute differing results to cross-linguistic differences in the way different languages encode phrasal prosody, at least two studies found that English-speaking children were able to exploit phrasal prosody to constrain parsing (de Carvalho et al., 2016b; Snedeker & Yuan 2008). Thus, at least part of the differences observed between studies seem to come from differences in the type of syntactic structures that were tested, and/or the reliability with which the prosodic structure reflected the syntactic structure. Our current findings appear to corroborate this hypothesis, since we saw that when tested with a different type of sentence, which is less often disambiguated through prosodic boundary information and is more complex, younger French children find it harder to disambiguate. However, we cannot rule out the possibility that the weaker results we found for the younger children were due to the methodological issue pointed out above, i.e., the higher complexity of the videos. In order to do so, we would need to create a control study where this methodological issue is addressed, perhaps by replacing the third character of the videos with an object, so as to make the videos less complex.

Preschoolers' success in our task however, along with previous studies, shows that young children already know a great deal about their native language's prosodic structure,

and can use this knowledge for many different aspects of sentence parsing (e.g., Dautriche et al., 2014; de Carvalho, et al., 2017, 2016a, 2016b, 2019; Hirose & Mazuka, 2017; Massicotte-Laforge & Shi, 2015, 2020; Snedeker & Yuan, 2008; Zhou, Crain & Zhan, 2012; Zhou, Su, Crain, Gao & Zhan, 2012). Our findings add new evidence to this literature, showing that young children can also use phrasal prosody to distinguish stripping from transitive sentences, and therefore to constrain their syntactic analysis and the interpretation they assign to each sentence.

To conclude, our results showed that only 3-to-4-year-olds were able to reliably use phrasal prosody to constrain their parsing of stripping sentences, whereas 28-month-olds had more trouble doing so. On the one hand, these results lend support to the hypothesis that phrasal prosody cues syntactic structure in early language development; on the other hand, they might suggest that children's success depends on the type of structure being tested, and on methodological factors such as the complexity of the task. Different types of structure might pose different challenges for children, as some structures might be more easily parsed and/or more easily disambiguated through prosodic boundary information than others. At the same time, different methodological choices might favor or hinder children's performance in tasks involving parsing and comprehension. Our study adds to the growing body of evidence on the role of prosody in constraining parsing in young children and lends support to the hypothesis that phrasal prosody is an important cue to syntactic structure during language acquisition.

## **5 Acknowledgments**

Jacques Mehler was certainly one of the first scientists in the world to get interested in the idea that babies could rely on prosodic information to discover some aspects of the

organization of their native language. Our study is thus indirectly the result of some of the very inspiring and challenging questions Jacques has asked throughout his long career. More directly, most of the authors of this paper are ‘kids’ or ‘grandkids’ of Jacques, who has a very large academic genealogical tree (<https://neurotree.org/neurotree/tree.php?pid=9240>). Jacques’ passion for research got transmitted, and we are proud to continue pushing the field of cognitive science further.

This research was supported by a PhD fellowship from CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Brazil) [Grants 88881.188983/2018-01 and 88882.329599/2018-01] to Leticia Kolberg; by a Research Grant from the Fyssen Foundation to Alex de Carvalho [Grant Eotp RH03J20AFO04\_EDENFCOGN° H03R8240]; and by grants from the Fondation de France, and the French national research agency ANR [grant numbers: ANR-13-APPR-0012 LangLearn, ANR-17-CE28-0007-01 LangAge, and ANR-17-EURE-0017 FrontCog].

## 6 Appendix

### APPENDIX A - Detailed stimuli and design



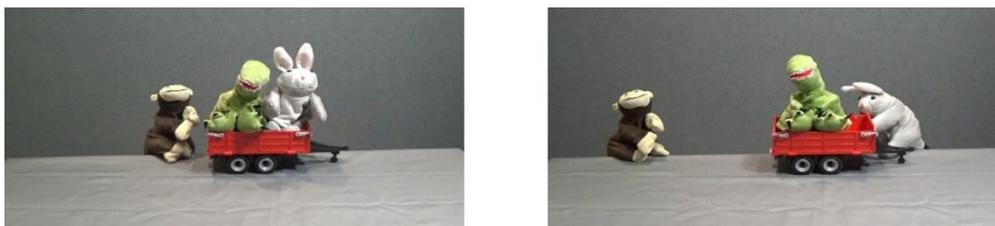
**Figure A.1.** "Carry" action. The one-agent video (left) showed the duck holding the dinosaur on top of one wing and a present on top of the other, while slowly swinging from side to side. The two-agent video (right) showed the duck and the dinosaur holding a present box each while slowly swinging from side to side.



**Figure A.2.** "Hit" action. The one-agent video (left) showed the tiger hitting the duck and the bunny alternately; the tiger hit the duck three times, then turned towards the bunny and hit it three times as well, repeating this until the end of the trial. The two-agent video (right) showed the tiger and the duck hitting the bunny at the same time; they hit three times, then paused for a second and hit another three times, repeating this pattern until the end of the trial.



**Figure A.3.** "Eat" action. The one-agent video (left) showed the tiger poking the dinosaur with a fork; he would poke the dinosaur five times, then stop and nod at the camera, then repeat this pattern until the end of the trial. The two-agent video (right) showed the tiger and the dinosaur poking the duck with a fork at the same time; they would poke it five times, then nod at the camera.



**Figure A.4.** "Push" action. The one-agent video (left) showed the monkey pushing the dinosaur and the bunny on a trolley from the left to the right of the screen. The two-agent video (right) showed the monkey and the bunny pushing the dinosaur on the trolley to each other alternately.

**Table A.1.** Description of test videos.

Action	Two-agent video	One-agent video
<b>Eat</b>	A tiger and a dinosaur repeatedly poke a duck (which is laying on a plate) with a fork	A tiger repeatedly pokes a dinosaur (which is laying on a plate beside a duck) with a fork
<b>Push</b>	A bunny and a monkey push a dinosaur that is sitting on a trolley back and forth to each other	A monkey pushes a bunny and a dinosaur that are sitting on a trolley from one side of the screen to the other
<b>Hit</b>	A tiger and a duck repeatedly hit a bunny with a stick at the same time	A tiger alternately hits a bunny and a duck with a stick
<b>Carry</b>	A duck and a dinosaur carry a present each while gently swinging from side to side	A duck carries a present over one wing and a dinosaur over the other while gently swinging from side to side

**Table A.2.** Video design. A single training phase (aimed at familiarizing the child with the idea that the lady would name one of the videos) was followed by four test phases (one for each of the actions above), which follow the same design as the one exemplified below. The order of the test phases, as well as the side in which each video appears is randomized.

Trial	Video stimuli		Audio stimuli	Time
	Left	Right		
Character introduction	Characters appear in the screen for 5s each		Characters were named once in sentences such as "Regarde ! C'est le tigre !"	30s
Interval	Laughing baby		Laughing baby	5s
Fixation circle	Fixation circle		no audio	User controlled (at least 1s)
Preview trial	Bunny jumping		" oh, regarde ! tu as vu ?"	5s
Interval	blank screen		no audio	1s
Preview trial		Bunny and monkey playing	" oh, regarde ! tu as vu ?"	5s
Fixation circle	Fixation circle		no audio	User controlled (at least 1s)
Contrast trial	Bunny jumping	Bunny and monkey playing	" oh, regarde ! tu vois ça ?"	5s
Audio prompt	Fixation circle		" Attention: le lapin va sauter !"	3s
Test trial	Bunny jumping	Bunny and monkey playing	" Oh, regarde ! Le lapin saute ! Tu vois ? Le lapin saute !"	12s
Pointing trial	Bunny jumping (paused)	Bunny and monkey playing (paused)	No audio	User controlled
Interval	Laughing baby		Laughing baby	5s

Young children can use prosodic information to interpret ellipsis

Trial	Video stimuli		Audio stimuli	Time
	Left	Right		
Character introduction	Characters appear in the screen for 5s each		Characters were named once in sentences such as "Regarde ! C'est le tigre !"	30s
Interval	Laughing baby		Laughing baby	5s
Fixation circle	Fixation circle		no audio	User controlled (at least 1s)
Fixation circle	Fixation circle		no audio	User controlled (at least 1s)
Intro test phase	Tiger eating duck with a fork		<i>"Regarde ! c'est le tigre ! Et le canard ! Oh, il le mange !"</i>	10s
Interval	blank screen		no audio	3s
Preview trial		Tiger and dinosaur eating duck	<i>"oh, regarde ! Tu vois ça ?"</i>	9s
Interval	blank screen		no audio	1s
Preview trial	Tiger eating dinosaur		<i>"oh, regarde ! Tu vois ça ?"</i>	9s
Fixation circle	Fixation circle		no audio	User controlled (at least 1s)
Contrast trial	Tiger eating dinosaur	Tiger and dinosaur eating duck	<i>" oh, et là ! Regarde ! Tu vois ça ?"</i>	9s
Audio prompt	Fixation circle		<i>"Attention: Le tigre va manger (.) le dinosaure aussi !"</i>	User controlled (at least 5s)
Test trial	Tiger eating dinosaur	Tiger and dinosaur eating duck	<i>" oh, regarde ! Le tigre mange (.) le dinosaure aussi ! Tu vois ? Le tigre mange (.) le dinosaure aussi !"</i>	12s
Pointing trial	Tiger eating	Tiger and dinosaur eating	no audio	User controlled

Test phase

Trial	Video stimuli		Audio stimuli	Time
	Left	Right		
Character introduction	Characters appear in the screen for 5s each		Characters were named once in sentences such as "Regarde ! C'est le tigre !"	30s
Interval	Laughing baby		Laughing baby	5s
Fixation circle	Fixation circle		no audio	User controlled (at least 1s)
	dinosaur (paused)	duck (paused)		
Interval	Laughing baby		Laughing baby	5s
Fixation circle	Fixation circle		no audio	User controlled (at least 1s)

## 7 References

- Abboub, N., Nazzi, T., & Gervain, J. (2016). Prosodic grouping at birth. *Brain and Language*, *162*, 46–59. <https://doi.org/10.1016/j.bandl.2016.08.002>
- Arunachalam, S., & Syrett, K. (2014). *Specifying event reference in verb learning* [Conference presentation]. 38th Boston University Conference on Language Development (BUCLD), Boston, MA, United States. <https://doi.org/10.7282/T3VQ34PV>
- Arunachalam, S., Syrett, K., & Chen, Y. X. (2016). Lexical Disambiguation in Verb Learning: Evidence from the Conjoined-Subject Intransitive Frame in English and Mandarin Chinese. *Frontiers in Psychology*, *7*, 1–14. <https://doi.org/10.3389/fpsyg.2016.00138>
- Ashby, W. J. (1988). The syntax, pragmatics, and sociolinguistics of left- and right-dislocations in French. *Lingua*, *75*(2–3), 203–229. [https://doi.org/10.1016/0024-3841\(88\)90032-0](https://doi.org/10.1016/0024-3841(88)90032-0)
- Bailey, T. M., & Plunkett, K. (1998). Learning prosody: Infants' learning of two word stress rhythms. *Infant Behavior and Development*, *21*, 172. [https://doi.org/10.1016/s0163-6383\(98\)91387-x](https://doi.org/10.1016/s0163-6383(98)91387-x)
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. <https://doi.org/10.1016/j.jml.2012.11.001>

- Barss, A. (Ed.). (2003). *Anaphora: A reference guide* (1st ed.). Orlando, FL: Wiley-Blackwell.
- Bernal, S., Lidz, J., Millotte, S., & Christophe, A. (2007). Syntax constrains the acquisition of verb meaning. *Language Learning and Development*, 3(4), 325–341. <https://doi.org/10.1080/15475440701542609>
- Choi, Y., & Mazuka, R. (2003). Young children's use of prosody in sentence parsing. *Journal of Psycholinguistic Research*, 32(2), 197–217. <https://doi.org/10.1023/a:1022400424874>
- Christophe, A., Dautriche, I., de Carvalho, A., & Brusini, P. (2016). Bootstrapping the syntactic bootstrapper. In J. Scott & D. Waughtal (Eds.), *Proceedings of the 40th Annual Boston University Conference on Language Development* (pp. 75–88). Somerville, MA: Cascadilla Press.
- Christophe, A., Gout, A., Peperkamp, S., & Morgan, J. (2003). Discovering words in the continuous speech stream: The role of prosody. *Journal of Phonetics*, 31(3–4), 585–598. [https://doi.org/10.1016/s0095-4470\(03\)00040-8](https://doi.org/10.1016/s0095-4470(03)00040-8)
- Christophe, A., Millotte, S., Bernal, S., & Lidz, J. (2008). Bootstrapping Lexical and Syntactic Acquisition. *Language and Speech*, 51(1–2), 61–75. <https://doi.org/10.1177/00238309080510010501>
- Cyrino, S. M. L., & Matos, G. (2002). VP ellipsis in European and Brazilian Portuguese: a comparative analysis. *Journal of Portuguese Linguistics*, 1(2), 177–195. <https://doi.org/10.5334/jpl.41>
- Dautriche, I., Cristia, A., Brusini, P., Yuan, S., Fisher, C., & Christophe, A. (2014). Toddlers Default to Canonical Surface-to-Meaning Mapping When Learning Verbs. *Child Development*, 85(3), 1168–1180. <https://doi.org/10.1111/cdev.12164>
- de Carvalho, A. (2017). *The role of phrasal prosody and function words in the acquisition of word meanings* [Unpublished MA dissertation]. Ecole Normale Supérieure, PSL Research University, Paris - France. Retrieved from <https://tel.archives-ouvertes.fr/tel-01943786>
- de Carvalho, A., Babineau, M., Trueswell, J. C., Waxman, S. R., & Christophe, A. (2019). Studying the Real-Time interpretation of novel noun and verb meanings in young children. *Frontiers in Psychology*, 10(274), 1–16. <https://doi.org/10.3389/fpsyg.2019.00274>
- de Carvalho, A., Dautriche, I., & Christophe, A. (2016a). Preschoolers use phrasal prosody online to constrain syntactic analysis. *Developmental Science*, 19(2), 235–250. <https://doi.org/10.1111/desc.12300>
- de Carvalho, A., Dautriche, I., Lin, I., & Christophe, A. (2017). Phrasal prosody constrains syntactic analysis in toddlers. *Cognition*, 163, 67–79. <https://doi.org/10.1016/j.cognition.2017.02.018>

- de Carvalho, A., Dautriche, I., Millotte, S., & Christophe, A. (2018). Early perception of phrasal prosody and its role in syntactic and lexical acquisition. In P. Prieto & N. Esteve-Gibert (Eds.), *The Development of Prosody in First Language Acquisition* (pp. 17–35). Amsterdam, Netherlands: John Benjamins Publishing Company. <https://doi.org/10.1075/tilar.23.02car>
- de Carvalho, A., He, A. X., Lidz, J., & Christophe, A. (2019). Prosody and function words cue the acquisition of word meanings in 18-Month-Old infants. *Psychological Science*, *30*(3), 319–332. <https://doi.org/10.1177/0956797618814131>
- de Carvalho, A., Lidz, J., Tieu, L., Bleam, T., & Christophe, A. (2016b). English-speaking preschoolers can use phrasal prosody for syntactic parsing. *Journal of the Acoustical Society of America*, *139*(6), EL216–EL222. <https://doi.org/10.1121/1.4954385>
- de Cat, C. (2007). *French Dislocation. Interpretation, Syntax, Acquisition*. Oxford, England: Oxford University Press.
- Delais-Roussarie, E. (1995). *Pour une approche parallèle de la structure prosodique: étude de l'organisation prosodique et rythmique de la phrase française* [Unpublished doctoral dissertation]. University of Toulouse 2. Retrieved from <https://www.theses.fr/1995TOU20097>
- Demuth, K., & Tremblay, A. (2008). Prosodically-conditioned variability in children's production of French determiners. *Journal of Child Language*, *35*(1), 99–127. <https://doi.org/10.1017/s0305000907008276>
- Dink, J. W., & Ferguson, B. (2016). eyetrackingR: An R Library for eye-tracking data analysis. Retrieved from <http://www.eyetrackingr.com>
- Féry, C. (2001). Focus and phrasing in French. In C. Féry & W. Sternefeld (Eds.), *Audiatur vox sapientiae. A festschrift for Arnim von Stechow* (pp. 153–181). Berlin, Germany: Walter de Gruyter.
- Fisher, C., & Tokura, H. (1996). Acoustic cues to grammatical structure in Infant-Directed speech: Cross-Linguistic evidence. *Child Development*, *67*(6), 3192. <https://doi.org/10.2307/1131774>
- Foley, C., Nuñez Del Prado, Z., Barbier, I., & Lust, B. (2003). Knowledge of variable binding in VP-ellipsis: Language acquisition research and theory converge. *Syntax*, *6*(1), 52–83. <https://doi.org/10.1111/1467-9612.00056>
- Gervain, J., & Werker, J. F. (2013). Prosody cues word order in 7-month-old bilingual infants. *Nature Communications*, *4*(1), 1490. <https://doi.org/10.1038/ncomms2430>
- Gout, A., Christophe, A., & Morgan, J. L. (2004). Phonological phrase boundaries constrain lexical access: II. Infant data. *Journal of Memory and Language*, *51*(4), 548–567. <https://doi.org/10.1016/j.jml.2004.07.002>

- Graf Estes, K., & Bowen, S. (2013). Learning about sounds contributes to learning about words: Effects of prosody and phonotactics on infant word learning. *Journal of Experimental Child Psychology*, *114*(3), 405–417. <https://doi.org/10.1016/j.jecp.2012.10.002>
- Gutman, A., Dautriche, I., Crabbé, B., & Christophe, A. (2015). Bootstrapping the syntactic bootstrapper: Probabilistic labeling of prosodic phrases. *Language Acquisition*, *22*(3), 285–309. <https://doi.org/10.1080/10489223.2014.971956>
- Hawthorne, K., & Gerken, L. A. (2014). From pauses to clauses: Prosody facilitates learning of syntactic constituency. *Cognition*, *133*(2), 420–428. <https://doi.org/10.1016/j.cognition.2014.07.013>
- Hawthorne, K., Mazuka, R., & Gerken, L. A. (2015). The acoustic salience of prosody trumps infants' acquired knowledge of language-specific prosodic patterns. *Journal of Memory and Language*, *82*, 105–117. <https://doi.org/10.1016/j.jml.2015.03.005>
- Hirose, Y., & Mazuka, R. (2017). Exploiting pitch accent information in compound processing: A comparison between adults and 6- to 7-year-old children. *Language Learning and Development*, *13*(4), 375–394. <https://doi.org/10.1080/15475441.2017.1292141>
- Hirsh-Pasek, K., Kemler Nelson, D. G., Jusczyk, P. W., Cassidy, K. W., Druss, B., & Kennedy, L. (1987). Clauses are perceptual units for young infants. *Cognition*, *26*(3), 269–286. [https://doi.org/10.1016/s0010-0277\(87\)80002-1](https://doi.org/10.1016/s0010-0277(87)80002-1)
- Höhle, B., Weissenborn, J., Schmitz, M., & Ischebeck, A. (2001). Discovering word order regularities: The role of prosodic information for early parameter setting. In J. Weissenborn & B. Höhle (Eds.), *Approaches to Bootstrapping: Phonological, Lexical, Syntactic and Neurophysiological Aspects of Early Language Acquisition* (pp. 246–263). Amsterdam, Netherlands: John Benjamins.
- Johnson, E. K. (2008). Infants use prosodically conditioned acoustic-phonetic cues to extract words from speech. *The Journal of the Acoustical Society of America*, *123*(6), 144–148. <https://doi.org/10.1121/1.2908407>
- Johnson, E. K., & Jusczyk, P. W. (2001). Word segmentation by 8-month-olds: When speech cues count more than statistics. *Journal of Memory and Language*, *44*(4), 548–567. <https://doi.org/10.1006/jmla.2000.2755>
- Johnson, E. K., & Seidl, A. (2008). Clause segmentation by 6-month-old infants: A crosslinguistic perspective. *Infancy*, *13*(5), 440–455. <https://doi.org/10.1080/15250000802329321>
- Jun, S.-A., & Fougeron, C. (2002). Realizations of accentual phrase in French intonation. *Probus*, *14*(1), 147–172. <https://doi.org/10.1515/prbs.2002.002>

- Jusczyk, P. W., Cutler, A., & Redanz, N. J. (1993). Infants' preference for the predominant stress patterns of English words. *Child Development*, *64*(3), 675–687. <https://doi.org/10.2307/1131210>
- Jusczyk, P. W., Hirsh-Pasek, K., Kemler Nelson, D. G., Kennedy, L. J., Woodward, A., & Piwoz, J. (1992). Perception of acoustic correlates of major phrasal units by young infants. *Cognitive Psychology*, *24*(2), 252–293. [https://doi.org/10.1016/0010-0285\(92\)90009-q](https://doi.org/10.1016/0010-0285(92)90009-q)
- Lehiste, I., Olive, J. P., & Streeter, L. A. (1976). Role of duration in disambiguating syntactically ambiguous sentences. *The Journal of the Acoustical Society of America*, *60*(5), 1199–1202. <https://doi.org/10.1121/1.381180>
- Lindenbergh, C., Van Hout, A., & Hollebrandse, B. (2015). Extending ellipsis research: The acquisition of sluicing in Dutch. In E. Grillo, K. Jepson, & M. LaMendola (Eds.), *BUCLD 39 Online Proceedings Supplement*. Boston, MA. Retrieved from <http://www.bu.edu/buclid/supplementvol39/>
- Mandel, D. R., Jusczyk, P. W., & Kemler Nelson, D. G. (1994). Does sentential prosody help infants organize and remember speech information? *Cognition*, *53*(2), 155–180. [https://doi.org/10.1016/0010-0277\(94\)90069-8](https://doi.org/10.1016/0010-0277(94)90069-8)
- Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190. <https://doi.org/10.1016/j.jneumeth.2007.03.024>
- Massicotte-Laforge, S., & Shi, R. (2015). The role of prosody in infants' early syntactic analysis and grammatical categorization. *The Journal of the Acoustical Society of America*, *138*(4), 441–446. <https://doi.org/10.1121/1.4934551>
- Massicotte-Laforge, S., & Shi, R. (2020). Is prosodic information alone sufficient for guiding early grammatical acquisition? *The Journal of the Acoustical Society of America*, *147*(3), 295–300. <https://doi.org/10.1121/10.0000887>
- Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition*, *29*(2), 143–178. [https://doi.org/10.1016/0010-0277\(88\)90035-2](https://doi.org/10.1016/0010-0277(88)90035-2)
- Millotte, S., René, A., Wales, R., & Christophe, A. (2008). Phonological phrase boundaries constrain the online syntactic analysis of spoken sentences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(4), 874–885. <https://doi.org/10.1037/0278-7393.34.4.874>
- Millotte, S., Wales, R., & Christophe, A. (2007). Phrasal prosody disambiguates syntax. *Language and Cognitive Processes*, *22*(6), 898–909. <https://doi.org/10.1080/01690960701205286>
- Morgan, J. L., & Demuth, K. (Eds.). (1996). *Signal to syntax: Bootstrapping from speech to grammar in early acquisition*. Mahwah, NJ: Lawrence Erlbaum Associates.

- Morgan, J. L., & Saffran, J. R. (1995). Emerging integration of sequential and suprasegmental information in preverbal speech segmentation. *Child Development, 66*(4), 911–936. <https://doi.org/10.2307/1131789>
- Nespor, M., & Vogel, I. (1986). *Prosodic phonology*. Dordrecht, Netherlands: Foris.
- Pinker, S. (1984). *Language learnability and language development*. Cambridge, MA: Harvard University Press.
- Postman, W., Foley, C., Santelmann, L., & Lust, B. (1997). Evidence for Strong Continuity: New experimental results from children's acquisition of VP-ellipsis and bound variable structures. *MIT Working Papers in Linguistics, 31*, 327–344. Retrieved from <http://mitwpl.mit.edu/>
- Ramachers, S., Brouwer, S., & Fikkert, P. (2017). How native prosody affects pitch processing during word learning in Limburgian and Dutch toddlers and adults. *Frontiers in Psychology, 8*, 1–20. <https://doi.org/10.3389/fpsyg.2017.01652>
- Rose, Y., & MacWhinney, B. (2014). The PhonBank Project: Data and software-assisted methods for the study of phonology and phonological development. In J. Durand, U. Gut, & G. Kristoffersen (Eds.), *The Oxford handbook of corpus phonology* (pp. 380–401). Oxford, UK: Oxford University Press.
- Ross, J. R. (1969). Guess who? In R. Binnick, A. Davidson, G. Green, & J. Morgan (Eds.), *Papers from the 5th Regional Meeting of the Chicago Linguistic Society* (pp. 252–286). Chicago, IL: Chicago Linguistic Society.
- Shattuck-Hufnagel, S., & Turk, A. E. (1996). A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research, 25*(2), 193–247. <https://doi.org/10.1007/bf01708572>
- Shukla, M., White, K. S., & Aslin, R. N. (2011). Prosody guides the rapid mapping of auditory word forms onto visual objects in 6-mo-old infants. *Proceedings of the National Academy of Sciences, 108*(15), 6038–6043. <https://doi.org/10.1073/pnas.1017617108>
- Snedeker, J., & Trueswell, J. (2001). *Unheeded cues: Prosody and syntactic ambiguity in mother-child communication* [Conference presentation]. Boston University Conference on Language Development (BUCLD), Boston, MA, United States.
- Snedeker, J., & Trueswell, J. (2003). Using prosody to avoid ambiguity: Effects of speaker awareness and referential context. *Journal of Memory and Language, 48*(1), 103–130. [https://doi.org/10.1016/s0749-596x\(02\)00519-3](https://doi.org/10.1016/s0749-596x(02)00519-3)
- Snedeker, J., & Yuan, S. (2008). Effects of prosodic and lexical constraints on parsing in young children (and adults). *Journal of Memory and Language, 58*(2), 574–608. <https://doi.org/10.1016/j.jml.2007.08.001>

- Soderstrom, M., Blossom, M., Foygel, R., & Morgan, J. L. (2008). Acoustical cues and grammatical units in speech to two preverbal infants. *Journal of Child Language*, 35(4), 869–902. <https://doi.org/10.1017/s0305000908008763>
- Soderstrom, M., Seidl, A., Kemler Nelson, D. G., & Jusczyk, P. W. (2003). The prosodic bootstrapping of phrases: Evidence from prelinguistic infants. *Journal of Memory and Language*, 49(2), 249–267. [https://doi.org/10.1016/s0749-596x\(03\)00024-x](https://doi.org/10.1016/s0749-596x(03)00024-x)
- Syrett, K., Arunachalam, S., & Waxman, S. R. (2014). Slowly but surely: Adverbs support verb learning in 2-year-olds. *Language Learning and Development*, 10(3), 263–278. <https://doi.org/10.1080/15475441.2013.840493>
- Syrett, K., & Lidz, J. (2010). 30-Month-olds use the distribution and meaning of adverbs to interpret novel adjectives. *Language Learning and Development*, 6(4), 258–282. <https://doi.org/10.1080/15475440903507905>
- Valleau, M. J., Konishi, H., Golinkoff, R. M., Hirsh-Pasek, K., & Arunachalam, S. (2018). An eye-tracking study of receptive verb knowledge in toddlers. *Journal of Speech, Language, and Hearing Research*, 61(12), 2917–2933. [https://doi.org/10.1044/2018\\_jslhr-1-17-0363](https://doi.org/10.1044/2018_jslhr-1-17-0363)
- van Ommen, S., Boll-Avetisyan, N., Larraza, S., Wellmann, C., Bijeljac-Babic, R., Höhle, B., & Nazzi, T. (2020). Language-specific prosodic acquisition: A comparison of phrase boundary perception by French- and German-learning infants. *Journal of Memory and Language*, 112, 1–16. <https://doi.org/10.1016/j.jml.2020.104108>
- Wellmann, C., Holzgrefe, J., Truckenbrodt, H., Wartenburger, I., & Höhle, B. (2012). How each prosodic boundary cue matters: Evidence from German infants. *Frontiers in Psychology*, 3, 1–13. <https://doi.org/10.3389/fpsyg.2012.00580>
- Wondershare. (2020). *Wondershare Filmora* [Computer software]. <https://filmora.wondershare.com/>
- Zhou, P., Crain, S., & Zhan, L. (2012). Sometimes children are as good as adults: The pragmatic use of prosody in children’s on-line sentence processing. *Journal of Memory and Language*, 67(1), 149–164. <https://doi.org/10.1016/j.jml.2012.03.005>
- Zhou, P., Su, Y. E., Crain, S., Gao, L., & Zhan, L. (2011). Children’s use of phonological information in ambiguity resolution: a view from Mandarin Chinese. *Journal of Child Language*, 39(4), 687–730. <https://doi.org/10.1017/s0305000911000249>