

Cross-linguistic evidence for the role of phrasal prosody in syntactic and lexical acquisition

Alex de Carvalho^{1*}, Leticia S Kolberg^{1*}, John Trueswell², Anne Christophe³

¹Laboratoire de Psychologie du Développement et de l'Éducation de l'Enfant (LaPsyDÉ),
Université de Paris, CNRS, F-75005 Paris, France

²Department of Psychology, University of Pennsylvania, Philadelphia, PA, USA

³Laboratoire de Sciences Cognitives et Psycholinguistique, DEC-ENS/EHESS/CNRS, Ecole
Normale Supérieure, PSL University, Paris, France

*joint first authorship

alex.de-carvalho@u-paris.fr, leticia.kolberg@gmail.com

Abstract

Phrasal prosody is believed to be crucial for syntactic and lexical acquisition, since prosodic boundaries in speech coincide with lexical and syntactic boundaries. However, studies on children's ability to use prosody for parsing have reported mixed results, with some providing positive evidence in French while many studies in other languages failed to observe this ability in children up to 6-years-old. To investigate whether the discrepancies between these studies are due to cross-linguistic or methodological differences, we adapted recent eye-tracking experiments showing that French children successfully use prosody to constrain parsing, to two different languages: English and Brazilian-Portuguese. In English, we used locally ambiguous sentences containing noun-verb homophones that could be disambiguated by prosody. In Brazilian-Portuguese, we tested preschoolers' ability to use prosody to interpret ellipsis (e.g., [The tiger is hitting!][The duck too!]) vs transitive sentences (e.g., [The tiger][is hitting the duck too!]). The results showed that 3-to-5-year-olds in the two languages successfully exploited prosodic structure to constrain their parsing. These studies provide cross-linguistic evidence for the role of phrasal prosody in parsing.

Index Terms: phrasal prosody; language acquisition; syntactic ambiguity resolution; parsing; eye movements; sentence processing.

1. Introduction

Because prosodic boundaries in speech correlate with lexical and syntactic boundaries [e.g., 1], the prosodic bootstrapping hypothesis claims that phrasal prosody is an important source of information for language acquisition, since it can assist young children in extracting word forms from the speech stream and in parsing sentences into syntactic constituents [e.g., 2-5]. Indeed, children have been shown to recognize the suprasegmental cues that signal prosodic phrase boundaries during their first year of life, and to reliably use this information for many different aspects of sentence parsing, such as word and sentence segmentation [e.g., 6-9].

However, studies investigating children's ability to exploit phrasal prosody to constrain sentence parsing have reported mixed results. While some studies provide positive evidence for this ability in young children (e.g., in French: [10-13]; and in English: [14-15]), other studies failed to observe this ability in

children up to 6-years-old (e.g., in English: [e.g., 16]; and in Korean: [17]).

Why do some studies fail to observe children's ability to use prosodic information for sentence parsing while others succeed? One possible explanation is that the discrepancies observed in these studies result from differences on how prosodic information signals syntactic boundaries across languages. English and Brazilian-Portuguese, for instance, are languages that have lexical stress and use prosody for topic and focus marking, whereas French, a language for which almost all previous studies have succeeded in showing children's use of prosodic information to constrain parsing, has no word-level stress and prosodic cues to mark topic and focus are rather optional and rarely used by adult speakers [e.g., 18]). So, during sentence processing, it is possible that prosodic cues could be more ambiguous to constrain parsing in young children learning English and Brazilian-Portuguese than for the ones learning French, where they are used mainly to cue phrasal prosodic structure. Learning how English and Brazilian-Portuguese use prosodic cues for different aspects of language (i.e., to signal prominence vs. the boundaries between prosodic units) could take extra developmental time compared to French, and so children learning these languages could take longer during sentence processing to identify which suprasegmental cues indicate phrase boundaries, whereas this could be easier and/or faster in languages such as French.

Alternatively, the discrepancy in results between French and other languages could be due to differences in the types of sentences tested and/or the methodologies used across studies. In a recent study testing American preschoolers' ability to use phrasal prosody to constrain parsing in English, [14] used the same type of locally ambiguous sentences containing noun/verb homophones than in their French study but with materials in English (e.g., [The baby *flies*] [hide in the shadows] vs [The baby] [*flies* his kite]). In an offline oral completion task, 4-to-5-year-olds successfully exploited the sentence's prosodic structure to assign the appropriate syntactic category to the ambiguous target word (e.g., interpreting 'flies' as a noun or a verb), mirroring previous results in French. This recent study may indicate that when presented with the same type of ambiguity and methodology that successfully showed French children's ability to use phrasal prosody to constrain parsing, English children perform as well as their French peers, suggesting that there is no difference in this ability for children learning English and French. However, since [14] did not test American preschoolers in an **online** task, it remains unclear

when the prosodic information was integrated during the parsing process in English and whether children learning languages that use prosodic information for many different linguistic aspects (e.g., lexical stress, prosodic prominence and phrasal boundaries, like in English and Brazilian-Portuguese) are as quick as their French peers in exploiting prosodic information to constrain parsing during real-time sentence processing. Since children were free to take as much time as they wanted to complete the test sentences in [14], the prosodic information could have been integrated relatively late during the parsing process in this task in English.

Here we investigated whether the discrepancies between previous studies are due to cross-linguistic or methodological differences, and whether children across different languages can integrate prosodic information in real-time. To do this, we adapted two recent eye-tracking experiments that succeeded in showing children's ability to use prosody in real-time to constrain parsing in French to two different languages: English and Brazilian-Portuguese. If American and Brazilian children tested on the same kinds of structures as in French (e.g., using noun-verb homophones in Experiment 1 with American children, and transitive vs. ellipsis sentences in Experiment 2 with Brazilian children) fail to disambiguate syntactic structures using prosodic information in real-time, we would conclude that the discrepancy in previous results might in fact be due to cross-linguistic differences. The transparency with which prosodic information reflects prosodic-syntactic relations may vary from one language to the next, and in languages in which this relation is less transparent (i.e., because phrasal prosody is used to cue different aspects of language) children may have more difficulty (or take more time) to use prosodic information to constrain real-time syntactic analysis. However, if American and Brazilian preschoolers, like their French peers, exploit phrasal prosody in real-time to constrain their interpretations of sentences, this will add cross-linguistic evidence to the role of phrasal prosody in constraining online syntactic analysis in young children.

2. Materials and Methods

2.1. Experiment 1: Testing preschoolers' ability to use prosody online to constrain parsing in English

Here we tested whether American preschoolers can exploit prosodic information in real-time to constrain their syntactic analysis. To do this, we used the same preferential looking task designed by [11] in French.

2.1.1. Participants

Twenty 3-to-5-year-old monolingual English-speaking children (*Age* = 52.8 months, range 41.1 to 61.5 months) were tested in a preschool in the Philadelphia area or in the Language Learning Lab at the University of Pennsylvania. Parents signed an informed consent form.

2.1.2. Materials

Six pairs of English noun-verb homophones (e.g., *flies*, *hand*, *watch*, *ring*, *duck*, and *clothes/close*) were used to create six pairs of experimental sentences and six pairs of images illustrating each of the homophone meanings (e.g., for the item 'flies', one picture showed a baby flying and the other a group of small flies around a big fly). Each pair consisted of a sentence in which the ambiguous word was used as a noun (hereafter the noun prosody condition, e.g., [The baby flies] [hide in the

shadows]) and a sentence in which the ambiguous word was used as a verb (hereafter the verb prosody condition, e.g., [The baby] [flies high up to the sky]). Although all these test sentences started with the same three words (e.g., the-baby-flies), they could be disambiguated by their different prosodic structures, reflecting their different syntactic structures. Utterances in the noun prosody condition contained a phonological phrase boundary after the target word (e.g., [the baby flies] ...), because in this case all the three words belonged to a single prosodic unit corresponding to the subject noun phrase. Utterances in the verb prosody condition had a phrase boundary before the target word (e.g., [The baby] [flies...]), corresponding to the boundary between the subject noun phrase and the verb phrase. A female native English speaker recorded the sentences in child-directed speech. In addition to the target sentences, six filler sentences and their respective images were created.

In order to make sure that prosodic information was the only cue that could be used to disambiguate the sentence beginnings in the test sentences, each sentence was cut off at the offset of the target word and its end was replaced by 1200ms of babble noise, which was obtained by superimposing the ends of all the filler sentences. Two counterbalanced lists of stimuli were used, each containing three noun targets, three verb targets, and six unambiguous fillers (three nouns and three verbs). The order of sentences within each list was randomized, with the constraint that there were no more than three target sentences in a row and no more than two consecutive test items from the same syntactic category.

2.1.3. Procedure

Participants were tested individually in the lab or in a quiet room in their preschools. They sat in front of a 23-in computer screen and wore headphones to listen to the audio stimuli. Participants' eye movements were recorded by an eye-tracker (Tobii TX300) placed below the screen and operating in a remote mode with a time-sample collected every 8ms. Toddlers listened to the beginnings of our ambiguous sentences while watching two images displayed side-by-side on a TV screen. For instance, when they heard the beginning of the sentence "the baby flies..." in either the noun prosody or in the verb prosody condition, they saw two pictures on the screen: one associated with the noun interpretation of the ambiguous target word in one side of the screen (e.g., some little flies) and another picture with the verb interpretation, on the other side (e.g., a baby flying). Their looking behavior was measured with the eye-tracker. If American children can rely on the prosodic structure of these sentences when conducting their syntactic computations in real-time, we expect them to look more often toward the noun picture when listening to sentences in the noun prosody condition than to sentences in the verb prosody condition.

2.1.4. Data analysis

The proportion of fixations toward the noun image was used as the dependent variable, because fixations to noun vs. verb image in this task are complementary (apart from the time spent looking away from the screen, which is not significantly different between conditions). To find the time-window(s) which exhibited a significant difference between conditions, a cluster-based permutation analysis was conducted (as in [11]; see [19], for a formal presentation of this analysis).

2.1.5. Results and discussion

Figure 1 shows the average proportion of looks toward the noun image in the noun prosody condition (red curve) and in the verb prosody condition (blue curve), time-locked to the beginning of the ambiguous word onset during the test trials. This reflects preschoolers' online interpretation of sentences as the linguistic input unfolds. Because only looks to the verb and noun images contributed to these proportions, below 0.50 indicates that participants preferred to look at the verb image (e.g., the baby flying) whereas above 0.50 indicates that participants preferred to look more to the noun image (e.g., the small flies). Participants initially looked more towards the verb image (which always contained humans). However, immediately after the end of the babble noise in the sentences (which happened on average around 1600ms after the onset of the ambiguous word), in the noun prosody condition, they switched their gaze toward the noun image more than participants in the verb prosody condition. A nonparametric cluster-based permutation test [19] revealed a time window (dark grey window in Fig1) containing a significant difference between the noun prosody and the verb prosody conditions ($p=.011$) starting about 1935ms after the ambiguous word onset and lasting until the end of the trials.

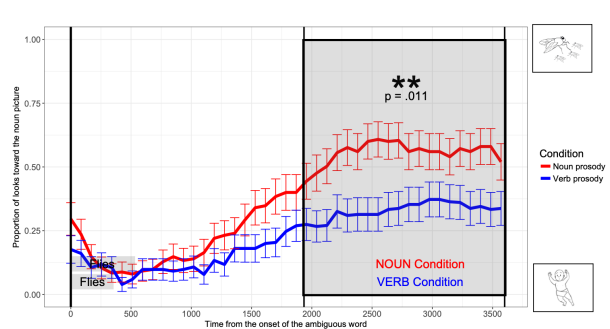


Figure 1: Results of Experiment 1.

These results show that English-speaking preschoolers, like their French peers, can exploit phrasal prosody in real-time during sentence processing to guide their syntactic interpretations, which in turn allowed them to identify the syntactic category of an ambiguous word (i.e., as a noun or a verb) in our task and access its meaning. However, we note that in the current experiment, the effect of prosody appeared later than it was observed in French (the significant difference between prosodic conditions in the eye-tracking data of [11] with 3-to-4-year-olds started at 226ms after the onset of the critical ambiguous word). This lends support to the idea that children learning English might take longer to integrate suprasegmental cues indicating phrase boundaries in their parsing decisions than children learning French. However, more studies are needed to directly compare the speed of prosodic processing to constrain parsing in French and other languages. Since our study in English used different target items and images, the differences in time might not be reflecting real differences in prosodic integration, but rather be related to differences due to methodological aspects of the task (e.g., maybe some ambiguous items were easier in French than in English). In Experiment 2, we further investigate children's ability to use prosodic information in real-time sentence processing by comparing French and Brazilian-speaking preschoolers in the same task, using the same video stimuli and ambiguous items.

2.2. Experiment 2: Testing preschoolers' ability to use prosody online to constrain parsing in Brazilian Portuguese

In this experiment, we used the same preferential looking task designed by [13], which recently showed that French 3-to-4-year-olds can use prosodic information to disambiguate ellipsis sentences (e.g., [*Le tigre tape*][*Le canard aussi*], “[The tiger is hitting][the duck too]!”), from transitive sentences (e.g., [*Le tigre*] [*tape le canard aussi*]!, “[The tiger] [is hitting the duck too]!”). While watching two videos side-by-side: one depicting the transitive interpretation of the sentences (e.g., a tiger hitting a duck and someone else), and another depicting the ellipsis interpretation (e.g., a tiger and a duck both hitting someone else), participants successfully exploited the prosodic information of each type of sentence to direct their attention to the video that best matched the sentence heard. In the present study, we tested Brazilian children with the exact same test sentences (translated to Brazilian-Portuguese) and the same test videos used in the French experiment. This allows us to compare the French and Brazilian results more closely, to investigate whether children learning these languages would present a similar speed of prosodic processing when presented with identical visual stimuli and similar sentences/ambiguities.

2.2.1. Participants

Fifty 3-to-4-year-old monolingual Brazilian-Portuguese-speaking children were tested in two preschools in the city of Campinas in Brazil. Twenty-five children were assigned to the transitive condition (*Age* = 41.3 months, ranging from 36 to 49.8 months), and 25 to the ellipsis condition (*Age* = 41.8 months, ranging from 34.6 to 50.2 months). Parents signed an informed consent form.

2.2.2. Materials

We presented children with four pairs of test videos showing animal puppets performing familiar actions (see full description of the videos and sentences in [13]). The actions used in the videos were *empurrar* (‘to push’); *cutucar* (‘to poke’); *comer* (‘to eat’); and *carregar* (‘to carry’). The videos and verbs were identical in the French and Brazilian experiment, except for the verb ‘to poke’, which was originally ‘to hit’ in the French experiment. The replacement was due to the fact that the verb ‘to hit’ would require a preposition in Brazilian-Portuguese. The sentences used in the experiment were recorded by a female native speaker using child-directed speech. The ellipsis sentences presented a phrasal boundary between the verb and the second noun (e.g., [*O tigre tá cutucando*][*O pato também*]! “[The tiger is poking][The duck too]!”), indicating that the second noun was not the object of the preceding verb phrase but rather two different sentences. The transitive sentences presented a phrasal boundary between the first noun and the verb phrase (e.g., [*O tigre*] [*tá cutucando o pato também*]!). For each pair of test videos, one video corresponded to the interpretation of the transitive sentences (e.g., one-agent video – a tiger poking a duck and a bunny with a stick alternately, for the ‘poke’ action), and the other video corresponded to the interpretation of the ellipsis sentences (e.g., two-agent video – a tiger and a duck both poking a bunny with a stick).

2.2.3. Procedure

Participants were tested in a quiet room in their own preschool. They sat in front of a 21.5-inch screen and listened to the test sentences through headphones. Participants' eye movements

were recorded by a Gazepoint eye-tracker placed below the screen and operating in a remote mode with a time-sample collected every 16ms in average. Children listened to four test sentences in either the ellipsis or the transitive condition, and at the same time, they watched the two videos showing each of two possible interpretations for each sentence side-by-side on the screen. The order of presentation was randomized and the side of correct videos in each trial (right vs left) was counterbalanced across participants.

2.2.4. Data analysis

Eye-tracking data was analyzed through a cluster-based permutation analysis as in Experiment 1, using the proportion of looks towards the two-agent action (i.e., ellipsis video) as the dependent variable. If Brazilian children can rely on the prosodic structure of our ellipsis vs. transitive test sentences when conducting their syntactic computations in real-time, we expect them to look more often towards the two-agent video when listening to sentences in the ellipsis condition than to sentences in the transitive condition.

2.2.5. Results and discussion

Figure 2 shows the time course of the proportion of looks to the two-agent video for participants in the ellipsis condition (red line) and in the transitive condition (blue line) through the entire test trial (12s). The cluster-based analysis found two clusters (gray rectangles) with a significant difference between conditions: one at the end of the first appearance of the test sentences during the trials (from 5270ms to 7310ms, $p=.046$) and another at the end of the second appearance of the test sentences (from 10540ms to 12000ms, $p=.03$).

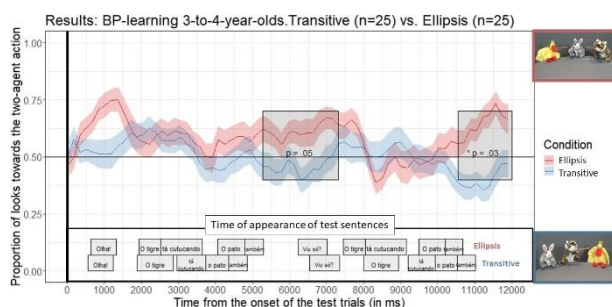


Figure 2: Results of Experiment 2.

These results show that Brazilian children, like their French peers in [13], can exploit phrasal prosody in real-time to guide their syntactic interpretations. However, we again notice a difference in the speed of integration of prosodic cues to guide parsing between the current results in Brazilian-Portuguese and the French results in [13]. The effect of prosody in French appeared earlier (first cluster found between 1920ms and 3820ms, and the second one between 3980ms and 8080ms, in the test trials of exact same duration, with the same video stimuli and similar time of appearance of word onsets in the test sentences). However, we acknowledge that one possible explanation for the difference between the Brazilian and French results might be due to the different eye-trackers used in the two studies. [13] used an Eyelink 1000, which has better spatial resolution and better temporal resolution (each time point was collected every 2ms) than the Gazepoint eye-tracker used in Brazil (each time point was collected every 16ms in average). The lower precision of Gazepoint could be at least partially

responsible for creating this apparent delay in speed of processing in the results of Brazilian children. Although these results point to a possible cross-linguistic difference in the ability to use phrasal prosody to guide real-time sentence processing, more studies are needed to support this finding, comparing children's performance in the same experiment, and using the same equipment, items, and visual stimuli in the two languages to be tested. This is certainly not an easy task to design.

3. Discussion

The results outlined in this paper show that 3-to-5-year-olds learning English and Brazilian-Portuguese successfully exploited the sentences' prosodic structures in real-time to constrain their syntactic analysis. In Experiment 1, children exploited prosodic information to constrain their interpretation of locally ambiguous sentences containing noun-verb homophones in English that could only be disambiguated by prosody. In Experiment 2, Brazilian-Portuguese-speaking children exploited the prosodic structure of sentences to disambiguate between ellipsis and transitive sentences. These findings mirror previous results observed in French, and challenge the failures observed in previous studies in other languages than French. We therefore provide cross-linguistic evidence for the role of phrasal prosody in lexical and syntactic acquisition in young children.

However, when comparing the present results to the French results, we also found that American and Brazilian children seem to take longer to integrate prosodic boundary cues in their parsing decisions than their French peers, since the significant differences between conditions were found later during the test trials when compared to the French results. As mentioned before, this could be due to methodological differences between studies, such as the different ambiguities in English and French, and the different eye-trackers used in each study in each country. These differences also prevent us from doing a statistical comparison of the results between language groups. Therefore, although our results suggest a possible difference in performance between French, American and Brazilian children, indicating that the two latter take longer to process prosodic information in real-time than their French peers, more studies are needed to corroborate this prediction, minimizing the experimental differences as much as possible by using the same stimuli, procedure, and equipment across studies.

4. References

- [1] M. Nespors and I. Vogel, *Prosodic phonology*. Dordrecht: Foris, 1986.
- [2] A. Christophe, S. Millotte, S. Bernal, and J. Lidz, "Bootstrapping lexical and syntactic acquisition.," *Lang. Speech*, vol. 51, no. 1–2, pp. 61–75, 2008.
- [3] A. Christophe, I. Dautriche, A. de Carvalho, and P. Brusini, "Bootstrapping the syntactic bootstrapper.," in *Proceedings of the 40th Annual Boston University Conference on Language Development*, 2016, pp. 75–88.
- [4] J. Mehler, P. Jusczyk, G. Lambertz, N. Halsted, J. Bertoncini, and C. Amiel-Tison, "A precursor of language acquisition in young infants," *Cognition*, vol. 29, no. 2, pp. 143–178, Jul. 1988.
- [5] J. Morgan and K. Demuth, *Signal to syntax: Bootstrapping from speech to grammar in early acquisition*, Erlbaum. Mahwah, NJ, 1996.
- [6] D. R. Mandel, P. W. Jusczyk, and D. G. Nelson, "Does sentential prosody help infants organize and remember speech information?," *Cognition*, vol. 53, no. 2, pp. 155–80, Nov. 1994.

- [7] S. Ramachers, S. Brouwer, and P. Fikkert, "How native prosody affects pitch processing during word learning in Limburgian and Dutch toddlers and adults," *Front. Psychol.*, vol. 8, no. SEP, pp. 1–20, 2017.
- [8] M. Shukla, K. S. White, and R. N. Aslin, "Prosody guides the rapid mapping of auditory word forms onto visual objects in 6-month-old infants," *Proc. Natl. Acad. Sci.*, vol. 108, no. 15, pp. 6038–6043, Apr. 2011.
- [9] M. Soderstrom, A. Seidl, D. G. Kemler Nelson, and P. W. Jusczyk, "The prosodic bootstrapping of phrases: Evidence from prelinguistic infants," *J. Mem. Lang.*, vol. 49, no. 2, pp. 249–267, Aug. 2003.
- [10] I. Dautriche, A. Cristia, P. Brusini, S. Yuan, C. Fisher, and A. Christophe, "Toddlers Default to Canonical Surface-to-Meaning Mapping When Learning Verbs," *Child Dev.*, vol. 85, no. 3, pp. 1168–1180, May 2014.
- [11] A. de Carvalho, I. Dautriche, and A. Christophe, "Preschoolers use phrasal prosody online to constrain syntactic analysis," *Dev. Sci.*, vol. 19, no. 2, 2016.
- [12] A. de Carvalho, I. Dautriche, I. Lin, and A. Christophe, "Phrasal prosody constrains syntactic analysis in toddlers," *Cognition*, vol. 163, 2017.
- [13] L. Kolberg *et al.*, "'The tiger is hitting! the duck too!' 3-year-olds can use prosodic information to constrain their interpretation of ellipsis," *Cognition*, no. January, 2021.
- [14] A. de Carvalho, J. Lidz, L. Tieu, T. Bleam, and A. Christophe, "English-speaking preschoolers can use phrasal prosody for syntactic parsing," *J. Acoust. Soc. Am.*, vol. 139, no. 6, pp. EL216–EL222, Jun. 2016.
- [15] J. Snedeker and S. Yuan, "Effects of prosodic and lexical constraints on parsing in young children (and adults)*," *J. Mem. Lang.*, vol. 58, pp. 574–608, 2008.
- [16] J. Snedeker and J. Trueswell, "Unheeded cues: Prosody and syntactic ambiguity in mother-child communication," in *Paper presented at the 26th Boston University Conference on Language Development*, 2001.
- [17] Y. Choi and R. Mazuka, "Young children's use of prosody in sentence parsing," *J. Psycholinguist. Res.*, vol. 32, no. 2, pp. 197–217, 2003.
- [18] C. Féry, "Focus and Phrasing in French," *Audiatur Vox Sapientiae. A Festschrift Arnim von Stechow*, no. October 1998, pp. 153–181, 2001.
- [19] E. Maris and R. Oostenveld, "Nonparametric statistical testing of EEG- and MEG-data.," *J. Neurosci. Methods*, vol. 164, no. 1, pp. 177–90, Aug. 2007.