The role of prosody in toddlers’ interpretation of verbs’ argument structure

Contents

1 What we know about words 3
  1.1 The Syntactic Bootstrapping Hypothesis 3
  1.2 Children’s use of syntactic information in sentence interpretation 5
  1.3 The Structure Mapping account 7

2 Multiple cues to early verb representation 10
  2.1 The link between prosody and syntax 10
  2.2 Early detection of prosodic units 11
  2.3 Prosody, a tool for syntactic analysis 11
  2.4 Dealing with multiple sources of information 12

3 Motivation and framework of the present study 14
  3.1 An overview of the linguistic accounts of Right-dislocated Sentences 14
  3.2 Dislocated structure frequency in Child Directed Speech 17
  3.3 Analyzing children’s production of dislocated sentences 21

4 Experiment 1 - Interpreting novel-verb argument structure 24
  4.1 Method 27
  4.2 Results 31

5 General Discussion and Perspectives 33

References 40

Appendix A Adult Study 44

Appendix B Acoustic Analysis 46

Appendix C Experiment 2 - Interpreting right-dislocated sentences 50
Abstract

Previous studies show that children are guided by the number of nouns in the sentence to interpret differently a novel verb occurring in a transitive structure (two nouns) or in an intransitive structure (one noun). We asked whether in some circumstances prosody can guide the structural representation of sentences even when the number of nouns is not informative of the verb’s argument structure. One of such case is right-dislocated sentences in French such as “il VERB, le bébé” (“he VERBing, the baby”) where prosody provides an unambiguous cue to the verb intransitivity despite the presence of two noun phrases. In one experiment, we presented 28-month-olds with dialogues introducing a novel verb (’daser’) in one of three conditions: transitive, dislocated and intransitive. Toddlers were then asked to look for ’daser’ while watching two videos displayed side-by-side: a two-participant action and a one-participant action. Children in the dislocated condition behaved like children in the transitive condition: they looked longer at the two-participant video than did those in the intransitive condition. This suggests that although children may process right-dislocated sentences, they rely on simple structural information leading them to interpret the sentence transitively.

Introduction

During the first years of life, children acquire the words of their language with astonishing speed and efficiency. By their sixth birthday they know about 8,000 root words [Carey, 1978] which means that from 18 months of age, they learn an average of 5 words per day. The question is: how do they do that? Intuitively, this problem can be solved in a simple way:

If we will observe how children learn languages, we shall find that ... people ordinarily show them the thing whereof they would have them have the idea, and then repeat to them the name that stands for it, as ‘white,’ ‘sweet’, ‘milk’, ‘sugar’, ‘cat’, ‘dog’. (Locke, Book 3, IX, 9, 1690)

Does this really work? One way to think about infants acquiring language is to picture yourself as a second language learner, surrounded by people not talking your mother tongue. Say that you grasped the word ”dax” from their weird language. What might ”dax” possibly mean given the number of possible objects and living beings surrounding you? The chair? The color of the chair? Perhaps the dog passing by? Or maybe just animal? Or hairy? Each object and each action can be described at multiple levels and from different perspectives. In these conditions, how does the child figure out which word in the sentence fits one of the ways one could think about things and events? The multiplicity of the possible word-to-world mappings makes the solution proposed by Locke very unlikely [Gleitman, 1990].
While the environment is for sure one of the first sources of information available, learning from observation alone seems to present considerable ambiguity. Yet, the environment is not the only source of information accessible to the child and some other sources may help her to reduce the number of potential referents in the world. Social-attentional cues [Tomasello, 2000a], representational biases [Markman, 1990] and linguistic information, as well as the combination of these cues [Nappa et al, 2009] are known to help children focusing on the relevant aspect of an observed scene.

In particular, linguistic information has been shown to be a particularly helpful cue, especially in discovering the meaning of verbs and other predicate terms [Gillette, Gleitman, Gleitman & Lederer 1999]. This is not surprising since syntactic structure describes the relationships between words in a sentence. Recent findings show that children are able to learn the combinatorial structure of the verb (i.e, its number of arguments) and use this information later on to constrain the verb’s meaning given two possible actions [Naigles, 1990; Yuan & Fisher 2009].

These results show that learning the meaning of words is facilitated if children have access to the syntactic structure of sentences (Syntactic Bootstrapping hypothesis). Intuitively, the way listeners build the syntactic structure of sentences is by first identifying words and retrieving their syntactic category, then grouping them into syntactic constituents to finally construct a complete syntactic structure of the sentence. But, how can children have access to such a representation with an incomplete lexicon? In other terms, what other sources of information could they rely on to build syntactic representations of sentences?

This is the question I propose to address in this study. More specifically, I will investigate the role of prosody, a cue that has been proposed to contribute to building the syntactic structure of sentences [Morgan, 1986]. In some circumstances, prosody can guide the structural representation of sentences. One such case is that of right dislocations in French, like “il, VERB, le bébé” (“he is VERBing, the baby = the baby is VERBing) where prosody provides unambiguous cues to the fact that ‘the baby’ is the agent of the verb, and therefore that the verb is used intransitively. In this study, I propose to investigate whether children are able to use the specific prosody of right dislocations to constrain their interpretation of a novel verb.
1 What we know about words

1.1 The Syntactic Bootstrapping Hypothesis

Gillette et al (1999) simulated observational word learning in adults. They showed adults muted videos of mothers talking to their children and asked them to guess the word uttered by the mother whenever a beep was heard. Beeps replaced the most frequent nouns and verbs of the mother’s speech. Adult participants saw 6 video clips in a row for each word, so that they had an opportunity for cross-situational observation to guide the interpretation of each word. They were told whether they had to guess a noun or a verb. The result of the experiment is dramatic: when adults rely only on the observation of the situation (visual context), they are able to correctly identify 45% of the nouns presented but they only guessed the correct verbs 15% of the time. These results suggest that the success rate in this task is dependent on the syntactic category of the word. But why is it more difficult to recover verbs than nouns?

The difficulty with verbs (and other predicate terms) is that they are highly abstract. A single event can be described by multiple verbs depending on the perspective the viewer is taking on the event. A cat is chasing a mouse could also be described as a cat is running, but also as they are playing and so on. It is commonly assumed that one should be able to distinguish between these verbs through the observation of multiple situations in which the verbs’ phonological forms are uttered. However, even in a cross-situational setting such as the one proposed by Gillette et al., adults, who have a fully developed sense of inference, find it difficult to guess the correct verb given a set of situations. If verbs cannot be extracted from observation alone, what are the possible sources of information available?

In a second experiment, Gillette et al. (1999) provided adults with linguistic information. They presented participants with a written set of 6 sentences extracted from the video clips containing the target word (ex: ’call’) but replaced all the content words (nouns, verbs, adjectives) with non-words and preserved most of the function words (determiners, prepositions) as in Figure 1. In such a condition where the sentence structure was preserved, adults correctly guessed 50% of the target verbs even without having access to information from the visual scene.

---

1Interestingly, a follow-up study showed a similar result when the category, noun or verb, was not given to the participants [Snedeker & Gleitman, 2004].
Frames in which the 6 utterances of CALL occurred:

<table>
<thead>
<tr>
<th>What does GORP mean?</th>
<th>Why don't ver GORP telfa?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GORP wastorn, GORP wastorn.</td>
</tr>
<tr>
<td></td>
<td>Ver gonna GORP wastorn?</td>
</tr>
<tr>
<td></td>
<td>Mek gonna GORP litch.</td>
</tr>
<tr>
<td></td>
<td>Can ver GORP litch on the fulgar?</td>
</tr>
<tr>
<td></td>
<td>GORP litch</td>
</tr>
</tbody>
</table>

**Figure 1:** Set of syntactic frames provided to adults in the Gillette et al. ’s study.

Thus, verb interpretation is guided by the structure in which they appear. Then, why does syntactic information guide semantic inferences so strongly?

This is a consequence of the systematic relationship between syntactic structure and meaning. Verbs appear in different syntactic structures (number and type of arguments) and these structures are consistently related to the semantic properties of the verb. For example, a self generated action such as ‘sleep’ will involve an intransitive structure with only one noun phrase (NP) (”Jane sleeps deeply.”). Actions involving the transfer of an object from one place to another such as ‘give’, may occur with three NPs (”Jane gives a cake to Marcella.”). Verbs describing mental states will typically appear with a clausal phrase complement (”Jane believes that Pete is wrong”) and so on. Moreover, many structures that appear with one type of verb are improbable for other types:

*Jane sleeps that Pete is wrong.

*Jane gives deeply.

*Jane believes a cake to Marcella.

Thus, verbs appear in specific syntactic structures because part of their meaning is encoded in the number and type of arguments involved. Indeed, experimental studies show that adults use the number and type of arguments required by verbs in order to predict distinctions in meaning among large classes of verbs [Fisher, 1994]. Hence it would be particularly useful for the learner to analyze the sentence structure in which a new verb appears, extract the argument-taking properties this new verb’s meaning involves and then use it to focus on the relevant aspect of the observed scene (Syntactic bootstrapping Hypothesis [Gleitman, 1990]). For example, if the learner hears ”the cat is running” while watching a scene of a cat chasing a mouse, then he can use the structure in which the verb appears (one NP) to focus on the running aspect of the scene which involves only one participant and eliminate the chasing aspect as it implies two participants.

However, even if the study of Gillette et al. shows that adults rely more on linguistic cues...
than on observation to find the target verbs, it does not follow that children would be able to
collect the same kind of strategy while learning novel verbs. The following section will present
some evidence showing that, indeed, children can use linguistic information to constrain the
meaning of verbs.

1.2 Children’s use of syntactic information in sentence interpreta-
tion

The plausibility of the syntactic bootstrapping hypothesis has been demonstrated by several
corpus studies of child directed speech where verbs have been shown to appear in distinctive
syntactic distributions depending on their semantic content [Lederer et al., 1995 for English].
Maternal speech thus provides sufficient syntactic information about common verbs for syntac-
tic bootstrapping to work.

Evidence that children do exploit this syntax-semantic relationship while interpreting sentences
has been found across numerous studies [Fisher, 1996; 2002; Naigles, 1990; Papafragou et al,
2007; Fisher et al. 2005]. I will review some of them below to support the syntactic boot-
strapping hypothesis and illustrate the kind of paradigm used to assess the comprehension of
pre-verbal children.

Causative sentences have been studied extensively from the perspective of syntactic bootstrap-
ping. A causative sentence includes a verb whose meaning incorporates the notion of causation;
for example ‘bring’ is a causative verb as it can be interpreted as ‘cause to come’. Quite a lot
of work has shown that transitive sentences are interpreted as causal and intransitive sentences
as non causal. For example, Naigles (1990) demonstrated that two-year-olds are able to asso-
ciate different meanings (causal ; non causal) with new verbs appearing in different syntactic
structures (transitive ; intransitive). Toddlers watched a video depicting a composite event: a
duck bending a bunny while both were moving their arm in a circle. Following this, the two
events were split into two videos displayed on different screens. Two-year-olds who heard the
novel verb in a transitive frame ("the duck is gorping the bunny") associated the new verb with
the causal action (the duck bending the bunny), whereas the children who heard ("the duck
and the bunny are gorping") looked at the other event (the duck and the bunny simultaneously
moving their arm in a circle).

In Papafragou et al. (2007), 4-years-olds were more likely to interpret novel verbs as referring
to a mental state when they occurred with a complement clause ("Matt DAXES that the
cookies are all gone") than when they appeared in a transitive sentence ("Matt DAXES a basket
of food!") especially while viewing salient scenes depicting mental state contents such as false

2Note that not all transitive verbs are linked with causality. "The dog heard the noise" does not mean that
the dog causes the noise to hear. However, a verb expressing causation cannot have less than two arguments,
thus in that sense, transitivity is a good predictor of causation.
beliefs situations. Children made use of the syntactic structure, in this case the occurrence of a verb in a sentence complement construction, to narrow down the range of possible interpretations to a set including mental verbs.

Children use the relationship between syntax and semantics to interpret not only novel verbs, but also other predicate terms such as prepositions. Fisher et al. (2005) showed that 2-year-olds can learn new prepositions from the sentence structure. While viewing a scene, children who heard (1) "This is a corp my box" interpreted 'acorp' as being the object's position relative to the noun 'my box', paying no attention the object's properties. In contrast, children who heard (2) "This is a corp" interpreted 'a corp' as being an object and ignored its location. Crucially, the sentences differed only in the additional noun 'my box' and informed the children that 'acorp' was a predicate in (1).

However, none of these experiments demonstrate that syntactic structure is informative by itself. Since sentences were always accompanied by scenes or events, it was perhaps these scenes that motivated the child to seek disambiguation in the structural information. If the syntactic structure is informative in itself, it should be possible for the child to gather the semantic structure of the verb (causative, transfer, etc.) from the linguistic contexts (number of arguments), store it in memory, then use it to retrieve the verb meaning from observational events experienced later on. The a priori knowledge of the verb's semantic structure should help the child to constrain the set of possible meanings a posteriori.

In a preferential looking paradigm task, Yuan and Fisher (2009) showed 28-month-olds a video of two women conversing and using a novel verb ('to blick') in a transitive frame ("Mary is blicking the cat") or in an intransitive frame ("Mary is blicking."). Crucially, there was no cue to the verb semantic content in the video. Then later on, toddlers watched two videos side-by-side: one with two participants and one with one participant. When asked to ("Find Blicking!") children who heard 'blick' in a transitive frame associated the new verb with a causal action involving two participants whereas those who heard the novel verb in an intransitive frame associated it with a one-participant action. In this study, children's interpretation of the novel verb was solely guided by the syntactic structure in which the verb appeared in the dialogue they heard.

All these studies demonstrate very convincingly that children between the age of 2 and 5 interpret differently novel predicate terms appearing in distinct sentence structures, even without visual information from a video scene.

However, how do children succeed in giving different interpretations to verbs based on the sentence structure they appear in? In the next section, we present a simple structural cue that

---

For discussion about why false belief contexts are facilitative for child word learning, see [p 8-9, Papafragou et al., 2007]

Children use the presence of the two NPs to infer that 'acorp' is an argument-taking predicate; however, this result does not tell us how they differentiate the syntactic category of these predicates (verb, preposition).
may allow children to build a partial semantic representation of the verb very early on: the number of nouns.

1.3 The Structure Mapping account

Since the syntactic structure specifies the relation between words in the sentence, it should be helpful to have access to the meaning of words in order to learn syntax. However, in the previous section, we just saw that syntactic information could be used to infer the meaning of new words and that children, indeed, can use it in this way. Hence, where do they start?

Fisher (1996) has argued that syntactic bootstrapping could originate from a bias to interpret each noun in the sentence as having a participant role in events, thus to be an argument of a predicate (Structure-mapping account of syntactic bootstrapping). Given this bias, the set of nouns in the sentence would be particularly meaningful to the child in constraining the kind of event described. In Fisher (2002), 28-month-olds were more likely to interpret a transitive sentence with two NPs such as "She is pilking her over there" as referring to a causal event with two participants than when they heard an intransitive sentence with one NP "She is pilking over there".

Interestingly, in Kannada\(^5\), transitive sentences are not the best predictor of causality. In this language, causality is expressed through a systematic verbal morpheme that can be attached to any verb. While transitivity does not always express causality in English or in French (see Note 2, p5), the causative morpheme is a very reliable cue to causative meaning in Kannada. Crucially, Lidz et al. (2000) showed that Kannada-learning 3-year-olds treated transitive structure but not morphology as a better indication of causality. On the contrary, Kannada adults were sensitive to the causative meaning of this morpheme, suggesting that this morpheme of Kannada is acquired later on by children.

These results provide further ground for the structure mapping account; the relation between the number of NPs in the sentence and the verb’s argument structure seems to be a recurrent property across languages that learners can use to bootstrap the meaning of unknown verbs.

Children’s bias to map the number of nouns in the sentence to the verb’s argument structure predicts two important results:

First, if there is such a bias, as soon as they recognize nouns, even toddlers with a limited lexicon would associate novel verbs with different representations depending on the number of nouns they co-occur with. In a series of experiments, Yuan et al. (in press) confirmed this prediction, since toddlers as young as 19 months used the number of co-occurring nouns to interpret novel verbs. A verb appearing in a transitive sentence featuring two NPs was interpreted more often as referring to a causal action between two actors than to a self-generated action performed by a single actor. Crucially the two videos presented during the test phase both dis-

\(^5\)A language spoken in South West India (Karnataka)
played two actors, but only the two actors in the causal action video had a participant-role in
the event. Thus, toddlers did not simply match the number of nouns to the number of actors in
the videos but they actually interpreted the verb as referring to a relational event with two core
participant-roles. This bias towards one-to-one mapping between nouns and participant-roles
may allow children to bootstrap their syntax from the few nouns they may have learnt from
observations.

Children’s one-to-one mapping bias between nouns and arguments of a predicate makes
a second important prediction. Obviously the number of nouns in a sentence does not al-
ways match the verb semantic representation. Drawing an interpretation of the verb by solely
counting the nouns can lead to predictable errors for toddlers when the number of NPs in the
sentence is not informative. For example, 19-month-olds failed to distinguish the conjoined
intransitive sentence (1) "the duck and the bunny are gorping") from the transitive sentence
(2) "the duck is gorping the bunny") [Hirsh-Pasek et al, 1996]. This failure may be due to the
difficulty of identifying the structure of complex sentences (as (1)) which require a more elab-
orated morpho-syntactic knowledge. To identify that (1) is an intransitive sentence despite
the presence of two NPs, children must have already learnt the meaning of the conjunction
’and’, that the morpheme ‘are’ is the marker of plural and finally that the canonical order of
causative sentences cannot be Noun1 Noun2 Verb in English. However this error of interpre-
tation vanished as soon as toddlers were able to combine these information in their sentence
representation. As we stated earlier, at 25-month-olds, toddlers succeeded in the same task
[Naigles, 1990].

More generally, a single utterance is often not sufficient to extract a verb’s argument struc-
ture. In many cases nouns can be dropped when they are obvious ("Jane is eating"), or there
 can be more nouns than minimally required by the semantic structure of the verb ("Jane saw
 a beaver on the pipple tree in the garden of the neighbor")6. In other words, a noun is not
necessarily an argument of the verb and there can be more participant-roles in the verb seman-
tic representation than the number of nouns in the sentence. However, as the study of Yuan
and Fisher (2009) suggests (see section 1.2), if children kept track of the verb’s distributional
properties across sentences, they would be able to come up with the most probable argument
structure for the verb.

To sum up, the structure-mapping account predicts that children are biased to interpret each
noun in the sentence as having a participant-role in the event described and this before acquir-
ing other morpho-syntactic cues. As a result, 19-month-old children successfully interpreted
novel verbs when the number of nouns in the sentence was informative [Yuan et al, in press]
but failed in complex sentences where the number of nouns was not informative of the verb
semantic representation [Hirsh-Pasek et al, 1996]. The number of nouns in the sentence seems
to provide a strong cue to uncover the verb semantic representation and this even if the link

---

6Note: we will not elaborate on the distinction between argument - or complement - of the verb and adjunct.
between the number of nouns and the number of arguments of the verb is not the most reliable cue in a given language [see for Kannada, Lidz et al, 2000].
2 Multiple cues to early verb representation

All the studies reviewed so far have shown that a distributional analysis of the number of NPs occurring with the verb is a reliable cue exploited by children to discover the verb semantic representation. However, a strict analysis based on the number of nouns in the sentence can mislead children about the verb meaning: any sentence with two NPs would be interpreted as transitive with a causative meaning even though this is not always the case. In this study, we explore whether prosody is taken into account to compute the argument structure of a novel verb in cases when the number of nouns is not informative.

2.1 The link between prosody and syntax

Acoustic characteristics of the speech input have been particularly studied in the field of language acquisition because of their potential to yield basic information about language structure. Prosodic properties of speech, such as rhythm and variations in pitch, are not randomly assigned parameters of the speech but help to structure discourse by grouping words into prosodic units. Prosodic units are organized hierarchically according to the depth of their acoustic markers (mainly pitch range, accent and tone [Hirschberg & Pierrehumbert, 1986]). Roughly, intonational phrases group words at the level of the sentence and can be broken down into smaller units: phonological phrases, which contain one or two content words along with their function words.

Crucially, phonological analyses revealed that these two kinds of prosodic boundaries correspond to boundaries of syntactic constituents such as clauses and phrases [Hirschberg & Pierrehumbert, 1986]. For example, the sentence "the little dog is eating a bone" contains two phonological phrases: "the little dog" which happens to be a noun phrase and "is eating a bone" which corresponds to a verb phrase. Phonological phrases are particularly interesting for syntactic analysis as they could provide the listener access to some of the syntactic constituents of the sentence. Crucially, adult listeners are sensitive to these units even when words are replaced by non-lexical units: they can make use of the prosodic characteristics of speech to find these groupings [de Pijper and Sanderman, 1994].

However it is worth noticing that if prosodic boundaries always correspond to syntactic boundaries, the reverse is not true. For example, the sentence "he is eating" is encapsulated in a single prosodic unit (brackets) whereas the pronoun is a clear syntactic constituent in the sentence ([he # is eating] where # symbolizes the syntactic constituent boundary). Thus, for all the sentences involving a subject pronoun, the major syntactic boundary separating the subject from the verb phrase does not coincide with a prosodic boundary. Hence, some syntactic boundaries will have no prosodic marker at all. But, conversely, all prosodic boundaries are systematically located at syntactic boundaries, which means that a prosodic analysis of the speech would provide the learner with reliable elements to guide his syntactic analysis of sentences.
2.2 Early detection of prosodic units

Prosody clearly plays a role in early speech perception. Critically, infants are sensitive to acoustic correlates of prosodic units and prefer to listen to sentences where artificial pauses are inserted at clause or phrasal boundaries rather than in the middle of a syntactic unit [Gerken et al, 1994; Etc]. Not only do they perceive prosodic unit boundaries, but it is also clear that toddlers use them to constrain lexical access. Gout, Christophe & Morgan (2004) showed that 13-month-old infants trained to turn their head for the word 'paper' did not recognize 'paper' if it spanned a phonological boundary as in [men with the highest pay][performs the most] whereas in the same sentence, they recognized the word 'pay'. Interestingly this result has been replicated with 16-month-olds French infants [Millotte, 2005] suggesting that phonological phrases boundaries can be used cross-linguistically to constrain lexical access online. So far, there is considerable evidence that children recognize prosodic units and use them to uncover several aspects of their language including word segmentation and lexical access. Because the prosody of a sentence depends on its syntactic structure as we saw previously, it has been proposed that children could rely on prosody to extract information about the syntax of their language [Christophe et al, 2003].

2.3 Prosody, a tool for syntactic analysis

While prosodic boundaries are often absent from short utterances, they are highly reliable when they are present and constitute a relevant predictor of syntactic structure. Millotte et al (2008) showed that adults are able to use phonological phrase boundaries to predict the syntactic category of a word. In "le petit chien mor[t/d]...", two interpretations are possible at that point of the sentence: the phonological form /mor/ could be interpreted as an adjective ("mort") or as a verb ("mordre"). Adults interpreted the ambiguous word as an adjective when it was followed by a phonological phrase boundary marking the noun phrase ([le petit chien mort] [...]) and as a verb when it occurred after the boundary ([le petit chien] [mord...]). Even though lexical information alone would lead to several syntactic parses, phrasal prosody allowed the listener to choose the correct interpretation online.

The effect of prosody has also been observed in resolving global ambiguity both in adults and 4-year-old children. In their study, Snedeker and Yuan (2008) presented 4-year-olds with ambiguous prepositional sentences such as "you can feel the frog with the feather" in a referential situation where both 'a frog with the feather' and 'a feather' are available. All prepositional sentences were headed by 'with' and could be interpreted as a modifier of the noun (the frog holding a feather) or an instrument (feel it using the feather). Crucially the verb did not bias the child to prefer any of these interpretations. However, children’s interpretation of sentences could be biased towards one or the other interpretation by manipulating the prosodic structure of the sentence. The insertion of an intonational boundary phrase after the verb favors the modifier interpretation ([you can feel] [the frog with the feather]) because 'the frog with the feather' is then grouped in a single prosodic unit ; while the insertion of an intonational phrase...
boundary after the noun phrase favors the instrument interpretation ([you can feel the frog] [with the feather]). Interestingly, not only does the prosody have an effect on the interpretation of sentences when the verb is neutral (no bias towards a modifier or an instrument interpretation) but also when the verb is semantically biased towards one or the other interpretation. For example, in a sentence such as "You can tap the frog with the feather", our semantic knowledge of 'tap' will favor an instrument interpretation. But if the prosody is manipulated as a modifier-like sentence ([You can tap] [the frog with the feather]), children (and adults) use the prosodic boundary to interpret the prepositional phrase as a modifier of the noun, overcoming the lexical bias of the verb.

Although the relation between phrasal prosody and syntax is not systematic, adults and children use prosodic information, along with lexical information, to parse sentences.

2.4 Dealing with multiple sources of information

Children have been shown to use multiple strategies while learning and processing words. As we just reviewed, sentence processing depends on referential, lexical and prosodic information. Thus it is expected that children are able to use jointly information coming from different sources. What kind of information is likely to be used by children in learning the various aspects of their language? Potential cues include information that is easily available, highly reliable and sufficiently frequent to be used.

While learning about a verb’s argument structure, sentential context, such as the number of nouns in the sentence, provides a simple structural cue that is highly reliable across sentences. If the probability of a given verb to appear with two noun phrases is high, then there is also a high probability that this verb is transitive. Thus it seems quite logical that the set of nouns in the sentence is one of the first cue used by children to constrain their early verb interpretation. However, there is ample evidence that children use multiple strategies while learning different aspects of their language. Regarding early verb representation, morpho-syntactic cues have been shown to overcome the initial structural bias. For instance, "the duck and the bunny are gorping" is not interpreted as transitive despite the presence of two NPs at 25 months of age [Naigles, 1990; Arunachalam and Waxman, 2010]. As we stated previously, the syntactic knowledge of the conjunction 'and', the plural 'are' and the word order (NP1 NP2 V) are available to the child at that age and thus, children included these new reliable sources of information in their syntactic representation of the sentences. However, this more sophisticated knowledge is a late-arriving cue and requires a complex processing.

In contrast, prosody is a source of information that is available very early on (at 9 months) from the speech signal itself without requiring any additional lexical processing. In addition, young children of 13 months exploit phrasal prosody to constrain lexical access. The role of phrasal prosody in syntactic analysis of sentences has been poorly investigated. However, since prosodic
boundaries always coincide with syntactic constituents boundaries, phrasal prosody may be a very reliable cue while analyzing sentences. Therefore, since phrasal prosody is universal, available early on and highly reliable across sentences, it makes sense to assert that phrasal prosody fulfills all the criteria to be a potential cue for the syntactic analysis of sentences.
3 Motivation and framework of the present study

Given what we have seen so far, there is ample evidence that the set of nouns in the sentence plays a substantial role in early verb learning. However, no study has tested the potential influence of prosody to constrain the meaning of verbs and this is precisely the question we will address in the following sections.

In some circumstances, prosody can guide the structural representation of sentences. One such case is the case of right-dislocated sentences in French:

(1) il_i mange le mouton_j

(2) il_i mange , le mouton_i

Sentence (1) means that someone, perhaps a wolf, is eating a sheep, while sentence (2) is stating that a sheep is eating. Crucially, the only information available to differentiate the meaning of these sentences is the specific prosody of sentence (2), with a perceived break between the verb phrase and the noun phrase ‘le mouton’ (re-transcribed here by a comma) and a special intonational contour on the dislocated noun phrase (see the acoustic analysis; Appendix B). This prosodic information is the only means by which the dislocated noun phrase can co-reference with the subject pronoun and thus, the only means by which listeners can infer that both NPs share the agent role assigned by the verb. So in this case, although the sentences are strictly the same regarding the lexical units and their order, the particular prosody of the dislocated sentence leads us to interpret it differently from the transitive sentence.

Using such sentences, we can test directly whether children use the prosodic information of right-dislocated sentences to constrain their interpretation of a novel verb and this, even when the number of nouns in the sentence is not informative of the verb’s semantic representation. In sentence (2), the prosody of the sentence suggests that the verb is employed intransitively since the two NPs correspond to the same participant-role. As we saw in section 1.2 and 1.3, children can distinguish between transitive and intransitive sentence structure. Thus we can use this result to test directly the effect of the prosody of right-dislocated sentences, on the semantic representations of verbs (transitive, intransitive) that children build.

However it is worth clarifying some linguistic properties of right-dislocated sentences before we start, and this is what the next section is about.

3.1 An overview of the linguistic accounts of Right-dislocated Sentences

In French, right-dislocated sentences (RD sentences) are typically associated with a resumptive pronoun (“il mange , le mouton”) or a clitic (“le loup l’a mangé , le mouton”) in the main clause. RD sentences are used almost exclusively in spoken language, thus, almost all
the studies on French dislocations use prosody as a diagnostic for these constructions [Rossi, 1999; De Cat, 2007; Avanzi (2010)]. Classically, the dislocated NP follows an intonational break and often has a low pitch. Regarding their interpretation, dislocated elements are widely acknowledged to be the topics\(^7\) of sentences. Right dislocations can also be used for clarification purposes, i.e. to make explicit a pronoun or a clitic that the speaker suddenly realizes that the audience may be unsure of. However this has been shown to be a secondary function of right dislocations [Reinhart, 1981].

We may ask now what the syntactic status of RD elements is. Evidence from word order supports the view that, in RD sentences, the dislocated phrase "doubled" by the clitic is an adjunct. However, a great body of the literature proposed RD sentences to be the result of a syntactic movement. According to Kaynes (1994), in RD sentences, the dislocated element doubled by the clitic is base generated in an argument position and moves to the left at Logical Form (LF) thus undergoing a \textit{covert} movement. For others [Cecchetto, 1999; Villalba, 2009], RD element is situated in the Topic Phrase (TopicP) just above the Verbal Phrase (VP) but appears to be at the right because the VP moves overtly higher in the tree. In another account of movement [Frascarelli, 2004], RD elements are in the specifier of a TopicP in the left periphery of the clause and become RD while the remnant of the clause (TP) is raised to the specifier of a phrase dominating the TopicP. Finally De Cat (2007) proposes that RD is not the result of a syntactic movement but is a right adjunction of an XP (see Figure 2 for a visual account of these 4 theories).

A comprehensive survey of all accounts of right dislocations is beyond the scope of the present work. However, this quick overview emphasizes the fact that right dislocations are a grammatical device and not a mere effect of sentence reparation.

More interesting for our purpose, these studies try to formalize the syntactic phenomenon of right dislocations cross-linguistically. Indeed, right dislocations are a common structure in Romance and Germanic languages: Italian, Spanish, Catalan, Dutch, French and to lesser extent English, but also Mandarin, Cantonese and some Bantu languages. Thus it is not a phenomenon restricted to French. However it is unclear whether all the languages use RD structures in the same way. For example, Cantonese right dislocation has been shown to be different from Romance right dislocation because Cantonese allows the dislocation of non-constituent elements [Law, 2003]. Similarly, while in French the resumptive pronoun seems obligatory in a RD sentence, this is not the case in Italian\(^8\) or in Cantonese. Hence, the interpretation of RD sentences may vary cross-linguistically.

Nethertheless, it seems that the prosody of right dislocations is similar across languages: an intonational phrase boundary is always found before the RD element [Dawning, 2010 for Bantu

\(^7\)\textit{Topics} here correspond to what the sentence is about relative to the discourse. See Reinhart (1981) for a more complete definition.

\(^8\)But see [Cardinalletti, 2002] for a distinction between right dislocations (with a compulsory clitic pronoun) and Marginalization (with an optional clitic pronoun).
To conclude, this section shows that, even though right dislocation is not an universal phenomenon, it is sufficiently represented in the world’s language to make its study interesting.

For toddlers to interpret correctly RD sentences, two important prerequisites should be met:

1. Toddlers should be able to detect prosodic patterns. In Section 2.2, we already reviewed some evidence from the literature showing that, indeed, toddlers are able to process phrasal prosody very early on.

2. Toddlers must get a sufficient number of dislocated sentences in the input to learn how to process this type of sentences. Importantly, this supposes that the input they receive is sufficiently non-ambiguous for them to learn the co-reference between the dislocated NP and the subject pronoun. This point will be addressed in Section 3.2, where we performed a corpus study of child-directed speech (CHILDES) and observed the high frequency and relative absence of ambiguity of dislocated sentences in parental speech. These analyses show that toddlers have access to all the necessary information to analyze and understand dislocated sentences. In addition, in Section 3.3, we looked at production data and found evidence that this structure is already known and used by children very early on.
3.2 Dislocated structure frequency in Child Directed Speech

For children to be able to interpret correctly dislocated sentences, they must have been exposed to a sufficient number of sentences exhibiting this construction. Dislocated sentences are very frequent in spoken French addressed to children, with as many as 32% of all utterances involving one or several dislocated elements [De Cat, 2007]. However, this number covers a wide range of dislocated elements including elements that are not regarded as the topic of the sentence such as locative or temporal adjuncts ("Maintenant, tu flottes"), infinitive clauses ("Tu ne penses qu'à ça, jouer à la poupée") or any clause modifiers. In these elements, there is no co-referential issue, thus they are not informative for children to understand who is the agent in sentences such as "il mange, le mouton".

To specifically study children’s access to sentences that may allow them to learn that the right-dislocated element is co-referential with a pronoun, we performed a more selective analysis of two corpora of child directed speech (CHILDES Marie and Tim, [Mac Whinney, 1985]). We extracted the frequency of informative dislocated sentences from the parental speech. Informative dislocated sentences for our experiment are sentences where the dislocated element is a noun phrase linked to the sentence nucleus via a clitic pronoun. The dislocated NP was either the subject (Subject Dislocation, SD) or the object (Object Dislocation, OD) and was located either at the beginning of the sentence (Left Dislocation, LD) or at the end of the sentence (Right Dislocation, RD). (see Table 1)

<table>
<thead>
<tr>
<th>Dislocation</th>
<th>Subject</th>
<th>Object</th>
</tr>
</thead>
<tbody>
<tr>
<td>Right</td>
<td>il mange la pâtée, le chien</td>
<td>le chien la mange, la pâtée</td>
</tr>
<tr>
<td></td>
<td>it is eating dog food, the dog</td>
<td>the dog is eating it, the dog food</td>
</tr>
<tr>
<td>Left</td>
<td>le chien, il mange la pâtée</td>
<td>la pâtée, le chien la mange</td>
</tr>
<tr>
<td></td>
<td>the dog, it is eating dog food</td>
<td>the dog food, the dog is eating it</td>
</tr>
</tbody>
</table>

Table 1: Possible NP dislocations in French

9I considered only utterances where the dislocated constituent is a full noun phrase and not a stressed pronoun as in "moi j’aime le chocolat" (Me, I like chocolate) or "j’aime le chocolat, moi" (I like chocolate, me) as there is no referential ambiguity to the subject pronoun in those. Moreover, the frequency of the "moi je" construction is so very high in spoken French (Culbertson (2010) evaluated that 70% of the sentences with a subject noun phrase in Child Directed Speech is subject to this "subject doubling") that I would rather think of it as a variant of the pronoun ‘je’, which would exclude it from the set of dislocated sentences.
To understand each type of dislocation, children have to learn that the dislocated element is coreferential with either the subject pronoun or the clitic object. Thus, all four types of dislocated sentences may contribute to the comprehension of co-reference. The number of dislocated items per type is summarized in Table 2:

<table>
<thead>
<tr>
<th>Dislocation</th>
<th>Objet</th>
<th>Subject</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left</td>
<td>118</td>
<td>615</td>
<td>733</td>
</tr>
<tr>
<td>Right</td>
<td>164</td>
<td>708</td>
<td>872</td>
</tr>
<tr>
<td></td>
<td>18%</td>
<td>82%</td>
<td>1605</td>
</tr>
</tbody>
</table>

**Table 2:** Frequency results for the four types of dislocation.

Taken together these four dislocation types represent 5% of all the sentences uttered by the parents of Marie and Tim through all the recorded sessions (Total number of sentences: 33491). This high representation of dislocated sentences in the parental speech should leave sufficient opportunity for children to learn to recognize and interpret these structures.

What strategies may children use to guide their identification of the correct referent? More specifically, how do they know that the dislocated NP refers to the subject or the object of the sentence?

In the first place, we can observe that the number of RD compared to the number of LD is balanced. However, interestingly, the majority of dislocated elements co-refer with the subject of the sentence (more than 80% of the dislocated sentences). Thus, given a dislocated sentence, there is a high probability that the dislocated NP co-refer with the subject pronoun and so, there is more chance that children interpret the dislocated NP as the subject of the sentence. In "il mange, le mouton", if children know that the prosody signals that the dislocated NP is co-referential with a pronoun, they would be more likely to interpret 'le mouton' as the agent, and not as an object, simply because they had more exposure to this kind of structure. Moreover, a simple cue to distinguish between the two possible referents (either subject or object) would be the presence (or the absence) of a clitic object. Indeed, the presence of a clitic object when the subject is realized as a full NP, is a cue that the dislocated NP must refer to

---

10 Total recorded time 92 hours from age 1 to 3 [Demuth and Tremblay, 2008]
11 After discarding utterances that cannot be potentially dislocated (such as one-word utterances and utterances without verbs).
the object of the verb. For example, in "le loup l’a mangé, le mouton", the only possible referent of 'le mouton' is the clitic 'l'.

However, this is not systematic when the subject of the sentence is a pronoun. For example, we can think about a situation where the object is the focus of previous sentences such as "Où est Paul ? - Il était dans sa chambre - Ah oui, mais il l’a renvoyé à l’école, Papa !" where we have a right dislocated element (‘Papa’) that co-refers to the subject pronoun because of the discourse context, but could also co-refer to the clitic object. However I did not find any of these situations in the corpus analyzed. So, despite the possible co-reference ambiguity of these sentences, I assume these situations to be not so frequent and thus to not weigh much in solving the co-referential issue. The same can be observed with more complicated sentences such as "il le mange, le serpent, le chien" where the co-reference is not straightforward due to the presence of two dislocated structures whose lexical information (number, gender) are similar and thus congruent with both the clitic object and the subject pronoun. Therefore the role occupied by each NP is ambiguous but, similarly to the previous example, such multiple dislocations are absent from the corpus analyzed, implying that children do not face these ambiguities frequently.

The distinction between SD and OD could potentially be derived from the presence of a clitic object. However, a sentence such as the one we propose to use "il mange, le mouton" (right dislocated subject) is still ambiguous with its transitive counterpart ("il mange le mouton") since without any previous knowledge of prosody, 'le mouton' could also be the verb’s object. How do children learn to use prosodic information to interpret the dislocated NP as the agent of the sentence and not as the object of the verb? Ideally, to assign the correct role to the dislocated NP, the child must be exposed to sentences such as "il VERB, le XXX" (he is VERBing, the XXX) in non-ambiguous contexts, to be able to compute that XXX is the agent of the sentence. Non-ambiguous contexts group all kinds of information (linguistic, visual, social) that the child may use to infer the agent of the action (this will be detailed later on).

"Il VERB, le XXX" could occur either if the verb is intransitive or if the verb is transitive but allows argument dropping such as in our example "il mange, le mouton". In order to investigate how often the child may receive this kind of structure and decide whether it is disambiguated by the context, we performed a detailed analysis on the set of sentences extracted from the two corpora (a total of 708 sentences).

First we looked at the type of verbs used and found that 87% (n= 622) of the sentences occur with être, avoir or faire with the great majority of sentences (46%) using an "être là" construction such as "il est là, le tambour". These three verbs always appear either with a complement ("Tu te rappelles qu’il fait du bruit, le cube ?") or with an adjective ("il est trop gros, Léon le bourdon.") thus they cannot appear in a structure such as "il VERB, le XXX". In total, only

---

12 Gender provides an excellent cue to pronoun resolution for young children [at 4-year-olds, Arnold et al, 2007] and could be available early on as [Martin, 1993] showed that 28-month-olds children are able to distinguish males from females. Thus, a sentence as "il la mange la souris, le chat" would be potentially less ambiguous as gender would constrain referent resolution in children.
5% of subject right-dislocated sentences have the form "il VERB, le XXX". (see Figure 3)

![Pie chart showing proportions of different sentence types](image)

**Figure 3:** Proportion of subject right-dislocated sentences as a function of the verb used in child-directed speech.

Given the small number of "il VERB, le XXX" sentences (n= 39), it was possible to analyze for each one whether the context could help the child to infer the agentive role of the dislocated NP. Different sources of information can help the child to solve the referential problem:

- **Knowledge of the verb meaning and its argumental properties**: If the verb is intransitive, the interpretation of the sentence is straightforward. In "elle tombe souvent, la tétine", 'la tétine' cannot be the object of the verb as the sentence would be ungrammatical ("elle tombe souvent la tétine" as it is not possible to 'fall a pacifier')

14 Note that it assumes that children acquired some knowledge of the argument structure of verbs and that they know in which sentence frames those verbs can or cannot appear in. However this assumption is on the line of the syntactic bootstrapping hypothesis (section 1).

15 It assumes also that children learn the semantic properties attached to the complement of a known verb. Recent study from [Yuan et al, 2011] show that 2-year-olds can learn about the semantic category of nouns serving as direct objects of a new verb from listening experience. Thus, its seems reasonable to postulate that children know what kind of complement can appear with a given verb and use this information to constrain the set of possible direct object.
• **Properties of the recent discourse.** For example in "Et qu’est ce qu’il sort du four ? Un gros gâteau. - T’as vu ? Il a gonflé, le gâteau", there is a focus on the agent of the verb 'le gâteau' that tends to make it mentally highly accessible for the resolution of the next dislocated sentences.\(^\text{16}\)

• **Access to a visual scene.** In CHILDES, where the parents are invited to play with the child, the utterances are often easily linked with the direct environment of the child. A sentence such as "il croque, le crocodile" uttered while looking at a picture book depicting a crocodile eating some animal may help the child to infer that 'le crocodile' is the agent of the sentence.\(^\text{17}\)

Finally, on the 39 "il VERB, le XXX" sentences, all fell within one (or more) of the above points and thus, can be considered as unambiguous. Even so, the restricted number of such constructions may not allow children to learn how to solve the referential problem of the dislocated NP. However, subject right-dislocated sentences using common verbs such as être, avoir or faire could also be useful to the child. Indeed, common verbs in RD sentences are highly frequent (2% of the total number sentences analyzed) and potentially non ambiguous as their object complement cannot be a NP ("il est beau, ce sanglier dis donc"). Thus, even if these sentences do not follow the pattern "il VERB, le XXX", they may be highly informative for the child to learn that the dislocated NP co-refers with the subject pronoun.

To sum up, dislocations are highly frequent in child-directed speech and may allow children to acquire the skill of reference resolution in the case of right-dislocated subject sentences. Even though children have very little exposure to sentences such as "il VERB , le XXX", their experience with dislocated structures using common verbs may help them to interpret the dislocated NP as the agent of the sentence. In contrast, in transitive sentences ("il VERB le XXX"), the absence of prosody signaling a dislocated element would lead children to interpret the final NP as the object of the verb.

### 3.3 Analyzing children’s production of dislocated sentences

An obvious start to investigate whether children interpret RD sentences correctly is to look at production data. To do so, we extracted children’s speech from three CHILDES corpora: Marie, Anaïs and Tim [Demuth and Tremblay, 2008] and analyzed their production from age 2 to 3. The first step is to examine whether there are any dislocated sentences and what kind of dislocation it is.

In the three corpora analyzed, we found 333 dislocated sentences across all children utterances.\(^\text{18}\)

\(^{16}\)Results from Song and Fisher (2005) suggest that children know that pronouns tend to refer to highly accessible entities.

\(^{17}\)We extracted the visual environment from the annotation in the CHILDES corpus or from its associated videos.

\(^{18}\)It is very difficult to compute the proportion of dislocation on the total number of sentences because the number of full sentences uttered varied considerably between 2 and 3 years old.
The proportion of sentences for each different structure (right or left dislocation; subject or object dislocation) is plotted in Figure 4.

![Figure 4: Proportion of Right vs Left and Subject vs Object dislocations in child-directed speech](image)

Following the adult data presented in Section 3.2, children have a tendency to produce more dislocated subjects than dislocated objects (94% of dislocated elements are subjects). This result supports our hypothesis that subject coreference is easier to learn than object coreference because of its frequency in the input.

Almost 70% of the dislocated sentences are right-dislocated in children’s speech whereas in adults’ speech, the number of right and left dislocations was balanced. However, this proportion of LD sentences may not reflect the reality because subject pronouns are easily dropped in children’s production. If a sentence such as "le chat, il veut du lait" was uttered as "chat, veut du lait", it would not be counted as a LD sentence in the absence of a resumptive pronoun. Analyzing the prosody of these sentences should permit to distinguish left dislocations from transitive sentences (see DeCat, 2007). Our frequency results are thus an underestimate of what children actually produce but it is interesting that dislocated structures can already be found in 2-year-old speech.

However, one must be careful when saying that children actually use dislocated sentences as adults do. In section 3.1, we described dislocation as a syntactic phenomena encoding the topic of the sentence, however we don’t know whether children already master the different operations that generate dislocation (see Section 3.1, Figure 2) and it is unclear whether they possess the pragmatic knowledge required to encode the dislocated NP as a topic. More importantly, although we inspected the local context of these sentences, it is quite difficult to exclude entirely the possibility that children’s production is not the mere copy of parental speech.

At any rate, both corpus analyses strongly suggest that children receive a sufficiently rich
input to learn the coreference implied by dislocated sentences, and that they use dislocated sentences productively as soon as they start uttering sentences (typically 2-3 years of age).
4 Experiment 1 - Interpreting novel-verb argument structure

So far, we saw that children have all the prerequisites to successfully interpret a right-dislocated sentence such as "il mange, le mouton". Children have been shown to detect prosodic markers very early in the time course of development (section 2.2) and to be exposed to a sufficient number of dislocations in non-ambiguous contexts to resolve the NP reference. Consequently, we propose to test whether this particular prosodic structure can guide toddlers to uncover the combinatorial properties of a novel verb independently of the number of NPs used in the sentence.

To do this, we used the preferential looking paradigm methodology developed by Yuan and Fisher (2009). Since 2-year-olds have a restricted productive skill, a comprehension paradigm allows a better assessment of their linguistic competence. The preferential looking method only requires the toddler to look at one of two simultaneous video events presented along with some recorded audio sentences. If the utterance is understood, then toddlers should focus on the scene consistent with that utterance. By analyzing the looking time to each video, we can assess the early syntactic knowledge of pre-verbal toddlers [Golinkoff et al. 1987].

Yuan and Fisher (2009) showed that linguistic information without any co-occurring event can induce toddlers to differentiate between transitive and intransitive verbs when the number of nouns is informative. Following their experimental design, we investigated whether 28-month-olds can assign different interpretations to novel transitive and intransitive verbs when the number of nouns in sentences is not informative, but the prosody is.

We presented toddlers with dialogues introducing a novel verb 'daser' in either one of three conditions (see Table 3 below for complete dialogues):

(1) transitive : il dase le bébé

(2) dislocated : il dase, le bébé

(3) intransitive : il dase

Then, toddlers watched two videos presented side by side: a two-participant event (one girl swinging another girl’s leg) and a one-participant event (a girl making circles with her arm) and asked them to look at the one that matched the action of the verb they just heard in the dialogue.

In order to check whether toddlers had an intrinsic preference for one or the other video of the test phase, we added a fourth condition: (4) control. In this condition, toddlers directly watched the pair of test videos with no prior dialogue exposure of the syntactic structure in which the novel occurs. Thus, they had no cue to infer the verb’s meaning from the dialogue’s
sentence structure but still underwent the same testing procedure: they had to look for the novel verb ‘daser’ while watching the one-participant and the two-participant videos. This condition gives children the same urging that children get in the three experimental conditions to choose one of the videos. It will show us what their baseline preference is.

<table>
<thead>
<tr>
<th>Condition (1) (Condition (2))</th>
<th>Condition (3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A: Hey il va daser (,) le bébé ! B: Ah bon, il va daser (,) le bébé ?</td>
<td>A: Hey il va daser ! B: Ah bon, il va daser ?</td>
</tr>
<tr>
<td>A: Oui, et en plus ils ont dasé (,) les canards B: C’est vrai, ils ont dasé (,) les canards</td>
<td>A: Oui, et en plus, ils ont dasé B: C’est vrai, ils ont dasé</td>
</tr>
</tbody>
</table>

Table 3: Dialogues used in the 3 testing conditions

Conditions (1) and (3) reproduce Experiment 1 of Yuan and Fisher’s study. Thus if children can learn the combinatorial structure of a verb without having access to its semantic content and can retrieve this structure while seeing the event, then the children in the transitive condition (1) should interpret the verb as describing an action relating two participants and thus look longer at the two-participant video, while the children in the intransitive condition (3) should interpret the verb as describing a self-generated action and thus look longer at the one-participant video19.

Conditions (1) and (2) directly test our hypothesis. Crucially, these two dialogues differ only in their prosody. As stated above, children should interpret the novel verb in (1) as describing an action involving two participants. For condition (2), if children do not interpret the prosodic boundary as a valid cue to compute the co-reference between the dislocated noun phrase and the subject pronoun, then they should probably rely on the number of nouns to learn the combinatorial structure of the new verb (see section 1.3) and thus interpret the novel verb as referring to an event relating two participants and look more at the two-participant video. (Scenario A - Figure 5)

On the contrary, if children use the prosodic boundary occurring before the dislocated NP to compute that ‘le bébé’ co-refers with the subject pronoun ‘il’ then they should consider the verb as occurring in an intransitive sentence as in (3). Thus, they should expect this verb

19 Note that “il dase” could also refer to the action of one participant in a two-participant event, as it is possible in many causal actions to drop the complement. For example, ‘chasser’ is a transitive verb but allows argument dropping such that “il chasse” is grammatical and can describe the same event as the fully specified transitive sentence “il chasse une souris”. However Yuan and Fisher showed that in an identical situation, toddlers looked longer at the one-participant event.
to describe a self-generated action and look longer at the one-participant video than at the 
two-participant video. (Scenario B - Figure 5)
A third possibility would be that toddlers consider the intonational boundary following the 
verb as an indicator of the end of the sentence and thus consider only the pre-boundary part 
of the sentence ("il dase") and drop the dislocated NP ("le bébé"). In that case they would 
process condition (2) as condition (3) and look more at the one-participant video (similarly 
see Scenario B - Figure 5).

Figure 5: Condition (1) (transitive): children should look longer at the two-
participant video. Condition (3) (intransitive): children should look longer at the 
one-participant video. Condition (2) (dislocated), Scenario A: Children in the dis-
located condition behave as children in the transitive condition: they interpret the 
novel verb as describing an action relating two participants and look longer at the 
two-participant video. Scenario B: Children in the dislocated condition behave as 
children in the intransitive condition: they interpret the novel verb as describing a 
one-participant event and look longer at the one-participant video.
4.1 Method

Participants

Sixty-four 28-month-olds (mean = 28.15; min = 27.13; max = 29.13; females = 30) participated in the study. Sixteen children were randomly assigned to each of the four experimental conditions (transitive, dislocated, intransitive and control). Six additional children were not included because of fussiness (n = 4), experimenter error (n = 1) or because they recognized one of the actor of the video (n = 1). All children were recruited from Paris and were acquiring French as their first language with less than 30% of exposition to another language. Parents completed the MacArthur form on vocabulary production [Kern S, 2007]. Vocabulary ranged from 71 to 614 words (mean = 392) and there were no differences in vocabulary between conditions.

Apparatus

Children sat on their parent’s lap 70cm away from a 42” television screen. Two videos of 30x30cm were displayed simultaneously on the right and left side of the screen and separated by 30cm. Children’s eye movements were recorded by a hidden camera placed on top of the screen and an eyetracker (Eyelink 1000). The caregiver and the experimenter were listening to acoustic masking throughout the test.

Materials

In the dialogue phase, toddlers viewed one of the three types of dialogue videos (transitive, dislocated and intransitive) showing a conversation between two women speaking in child directed speech (highly exaggerated to attract children’s attention). During the recording of the dislocated condition, the actors were asked to speak without any break in the sentence between the dislocated phrase and the rest of the sentence, thus no pause were observed in the acoustic signal of these sentences. In order to compare the prosodic difference between the transitive and the dislocated conditions, we performed an acoustic analysis of these sentences. This analysis revealed a significant lengthening of the last syllable of the verb in dislocated sentences ($t(14)= -10.31, p<0.001$) as well as a pitch discontinuity ($t(10) = 4.01, p<0.01$) at the left boundary of the dislocated constituent, showing that prosody alone differentiates between dislocated NP and object NP (detailed analyses are available in Appendix B).

In order to check whether these acoustic patterns are salient enough, we ran an adult study. For each sentence of conditions (1) and (2) (16 sentences), we asked twenty adults to decide who was performing the action of the sentence. For example, for "il dase le bébé", if prosody is informative, we expect adult participants to answer that the agent is le bébé while in "il dase le bébé", they should answer that the doer of the action is an external agent realized as a pronoun ("il"). Adults were correct 90% of the time in their choice of the agent, with a highly significant effect of prosody for both the by-subjects ($F_1(1, 19)= 355, p<10^{-13}$) and by-items analyses

\[^{20}\text{Only on the 12 declarative sentences, see Appendix B}\]
(F_2(1, 14) = 393, p < 10^{-10} **). Detailed method and results are available in Appendix A. The difference in the prosodic patterns observed in the acoustic analysis performed constitutes a reliable cue for adults to successfully interpret the sentences from dialogue conditions (1) and (2). Therefore, both experimental and acoustic results show that the dialogues are non-ambiguous and thus, suitable to be used to test young children.

In the test phase, children watched videos of actors performing actions. We used the same videos as in [Yuan and Fisher, 2009]. The audio material were recorded by a native French speaker.

**Procedure**

Each child saw three trials: two practice trials with familiar verbs (one with transitive verbs: *pousser/to push or porter/to carry*) and one with intransitive verbs: *danser/to dance or marcher/to walk*) followed by the critical test trial involving the novel verb (*daser/to dase*). The practice trials consisted of two 8s test events in which a synchronized pair of videos was presented on both sides of the screen along with a single soundtrack. The test trial consisted of two phases: a dialogue phase and a testing phase. During the dialogue phase, children heard 8 sentences involving the novel verb in either one of the three conditions and the test phase was composed of three 8s test events following the same procedure as the practice trials.

In order to give the children sufficient time to inspect the action of the videos, a preview period preceded each testing phase. Each video of the test event was seen alone during 5s before the actual test event started. The complete procedure and timing of the experiment is described below (see also Figure 6).

The first practice trial featured the intransitive verb (*marcher/walk*). The target video depicted one woman walking from left to right and the distractor video displayed an other woman dancing from left to right at the same pace. This first trial started with a preview period, one of the video appeared on one side of the screen for 5s accompanied by a neutral sentence ("Hey regarde ! Tu l’as vue ?"/"Hey look ! Did you see her ?") followed by a black screen of 3s. The other video was displayed in the same way on the other side of the screen. These previews were followed by a 4s black screen interval during which children heard an other audio prompt ("Hey, elle marche"/"Hey she is walking"). Then, in the first test event, the 2 videos were displayed during 8s simultaneously while children heard ("Tu la vois qui marche ? Regarde celle qui marche !"/"Do you see the one walking ? Look at the one walking"). Following a 3s black screen accompanied by the prompt ("Elle marche !"/"She is walking"), the same event was repeated a second time ("Elle est où celle qui marche ? Regarde celle qui marche !"/"Where is the one walking ? Look at the one walking !").

After a 4s black screen, the second practice trial proceeded in the same way with one of the transitive verbs (*pousser/to push or porter/to carry*). Crucially, the two videos shown during each practice trial had the same number of participants matching with the target verb: either transitive (two participants) or intransitive (one participant). This was to ensure that the children worked on the meaning of the verb and not on the number of participants involved. The
practice trials were designed to familiarize the child with the experimental procedure, showing them that the audio prompt matched one of the videos.

The test trial started with the dialogue phase (except for the control condition). Two four-sentence video clips of 28s separated by a 3s interval were presented simultaneously on both sides of the screen. Thus, each child was exposed to 8 transitive sentences (e.g. "il dase le bébé") or 8 dislocated sentences (e.g. "il dase , le bébé") or 8 plain intransitive sentences (e.g. "il dase"). After a 4s black screen, following the same procedure as the practice trials, children watched two novel actions: one woman making a circle with her arm on one side (one-participant action) and one woman swinging the leg of an other woman on the other side (two-participant action). During this testing phase, children saw the videos while hearing a series of sentences featuring the novel verb in an intransitive construction in all conditions ("Hey elle dase ! Tu la vois qui dase ? Regarde celle qui dase ! Elle est où celle qui dase ?").

In the control condition, children saw the two practice trials directly followed by a test trial with no dialogue exposure.

The left/right position of the target video was counterbalanced across participants, within each of the four conditions.

**Coding**

We used both a camera and an eye-tracker system (Eyelink 1000) to record children eye movements. Each child had a sticker on the forehead for the eye-tracker to work in a remote mode. We used a 5-points calibration routine with customized calibration points (a little dog reducing to a small dot) to make the calibration more 'child-friendly'. Children sat 70cm away from the screen, but, due to the large size of the screen used, it was impossible to prevent head movements at such a small distance. Thus we could only get 25% of data from the eye-tracker.

From muted video, we coded where children were looking (Right, Left, Away) frame by frame using SuperCoder [Hollich, 2005]. Reliability was assessed for 9% of the data (6 children), the two coders agreed on 97% of the frames (Cohen’s Kappa : 0.94).

---

21 Note that we checked the reliability of the eye-tracking system versus the frame-by-frame coding for two children. The two coding methods produced similar results (same fixation location for 92% of coded frames ; Cohen’s Kappa = 0.87)
Dialogue Phase
4 sentences (2 x 28s)

Testing Phase
Preview (5s):
Hey, regarde là !

Preview (5s):
Et là, tu vois ça ?

Black screen (4s):
Hey, elle dase !

Test events (3 x 8s):
Tu la vois qui dase ?
Regarde celle qui dase !

Figure 6: Sequence of events composing the test trial of Experiment 1.
4.2 Results

Looking times to the two-participant video, to the one-participant video, and away from the videos were averaged across the three test events. Data from test events where the child looked away more than half of the event duration (4s) were considered as missing (3 test events overall).

As Figure 7 shows, looking times are affected by dialogue condition. A by-subject ANOVA revealed an effect of condition on looking time to the 2-participant video ($F(3,60) = 3.56$, $p = 0.019^{*}$).

![Figure 7: Average looking-time at each video averaged across the three 8s test trials.](image)

The pattern found for conditions (1) and (3) reproduces the results of Yuan & Fisher (2009). Children who heard the novel verb in the transitive dialogue (1) looked longer at the two-participant event than children who heard it in the intransitive dialogue (3) ($F(1,30) = 7.77$, $p < 0.01^{**}$). They also looked reliably less at the one-participant event than did the children in the intransitive condition ($F(1,30) = 9.44$, $p < 0.01^{**}$). Children are able to use the syntactic structure in which the verb appears to infer how many participants are involved in the relation it describes and subsequently use this information to retrieve the verb meaning given two scenes that differ in the number of actors involved. The structural information relevant to the verb’s semantic structure can be retrieved without access to the verb’s semantic content.

Regarding the critical condition, children who heard the dialogue using dislocated sentences (2) behaved like children in the transitive condition (1) ($F(1,30) < 1$). They looked reliably longer at the two-participant video than the children who heard the intransitive dialogue (3) ($F(1 , 30) = 11.38$, $p <0.01^{**}$) and reliably less to the one-participant video ($F(1 , 30) = 9.71$, $p < 0.01^{**}$).
Although children in condition (2) seem to look more at the two-participant event than children in condition (1), the observed difference was not reliable ($F < 1, p > 0.5$).

As for the control condition (4), children who did not hear any dialogue introducing the novel verb in a given sentence structure looked equally long at both videos, showing no intrinsic preference for one event over the other. This result suggests that in the three experimental conditions, children attached the dialogue information to the novel verb and thus interpreted the verb according to the syntactic structure in which it appeared in the dialogue.
5 General Discussion and Perspectives

In this study, we tested whether toddlers exploit prosody when analyzing the argument structure of sentences that contain a to-be-learnt verb. Following the preferential looking paradigm of Yuan and Fisher (2009), we presented 28-month-olds with dialogues introducing a novel verb (’daser’) in one of three conditions: transitive, dislocated and intransitive. Toddlers were then asked to look for ’daser’ while viewing two videos displayed side-by-side: a two-participant action and a one-participant action.

Children who heard the novel verb in the transitive dialogue looked longer at the two-participant video than did children who heard it in the intransitive dialogue (a replication of Yuan and Fisher’s results). Regarding the critical condition, children who heard the dialogue with right-dislocated sentences behaved like children in the transitive condition. Thus, the prosodic structure of right-dislocated sentences was not sufficient to prevent children from assuming that the final noun phrase in ”il dase , le bébé” was the patient of the action.

Crucially, although the dislocated NP followed an intonational phrase boundary, children did not consider it as a separate fragment of the sentence. It has been shown that prosody universally constrains word segmentation by preventing lexical activation of words spanning prosodic boundaries [Shukla et al, 2010; Gout et al, 1994]. It is unlikely that this prosodic chunking would operate in the same way with regard to syntax. Indeed, syntactic calculations require children to compute relations across constituents. In section 2.1, we saw that prosodic boundaries mark syntactic constituent boundaries. In Snedeker and Yuan’s (2008 ; reviewed in Section 2.3), an intonational phrase boundary precedes the instrument prepositional phrase in [you can feel the frog][with the feather]PP. When tested with this sentence, children performed an instrumental action (they used a feather to feel the toy frog). Thus, although the first prosodic unit (’you can feel the frog’) is a syntactically correct sentence and is prosodically self-sufficient, children did not stop their syntactic analysis of the sentence at the intonational phrase boundary; they included the prepositional phrase into their sentence representation and attached it to the preceding verb phrase.

Prosodic phrase boundaries help in retrieving the syntactic constituents of the sentence but do not limit syntactic parsing inside a single prosodic unit. As a result, children may not limit their syntactic analysis within a prosodic chunk or unit but rather form an interpretation that takes all the units into account.

Moreover right-dislocated sentences are not acoustically split into two distinct units (il dase ! Le bébé !), rather, the particular intonational contour of the dislocated noun phrase makes it an integral part of the sentence (see Appendix B). Thus, the dislocated noun phrase is not a constituent that children can eliminate from their syntactic analysis of sentences, they have to integrate it and to find how it is related to the rest of the sentence. Crucially, our results show that children, like adults, attempt an interpretation of the sentence taking into account the dislocated NP.
Thus, 28-month-olds already know that the second noun phrase is a constituent that belongs to the sentence. However, it is unclear how toddlers link it to the rest of the sentence. The present results show that 28-month-olds behave similarly for right-dislocated sentences and for transitive sentences. This suggests that, contrary to our adult participants (see Appendix A), children don’t yet consider the second noun phrase in right-dislocated sentences as coreferential with the subject pronoun, but interpret it as the patient of the action.

A dislocated constituent follows an intonational phrase boundary and has a particular prosodic contour: either flat, low-pitched and de-stressed or a copy of the pitch contour of the end of the sentence nucleus. If children did not learn that this particular prosody characterizes a structure in which the final noun phrase and the subject pronoun are co-referent, then they would rely on another interpretative strategy that leads them to interpret the sentence as transitive. On the contrary, if children are already able to interpret right-dislocated sentences correctly, then they have to weigh this information relative to other sources of information available in the sentence. Thus, although children may know that the final noun phrase in a right-dislocated sentence is the agent of the action, this is not the only possible interpretation: their analysis of the prosodic information would be in competition with other interpretative strategies favoring the transitive interpretation of the sentence.

In the following, we propose a follow-up experiment to test children’s understanding of right-dislocated sentences and discuss the interpretative strategy adopted by children in our experiment in the light of the current literature.

Our CHILDES analyses (section 3.2 and 3.3) suggest that children between 2 and 3 years of age produce right-dislocated sentences. However, as we stated in section 3.3, it is unclear whether children can actually understand that the dislocated noun phrase is co-referent with the subject pronoun. To investigate this, we are currently testing 28-month-olds’ interpretation of right-dislocated sentences featuring known verbs such as ‘porter’, ‘pousser’, ‘taper’ and ‘manger’. Those four verbs are transitive but can be used intransitively such that, as in Experiment 1, it is possible to form pairs of transitive and intransitive sentences which feature the same words in the same order but differ in their prosodic pattern:

(1) **transitive:** il, va manger le canard, (Someone will eat a duck)

(2) **dislocated:** il, va manger, le canard, (The duck will eat)

A simple way to investigate whether children understand that ‘le canard’ co-refers with the subject pronoun ‘il’ in (2), is to test children’s understanding of who the agent of the action is in each sentence. To do so, we will use a preferential looking task. We will present children with pairs of action videos: a causal action where the second noun phrase (‘le canard’) is the patient of the action (such that a tiger is eating a duck) and a solo action where the second NP is the actor (a duck eating some bread). Each pair is accompanied by a soundtrack introducing the familiar verb (‘manger’) in either one of the two conditions (transitive or dislocated). This experiment is currently in progress and its detailed method is presented in Appendix C.
Such a design will allow us to test directly whether toddlers can exploit online the prosody of right-dislocated sentences to interpret the coreference between the two noun phrases. If children fail to interpret differently right-dislocated and transitive sentences featuring known verbs, it is obvious that they won’t be able to use this information with novel verbs and this explain their failure to exploit prosody in our main experiment. As reviewed in section 2.2, 9-month-olds are sensitive to phonological phrase boundaries that are acoustically less marked than the intonational phrase boundaries characterizing right-dislocated structure. Thus, it is clear that 28-month-olds must perceive that there is something special with the prosodic pattern of dislocated sentences. However, even though children may not yet be aware of the semantic representation involved by the right-dislocated prosody, they still attempt to give an interpretation of the sentence that takes all the constituents into account.

On the contrary, if our second experiment shows that children know how to interpret the coreference in right-dislocated sentences featuring known verbs, we can then interpret the results of the first experiment as showing that this prosody-induced interpretation is superseded by other sources of information leading children to interpret the sentence transitively. In both cases, to interpret right-dislocated sentences, children rely on an other interpretative strategy which is supposedly available early on in language development and that overcomes the representation of right-dislocated sentences they might have.

Relying on the set of nouns (or the number of nouns) has been shown to be an active strategy for toddlers to interpret novel verbs (Section 1.3). Children are biased to interpret a two-noun sentence as a relation between two participants and a one-noun sentence as an action including only one participant. For example, 19-month-olds failed to distinguish between intransitive conjoined subject sentences (“the duck and the bunny are gorping”) and transitive sentences [Hirsh-Pasek et al, 1996]. However, older children of 24 months succeeded in giving different interpretations to these sentences [Naigles, 1990] because they acquired a more sophisticated syntactic knowledge of their language, such as word order. Indeed, 21-month-olds have been shown to know about the word order pattern of their language [Gertner and Fisher, 2006]. In transitive sentences featuring a novel verb (“the duck is gorping the bunny”), children interpreted the noun preceding the verb as the agent of the action and the noun following the verb as the patient. Thus, from 21 months of age, children possess an abstract knowledge of word order and of the semantic role attached with nouns’ position.

In right-dislocated sentences, the second NP occurs after the verb, and thus occupies the position that the direct object of the verb should occupy. Hence, in our experiment, children may have relied on what they have learnt about word order to interpret right-dislocated sentences and therefore, interpreted the final noun phrase as the object of the verb.

In support of this interpretation, Gertner and Fisher (2012) recently showed that 21-month-olds interpreted in the same way intransitive sentences with a two nouns conjoined subject ("The boy and the girl are gorping") and transitive sentences ("The boy is gorping the girl") (following the results obtained by Hirsh-Pasek et al, 1996). In addition, they showed that children give
different interpretations to intransitive sentences featuring a different order of the nouns (*"The boy and the girl are gorping"* vs *"The girl and the boy are gorping"*). When presented with a causal event (a boy swiveling the girl in a chair) and a simultaneous action (the boy and the girl twirling ribbons), children who heard *"The boy and the girl are gorping"* behaved just like children in the transitive condition; they looked more at the causal action than children who heard *"The girl and the boy are gorping"*. Although noun order does not convey an agent-patient information in the case of a conjoined noun phrase subject, 21-month-old children behaved as if it did; they interpreted the first noun in *"The boy and the girl are gorping"* as the agent of the sentence and thus, interpreted the verb as referring to the causal action where a boy was acting on a girl. This study suggests that in the absence of exploitable cues (such as conjunction or morphological knowledge), children are heavily biased to interpret the first of two nouns as an agent and the second as a patient, even when the sentence does not follow a 'noun-verb-noun' order.

The same holds for our study: if children don’t know how to interpret the coreference implied by RD sentences, they would be biased to build the most likely sentence representation given their knowledge; and thus apply an agent-patient ordering leading them to interpret RD sentences transitively. Interestingly, Ferreira (2003) showed that English-speaking adults experienced more difficulty to parse sentences that do not follow the canonical order *Noun Agent-Verb-Noun Patient* than sentences following this pattern. They tended to interpret the first noun of a patient-before-agent sentence as the agent and not as the patient and the second noun as the patient and not as the agent. Thus, although adults possess a full syntactical knowledge, they still relied on this simple syntactic representation (noun-verb-noun) to interpret sentences in the first place. Similarly, children may be influenced by a simpler interpretative strategy based on word order (the noun-verb-noun pattern) even if they possess more sophisticated knowledge of their language such as RD sentences interpretation.

To sum up, whether or not children are able to interpret online right-dislocated sentences featuring known verbs, they are still highly influenced by this agent-patient ordering. Since the two noun phrases occupy the traditional agent and patient positions in the sentence, children interpret the sentence transitively even when some other sources of information are available to them. Our results show that children are biased to perform a syntactic analysis of sentences following this simple strategy. Like adults, when children encounter a sentence, they will tend to interpret the first noun encountered as the agent of the sentence and the second noun as the patient.

Children have been shown to use right-dislocated sentences productively (Section 3.3) and may potentially use them to interpret sentences with known verbs (see Experiment 2 ; Appendix C). Then, a question for future research is: why can’t 28-month-olds use the prosodic information of right-dislocated sentences to interpret novel verbs?
In our main experiment we pre-exposed children to dialogues introducing the novel verb within a given syntactic structure (transitive, dislocated or intransitive) before presenting the two video events. The success in this task requires children to build a partial semantic representation of the verb (a two-participant relation or a one-participant action) while listening to the dialogue, to be able to focus later on, on the relevant event matching this representation. While we are currently testing children on their online comprehension of right-dislocated sentences, our main experiment requires a more complex processing: for each sentence of the dialogue in the dislocated condition, children must first apply the coreference and then weigh this information relatively to their knowledge of word order. Finally, they need to integrate the syntactic information provided by all individual sentences of the dialogue to build a partial representation of the verb meaning. Thus, right-dislocated sentences add an additional cost to the overall processing of dialogues which could prompt children to look for a less costly interpretative strategy and thus follow the agent-patient information conveyed by the ‘noun-verb-noun’ pattern.

Another possibility could be that toddlers simply need more information to access the information conveyed by the prosody of right-dislocated sentences with unknown verbs. In our dialogue, all the sentences used are right-dislocated, while in natural discourse situations, it seems unlikely that a novel verb would appear only in right-dislocated structures. Thus it is questionable whether children would use the right-dislocated pattern alone to build a conceptual representation of the verb without evidence that the verb is intransitive or having some referential or discourse evidence. By definition, the dislocated constituent is the topic of the sentence and thus, should be easily recoverable in the discourse or referential context. However, one essential property of our experiment is that right-dislocated sentences in the dialogue phase are heard in a non-referential context, and switch from a topic to another without introduction of the new referent.

A way of getting rid of the complex processing involved by the dialogue-test paradigm and to integrate some referential information is to test toddlers’ comprehension of right-dislocated sentences introducing a novel verb in an online fashion. We could use the experimental design of our experiment testing right-dislocated sentences’ comprehension with familiar verbs (see Appendix C). Children would hear "il dase, le bébé" while watching two videos: one with a baby doing a self-generated action (one participant) and the other of a grown-up man acting on a baby (two participants). If the child interprets the sentence intransitively then she should look more towards the one-participant action. However, to succeed in this task, it is sufficient for the child to extract the agent from the right-dislocated sentences (‘le bébé’) and to match it to the agent of the action; there is no need to analyze the verb’s argument structure. Although introducing some useful referential context such as the actors of the action, such a design won’t allow us to check whether children are able to use the prosody of right-dislocated sentences to compute the verb’s argument structure.

Thus the dialogue-test paradigm seems to be the right testing option. Although referential
information is absent from the dialogue phase, it is possible to modify these dialogues in order to include some discursive information that will not only produce more natural dialogues but also help children to access the topic of right-dislocated sentences:

<table>
<thead>
<tr>
<th>Old dialogue</th>
<th>New Dialogue</th>
</tr>
</thead>
<tbody>
<tr>
<td>A: Hey il va daser , le bébé !</td>
<td>A: Hey il va daser , le bébé !</td>
</tr>
<tr>
<td>B: Ah bon, il va daser , le bébé ?</td>
<td>B: Ah bon, il va daser , le bébé ?</td>
</tr>
<tr>
<td>A: Oui, et en plus ils ont dasé , les canards</td>
<td>A: Oui! Et les canards?</td>
</tr>
<tr>
<td>B: C’est vrai, ils ont dasé , les canards</td>
<td>B: Ils vont daser , les canards !</td>
</tr>
<tr>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>A: Tu sais quoi ? Ils ont dasé , les chevaux !</td>
<td>A: Tu sais quoi ? Ils ont dasé , les chevaux !</td>
</tr>
<tr>
<td>B: C’est vrai ? Ils ont dasé , les chevaux ?</td>
<td>B: Ah oui? Et la maman?</td>
</tr>
<tr>
<td>A: Oui, et elle va daser , la maman !</td>
<td>A: Elle va daser , la maman !</td>
</tr>
<tr>
<td>B: Waouh, elle va daser , la maman !</td>
<td>B: Waouh, elle va daser , la maman !</td>
</tr>
</tbody>
</table>

Table 4: Possible dialogue to use in a follow-up experiment. The sentences in **bold** introduce the referent of the next RD sentence.

The dialogue presented in Table 4 introduces the new discourse referent of the next sentence ("Et les canards ?") and enables the listener to focus on the new topic of the upcoming sentence. By giving toddlers some discursive hints about the topic of right-dislocated sentences, it may be possible to reduce the complexity of processing the dialogue. Moreover, since right-dislocated sentences are discursive structures, it seems reasonable to add a suitable discursive environment to make their interpretation possible.

The proposed experiment would allow us to clarify the potential need for other sources of information to use the prosodic pattern of right-dislocated sentences as a cue to interpret a novel verb’s argument structure.

Finally, it is important to integrate the results exposed in this dissertation in a more general framework. Some aspects of phrasal prosody, such as intonational or phonological phrase boundaries, are universally marked, systematic and highly relevant to determine syntactic constituent boundaries. Right-dislocations cannot be restricted to a simple universal prosodic cue; they are a type of syntactic structure that manifests itself through an intonational phrase boundary as well as a specific (and language-specific) intonational contour on the dislocated element. Although highly frequent in natural speech (at least in French, and potentially in Romance languages), they do not systematically occur. Our results show that the prosodic pattern of right-dislocated sentences is outweighed by the information deployed by a simple ‘noun_**agent**-verb-noun_**patient**’ heuristic. However, even though the prosody right-dislocated sentences does not constrain a novel verb’s interpretation in 28-month-olds, this does not mean
that phrasal prosody does not constrain syntactic analysis of sentences in toddlers. As we saw in section 2.3, adults are able to use phonological phrase boundaries to predict the syntactic category of a word [Millotte et al, 2008]. Moreover, work in computational modeling suggests that the categorization of syntactic phrases (NP, VP) is sensibly improved on a corpus augmented with phrasal prosody [Gutman et al, in preparation]. Thus, phrasal prosody can in some circumstances directly inform syntactic processing; further experiments should explore its role in toddlers.

Summary and conclusions

Strikingly, our results stressed the substantial importance of structural biases in word learning. Following the structural account of syntactic bootstrapping (Section 1.3), children tend to interpret a noun as having a participant-role in the relation described by the verb. This rudimentary knowledge would allow children to construct a partial representation of sentences and of predicates without any other morpho-syntactic knowledge. Even if children potentially know about right-dislocated sentences’ structure, they still rely heavily on more simplistic representations of sentences.

We conclude that, early in life, simple strategies, such as the agent-patient information conveyed by the noun-verb-noun order, is prepotent when analyzing sentences’ syntactic structure. In contrast, adults have adjusted their syntactic representation to take into account other cues, such as the prosody of right-dislocated sentences, that they derived from a long-standing exposure to their language.
References


42


A Adult Study

Before testing children, we first ran a study with adults to check whether the prosody of the sentences in dialogues (1) and (2) (Table 3, section 4) was adequate. In these sentences, critically, the prosody of right-dislocated sentences leads to a different interpretation regarding who is the agent of the verb *daser*, thus for each sentence we asked adults to decide who is the agent of the sentence. For example, for "*il dase*, le bébé", if the prosody is informative, we expect the adult participants to answer that the *doer* of the action is *le bébé* while in "*il dase le bébé*", they would answer that the *doer* is an external agent represented by the pronoun 'il'.

Method

Participants

Twenty French adults participated in the study. One adult was not included because French was not his mother tongue. All were volunteers recruited in a public library in Paris and in Paris V university.

Materials

We used the sentences from the dialogue of the condition (1) and (2) thus a total of 16 sentences featuring the non-verb *daser*: 8 transitive sentences (as "*il dase le bébé*”) and 8 right-dislocated sentences (as "*il dase le bébé*”) recorded by two native female speakers.

Procedure

Adults were tested individually. They were given earphones and faced a laptop playing the sentences one at a time. The experiment started with a training of 4 sentences with real verbs to familiarize the participant with the testing procedure. Then, the test session started. Each sentence was repeated twice, thus a total of 48 sentences were played in random order with no repetition of sentences with identical lexical material. Participants were told that they didn’t need to understand all the words in the sentence to succeed in the task. For each sentence, we used a forced-choice paradigm. On the screen, the same prompt appeared "Qui est-ce qui dase ? / who is dasing ?” along with two options: AUTRE (OTHER) or the noun phrase used in the sentence. For example, for the sentence "*il dase , le bébé*”, the participant had to choose between AUTRE and LE BEBE. Participants were told that AUTRE (OTHER) refers to any external character who is not 'le bébé’ but represented by the pronoun 'il'. Once the participant clicked on one of the answers on the screen, the next sentence started to play. No time limit was set. Finally, the order of sentences was random and different for each participant.

22Training sentences used:
"Il joue, le chat.”
"Elle appelle maman.”
"Elle est cassée, la branche.”
"Elle mange une souris.”
Experimental Results

As Figure 8 shows, adults used the prosody of the sentence to figure out whether the post-verb noun phrase was the agent of the sentence.

**Figure 8:** Adults chose 'le bébé' (or the corresponding noun phrase for that sentence) 91% of the time when they heard a dislocated sentence (e.g. "il dase , le bébé") and only 11% of the time when they heard a transitive sentence (e.g. "il dase le bébé")

Adults interpreted the post-verb noun phrase as the agent of the sentence when they heard a dislocated sentence whereas they did not when presented with a transitive sentence. An ANOVA revealed a significant effect of prosody for both by-subjects ($F_1(1,19)=355, p<10^{-13} \ast \ast \ast$) and by-items analyses ($F_2(1, 14)=393, p<10^{-10} \ast \ast \ast$).

In order to check whether the final performance of participants is due to a training effect to the fact that they heard repetitively the contrast between dislocated and transitive sentences, we analyzed their five first answers. There was no significant difference in their performance on the five first test sentences compared to their overall performance ($M = 85\%$ of correct answers; $SD = 4.78\%, p>0.3$).

To sum up, since the sentences of this experiment differ only in their prosodic structure (conditions(1) and (2)), this pre-test demonstrates that adults are able to use the particular prosody of dislocated sentences to infer the correct agent of the event without relying on the meaning of the verb in the sentence. Moreover, their high performance on the first items shows that adults are capable of interpreting correctly the test sentences in the first instances, without need to contrast the prosody of transitive and dislocated sentences. To conclude, the results show clearly that adults can interpret differently the experimental sentences from conditions (1) and (2).
B Acoustic Analysis

How do dislocated sentences differ acoustically from transitive ones?

It is interesting to note that both speakers were asked to record the dislocated sentences such that there was no audible break between the sentence nucleus and the dislocated constituent. Thus, no pause can be observed between the verb phrase and the dislocated noun phrase on the acoustic spectrum\(^{23}\). Hence no break can be used to distinguish between a transitive and dislocated sentence. However, the literature on dislocated sentences report two main prosodic characteristics of these sentences that are not common with transitive sentences:

- An intonational phrase boundary is following the verb. Intonational phrase boundaries are marked by a final-lengthening of the preceding syllable, a pitch discontinuity and sometimes a pause.

- The sentence nucleus (such as 'Elle va daser' in "Elle va daser, la maman") in isolation carries meaning by itself. Since the dislocated NP is referring to the subject pronoun, the sentence "Elle va daser" without its dislocated part is a complete declarative sentence and thus is prosodically well formed [Rossi, 1999]. Therefore the prosodic contour of the sentence nucleus will be very similar to a sentence’s prosodic contour and thus, would be prosodically self-sufficient. However, the prosody of the dislocated constituent is highly dependent on the prosodic contour of the sentence nucleus. Rossi (1999) argues that the prosody of the dislocated NP is dependent of the rest of the sentence: the prosodic contour of the dislocated NP is the copy of the final contour of the sentence nucleus with a decreased pitch. This principle of ‘copy’ is less visible for declarative sentences than for interrogative sentences. The literature describing the acoustic properties of declarative RD sentences observed that right-dislocated NPs can also be destressed with a lower pitch than the intonation of the sentence nucleus [Ashby, 1994]. By “intonation”, Ashby means the most salient piece of information in the sentence.

In order to evaluate the prosodic differences between the transitive condition and the dislocated condition in our test dialogues, we performed an acoustic analysis on the 16 sentences that constitute the experimental material (8 transitive and 8 dislocated sentences) using Praat.

To observe the intonational phrase boundary characterizing RD sentences, we analyzed the duration pattern of the pre-boundary word (i.e 'daser') according to the type of sentences (transitive or dislocated):

\(^{23}\)Ashby, 1994 and Rossi, 1999 have shown that this pause is not obligatory in RD sentences.
<table>
<thead>
<tr>
<th>Segment</th>
<th>Avg Duration Transitive (ms)</th>
<th>Avg Duration Dislocated (ms)</th>
<th>Lengthening (%)</th>
<th>t (14)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>d</td>
<td>184</td>
<td>91</td>
<td>-51</td>
<td>-3.13</td>
<td>&lt;0.01**</td>
</tr>
<tr>
<td>a</td>
<td>217</td>
<td>228</td>
<td>5</td>
<td>0.69</td>
<td>0.50</td>
</tr>
<tr>
<td>s</td>
<td>91</td>
<td>148</td>
<td>62</td>
<td>3.64</td>
<td>&lt;0.01**</td>
</tr>
<tr>
<td>er</td>
<td>135</td>
<td>485</td>
<td>258</td>
<td>10.98</td>
<td>&lt;0.001***</td>
</tr>
</tbody>
</table>

Figure 9: Average length of the different segment constituting the two syllables of “daser”

In dislocated sentences, the final syllable of "daser" is longer, mostly due to the fact that the syllable nucleus is 258% longer compared to transitive sentences (see Figure 10). Similarly to what is described in the literature, we observe that an intonational phrase boundary precedes the dislocated NP.

Figure 10: On top: example of a dislocated sentence with a lengthening of the last syllable before the dislocated constituent ("Elle va daser, la maman"). Below: a transitive sentence in which the final syllable of the verb is shorter than the first one ("Elle va daser la maman"). Note that the speech is continuous in both cases.

---

24 There is also a shortening of the first syllable of "daser", the onset of the syllable and especially the closure time is significantly shorter by 51% in dislocated sentences compared to transitive sentences. However this is probably not a demonstration of the prosody of RD. Indeed, speakers may have added some focus on the verb in transitive sentences to make them easily recognizable from dislocated sentences.
To observe whether the prosodic contour of the dislocated NP reproduces the contour of the sentence nucleus as suggested by Rossi (1999), we looked at non-declarative sentences where this pattern might be easier to observe.

If the principle of copy holds, then for interrogative sentences (4 in our dialogues, 2 in each condition), we should observe two interrogative pitch rises: one at the end of the sentence nucleus and one at the end of the dislocated constituent, see below Figure 11.

![Figure 11: Variation of Pitch for: Top: "il va daser, le bébé ?". Below: "il va daser le bébé ?". The pitch rise characterizing interrogative sentence is circled in red](image)

In the transitive sentence, the pitch characterizing interrogation is observed on the nucleus of the final syllable (the second "é" of bébé). Interestingly, in the dislocated sentence we observe as expected by the theory, two rising pitches, one at the end of the sentence nucleus and one at the end of the dislocated NP.

For declarative sentences, it is expected that the dislocated NP should be destressed with a lower pitch than the maximum intonation of the sentence nucleus. Thus, I compared the pitch difference between the dislocated part (or the object part) and the intonation peak of the sentence (situated for all sentences on the verb "daser").

<table>
<thead>
<tr>
<th>Sentence</th>
<th>Avg Pitch drop (Hz)</th>
<th>Std Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transitive</td>
<td>34</td>
<td>7</td>
</tr>
<tr>
<td>Dislocated</td>
<td>82</td>
<td>9</td>
</tr>
</tbody>
</table>

![Figure 12: Average pitch difference between the main intonation peak of the sentence nucleus (on daser) and the pitch of the NP first segment.](image)

As expected, there is on average a greater pitch difference between the verb and the dislocated constituent than between the verb and its object (M = 48 ; t(10) = 4.01, p < 0.01**).

**25**Only declarative sentences have been considered here (thus 12 sentences) as we saw that interrogative dislocated sentences are subject to a pitch rise in its dislocated part.
To conclude, the analysis conducted on the test sentences shows that prosody alone allows to differentiate between RD and transitive sentences. Left boundaries of dislocated constituents are marked by syllable lengthening and dislocated constituents with a destressed prosodic contour or a copy of the contour of the sentence nucleus. This particular prosodic pattern constitutes a reliable cue leading adults to successfully interpret the right-dislocated sentences from the dialogue. Therefore, both experimental and acoustic results show that the test sentences are non ambiguous and thus, suitable to be used to test young children.
C Experiment 2 - Interpreting right-dislocated sentences

We tested 28-month-olds’ interpretation of right-dislocated sentences featuring known verbs such as ‘porter’, ‘pousser’, ‘taper’ and ‘manger’. Those four verbs are transitive but can be used intransitively such that, as in Experiment 1., it is possible to form pairs of transitive and intransitive sentences which feature the same words and the same order but differ in their prosodic pattern:

(1) transitive: il mange le canard (Someone will eat a duck)

(2) dislocated: il mange, le canard (A duck will eat)

A simple way to investigate whether children understand that ‘le canard’ co-refers with the subject pronoun ‘il’ in (2), is to test children’s understanding of who the agent of the action is in each sentence. To do so, we used a preferential looking task. We presented children with pairs of action videos: a causal action where the second NP (‘le canard’) is the patient of the action (such as a tiger eating a duck) and a solo action where the second NP is the actor (a duck eating bread). Each pair was accompanied by a soundtrack introducing the familiar verb (‘manger’) in either one of the two conditions (transitive or dislocated) following the procedure presented in Figure 16.

If children in the dislocated condition (“il va manger, le canard”) successfully compute that the dislocated NP and the subject pronoun co-refer, then they should interpret the duck as the agent of the sentence and look more at the video event where the duck is the agent of action ‘manger’ (the NP-agent video). In contrast, children who hear the verb in a transitive frame (“il va manger le canard”) should interpret the duck as the object of the action and look more towards the video event depicting a tiger eating a duck (the NP-patient video). If children in the dislocated condition behave like children in the transitive condition then it will be necessary to run a baseline condition in order to check whether this is not the result of a mere effect of preference towards the two-participant video event.26

Method
Participants

Seventeen 28-month-old children already participated in the study, 9 in the transitive condition and 8 in the dislocated condition. Three additional children were not included because of fussiness. We are planning to test sixteen children in each experimental condition.

Apparatus

Children sat on the parents laps 70 cm away from a 42” television screen. Two videos of 30x30cm were displayed simultaneously on the right and left side of the screen and separated

---

26Such a control condition would follow the same procedure with the intransitive test sentence "Le canard va manger"
by 30cm. Children’s eye movements were recorded by a hidden camera placed on top of the screen. The caregiver and the experimenter were listening to acoustic masking throughout the test.

Materials

Children watched pairs of videos showing puppets: a rabbit, a duck, a tiger and a monkey. Four pairs of video were created for the following verbs: 'pousser', 'porter', 'manger' and 'taper'. The order of the video pairs and the left-right position of the videos were counterbalanced across subjects, within each condition.

The soundtrack was recorded by a female French native speaker speaking in child-directed speech. We pretested our test sentences with 10 adults (following the same procedure as the adult study presented in Appendix A) and ensured that each sentence used in the experiment was appropriately considered either as transitive or dislocated more than 90% of the time.

Procedure

The experiment started with 4 characters identification videos, one for each puppet used in the subsequent actions. Each puppet was presented in a 5s video waving its arm and labelled one time ("Oooh, regarde, c’est un lapin !" / 'Oooh, look ! It’s a rabbit !”). A 2s black screen interval was inserted between the identification videos.

Next, children saw one practice trial (‘sauter’) in the same structure as the subsequent test trials (see Figure 16). In the target event a bunny was jumping and in the distractor event a bunny and a monkey were playing.

Each video was first played individually on one side of the screen and accompanied by a neutral soundtrack (‘Hey regarde ! Tu vois ça?’) to encourage children to look at the displayed event. Next, during a 6-s black screen interval, children heard one sentence containing the target verb (‘Attention, regarde ! Le lapin va sauter !”). Then, both video events were displayed side-by-side during 8s accompanied by a test sentence (‘Le lapin va sauter ! Regarde !”). This first test trial was followed by a 6-s black screen interval in which an additional test sentence containing the target verb was played ( ‘Tu as vu ça ? Le lapin a sauté !’). A second presentation of the two video events (8s) introduced the verb one more time (‘Oui ! Le lapin a sauté ! Regarde encore !’). In total children heard 4 instances of the verb. This trial familiarized children with the procedure and showed them that the soundtrack matched one of the videos.

Finally, 4 test trials using familiar verbs were presented following the same procedure as the practice trial (see Figure 16). For each verb, a pair of videos was played such that the action of the verb was performed on both videos. Videos were paired such that the agent in one video is the patient of the other one. For 'manger', the NP-patient video showed a tiger eating a duck and the NP-agent video a duck eating bread such that while hearing "il mange (,) le canard”, the NP 'le canard’ can be interpreted either as the agent (dislocated condition) or as the patient (transitive condition) of the action.

51
Thus, in total each child saw 5 test trials: 1 practice trial and 4 test trials in the dislocated condition or in the transitive condition.

**Preliminary results**

Looking times to the NP-patient video, to the NP-agent video and away from the screen were averaged across all the four test trials (pousser, taper, manger, porter). As Figure 13 shows, looking times seem to be influenced by the sentence condition. The nine children in the transitive condition behaved as if they interpreted 'le canard' in "il mange le canard" as the object of the verb and thus looked more at the video where the duck is the patient of the action (the NP-patient video) than at the video where the duck is the agent of the action (the NP-agent video). In contrast, the eight children in the dislocated condition looked equally long at both videos.

![Figure 13: Average looking-time at each video averaged across the two 8s test trials.](image)

Figure 13 shows children’s behavior for each verb (pousser, taper, manger, porter). The looking-time pattern for 'taper’ does not seem to be affected by dialogue condition. Indeed, we suggest that children are heavily biased towards the NP-patient video which shows a monkey (happily) hitting a tiger. This video event is thus certainly more eye-catching than the NP-agent video where the poor tiger is hitting a cup.
Since children showed a bias for the NP-patient video from the data we analyzed, we excluded this test trial from the average looking-time results. As Figure 15 shows, children in the transitive condition tended to look more at the patient-NP video than children in the dislocated condition (as the Figure 7 also suggested). Crucially, children in the dislocated condition tended to look more at the NP-agent video than children in the transitive condition. Thus without the taper-trial which biased them to look at the patient-NP video, children in the dislocated condition seems to know how to interpret right-dislocated sentences since their behavior suggests that they interpreted ‘le canard’ in ‘il va manger, le canard’ as the agent of the sentence.
Figure 15: Average looking-time at each video averaged across the two 8s test trials for each verb (pousser, taper, manger, porter) in both conditions: transitive (top) and dislocated (bottom).

Although we only tested seventeen children out of the thirty-two we are planning to test, our preliminary results suggest that children are able to interpret right-dislocated sentences with familiar verbs when they have access to the referential context described by the sentence. In order to document the biases children may have experienced with the videos, we propose to ask adults to evaluate the goodness of the action of each video.

The final results of this experiment will be particularly informative to understand the behavior of children in Experiment 1.
Figure 16: Sequence of events composing the test trial of Experiment 2.