Eye-tracking children’s acquisition of three types of recursive computations using PPs in Brazilian Portuguese

Aleria Lage*, Aniela Improta França*, Marije Soto **, Mayara de Sá Pinto*, Nathacia Lucena Ribeiro*/***

Introduction

A significant theme in current linguistic research is the processing of recursion. Assuming the contrasts in Roeper and Oseki (2018) about the difference in the level of computational complexity, there are three types of recursion: (i) Direct Unstructured Recursion (DUR); (ii) Direct Structured Recursion (DSR); and (iii) Indirect Recursion (IR). DUR “adds an infinitely large number of XPs in a linear manner without hierarchical structures generated” and, thus, yields a conjunctive reading, semantically (Roeper and Oseki, 2018, p. 269). For example, in (i) there is an apple on the table, and another on the tablecloth, and another on the plate. DSR is compositional and hierarchical; XPs are sisters in the saturation of the argument. Roeper e Oseki claim this is achieved by Feature Sharing (Chomsky, 2013), such that any XPs following a sister XP under the same node will be labeled according to the feature they share. Semantically, the first XP summarizes information contained in the following XPs. For instance, in (ii) the apple is on the table, and the two following PPs, ‘on the tablecloth’, and ‘on the plate’ expand on the exact localization of the apple on the table. In that sense, DSR is compositional, given that the order of the XPs seem to parallel a pragmatically motivated order going from the more salient information, expanding on it subsequently. Diagnostic evidence from syntactic behavior is that XPs can be extracted out of DSR generated structure, which means that XPs are embedded in a hierarchical organization. For example in a Wh-question: ‘<Where> is the apple on the table _______?’.

Both DSR and DUR may stack identical XPs, without requiring an intervening category YP. This is in contrast to IR, which is fully compositional and hierarchical in nature, accepting an intervening category YP, such that XP -> X YP, and YP -> Y XP (Roeper and Oseki, 2018). In (iii), for instance, PPs consecutively modify previous NPs, such that the apple is on the plate, which is on the tablecloth, which is on the table. The mastery of these three types of recursion requires distinct cognitive resources and should come at different stages in language acquisition.

1* UFRJ (Federal University of Rio de Janeiro)
** UERJ (State University of Rio de Janeiro)
*** CUNY (City University of New York)
In fact, there is plenty behavioral evidence advocating that DUR is a simpler computation since it appears earlier in language acquisition, both in children’s comprehension and in production. Also both DSR and IR do not seem to be present in children’s earliest utterances (Roeper and Snyder 2004, 2005; Roeper, 2011; França et al., 2014). This bears a consonance with evolutionary studies that show that while monkeys (i.e. cotton-top tamarins) do not seem to incorporate grammars capable of building or analyzing hierarchical structure typical of phrase-structure grammars, they do master simple organizational principles, that control regularities limited to neighboring units, such as linear adjunction (Hauser, Weiss, and Marcus, 2002; Fitch and Hauser, 2004.)

Findings contrasting the two extremes of this continuum (DUR and IR) are also supported by evidence from recent image and electromagnetic studies that purport that these computations are processed in different neural systems within the human brain. Two perisylvian language pathways, that differ both phylogenetically and ontogenetically, seem to be specialized to process different levels of linguistic complexity. The ventral pathway, much older in the species, is available since birth, and underlies the mechanisms for acquisition of lexical items and possibly local syntactic phrase building (cf. Friederici, 2012). Contrastingly, the dorsal pathway, which greatly increases in dimension and connectivity during the first years of infancy, underlies processing of asymmetrical structures. This dorsal pathway is also newer in the species: it appeared 50,000 years ago (Friederici, 2004; Friederici and Brauer, 2009; Berwick et al., 2013). This might explain, in part, third factor restrictions on the faculty of language as described by Chomsky (2005, p.6) that states that: “developmental constraints
enter into canalization, organic form, and action over a wide range, including principles of efficient computation”. Thus, this suggests a relation between developing neuroarchitecture and computational capacities, which may impact the computation of more complex syntactic structures throughout acquisition. This further motivates claims that the acquisition path follows the putative increasing syntactic complexity of the proposed typology of recursion (DUR->DSR->IR) (Roeperto Oseki, 2018).

However, it is still under debate which of the restrictions, representative or performance, are most relevant in the limitations on comprehension and production of these structures during acquisition. Chomsky remarks on this matter: “It could be that unbounded Merge, and whatever else is involved in UG, is present at once, but only manifested in limited ways for extraneous reasons (memory and attention limitations and the like)” (Chomsky, 2005, p. 12).

An important aspect in this discussion is the different depths of embedding. A study in Japanese, a right branching language, compared eye tracking fixation durations of a relative clause modifying the matrix object with those of two recursively embedded relative clauses and found that the greater the depth of embedding, the harder it is to process the sentence (Mazuka et al., 1989).

Nevertheless, there is opposition to this one-to-one relationship between embedding and progressive computational complexity. Based on results of modeling studies, Stabler (2014) argues that the recursive depth of a structural analysis “does not correspond in any simple way to depth of the calculation of that structure in linguistic performance.” (p.159). Maia et al (2018) reports a series of experiments with adult participants in Portuguese, English and Karajá, an indigenous Brazilian language of the Macrojê family, using different online methodologies, which unveiled a cognitive advantage of embedding PPs after the second layer, and a progressive difficulty of conjoining them.

This paradox can be explained by considering interface dynamics. Linearly adjoined XPs can only be sent to the conceptual interface after every XP is merged. So, if the list of XPs is too long it exhausts memory capacity within the phase, and the derivation crashes. It has long been known that the DUR type structures should be limited to a few items. When memory is challenged, a way to dissolve the informational bottleneck is by using syntax to chunk the items (Miller, 1956). Rummer, Engelkamp and Konieczny (2003) provide relevant independent experimental evidence from adult language that IR constructions are kept in working memory more easily than DUR constructions. There is also neurophysiological
evidence that working memory benefits from interacting with syntax in Broca’s area to process complex hierarchies (Cooke et al., 2001, Santi and Grodzinsky, 2007).

As mentioned before, between direct and indirect recursion processing, children seem to often prefer the latter. However, evidence from online comprehension studies are sparse and inconclusive. One suggestion is that children specifically struggle with the restrictive semantic interpretation of IR structures, and, therefore, fall into a non-restrictive reading available in direct recursion. As such, DSR could be a “stepping stone”, going from DUR to IR, in the acquisition of IR computation (Oseki and Roeper, 2018). Note that this could also be characterized, representationally, by considering that functional nodes licensing direct and indirect structures have different scopes. Therefore, differences might lie in the semantic interpretation of these nodes (Di Sciullo, 2015).

Roeper (2011) suggests that children initially build structures with complements whose phrase structure rules are informed by lexical restriction (ex. think may be complemented with a sentence (NP VP): John thinks it is raining). These nodes are limited from recursion, in the sense that they do not further project a maximal projection, which would allow for unlimited recursion (ex. Johns thinks Mary said George thought it was going to rain). Once children acquire this functionally differential node (one that does allow for recursive projection), they substitute the limited, lexically specified nodes, for functionally recursive ones. One of the reasons, children may do this later in their acquisition path, Roeper suggests, is because they are exposed relatively little to expanded recursive structures, and, thus, may be sparsely triggered by these type of structures.

In order to contribute with more acquisition data to this theme, we decided to verify children’s processing of embedding and conjoined structures in a series of oral sentence picture matching experiments. In convergence with Roeper (2011), in an earlier study (França et al., 2014), we focused on locative prepositional phrases (PP). In a picture matching task, two groups of children (3 y-o and 4 y-o) heard sentences with to two types of recursion: (iv) DUR and (v) IR. These sentences could be matched with an image representing a distributed reading (there is one frog on the leaf, one on the stone, and another on the sand) or an image representing an “embedded” reading (there is a frog is on the leaf, which is on the stone). Prosody was manipulated to be equal for both types of sentences, such that only the conjunctive ‘and’ overtly marked stimuli as being coordinate. A specificity of Brazilian Portuguese is that a distributive reading of a sentence without the overt conjunctive marking ‘and’ (as in v) is also possible. This is due to the fact that it is possible to use a bare noun in the singular with the verb to have (‘ter’) with an existential reading for both singular and
plural meanings. Thus, both an “embedded” and distributive reading are both felicitous for (v).

(iv) Tem sapo na folha, e na pedra, e na areia

there is frog on the leaf and on the stone and on the sand

*There is a frog/are frogs on the leaf and on the stone and on the sand*

(v) Tem sapo na folha, na pedra, na areia

there is frog on the leaf on the stone on the sand

*There is a frog/are frogs on the leaf on the stone on the sand*

Given this duality, children may favor one of the two interpretations depending on the recursive operations available to them throughout the acquisition process. If the sentence is computed following the DUR type of recursion, the noun is merged with adjoined locative PPs, for which the result is a distributive reading; but if IR is triggered, the result is an “embedded” reading. In fact, the three to four year-olds in the study preferred to match this type of sentence (v) to images representing a distributive reading rather than to an image representing an “embedded” reading, suggesting they engage in DUR. The “embedded” interpretation only appeared to be meaningful – and to a lesser extent – at the age of 4. The number of recursive layers was not a statistically significant factor.

It was difficult to ascertain if their preference for a distributive reading was rooted in a computational restriction, given that syntactically more complex structures such as IR (and thus the “embedded” reading) were not yet available to them; or that they simply preferred the distributive reading, since direct recursive structures are semantically less restrictive in their interpretation. Therefore, we decided to work on new experiments that would create the possibility to investigate whether children engage in a type of processing that would bridge the gap between DUR and IR processing.

In the current study, we used a more engaging task, invoking the participants’ executive functions needed in a role-playing game. It is a game featuring elusive blue ETs named Zukis. Before the game we showed a video with the Zukis story and the participant’s mission: to rescue the Zukis. We put together two new experiments in this fashion, and we also adjusted the age bracket of participants to 4.5-6.5. Furthermore, we excluded the prosodic contour of the stimuli (now spoken by one of the characters in the story, the robot) and monitored the child’s gaze with an eyetracker. The new methodology could give us a more
precise control of the different steps, that is, what visual information was being accessed before the longer fixation which decides which picture matches the stimulus.

In Experiment 1 we tested two and three PPs in direct and indirect recursive structures, similar to the previous study described in examples (iv) and (v) (França et al., 2014).

In Experiment 2, which was tested in the same role-playing mode, there were two conditions, each mixing DSR and IR types in the same sentence. In these instances, which can be seen in (vi) and (vii), the phrase structure rule is binary and it generates both linear and hierarchical structures. In (vi), for example the first PP (‘on the fish’) is aligned linearly with the second complex PP (‘on the plant in the tank’) achieved through a shared feature between XPs (PPs) (Chomsky, 2013:46). Within the second PP, there is a hierarchical structure in which the PP, ‘in the tank’, modifies the head DP, ‘the plant’. However, note that in these constructions the PP attachment point is ambiguous between local and distant attachments, depending on the parsing strategy followed.

(vi)    Tem Zuki No peixe, e na planta, e no aquário
(local) there is a zuki [pp on the fish] and [pp on [dp the plant] [pp in the tank]].
(distant) There is a Zuki / Zukis [pp on [dp the fish] (and on) [dp the plant]] [pp in the tank]].

(vii)   Tem Zuki na planta e no peixe no aquário.
(local) there is a zuki [pp on the plant] and [pp on [dp the fish] [pp in the tank]].
(distant) There is a Zuki / Zukis [pp on [dp the plant] and on) [dp the fish]] [pp in the tank]].

In order to tap into these potentially ambiguous interpretations, we will consider not only the representational status of these structures, but also the parsing strategies that may guide children in processing. Taking (vi), for instance, participants could parse according to a depth-first strategy making immediate commitments involving the first DP and the first PP. This strategy is known as Early Closure: Immediately close a structure, that could potentially remain open for the association of more lexical input material (Frazier and Rayner, 1982). Conversely, participants could choose not to commit instantly and wait until the end of the sentence. In this case, the sentence would be parsed bottom-up: the Zukis on the plant in the tank, and the fish and the Zukis on the fish in the tank, following a parsing strategy named Late Closure: if grammatically allowed/licensed, attach new items into the clause or phrase
currently being processed, that is the phrase or clause postulated most recently (Frazier, 2002: 3).

Additionally, to manipulate children’s preference we inserted a pragmatic bias in (vi). The distant PP (fish) should be more canonically placed as an element contained in the local PP (tank) than it would be as an element in an independent location (the fish hanging on the wall). So (vi) is biased for Late Closure while (vii) the neutral condition should favor Early Closure.

According to what we have laid out in previous sections, we expect the feature-sharing structures in (vi) and (vii) (DSR) to be a little harder to compute than linearly organized structure (DUR), as in (iv), and a little less complex than indirect recursion (IR) as in (v). This might be due to the fact that DSR structures, in comparison to DUR structures, are more complex syntactically, and are also more restrictive in their semantic interpretation; while IR structures are syntactically more embedded than DSR structures, and even more restrictive semantically speaking (Roeper and Oseki, 2018: 270).

In accordance with the Eye-mind Hypothesis, we foresee a correlation between the duration of fixation times on target and non-target images to the mental process taking place (Just and Carpenter, 1980). Therefore, we expect the eyetracking data to show the difference in cognitive effort with progressively longer fixation times for DUR<DSR<IR conditions, as well as longer fixation times for 3 PP structures than for 2 PP structures. We also expect that the proportion of fixations on a given image, especially towards the end of the trial, may reflect a preference for a given interpretation (distributive or more embedded, for example). We expect DUR and DSR conditions to yield higher proportions of fixations on target images as compared to IR conditions. The same measure may indicate children’s parsing preference and their tendency to be guided or not by semantic-pragmatic bias during parsing.

The eye-tracking experiments

The capsule review: In Experiment 1, we looked at the processing of direct unstructured recursion (DUR) versus indirect recursion (IR) and whether and how an increasing amount of linguistic information, two prepositional phrases (PPs) versus three PPs, affected processing in young children. In Experiment 2, we investigated direct structured recursion (DSR) with ambiguous attachment points and we manipulated semantic-pragmatic contexts to evaluate if it would affect the direction of the gaze (e.g. fish-tank).
Materials and Methods

Participants:

Twenty children (5 girls) with a mean age of 5 years, 6 months (range: 4;5 to 6;4) participated in the study. All children were raised monolingually with Brazilian Portuguese as their native language and were typically developing according to the information obtained from their school, a public pre-school program. They had no known visual or auditory deficits. All parents or legal guardians had given permission previously, signing a consent form (Ethics Committee process # 983_18). Data from 2 children had to be excluded because of inattention during the eye tracking session, such that 18 datasets were analyzed.

Experimental setting:

We used the eye-tacking facilities at LER Lab (Laboratory of encephalography and eye-tracking for language studies) at the Federal University of Rio de Janeiro. The oral sentence-picture matching protocol was conducted in a TOBII TX300Hz eye-tracker. Each participant sat in a naturally lit room, at a distance of 0.9m to a 24 inch screen, always accompanied by a member of the research team. Before the oral sentence picture-matching experiment started, we played an introductory cartoon of two minutes to acquaint the participants with the story of Zukis. In short, the story informs participants that the Zukis are naughty blue ETs who manage to get on a rocket that is accidently set to travel to Planet Earth. A robot that is in the control room talks to the participants involving them in the rescue of the characters back to their planet. The rocket landed on earth on a house’s front-yard and the Zukis are all over the house. The robot explains that in order to capture the Zukis it is necessary to look at them fixedly, but warns that the Zukis can make clones of themselves, and only the robot knows how to tell a real Zuki from a clone. If participants want to help capture the Zukis and send them back to their planet, they should direct their gaze to the picture with the real Zukis, following the robot’s instruction. So the plot of the cartoon helps to motivate the experimental task which is gazing at the picture that matches the oral sentence. Also the robot’s natural monotone voice turns the artificially flattened prosodic contour more natural.

Thus, in a game-like setting, main characters (the Zukis) and context (their escape to earth) were set (see Fig.1).
For each trial, the robot tells the children where to look in order to ‘catch’ the real Zukis with participant’s ‘magic’ gaze. Thus, the context of the game directly motivates children to direct their gaze to one of two images that are presented side by side as they hear one of the two following sentence types: (i) *There are Zukis in the box and in the drawer*, (ii) *There are Zukis in the box in the drawer*. But only one of the images can be paired with the sentence they heard (see Fig.2).

![Fig. 2. The protocol for test trials](image)

The use of the robot voice allowed for a flat prosody unbiased for interpretation of recursive type. Therefore, children would have to match sentences and target images based solely on structural information. As subjects performed this task, their eye movements were monitored through the TOBII TX300Hz eye-tracker.

The introductory cartoon that set the context for the role-playing mission was put together at Camelo Azul Studio in Rio de Janeiro and the cartoon and stimulus images were all prepared in experimenters’ lab. Voice recording for the introductory cartoon and for all stimulus sentences in the experiment was done in a studio, and subsequent pitch manipulation and editing was done in Praat (version 6.0.40). Image creation and editing was done in Adobe Photoshop (version CC 2017). The generated clips were then inserted in Tobii Studio to command the eye-tracking session. In the first two and a half minutes, context and task were explained. Then, in a playful manner, eyetracker calibration took place during which participants were asked to follow a visual object (the robot’s hands) moving on the screen. Also, participants went through two training trails to ensure comprehension.
The presentation protocol of the test trails was as follows: during the first 2 seconds, two pictures were presented, then a sentence was played (average duration: 2.5 to 3.5 seconds), then after 2 seconds, a short clip for positive feedback appeared. Participants could visualize the effect of their rescue: Zukis being sent back home in a rocket. On average, total trial duration was 7 seconds, and the feedback lasted for an additional 7 seconds (see Fig. 2). Each participant saw 40 trials in total. Total duration of the experiment was 12 minutes on average.

The spatial position of the target image (left or right) was varied pseudo-randomly among trials. The number of Zukis was always the same in both the target and non-target images. Stimuli were equally distributed over four spatial contexts: the living room, the bedroom, the kitchen and the yard. Spatial elements which appeared in the images, and related to the sentences (e.g. in the box), were also equally distributed among stimuli conditions.

**Experimental design: Experiment 1**

The independent variables were (1) recursion type (direct unstructured or indirect), and (2) number of PPs (two or three), resulting in a full factorial design. The experimental design resulted in 24 sentences for each condition (96 in total), which were pseudo-randomized and distributed over 6 lists, such that each participant heard 4 sentences for each condition (16 in total) (see Table 1). A total of 10 distractor sentences were added to each list (e.g. There is/are Zuki playing in the yard with the dog; ‘Tem ZUKI no quintal brincando com o cachorro’).

<table>
<thead>
<tr>
<th>Recursion type</th>
<th>Direct Unstructured Recursion (DUR)</th>
<th>Indirect Recursion (IR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Two</td>
<td>Tem Zuki no peixe e no aquário (n-24) (There is/are Zuki(s) on the fish and in the tank)</td>
<td>Tem Zuki no peixe no aquário (n-24) (There is/are Zuki(s) on the fish in the tank)</td>
</tr>
<tr>
<td>Three</td>
<td>Tem Zuki no peixe e na planta e no aquário (n-24) (There is/are Zuki(s) on the fish and on the plant and in the tank)</td>
<td>Tem Zuki na planta no peixe no aquário (n-24) (There is/are Zuki(s) on the fish on the plant in the tank)</td>
</tr>
</tbody>
</table>

**Experiment 2: ambiguous interpretation (DSR)**

<table>
<thead>
<tr>
<th>with pragmatic-semantic bias</th>
<th>without pragmatic-semantic bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tem Zuki no peixe e na planta no aquário (n-24) AWPB (There is/are Zuki(s) on the fish and on the plant in the tank)</td>
<td>Tem Zuki na planta no peixe noaquário (n-24) ANPB (There is/are Zuki(s) on the plant and on the fish in the tank)</td>
</tr>
</tbody>
</table>

Table 1: Experimental design and stimulus samples
Experimental design: Experiment 2

The independent variable of this test was a pragmatic-semantic bias favoring a Late Closure parsing strategy. The sentence stimuli for this variable were ambiguous for the picture matching task, as both images were a possible target. The objective of this test was to investigate if a semantic and pragmatic bias may lead to a specific parsing strategy. Thus, the sentences could be parsed following one of two parsing strategies: (i) Late Closure (e.g. [[[on the fish] and [on the plant]] in the tank]); or (ii) Early Closure (e.g. [[on the fish] and [on the plant] in the tank]). The Late Closure strategy would lead to an interpretation in which the 1st PP is only fully processed after the 2nd and 3rd PPs are heard, yielding a structure in which the 1st and 2nd PP are coordinated in a larger PP, and are both modified by the 3rd PP, placing both the fish and the plant in the tank. According to the Early Closure strategy, minimal structure building favors the closing of the 1st PP as soon as possible, leaving only the 2nd PP to be modified by the 3rd PP, thus, placing only the plant in the tank. In the sentences with pragmatic-semantic bias, there is a critical word in the first PP, such as fish in the following example: There are Zukis on the fish and on the plant in the tank. In the example, the word fish may bias the participants to adopt a Late Closure strategy, due to the semantic and pragmatic association between fish and tank.

The other condition presents a sentence without bias, such as in the following example: There are Zukis on the plant and on the fish in the tank. In this example, since there is no obvious association between plant and tank, the preferred parsing strategy might be biased towards Early Closure (see Fig. 3). The experimental design resulted in 24 sentences for each condition (48 in total), which were pseudorandomized and distributed over 6 lists, such that each participant heard 4 sentences for each condition (8 in total). A total of 6 distractor sentences were added to each list (see Table 1).

Fig. 3: Experimental conditions with and without pragmatic-semantic bias
Data collection and analyses

Duration and position of fixation and movement of saccades were recorded by Tobii Studio (version 3.4.2). Two areas of interest were determined: the target image and the non-target image. The dependent variables used for the analyses of this study were total fixation duration in each image and the accuracy rates in terms of the number of fixations and the duration of fixations on the target images in proportion to the total number and duration of fixations. The collected data were analyzed in R (version 3.4.4), applying a multi-level ANOVA analysis.

Results

Experiment 1 Total Fixation Duration

![Graph 1: Total fixation duration per condition on target images](image)

The total fixation duration in an eye-tracking test is a measurement that reflects the Eye-mind hypothesis, which predicts that fixations reveal facts about attentional mechanisms, end of processes and choices made (Just and Carpenter, 1980). Relatively longer fixation durations correlate the visual object of the gaze to the mental process taking place.

There was a main effect correlating accuracy and type of recursion \((F(1,92) = 14.9\), \(p<0.000214\) \(SS=10.38\) \(MSe=0.70\)). This main effect indicated that the proportion of target fixations correlates to the recursion type, direct or indirect. There is also slight effect, which nearly reaches significance, for number of PPs \((F(1,92) = 3.05\) \(p<0.084013\) \(SS=2.85\) \(MSe=0.91\)).
MSe=0,93). This is just a tendency because, in fact, as verified in previous experiments (França et al 2014; Maia et al, 2018) the time increase related with the number of items is only significant in direct recursion. The depth in indirect recursion is not statistically significant.

As can be observed in Graph 1, direct recursion with two and three PPs are statistically different measures ([DUR2PPs]vs[DUR3PPs] t(46)=2,78 \( p< 0,0078 \)). Three coordinated PPs take longer to process than two, while fixation times for indirect recursion with two and three PPs are not statistically significant ([IR2PPs]vs[IR3PPs] t(46)=0,95 \( p< 0,3491 \)).

**Number of fixations**

For the analysis of the number of fixations, we selected a time interval covering the last two seconds of each trial. Then we calculated the number of fixations on target and non-target images as proportional to the total number of fixations. The focus on this time slot was warranted, because as we compared the proportion of fixations on the target image during and at the end of the trial, there was a main effect for time F(1, 544)=5.71, \( p=0.017 \). This main effect indicated that the proportion of target fixations increased over time, suggesting that the actual selection of the target image by participants takes place at the end of the trial (see Graph 2). There was also a mean effect of recursion F(1, 544)=53.30, \( p<0.000 \), reflecting a generally higher proportion of target directed gazes for direct recursion type stimuli.
In Graph 3, we can see the proportion of fixations on target and non-target images in the last time interval. For direct unstructured recursion the proportions of looks was above chance level, both for 2 and 3 level coordination (62.57% and 61.88%, respectively). For indirect recursion, there was a slight tendency to look more at the non-target images (52.3% and 53.80%); however, performance remained at chance level. Overall, there was an effect of recursion $F(1,270)=19.19$, $p<0.000$ with a higher percentage of fixations for the target image for direct unstructured recursion as compared to indirect recursion (DUR: 62.57% (SD:25.42%) and 61.88% (SD:25.15%); IR: 47.70% (SD:32.50%) and 46.20% (SD:32.51%)), and no relevant differences or interactions for number of PPs.

**Total count of fixations**

<table>
<thead>
<tr>
<th></th>
<th>Two PPs</th>
<th>Three PPs</th>
<th>Two PPs</th>
<th>Three PPs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct Recursion % at target image</td>
<td>62.57%</td>
<td>61.98%</td>
<td>52.30%</td>
<td>53.80%</td>
</tr>
<tr>
<td>Indirect Recursion % at non-target image</td>
<td>37.43%</td>
<td>38.12%</td>
<td>47.70%</td>
<td>46.20%</td>
</tr>
</tbody>
</table>

Graph 3: Proportion of the number of fixations on target and non-target images

The statistical analysis of the absolute number of fixations in this interval confirms the afore mentioned data (see Graph 4). In this analysis, fixation counts were separated for all conditions and for target and non-target images. There was a main effect de accuracy $F(1,526)=12.47$, $p<0.000$, with a strong interaction with recursion type $F(1, 526)=19.98$, $p<0.000$. This means that for Direct Unstructured Recursion, there were significantly more fixations on target images as compared to non-target images, for both two PPs and three PPs (Target 2PPs: 7.58 (SD:4.23) vs. Non-target 4.51 (SD:3.29), $p$-value $< 0.000$; Target 3PPs: 8.03(SD:5.10) vs. 5.07(3.85), $p=0.001$). For Indirect Recursion, on the other hand, there were no significant differences in the number of fixations for target and non-target images,
indicating that performance was at chance. There was no main effect for number of PPs (F1,526)= 2.12, p=0.14.

![Graph 4: Fixation count in absolute numbers in the last interval](image)

**Results:** **Experiment 2**

**Total Fixation Duration**

Ex. There is Zuki on the plant and on

Ex. There is Zuki on the fish and on
Graph 5: Total fixation duration: (A) for the non-biased condition, (B) for the biased condition.

In Graph 5, we can see that the duration of the fixations are longer for the images favoring a Late Closure parsing strategy, for both the non-biased and the biased condition, with $p=0.049$ and $p=0.025$, respectively. This effect increases slightly for the pragmatically biased condition, but can be considered a general preference in the processing of both biased and non-biased ambiguous sentences.

**Number of fixations**

In Graph 6, the absolute number of fixations in the end of trial interval is presented.

There was no main effect for attachment preference $F(1, 265)=0.33$, $p=0.57$, nor for pragmatic bias $F(1, 265)=1.043$, $p=0.31$. There was also no interaction $F(1,265)=0.36$, $p=0.55$. Different from what the total fixation duration measures show, these data suggest that there was no clear preference for either type of parsing strategy, and that the pragmatic bias did not alter the lack of preference.

However, if we zoom in on the final gaze of the trial, we can see a tendency towards Late Closure. In Table 2, we can see last gaze preference expressed as a percentage of total preference for all trials per condition. There is a main effect for attachment preference $F(1,58)=4.45$, $p=0.039$. This preference seems to be more expressive for stimuli with a
pragmatic bias. In 60.0% of the trials, there was a preference for Late Closure as compared to Early Closure (40.0%). For sentences without pragmatic bias, there was a 54.69% preference for Late Closure as compared to 45.31% for minimal attachment. Although there seems to a bias favoring Late Closure for sentences with a pragmatic bias, there is no statistical significance for this effect.

We can also observe a visual bias towards the image placed at the left with a 64.62% and 62.32% preference for the left side for the sentences with and without pragmatic bias, respectively. Given that conditions were equally distributed for left and right placement, no interaction is expected for this visual bias.

Finally, observing fixations within the images in more detail, we can conclude that in 83.33% and 89.06% of the cases, the final fixation is on the first PP. Statistical analysis yielded a strong main effect for PP preference: F(1, 58)=248.10, p<0.000. This preference is slightly higher for non-pragmatic conditions, but analysis showed no effect for pragmatic bias ((F1,25)=1.99, p=0.15)), nor for attachment preference (F(1,25)=2.11, p=0.16). This indicates that there is an overall preference to finish parsing by the PP that was heard first, instead of the PP that was heard last, a true Late Closure visual inspection.

<table>
<thead>
<tr>
<th>Parsing preference</th>
<th>Location of last fixation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Late Closure</td>
<td>1st PP</td>
</tr>
<tr>
<td>Early Closure</td>
<td>2nd PP</td>
</tr>
<tr>
<td>Left Image</td>
<td></td>
</tr>
<tr>
<td>Right Image</td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Last gaze, preference expressed as a percentage of all participants

Conclusions: Experiment 1

In the eyetracking literature there are classical findings relating increased fixation duration times to increased syntax complexity (cf. Frazier and Rayner, 1982; Clifton, Staub and Rayner, 2007; Maia, 2016).

In Experiment 1, total fixation duration times clearly revealed that direct recursion yields longer fixation times, thus, correlating accuracy and type of recursion (F(1,92) = 14.9, p<0.000). Thus, we may conclude that overall, direct recursive structures are easier to process for children at the mean age of 5. The number of fixations on the target images in proportion to the total number of fixations towards the end of the trial also showed that there was a main
effect of recursion \( (F(1,270)=19.19, \ p<0.000) \), pointing to a higher percentage of target directed fixation for direct recursive structures as compared to indirect recursive structures.

However, fixation duration times, but not number of fixations, showed an additional effect: namely, the nearly significant effect for number of PPs \( (F(1,92) = 3.05 \ p<0.084) \), indicating that only for direct recursive structures, the 3\(^{rd}\) PP poses an additional burden on processing. Indeed, the interpretation of direct recursive structures with 3 PPs lead to longer fixation times compared to 2 PPs ([DR2PPs] vs [DR3PPs] \( t(46)=2.78 \ p<0.0078 \)). For indirect recursion structures, on the other hand we see that the difference in depth (i.e. 2 or 3 embedded PPs) does not correlate to a significant increase in fixation times ([IR2PPs] vs [IR3PPs] \( t(46)=0.95 \ p<0.3491 \)). This replicates earlier findings by França et al. (2014) and Maia et al. (2018), which leads us to conclude that there is a natural structural gain with indirect recursion, once the child has acquired this computation. This result is even more striking, when we take into account that all stimuli were manipulated to minimize variations in pauses and intonation. Even so, participants could still infer the embedding syntax and the number of embedded PPs did not become increasingly harder to compute; whereas, conjoining becomes a heavier burden as the number of items increases.

**Conclusions: Experiment 2**

For sentences with a potentially ambiguous interpretation, we might expect general parsing strategies to guide preference for one of the images. In the condition without pragmatic-semantic bias (e.g. *There are Zukis on the plant and on the fish in the tank*), we may contend that only parsing strategies – and not pragmatic bias - guide listeners’ interpretation. A Late Closure strategy (LC) would foresee suspending the full processing of the 1st PP, until the moment at which incoming linguistic material can be organized in a structure with a minimal number of nodes (e.g. both the *plant* and the *fish are in the tank*). Another possible strategy would be the Early Closure (EC) route, according to which the 1st PP is immediately processed, whereas the last and 3\(^{rd}\) PP is expected to be locally attached to the 2nd PP (e.g. only the *fish is in the tank*). In terms of the visual behavior, we might predict that if the LC strategy prevails, the eyes will fixate on the visual item represented in the 1st PP last, given that full processing of this structure is suspended up until the very last moment - when Late Closure can occur as soon as the final and 3\(^{rd}\)PP has been heard. If EC is the preferred parsing strategy, we would expect that fixations follow the order of interpretation: the first item presented in the 1st PP, given its immediate availability for interpretation, and
later the item mentioned in the 2nd PP modified by the final and 3rd PP. Notice that both of these hypotheses attribute a strong influence of syntactic structure on comprehension processes.

Another hypothesis altogether would be that it is mainly the lexical content rather than syntactic structure that drives linguistic processing (i.e. The Good Enough hypothesis; Ferreira et al., 2002), and consequently, that steers the eye. If this is indeed the case, we would expect the eye to fixate on items following the direct sequential order in which they appear in the sentence (1st PP, 2nd PP, and, finally, 3rd PP) without considering the syntactic hierarchy in which they are embedded.

Interestingly enough, the data show that, by the end of the trial, most participants fixated on the item related to the 1st PP, either in the image that favored LC (83.33%) or in the image that favored EC (89.06%). This points to the possible influence of a LC parsing strategy, and discards the option of Good Enough processing. After all, what reason would there be to return to the item that is in fact least recent in memory?

On the other hand, the data show no statistical evidence for overall preference for the image favoring the LC interpretation (for sentences with pragmatic bias: 60.0%, without pragmatic bias 54.69%). Although there seems to be a tendency for pragmatic bias, great variance among participants may have occurred, which would explain why this tendency is not robust enough. This might be accounted for if we consider that, although a given image may favor LC interpretation over an EC interpretation, neither of the images are, in fact, incompatible with either interpretation. Furthermore, other language-independent aspects, such as general visual processing strategies may have affected image preference somewhat. We noticed, for example, a slight tendency for final fixations on the image on the left side (63.47%) (the position of images was counterbalanced throughout the experiment).

On the whole, the pragmatic-semantic bias altered preferences very little. For example, for sentences without pragmatic bias, participants’ last fixations concentrated on the images favoring LC reading in 52.86%. In comparison, for sentences in which the pragmatic and semantic qualities of the noun contained in the 1st PP may have established a strong semantic association to the 3rd PP (e.g. A fish belongs in an tank) preference for LC reading increased only slightly to 56.06%.

The fact that the participants’ visual behavior revealed a general tendency for participants to revisit the 1st PP at the end of the trial is thus a strong indication that it is the order of syntactic processing that determines the saliency of a given item. Although there is no direct evidence in the data for an LC or EC preference, we may infer that, under the EC
parsing strategy, we would expect participants to look first at the item mentioned in the first PP, and later to look at the item mentioned in the 2nd PP; whereas the LC strategy predicts the opposite, which ended up being in accordance with the observed data.

**Overall Discussion**

In two eye-tracking experiments, we were able to investigate online parsing strategies of children of mean age 5. We looked at processing of three types of recursion: (i) Direct Unstructured Recursion (DUR); (ii) Direct Structured Recursion (DSR); and (iii) Indirect Recursion (IR) (Roeppe and Oseki, 2018). In the Experiment 1 we compared DUR versus IR and whether and how an increasing amount of linguistic information, two prepositional phrases (PPs) versus three PPs, affected processing in young children. Results pointed to the fact that, at mean age 5, participants were more accurate towards direct than indirect recursion. Although indirect recursion seems harder to parse, the increase in the number of PPs affects the processing time of DUR more than that of IR. In Experiment 2, we investigated DSR, with ambiguous attachment points and we manipulated semantic-pragmatic contexts to evaluate if they would affect the direction of the gaze (e.g. *fish-in the tank*). Participants’ displayed an overall preference (56.06%) for Late Closure a choice backed up by a final gaze preference of nearly 90%, toward the distant item. This means that they included the distant item in the container that was the last PP in the sentence. For example: examining children’s gaze pattern, we could perceive that the stimulus “There is Zuki on the fish and on the plant in the tank” was paired with the picture that showed that Zuki, fish and plant were in the tank, and the fish was the last object they looked at.

Although DUR and IR were tested in a separate experiment from DSR, it is striking that the Late Closure preference approximates the parsing to that of indirect recursion. As attested in Experiment 1, both direct and indirect recursion types are available at mean age 5. So, participants had the syntactic resources to choose between an immediate commitment (Early Closure) or wait until the end, which implied in a more nested calculus following a bottom-up course. Then, the Late Closure course preferred by participants does not seem to be a strictly linguistic choice, but rather one deriving from memory restriction that parses first the item heard last. This heuristics, which can be viewed as a performance or third factor phenomenon, turned the semantic bias almost unnoticeable, since the items without the bias were also interpreted according to a bottom-up course. Interpretation both of pragmatically neutral and biased sentences was that all the Zukis ended up being placed in a single location (for instance, in the tank). Notice that when they listened to the DSR
sentence, they did not jump into a DUR distributive reading. They made a compositional reading instead, but going from the last item to the first, due to memory restrictions.

Note that Late Closure is semantically more related with IR that assigns a nested reading to items and it also seems to inherit the structural advantage of this type of recursion, observed in Experiment 1: once engaged in the syntactic recursive algorithm, subsequent embeddings are facilitated. In fact, total fixation duration for the DSR was smaller than those of DUR, which seems to support Roeper and Oseki (2018: 270) hypothesis that feature-sharing structures could be a little harder to compute than direct recursion and a little less complex than indirect recursion, being the intermediate type.
References


