

The processing of Topic-Comment Structures in Brazilian Portuguese: a visual world paradigm study

Processamento de estruturas de tópico-comentário no português brasileiro: um estudo de paradigma do mundo visual

Lorrane da Silva Neves Medeiros Ventura 

Universidade Federal do Rio de Janeiro – UFRJ – Rio de Janeiro – Brasil



Abstract: The aim of this study is to investigate topic-comment and subject-predicate structures in Brazilian Portuguese (BP), through the experimental psycholinguistics approach, in order to bring empirical support to the hypothesis that BP would be a mixed language (Li & Thompson, 1976), in which both, topic and subject ideas are important. In other words, BP would be both a discourse-oriented and sentence-oriented language, unlike English, that is, undoubtedly, a sentence-oriented language. Base-generated topic structures will be investigated, because such structures are prototypical in a topic-prominent language, as Chinese, in which the topic notion is central, but absent in a language in which the subject notion is predominant, as English (Yuan, 1995), in addition to the topic structures generated through syntactic movement (topicalization). The visual world paradigm experiment found that Brazilian Portuguese speakers are sensitive to prosodic cues present in auditory stimuli. That is, when they received an auditory input that favors a subject, the speakers focused more on the image that contained a subject NP and chose the option that represented a Subject-Predicate sentence. On the other hand, when listening to an auditory input that favored a topic element, the speakers focused more on the image that contained a topic NP and selected the option that contained a topic-comment Sentence. These results suggest a classification for BP as a mixed language.

Keywords: Brazilian Portuguese; topic; mixed language; visual world paradigm.

Resumo: O objetivo deste estudo é investigar as estruturas de tópico-comentário e sujeito-predicado no português brasileiro (PB), por meio da psicolinguística experimental, a fim buscar sustentação empírica em favor da hipótese de que o PB seria uma língua mista (Li & Thompson, 1976), onde tanto a noção de tópico quanto a de sujeito são importantes. Em outras palavras, o PB seria uma língua orientada para o discurso e para a sentença, diferentemente do inglês, que é, sem dúvida, uma língua orientada para a sentença. Estruturas de tópico geradas na base serão investigadas, pois tais estruturas são prototípicas em uma língua onde a noção de tópico é predominante, como o chinês, mas ausente em uma língua onde a noção de sujeito é predominante, como o inglês (Yuan, 1995), além das estruturas de tópico geradas via movimentação sintática (topicalização). O experimento de paradigma do mundo visual evidenciou que os falantes de português brasileiro são sensíveis às pistas prosódicas presentes nos estímulos auditivos. Ou seja, ao receberem um *input* auditivo que favorece um sujeito, os falantes focaram mais na imagem que continha um SN sujeito e escolheram a opção que representava uma sentença de Sujeito-Predicado. Por outro lado, quando ouviam um *input* auditivo que favorecia o elemento tópico, os participantes focaram mais na imagem que continha um SN tópico e selecionavam a opção que continha uma sentença de tópico-comentário. Esses resultados sugerem uma classificação para o PB como língua mista.

Palavras-chave: Português brasileiro; tópico; língua mista; paradigma do mundo visual.

Introduction

The present study investigates the processing of topicalization constructions (1), compared to Chinese style topics (2) and subject-predicate sentences (3) in Brazilian Portuguese (BP), as follows:

(1) A menina cega, o amigo ajudou a atravessar a rua.

“The blind girl, the friend helped (her) to cross the street”

(2) A menina cega, o amigo chegou quando o trânsito abriu.

“The blind girl, the friend arrived when the traffic opened”

(3) A menina cega o amigo com a faixa para brincarem.

“The girl blinds her friend with a bandana to play.”

According to our analyses, the topic in sentence (1) results from syntactic movement, that is, it is in the position of the head of a chain whose foot is the empty category in direct object position of the verb within the comment. In sentence (2) we have a different configuration, since there is no possibility of syntactic movement. Chinese style topics are topics that are not related to a syntactic position in the comment. In fact, this kind of structure is a very productive phenomenon in BP, especially in the oral modality. The example in (2) shows this topic-comment construction, where, apparently, there is no syntactic relation between the topic and the comment. In fact, Chinese style topics would be related only semantically to their predicates, since, in the comment, there is no site where the topicalized element could have been generated and extracted. In contrast, we have (3) in the subject-predicate configuration. In Brazilian Portuguese, it seems that topic constructions are commonly used by speakers, according to several studies on the subject (among others Pontes, 1987; Negrão, 1999; Orsini, 2003; Kato, 2006; Kenedy, 2002, 2014; Berlinck et al, 2015).

Li & Thompson (1976) emphasize that topic constructions are found in all natural languages. However, there are different levels of use considering a particular language. The authors use the subject prominent/topic prominent parameter to propose that the languages of the world may be classified in four types, namely, (i) subject prominent languages (e.g., English); ii) topic prominent languages (e.g., Chinese); iii) mixed languages (e.g., Japanese); iv) neither subject nor topic prominent languages (e.g., Tagalog). Brazilian Portuguese has been the object of dispute as to which of these types would provide the best classification for it.

The purpose of this study is to verify if Brazilian Portuguese native speakers can be guided by prosodic cues during the hearing of a topic and a subject element during processing, to contribute to a clearer classification as to what the typological status of BP would be and to provide contributions of psycholinguistic processing to the debate.

1. Theoretical Framework

Since Pontes (1987), linguists in Brazil have investigated topic constructions in BP, in order to try to find the best typological classification for the language. On the one hand, some linguists claim that BP is a subject prominent language (among others Duarte, 1996; Kenedy, 2002, 2014), where the predominant word order is Subject-Predicate. On the other hand, some researchers have claimed that BP should be classified as a topic prominent language (among others Negrão, 1999; Kato, 2006), in which the predominant word order is given by topic-comment. Finally, a mixed typological characterization (cf. Pontes, 1987; Orsini e Vasco, 2007; Berlinck et al, 2015) about BP is also advanced given, where both, subject prominence and topic are predominant. Given this controversy, it is necessary to investigate further on this issue, in order to contribute to a clearer classification about what the typological status of BP would be. Therefore, this study has a direct impact on the typological characterization of Brazilian Portuguese. In this sense, in order to provide psycholinguistic evidence to the debate and to support the hypothesis that Brazilian Portuguese would

be, if not a topic prominent language, at least, a mixed language, we conducted a visual world paradigm experiment.

It is important to note that the studies carried out on this issue about BP are supported by *corpus* analyses or by linguist's intuition (cf. Pontes, 1987; Negrão, 1999; Duarte, 1996; Kato, 2006; Orsini e Vasco, 2007), except for Kenedy's study (2014). He has investigated the subject through an experimental study. Kenedy (2014) aimed at testing the hypothesis that BP would be a topic prominent language using a grammaticality judgment task and a self-paced listening task. He concluded that BP would be classified as a subject prominent language. Kenedy (2014) argued that BP is a subject prominent language only and did not consider the occurrence of the innumerable topic constructions, mainly Chinese-style topics, which are prototypical structures in a topic prominent language. Furthermore, the offline technique used in the first experiment did not allow a direct measurement of processing. Our hypothesis is that BP should be classified at least as a mixed language, according to Li & Thompson's (1976) proposal. Moreover, this study uses an online technique, which has the capacity to measure the processing at the moment it happens, through VWP (visual world paradigm).

According to Yuan (1995), a topic-prominent language, as Chinese, allows base generated topics, which are not created through movement. They are base-generated rather than derived by movement, from some other position in the sentence. In a subject-prominent language, as English, some cases of base-generated topics have to be introduced by expressions like *as for* to be acceptable. BP allows topics derived by movement and also based generated topics and there is no need to introduce a base-generated topic by an expression like *as for, when it comes*, etc. (Medeiros, 2021).

Yuan (1995) argues that English does not allow base-generated topics, like Chinese-style topics. It is said that English is a subject prominent language governed by its rigorous grammatical rules and its topic constructions are not a core phenomenon in the

language. So, topic constructions are peripheral in English and are usually formed by syntactic movement (e.g., John, I did not like him). In other words, the subject in English is obligatory as a sentence constituent, following the SVO word order, unlike Brazilian Portuguese, which tends to be more flexible in its word order.

2. Methodological Framework

According to our hypothesis, there are two sentence construction strategies that are equally important in BP, that is, topic-comment and subject-predicate. The purpose of the experiment is to verify if the speakers are able to anticipate which structure will be constructed from the beginning of an auditory input that pre-activates a topic or a subject-predicate sentence. In this way, speakers will use characteristic features of each type of structure, in this case from given prosodic cues, and will recognize the natural structures in their language, if both structures are natural in the language. The speakers will only be able to anticipate the type of structure that will still come, if for them the tract contained in the auditory input is natural in their language. Moreover, by identifying in the input characteristic traits of a given structure, the listener should anticipate that the continuation of the sentence would follow this pattern, and when the pattern is not correct, there will be strangeness.

The logic of a visual world paradigm experiment is that the ocular behavior displayed by a listener during image tracking, when receiving an auditory input, can be very useful for us to understand how human language is processed in our mind/brain, during the time course. According to Salverda & Tanenhaus (2017):

Participants' eye movements are monitored as speech unfolds. Of interest is at what point in time with respect to some acoustic landmark in the speech signal (e.g., the onset of a word) a shift in the participant's visual attention occurs, as measured by a saccadic eye movement to an object or picture. (Salverda & Tanenhaus, 2017, 4)

The visual world paradigm technique is a tool that shows how is it the immediate integration of visual

and linguistic information at the time of word recognition, syntactic processing etc. By recording the eye movements during the processing of the visual and auditory stimuli we expected to observe whether participants can be guided by prosodic cues, while viewing images.

2.1 Method

In this section we describe the participants, the materials and design as well as the testing procedures used in the experiment.

2.1.2 Participants

Forty-eight Brazilian Portuguese native speakers, undergraduates from the Federal University of Rio de Janeiro, volunteered to serve as participants. They all spoke BP as their first language and they had normal hearing and normal or corrected vision.

2.1.3 Materials and Design

The independent variables were the factors Type of initial input, with two levels (Topic X Subject) and Type of complement, with three levels (Topicalization commentary X Chinese style topic commentary X Predicate). We conducted a visual world paradigm experiment with a 2x3 type of design, between subjectsⁱ. Therefore, we divided the

experiment in 2 groups: Match and Mismatch. The match group tested sentences containing input and the remainder of the sentence equivalent, that is, if the participant received a subject input, the remainder of the sentence would be a predicate. If they received a topic input, the remainder of the sentence would be a comment (it could be topicalization or Chinese-style topic). The mismatch group tested sentences that had no sync in their structure, that is, subject input and comment complement or topic input and predicate complement. This procedure was adopted to verify if Brazilian Portuguese speakers are able to identify, even in the initial processing of the structure, prototypical characteristics of a topic or a subject structure, perceiving when there is a mismatch between the prosody of the initial NP and the complement of the sentence.

The experimental set was made up of 6 conditions with 12 sentences per condition (see TABLE 1, as an example of the set). Each participant was presented to 12 experimental sentences embedded in an extra set of 24 distractor sentences and 4 practice sentences. The distractor sentences were unrelated grammatical constructions. All the sentences were presented in audio files. For each audio file, the participant viewed two images on the screen, one representing a topic-comment sentence and the other one representing a subject-predicate sentence.

Table 1: Experiment conditions

Sentences for the Match experiment

[A menina cega.]subject [o amigo com a faixa para brincarem.]predicate (Suj_Pred)

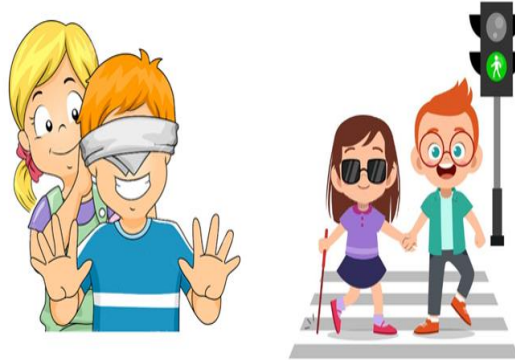
“The girl blinds her friend with a bandana to play.”

[A menina cega,]topic [o amigo ajudou a atravessar a rua.]comment (Top_C_Top)

“The blind girl, the friend helped (her) to cross the street”

[A menina cega,]topic [o amigo chegou quando o trânsito abriu.]comment (Top_C_Chin)

“The blind girl, the friend arrived when the traffic opened”



Interpretative question:

- a) The girl is blind b) The girl blinds someone

Sentences for the Mismatch experiment

[A menina cega,] subject [o amigo ajudou a atravessar a rua.]comment (Suj_C_Top)

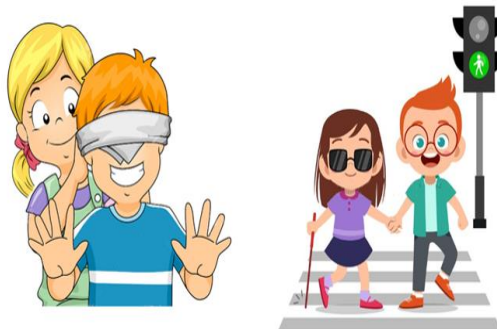
“The blind girl, the friend helped (her) to cross the street”

[A menina cega,] subject [o amigo chegou quando o trânsito abriu].comment (Suj_C_Chin)

“The blind girl, the friend arrived when the traffic opened”

[A menina cega]topic [o amigo com a faixa para brincarem.]predicate (Top_Pred)

“The girl blinds her friend with a bandana to play.”



Interpretative question:

- a) The girl is blind b) The girl blinds someone

The audio files were recorded by a Brazilian Portuguese native speaker. The *prosodic cues we used were the pitch accent, duration and pause. In more details, the topic structures (Chinese-style topics and topicalization sentences) were prosodically divided in two parts: the topic and the comment. The topic sentences present an intonation phrase boundary (I) between the topicalized initial NP and the rest of the sentence (Nespor & Vogel, 2007). The topicalized NP showed a pitch accent L+H* on the last phonological word and low boundary tone L%. The topicalized NP also presents a stretching of the stressed syllable at the boundary position of I. In addition, the comment presents a neutral declarative sentence pattern. Between the topicalized DP and the comment sentence there was a 200-millisecond manipulated pause. The Subject-Predicate sentences form a single I, with a pitch accent H+L* in the initial NP and a final low boundary tone (L%). Concerning durational measurements, there were no lengthening of the nuclear and post-nuclear syllables in the first NP.*

The dependent variables were the proportion of looks at the target and the distractor images on specific time windows (online) and the choice of the participants about the image interpretations (offline). According to the hypothesis adopted, the following predictions are made for the **Match group**: Brazilian Portuguese native speakers, upon hearing a topic input, will predict a comment and look at the image representing a topic-commentary sentence. That is, upon hearing the initial NP of the sentence, which is a topic in this condition, the match group will interpret the word that follows NP1 as an adjective. Thus, when they hear the NP with topic prosody, they will look more at the image that shows a blind girl, for example. On the other hand, when they hear the NP with a subject prosody, they will look more into the picture that shows the girl blinding someone. In this case, when they hear the critical region they will interpret it as a verb of the sentence and they will predict a verb + complement, so they will look at the image representing a subject-predicate sentence. For the **Mismatch group**, upon hearing the conditions where initial input and complementation are not synchronous, participants will

have difficulty during the online processing, as they are able to differentiate topic and subject prosodic features. Participants will be able to perceive when there is a mismatch between the prosody of the initial NP and the complement of the sentence, considering that both topic and subject are important concepts in the mental grammar of BP. The counter-expectation will be observed through a differentiated pattern of ocular fixations, indicating great confusion during the incremental processing of the structure.

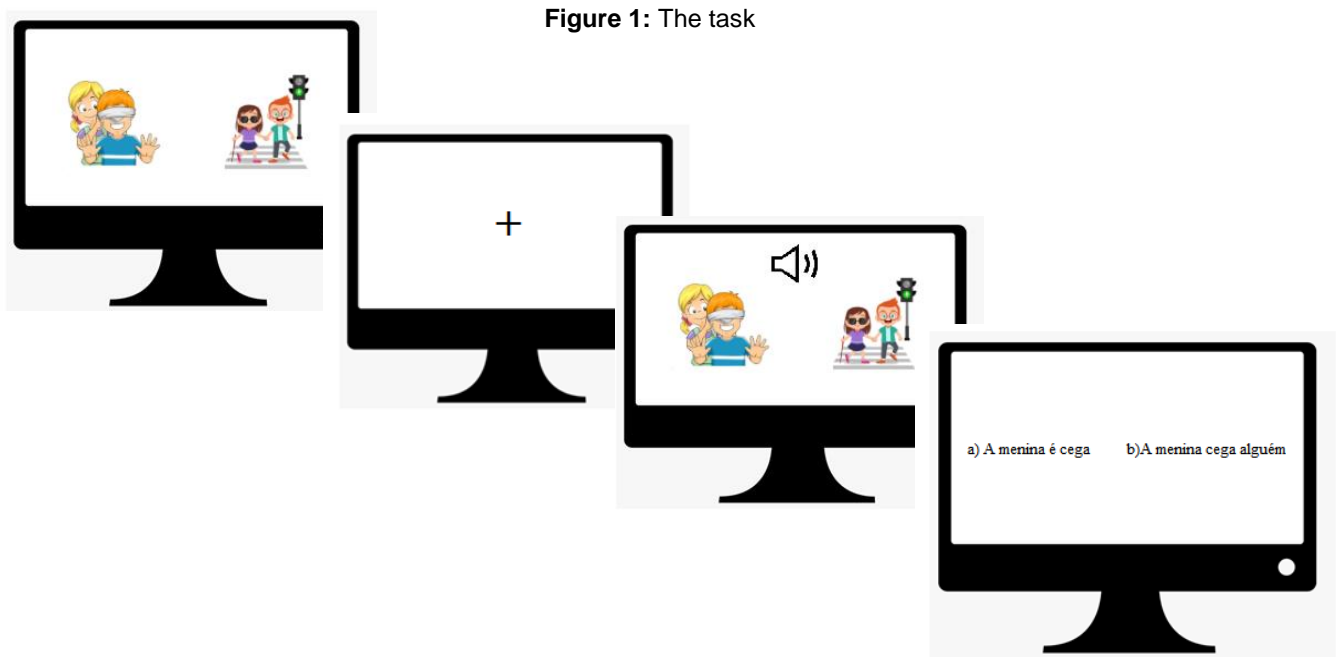
2.2 Procedures

The proposed questions were investigated through the eye-tracking technique. We used an EyeLink 1000 (SR Research) eye tracker, with a high-precision camera with 1000Hz accuracy, configured for monocular recording, coupled to a 32-inch screen, with 1920x1080 px resolution. A chin rest and nasal clip were used during the task, while the participants' eye movements were recorded, as they viewed the images and listened to the sentences.

The task asked of each participant was to listen to the audios and to visualize the images, while the eye tracker monitored their eye movements. After explaining the task to the participant, the experimenter started the participant's pupil calibration and validation process, in order to check the ocular fixations. After the calibration and validation process, the practice session began to familiarize them with the test. The participant was instructed to view images on the screen freely for 3 secondsⁱⁱ, then to listen to the sentences and simultaneously look at the image that represented the audio that was being heard and, later, to answer an interpretive question about the sentence heard. The participants were exposed to four practice trials as training, before performing the actual experiment. Participants freely visualized, for 3 seconds, the two images, which could represent two possible syntactic structures: a topic-comment structure and a subject-predicate structure. After the previous screening of the scenes, the images stopped appearing, and then a screen with a white background appeared with a fixation crossⁱⁱⁱ in the center of the screen, for one

second. After viewing the fixation cross, the auditory and the visual stimuli screen automatically appeared after one second. The audio started playing as soon as the image was loaded onto the screen, that is, audio and image appeared simultaneously. The participant was instructed to listen to the audio and to look at the corresponding image during the processing. After

finishing the audio, the participant should answer an interpretative question about the audio heard, as below:



Each session lasted between 30 minutes, depending on each volunteer and the good calibration of each one.

3. Results

For data analysis, we selected three regions of interest for the online measure, namely, **ROI1**, **ROI2** and **ROI3**. The same regions of interest were applied for both Match and Mismatch groups. The image that was related to the auditory input was the **target** and the image that was not related to the auditory input was the **distractor**. As we said, the dependent variable was the Proportion of looks to the images, during the time windows related to ROI1, ROI2 and ROI3.

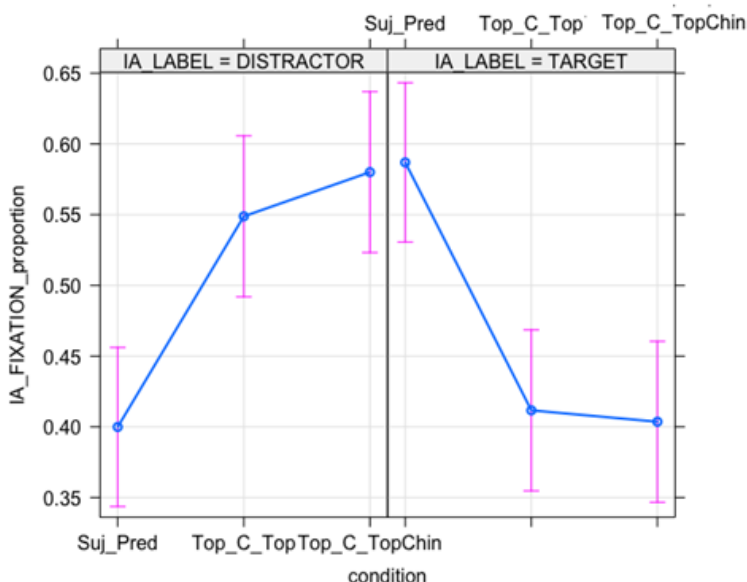
ROI1 refers to the time window that starts from the onset of the sentence to the offset of the ambiguous word, or duration of NP1. For example, in the sentence "The girl blinds her friend with the band to play with", in the Suj_Pred condition, the ROI1 is the excerpt "The blind girl". In the conditions Top_C_Top and Top_C_Chin, in the sentences "The blind girl, the friend helped to cross the street" and "The blind girl, the friend arrived when the traffic opened", respectively, the ROI1 is the excerpt "The girl blind, ...", after the pause contained after the topicalized element in topic-comment structures. By measuring ROI1, we will be able to verify whether, even in the initial auditory input processing, participants are able to identify the prosodic cues that disambiguate the NPs. Thus, when hearing the sentence "The blind girl" in the Suj_Pred condition, for example, if the participants are guided by the prosodic features contained in the auditory input at this point of processing, it is expected that the participants will focus more on the image that

represents the girl who is not blind, because, in this way, the listeners would be interpreting the word "blind" as a verb, which needs a complement. Under the conditions Top_C_Top and Top_C_Chin, participants are expected to focus their eyes on the image where a girl who is blind appears, because in this way, listeners would be interpreting the word "blind" as an adjective. The same prediction occurs for the both Match and Mismatch group. ROI2 measures the duration of NP2, "the friend". ROI2 is important, because it is at this point in the audio that listeners will complete their interpretation of the NP1. That is, if the listeners understood that the NP1 is a subject, they are expected to look at the image that shows a complement to this subject in an SVO structure, in this case the image that shows that "the girl blinds the friend", interpreting the NP2 as the complement of the verb. If the listeners understood NP1 as a topic, with the word "blind" being an adjective, it is expected that they will look to the image that shows "the blind girl, the friend...". In the Mismatch group, at this point, we will be able to observe if the participants are able to identify a mismatch in the auditory input, because if they hear what they do not expect as a complement to ROI1, we will observe a change in the pattern of fixations, indicating counter-expectation. ROI3 starts after the offset of NP2 until the end of the sentence. This part of the stimulus was important because the participants would complete the interpretation about they just heard incrementally at this point.

3.1 Results – Match group

GRAPH 1 shows de proportion of looks to the images (Target and Distractor) during the first region of interest (ROI1):

Graph 1: Proportion of looks at the images (Early looks) – ROI1
Match group



These data indicate that, at is point of the processing, participants seem not to use the prosodic cues of topic structures, because they looked more to the distractor image (the subject not the topic) when they received an input of topic in the **Top_C_Top** and **Top_C_Chin** conditions. In other words, it seems that the participants did not recognize the NP's as a topic. When the prosodic cue was a subject, the participants looked more to the target image (**Suj_Pred**), indicating that they recognize the NP as a subject. GRAPH 1 shows that the subject-predicate condition had a low percentage of looks to the distractor image, during ROI. We conducted a linear regression analysis in order to verify statistical differences between the conditions. We selected the **Suj_Pred** condition and the **Target** image as the Intercept (reference), thus, the other conditions were compared to this one, observing if the Target images could be more fixed in a certain condition compared to others. The result can be seen in the table below:

Table 2: Linear regression test - Match Group – ROI1

Conditions	Estimate	Std. Error	t value	Pr(> t)
Suj_Pred_Target X Top_C_Top_Target	-0.17530	0.04074	-4.303	2.01e-05 ***
Suj_Pred_Target X Top_C_Chin_Target	-0.18336	0.04074	-4.501	8.33e-06 ***

The test indicated a significant difference between the **Suj_Pred** condition and the **Top_C_Top** condition and also between the **Suj_Pred** condition and the **Top_C_Chin** condition, which confirms that the Target image of the **Suj_Pred** condition was the one that received the most fixations during the ROI1. We performed a *post-hoc tukey* test, in order to compare all pairs, in relation to the fixations towards both the Target images and the Distractor images, to verify, within each condition tested, which image was more fixed by the participants of the experiment, when listening ROI1. This result can be verified below:

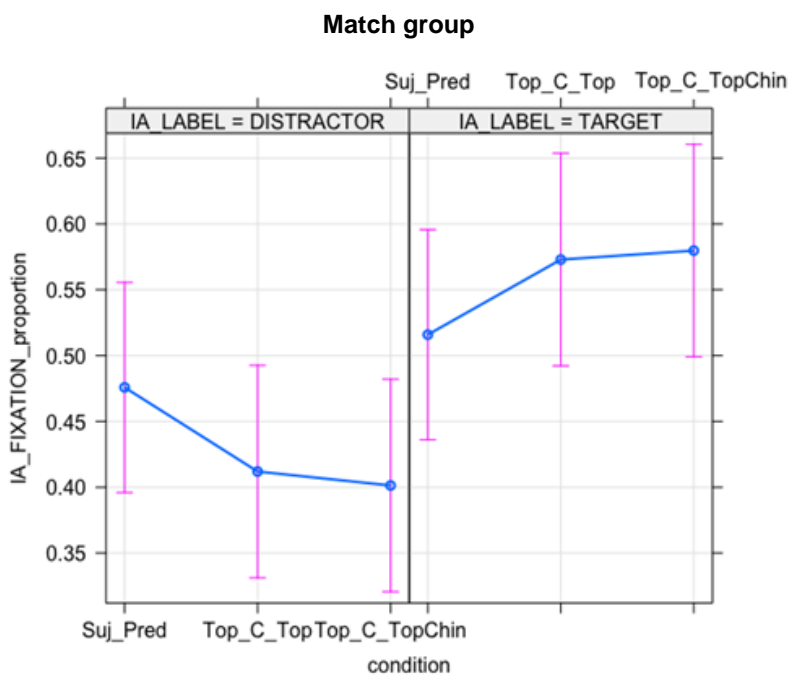
Table 3: Linear regression test (*Post-hoc tukey*) - **Match Group – ROI1**

Conditions	Estimate	Std. Error	df	t.ratio	p value
Suj_Pred_Target x Suj_Pred_Distractor	0.18704	0.0405	526	4.617	0.0001
Top_C_Top_Target x Top_C_Top_Distractor	-0.13720	0.0410	526	-3.349	0.0112
Top_C_Chin_Target x Top_C_Chin_Distractor	-0.17649	0.0410	526	-4.308	0.0003

When comparing the **Target x Distractor** images, the *Post-hoc tukey* test indicated a significant difference between the three conditions. In the **Suj_Pred** condition, the Target images were proportionally more fixed than the Distractors at the time of hearing the ROI1. However, in the topic-comment conditions, Target images were not proportionally more fixed than Distractors. The *post-hoc tukey* tests indicated that the Distractor images, that is, images that represented a subject-predicate structure, were more fixed. These data suggest that participants do not seem to use the prosodic cues of topic-comment structures as a guide at this early point of the online processing.

GRAPH 2 shows de proportion of looks to the images (Target and Distractor) during the second region of interest (ROI2):

GRAPH 2: Proportion of looks at the images (Late looks) – **ROI2**



There is an increase in fixations towards the Target image, under topic-comment conditions, with these fixations having more fixations than the Distractor images. This may suggest that, at this point in the auditory stimulus, the participants seem to relate correctly the prosodic cues to the visual stimulus, under these conditions. In the subject-predicate condition, there seems to have been a shift in the attentional process, in relation to the previous region, as it seems that the ocular fixations are not still focused on the Target image. In the graph, it is possible to verify that the Distractor images presented a low percentage of fixations, mainly in the topic-comment conditions, (Top_C_Top or Top_C_Chin). In the subject-predicate structure, on the other hand, we observed a percentage of fixations close to the margin in the Distractor image, during ROI1. The linear regression analysis has shown that the fixation proportions for the Target image, comparing the Suj_Pred (Intercept) condition with the other conditions, Top_C_Top and Top_C_Chin, were not statistically different. This result can be seen below:

TABLE 4: Linear regression test - Match Group – ROI2

Conditions	Estimate	Std. Error	t value	Pr(> t)
Suj_Pred_Target X Top_C_Top_Target	-0.06389	0.05780	-1.105	0.2695
Suj_Pred_Target X Top_C_Chin_Target	-0.07449	0.05780	-1.289	0.1980

This result is consistent, since, as it can be seen in GRAPH 2 above, in the subject-predicate condition, the Target and Distractor images were fixed in the same proportion.

We performed a *post-hoc tukey* test, comparing all pairs, in relation to the fixations towards both the Target images and the Distractor images. This result can be seen below:

TABLE 5: Linear regression test (*Post-hoc tukey*) - Match Group – ROI2

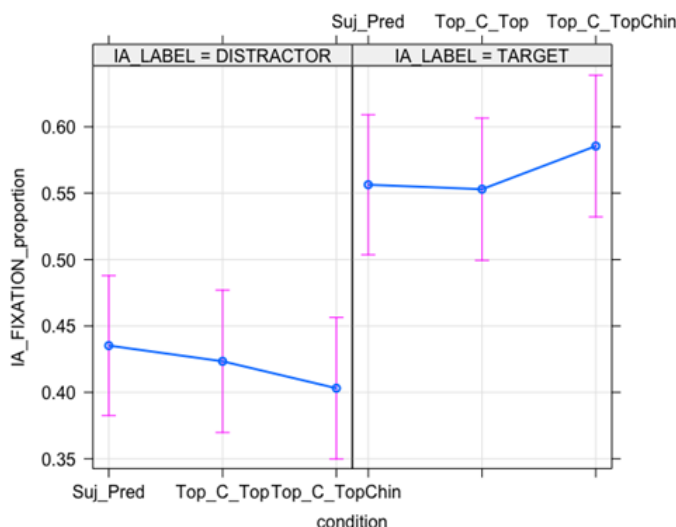
Conditions	Estimate	Std. Error	df	t.ratio	p value
Suj_Pred_Target x Suj_Pred_Distractor	-0.04003	0.0575	526	-0.697	0.9823
Top_C_Top_Target x Top_C_Top_Distractor	-0.16099	0.0581	526	-2.770	0.0640
Top_C_Chin_Target x Top_C_Chin_Distractor	-0.17841	0.0581	526	-3.070	0.0272

When comparing the Target x Distractor images, the statistical tests did not indicate a significant difference in the Suj_Pred condition ($p < 0.9823$). In the Top_C_Top condition, we did not obtain a significant difference ($p < 0.0640$), although we observed a visual difference towards a greater fixation in the Target images (58%) in contrast to the Distractors (41%). The Top_C_Chin condition presented a significant difference ($p < 0.0272$), with the Target image being more fixed than the Distractor image at this moment of the audio, referring to the hearing of the temporal window for region 2. The results of the analysis of ROI2 suggest that, at this point in the hearing of the stimulus, that is, during the hearing of NP2, when this element is placed as the subject of the comment, the participants seem to have understood NP1 as a topic in the Top_C_Chin condition, since there was a higher proportion of fixations in the Target image, which at an earlier point of hearing (during NP1) of the stimulus it was not possible to capture. In the condition of Top_C_Top, despite not having opened statistical significance in the comparison, the results were very close to what was observed in the condition of Chinese-style topic. In the Suj_Pred condition, at this point of the stimulus, when NP2 is placed as the complement of the verb, there seems to have been a shift in attention, since the greater proportion of fixations in the Target image was not maintained. However, we cannot say that listeners no longer consider such input as a subject.

GRAPH 3 shows the proportion of looks to the images (Target and Distractor) during the third region of interest (ROI3). In the graph, it is possible to verify that the Distractor images presented a lower percentage of fixations, in the three experimental conditions. Consequently, we observed a higher percentage of fixations in Target images, as below:

GRAPH 3: Proportion of looks at the images (Late looks) – ROI3

Match group



These data seem to indicate that, at the time of hearing the full sentence, the participants made the correct audio-image association during the online processing. Participants recognized and differentiated the different structures. The linear regression analysis was not statistically different, as the Target images were more fixed in the three conditions. Table 6 shows this result:

TABLE 6: Linear regression test - Match Group – ROI3

Condições	Estimate	Std. Error	t value	Pr(> t)
Suj_Pred_Target X Top_C_Top_Target	-0.011827	0.038251	-0.309	0.75730
Suj_Pred_Target X Top_C_Chin_Target	-0.032093	0.038141	-0.841	0.40048

After comparing the Intercept condition with the other conditions, we performed a post-hoc tukey test. This result can be seen below:

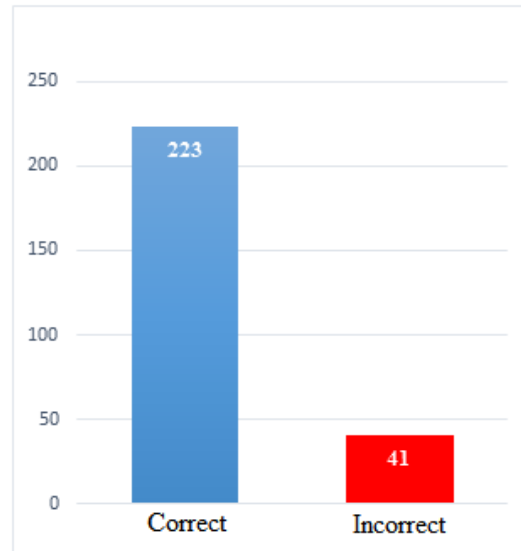
TABLE 7: Linear regression test (Post-hoc tukey) - Match Group – ROI3

Conditions	Estimate	Std. Error	df	t.ratio	p value
Suj_Pred_Target x Suj_Pred_Distractor	-0.12112	0.0379	524	-3.194	0.0185
Top_C_Top_Target x Top_C_Top_Distractor	-0.12958	0.0386	524	-3.359	0.0108
Top_C_Chin_Target x Top_C_Chin_Distractor	-0.18232	0.0384	524	-4.754	<.0001

There was a significant difference in the comparisons of the three experimental conditions, Suj_Pred ($p < 0.0185$), Top_C_Top ($p < 0.0108$) and Top_C_Chin ($p < 0.0001$), in the sense that the Target images were proportionally more fixed than the Distractors images, in ROI3. This result confirms that the participants correctly related audio and image, being able to differentiate the structures from each other.

The graph below shows the result of the offline measurement, referring to the response rate of the interpretative questions, after viewing and listening to the stimuli. This measure is a confirmation of the participants' interpretation of NP1, when judging the linguistic structure.

GRAPH 4: Response rate - Match group

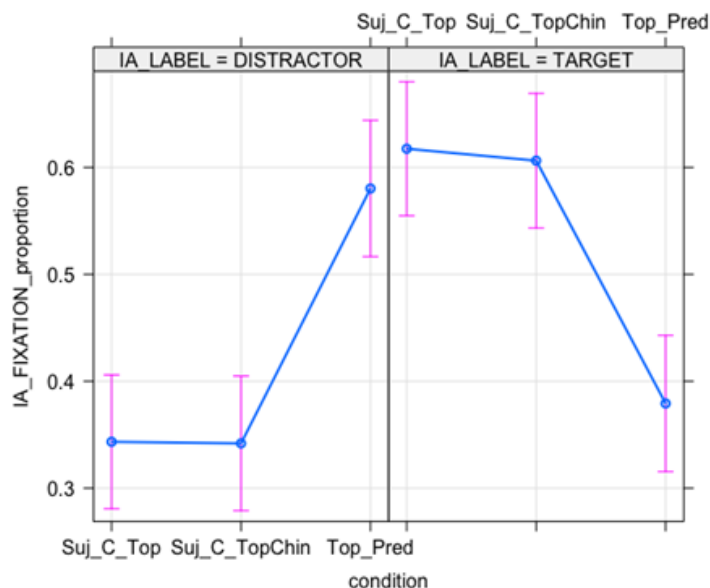


The graph above indicates a low rate of errors, in general. The experiment had 264 interpretative questions, 88 in each experimental condition. The **Suj_Pred** condition had the lowest occurrence of errors (7%). The **Top_C_Chin** was the condition that got the most errors (51%), then the **Top_C_Top** condition (42%). The Chi-square statistical test indicated significance among the three experimental conditions (X-squared = 13.851, df = 2, p-value = 0.0009825)

3.2 Results – Mismatch group

GRAPH 6 shows the proportion of looks to the images (Target and Distractor) during ROI1:

GRAPH 6: Proportion of looks at the images (Early looks) – ROI1
Mismatch group



The Distractor images presented a higher percentage of fixations only in the condition where the initial NP has a topic prosody, which is the Top_Pred condition. The structures that have the initial NP with a subject prosody, on the other hand, have a high percentage of fixations in the Target images. Such observation goes in the same direction as what was verified in the Match group, in the same region of interest. It seems that the fixations were directed to the image that represents a subject structure, because in all conditions there is a greater concentration of fixations in the image in which the NP represents a subject and not a topic. The linear regression test shows that the fixation proportions for the Distractor image only indicated a significant difference in comparison with the Top_Pred condition.

TABLE 8: Linear regression test - Mismatch Group – ROI1

Conditions	Estimate	Std. Error	t value	Pr(> t)
Suj_C_Top_DistractorX Suj_C_TopChin_Distractor	-0.001496	0.045178	-0.033	0.974
Suj_C_Top_Distractor X Top_Pred	0.236913	0.045444	5.213	2.69e-07 ***

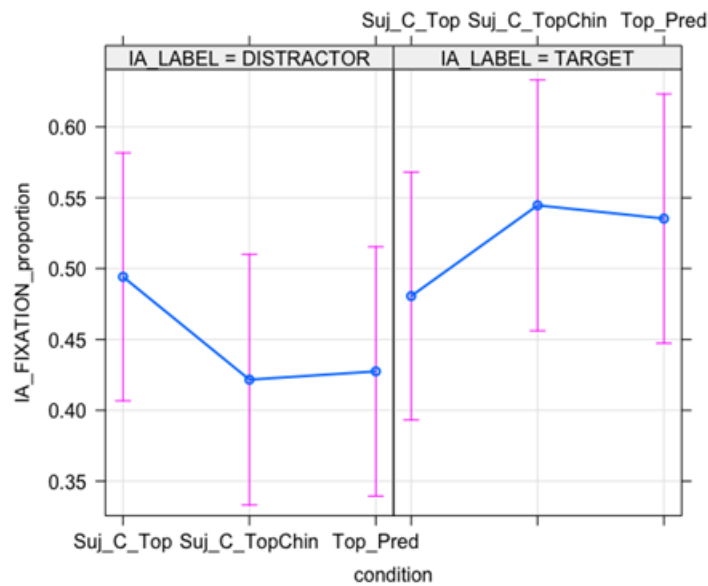
This result confirms that the participants, during the audition of ROI1, looked more at the Distractor image, when the input was a topic. These results are consistent with the data obtained in the Match group, which verified the same phenomenon at this point of the online processing. When comparing the Target x Distractor images in a post-hoc tukey test, we have a significant difference in the three experimental conditions. This result can be seen below:

TABLE 9: Linear regression test (*Post-hoc tukey*) - Mismatch Group – ROI1

Conditions	Estimate	Std. Error	df	t.ratio	p value
Top_Pred_Target x Top_Pred_Distractor	0.2011	0.0458	514	4.387	0.0002
Suj_C_Top_Target x Suj_C_Top_Distractor	-0.2741	0.0450	514	-6.084	<.0001
Suj_C_Chin_Target x Suj_C_Chin_Distractor	-0.2644	0.0453	514	-5.837	<.0001

Target images were proportionally more fixed than Distractors at the time of hearing ROI1, except in the Top_Pred condition, where NP1 had a topic prosody. These results suggest that participants do not seem to use the prosodic cues of topic-comment structures as a guide, at this starting point of online processing, as in the Match group.

GRAPH 7 shows de proportion of looks to the images (Target and Distractor) during the second region of interest (ROI2).

GRAPH 7: Proportion of looks at the images (Early looks) – ROI2**Mismatch group**

In the graph above, it is possible to verify that the Target and Distractor images, in the three experimental conditions, have a similar fixation distribution pattern, which will be confirmed in the statistical tests. There seems to be a similar distribution of fixations between the two images, as the fixation percentages, both for the Target image and for the Distractor image, were close. The linear regression test have shown that the proportions of fixations for the Distractor image were not different.

TABLE 10: Linear regression test - Mismatch Group – ROI2

Conditions	Estimate	Std. Error	t value	Pr(> t)
Suj_C_Top_Distractor X Suj_C_TopChin_Distractor	-0.07256	0.06332	-1.146	0.252
Suj_C_Top_Distractor X Top_Pred_Distractor	-0.06674	0.06314	-1.057	0.291

This result indicates that the Distractor images were fixed in the same proportion, in the three conditions tested. The post-hoc tukey tests did not indicate a significant difference in any of the conditions. That is, the fixations in the Target images were proportionally equal to the Distractors, at the moment of hearing the ROI2, in the three conditions. This result can be seen below:

TABLE 11: Linear regression test (Post-hoc tukey) - Mismatch Group – ROI2

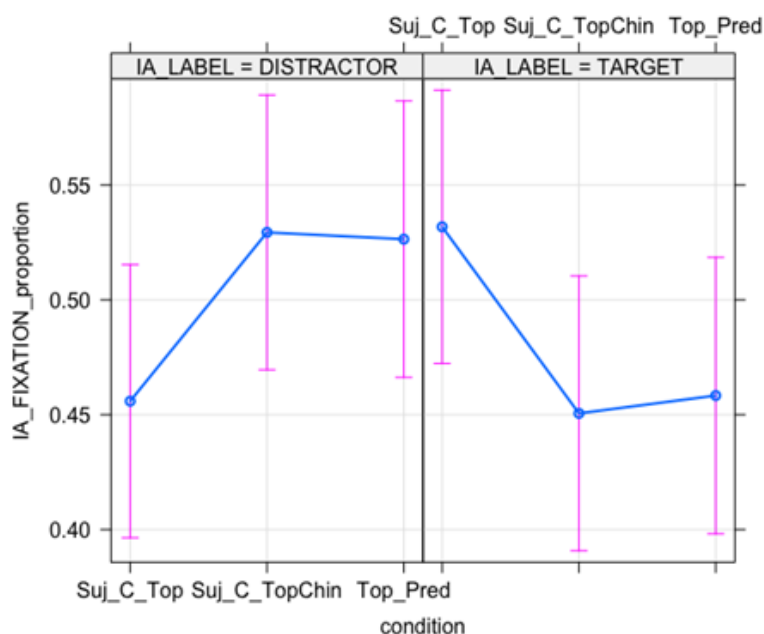
Conditions	Estimate	Std. Error	df	t.ratio	p value
Top_Pred_Target x Top_Pred_Distractor	-0.10785	0.0633	504	-1.703	0.5303
Suj_C_Top_Target x Suj_C_Top_Distractor	0.01357	0.0630	504	0.216	0.9999
Suj_C_Chin_Target x Suj_C_Chin_Distractor	-0.12302	0.0637	504	-1.931	0.3842

As verified above, there was no statistically significant difference between the proportion of fixations between the Target and Distractor images in any condition. The result of the analysis of the ROI2 suggests that, at this point of hearing the stimulus, that is, while listening to NP2, the participants identified problems in complementing NP1, as we verified that both images were fixed in the same proportion. This suggests that there was a strangeness generated by the participants' counter-expectations regarding what was being heard.

In the Match group, participants were able to associate image and audio, at this point in the processing under topic-comment conditions, because input and complement (NP1 + NP2) were synchronous. In the case of the Mismatch group, as input and complement were not synchronous, there was a break in expectation, observed by proportionally equal fixations in the two images.

GRAPH 8 shows de proportion of looks to the images (Target and Distractor) during ROI3.

GRAPH 8: Proportion of looks at the images (Early looks) – ROI3
Mismatch group



There is a pattern similar to what was seen in ROI2: the distribution of fixations in the Target and Distractor images was similar, which indicates an oscillation between the two images. This observation will be confirmed by statistical tests, which maintains the pattern observed during ROI2. The linear regression test below shows that the Distractor images were fixed at the same proportion, in the three experimental conditions.

TABLE 12: Linear regression test - Mismatch Group – ROI3

Conditions	Estimate	Std. Error	t value	Pr(> t)
Suj_C_Top_Distractor X Suj_C_TopChin_Distractor	0.07346	0.04293	1.711	0.0877
Suj_C_Top_Distractor X Top_Pred_Distractor	0.07051	0.04306	1.638	0.1021

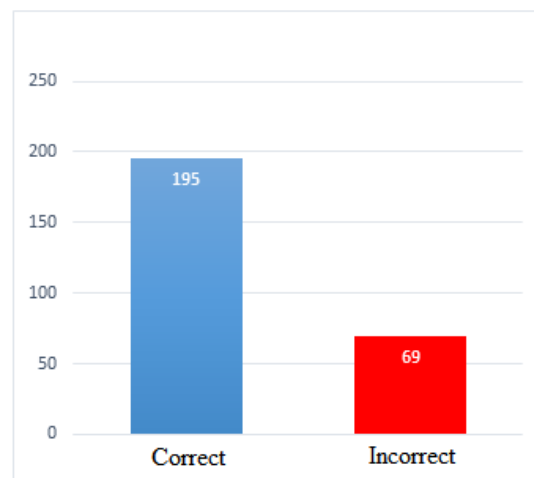
The post-hoc tests did not indicate significant difference in any of the three conditions, comparing the Target and Distractor images:

TABLE 13: Linear regression test (*Post-hoc tukey*) - **Mismatch Group – ROI3**

Conditions	Estimate	Std. Error	df	t.ratio	p value
Top_Pred_Target x Top_Pred_Distractor	0.06808	0.0433	510	1.572	0.6175
Suj_C_Top_Target x Suj_C_Top_Distractor	-0.07588	0.0428	510	-1.773	0.4845
Suj_C_Chin_Target x Suj_C_Chin_Distractor	0.07878	0.0431	510	1.830	0.4475

The result of the analysis of ROI3 maintains what was verified in ROI2, that is, the participants showed strangeness regarding the correct complementation of the structure, caused by the lack of synchrony between the beginning of the input and its continuation. This strangeness was captured by verifying the proportion of fixations equally towards both images, which represent the two possible structures. Here we confirm more fully the issue of breach of expectation. It seems that the participants are able to perceive that something is not correct in the prosodic formation of the structure.

GRAPH 9 shows the result of the offline measurement, referring to the response rate of the interpretative questions, after viewing and listening to the stimuli. The experiment had 264 interpretative questions, 88 in each experimental condition. It can be seen that the number of errors was about 1/3 of the general number of questions, an increase in relation to the number of errors in the Match group.

Graph 9: Response rate - **Mismatch group**

The Suj_C_Chin was the condition that got the most errors (55%), then the Suj_C_Top (27%) condition. The Top_Pred condition was the condition that had the lowest occurrence of errors (22%). The Chi-square statistical test indicated significance between the three experimental conditions ($\chi^2 = 12,639$, $df = 2$, $p\text{-value} = 0.001801$).

4. Discussion

In summary, the analysis of the Match group revealed that the prosodic cues present in the auditory

inputs guided, in part, the participants during the processing at the beginning of the structure. That is, during NP1 listening, the prosodic cue seems not to have influenced the participants to recognize the NP as a topic when listening to the audio. However, it is not possible to state that the participants did not recognize the elements as a topic, because in the immediately subsequent NP, the participants made the correct association with the images that represented the topic-comment structures, demonstrating that they recognized the structures as such, therefore in NP1, despite not having seen the corresponding images at

that time. At the end of the sentence, the participants recognized the three structures and correctly associated them with the corresponding images. The result of the offline measurement confirmed the correct identification of the SN1 structures by the participants, through the low rate of errors in the interpretative questions. The analysis of Mismatch group revealed that, although we did not capture it, during the NP1, the participants' use of the prosody clue to recognize the element as a topic, the NP2 processing revealed that the participants had this awareness. This observation was consistent with the findings of the Match group. During NP2 listening, we observed an oscillation of fixations between the two images in the Mismatch experiment, which may have been due to a break in expectations regarding NP1 complementation. In the Match experiment, participants looked for images corresponding to the topic-comment structure, being able to correctly associate audio and image by identifying the NP2 within the topic-comment structures. In the Mismatch group, we observed a large oscillation between the two images, caused by the lack of synchrony in the parts that make up the structures, which allows us to infer that the participants recognized, even in NP1, the relevant prosodic clues, although they did not demonstrate it, by viewing the images. At the end of the sentence, similarly to the eye fixation pattern identified during NP2 listening, the participants keep an oscillation between the two images, indicating again, a break in expectation. Participants can perceive a lack of harmony in the structure's prosodic formation. This is also reflected in the offline measure, which shows an error rate of 1/3 of the total number of questions and an increase in the number of errors, relative to the Match group.

5. Conclusions

This study has shown that Brazilian Portuguese speakers are sensitive to prosodic cues present in auditory stimuli. That is, when they received an auditory input that favored a subject element, the speakers focused more on the image that contained a subject NP (Early looks) and chose the option that represented a subject-predicate sentence (Late looks).

On the other hand, when listening to an auditory input that favored a topic element, the speakers focused more on the image that contained a topic NP and selected the option that contained a topic-comment sentence. Participants in the Mismatch group demonstrated higher cost during the choice of the image when input and complement were not synchronous, indicating that they were able to perceive the mismatch sentences. These results suggest a classification for BP as a mixed language, since the native speakers of BP were able to recognize and differentiate both structures, topic-comment and subject-predicate, as natural in their language, being able to perceive when there is something wrong in the structures during the online processing.

References

- Berlinck, R.; Duarte, M. E. L.; Oliveira, M. (2015). Predicação. In M. A. Kato & M. Nascimento (Eds.), *Gramática do Português Culto Falado no Brasil: a construção da sentença* (v. II pp. 81-149). Contexto.
- Duarte, I. (1996). A topicalização no português europeu: uma análise comparativa. In I. Duarte & I. Leiria (Eds.), *Actas do Congresso Internacional sobre o Português*. APL/Colibri.
- Kato, M. (2006). Comparando o português da América com o português de Portugal e com outras línguas. Museu da língua Portuguesa. <https://www.museudalinguaportuguesa.org.br/biblioteca/comparando-o-portugues-da-america-com-o-portugues-de-portugal/>
- Kenedy, E. (2002). *Aspectos estruturais da relativização em português: uma análise baseada no modelo raising*. UFRJ.
- Kenedy, E. (2014, jun./dez). O status tipológico das construções de tópico no português do Brasil: uma abordagem experimental. *Revista da ABRALIN*, 13, 2, 151-183. <https://revista.abralin.org/index.php/abralin/artic le/view/1190>
- Li, C-N. & Thompson, S. (1976). Subject and topic: a new typology of language. In LI, C-N (Ed.). *Subject and Topic* (pp. 457-489). Academic Press.

- Negrão, E. V. (1999). *O português brasileiro: uma língua voltada para o discurso*. USP: SP.
- Nespor, M; Vogel, I. (2007). *Prosodic Phonology: with a new foreword*. Mouton de Gruyter.
- Orsini, M; Vasco, S. (2007). Português do Brasil: língua de tópico e de sujeito. In: *Diadorim – Revista de estudos linguísticos e literários*, 2, 83-98. UFRJ, Programa de Pós-Graduação em Letras Vernáculas.
- Pontes, E. (1987). *O tópico no Português do Brasil*. Ed. Pontes.
- Salverda, A.P; Tanenhaus, M.K. (in press). (2017). The visual world paradigm. In A.M.B. de Groot & P. Hagoort (Eds.), *Research methods in psycholinguistics: A practical guide*. Wiley-Blackwell.
- Medeiros, L. da S. N. (2021). *Processamento de construções de tópico-comentário e sujeito-predicado no português no Brasil: língua orientada para a sentença, para o discurso ou mista? Tese*. Programa de Pós-Graduação em Linguística, UFRJ.
- Yuan, B. (1995). Acquisition of base-generated topics by English-speaking learners of Chinese. *Language Learning*, 45, 567–603.

ⁱ Such a design is justified because we tested Noun-Verb homographs like *cega* (blind), which means “someone who is blind” or “to blind someone” in BP, depending on the context, and there are not many words available to use for a bigger set of materials.

ⁱⁱ This procedure was done so that participants could perform a quick screening of the scenes beforehand, to know and become familiar with the scenes, before listening to the audios, since, at the time of hearing the stimulus, they could look directly and simultaneously at the image corresponding to the audio. If the participants did not know the scene beforehand, their gaze could be “lost on the screen”, that is, we would not be able to say for sure whether the look in a certain region was for the purpose of identifying and associating with what they were listening to, or if it was merely in order to carry out a screening for understanding/familiarization of the scenes. In addition, participants could not pay attention to the auditory stimulus and pay more attention to the visual stimulus for scene recognition, if they were not previously aware of this stimulus, having to choose whether to hear the audio or see what the image was about.

ⁱⁱⁱ This was done to ensure that after tracking the scenes, the participants’ gaze was attracted to a neutral region, outside the two images that represented the structures that would be heard, so that when the audio appeared, they could look directly at the proper image.