Lexical prediction mechanisms in Brazilian Portuguese: addressing methodological issues

Kaylee Fernandez

Lexical prediction mechanisms in Brazilian Portuguese: addressing methodological issues

# 1 Introduction

* The present study is a self-paced reading experiment examining whether grammatical gender markings in Brazilian Portuguese are used to make predictions during language comprehension.

# 2 Research Questions

**RQ 1.** Can adult Portuguese speakers predict specific lexical items during reading using grammatical gender cues?

# 3 Hypotheses

Regarding RQ1, it was predicted that participants would use grammatical gender markings to make predictions during self-paced reading.

# 4 Methods

## 4.1 Participants

There were 339 participants between 18 and 30 years old included in the study. Originally, there were 343 participants, but 2 participants were excluded for having completed the experiment more than once and 2 participants were excluded because they reported not having completed high school. All participants reported Brazilian Portuguese as a first language. Participants were recruited within the academic communities of the Federal University of Rio Grande do Norte and the Federal University of Minas Gerais through an advertisement sent by email to institutional emailing lists.

## 4.2 Materials

There were 20 experimental items with two conditions in the self-paced reading task: predictable and unpredictable. The predictable condition contained an expected critical noun while the unpredictable condition contained a plausible but unexpected critical noun. Nouns always had different grammatical genders across conditions and were classified as predictable or unpredictable on basis of the results of a previous experiment. Item dispalay order was randomized during the experiment.

An example item in both conditions is provided in (1).

* 1. Predictable condition **Marina estava dirigindo e percebeu que havia passado por cima de um prego. Ela parou e pegou as ferramentas para trocar om velhom e gastom pneum o mais rápido possível.** “Marina was driving when she realized that she had run over a nail. She pulled over and got the tools to change them old(m) and worn-out(m) tire(m) as quickly as possible”
  2. Unpredictable condition **Marina estava dirigindo e percebeu que havia passado por cima de um prego. Elaparou e pegou as ferramentas para trocar af velhaf e gastaf rodaf o mais rápido possível.** “Marina was driving when she realized that she had run over a nail. She pulled over and got the tools to change the(f) old(f) and worn-out(f) wheel(f) as quickly as possible”

It was predicted that unpredictable nouns would have a longer reading times. It was also anticipated that if critical nouns were activated before they were shown, this effect would show up before the noun because the preceding words indicated the gender of the upcoming noun.

## 4.3 Procedure

Once participants clicked the link in the advertisements, they were directed to the experiment page on the Ibex Farm platform (Drummond, 2017). After reading the instructions, indicating consent and filling out a general demographic form, participants completed a practice round before the actual experiment began.

During sentence reading, only one word was revealed at a time, and participants used the space bar to move from one word to the next. After reading each sentence participants were asked a comprehension question, which could be answered by either pressing “1” or “2” on the keyboard or clicking the chosen answer, which was either “yes” or “no.” Comprehension questions were balanced so that half would have “yes” as the correct answer. Responses were recorded and later used in data analysis to control for attention to the task. After participants answered a question, a message informed them if they had chosen the correct answer or not. Following the completion of the experiment, participants could leave a comment about the task by filling out an embedded Google Form.

## 4.4 Data analysis

The data were analyzed in R using a generalized linear mixed-effects model with a binomial linking function. The model included *frequency of codeswitching* as the dependent variable and *proficiency in Romanian*, *generation status* and *participants’ gender* as fixed factors. High frequency of CS was coded as “1” and low frequency of CS was coded as “0.” Significance of main effects and all possible interactions were examined using hierarchical partitioning of the variance via nested model comparisons.

Lastly, I used R [Version 4.0.3; R Core Team (2020)] and the R-package *papaja* [Version 0.1.0.9997; Aust and Barth (2020)] for all the analyses.

# 5 Results

The two panels in *Figure 1* show the CS frequency of participants as a function of Romanian language proficiency, as well as generation (G1 or G2) and gender (female or male). The model that examined frequency of CS as a factor of proficiency provided the best fit (with Pr(>Chi) = 0.0007759). A visual analysis of the plots shows that for G1 participants, frequency of CS decreased as proficiency increased for both female and male participants. However, for G2 male participants, frequency of CS increased with proficiency, while for G2 female participants it decreased.

# 6 Discussion

This study looked at Romanian-Spanish bilinguals’ frequency of codeswitching and the potential influence of proficiency in Romanian, gender identity, and generation status.

# 7 References

Aust, F., & Barth, M. (2020). *papaja: Create APA manuscripts with R Markdown*. Retrieved from <https://github.com/crsh/papaja>

R Core Team. (2020). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>