

Lexical prediction mechanisms in Brazilian Portuguese: addressing methodological issues

Neemias Souza Filho (Federal University of Rio Grande do Norte) & Mahayana Godoy
(Federal University of Rio Grande do Norte)
neemiasfilho@ufrn.edu.br

A series of studies from the early 2000s found evidence suggesting that specific lexical items and some of their features (e.g., word form and grammatical gender) can be predicted during language comprehension (DeLong et al., 2005; Van Berkum et al., 2005). These studies consisted mainly of electroencephalography (EEG) experiments, but the work conducted in Dutch by Van Berkum et al. (2005) included a self-paced reading experiment, in which the authors reported larger reading times for sentences with unlikely but plausible nouns. Such effect was registered before the presentation of the noun itself in the stimuli.

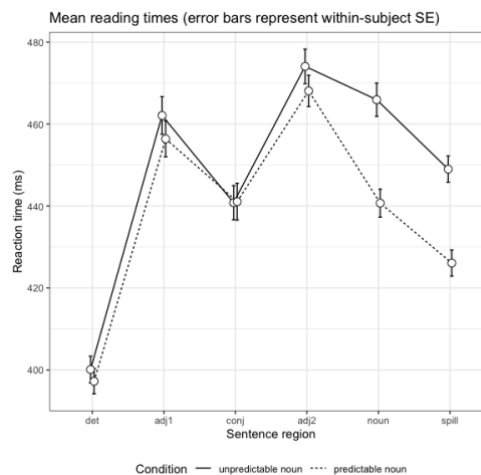
In an attempt to replicate these findings in Brazilian Portuguese (BP), we conducted a self-paced reading experiment with constraining contexts of high cloze probabilities, as well as controlled spillover effect regions both before and after critical nouns. Participants ($N = 339$) read 20 experimental items such as 1a and 1b (see second page for example stimuli), which, in their second sentences, contained likely and unlikely (though plausible) nouns, respectively. Nouns were categorized as likely or unlikely based on a previously conducted cloze test with 311 respondents.

The logic of the experiment lies in the fact that BP nouns have a binary grammatical gender, and accompanying articles and adjectives receive an agreeing inflection. This is indicated by the subscript notations in the example stimuli. Our hypothesis was that sentences such as 1b, which contain unlikely nouns, would have larger reading times for both the critical noun ('note') and the preceding inflected words ('the long and detailed'), since these indicate the gender of the upcoming noun. Such results would suggest that not only was it harder to process an unexpected noun, but that this difficulty could be attributed to the pre-activation of a different noun (or of its features), since it emerged prior to encountering the noun itself.

Analysis of critical words' log-transformed reading times was performed using linear mixed models with fixed effects of item condition (likely vs. unlikely), as well as random effects for items and participants whenever possible. Significant differences in the expected direction were observed for the critical noun ($p = 0.0017$, $b = -0.04$) and its following word ($p = 0.002$, $b = -0.05$) (see Figure 1), but not for any of the words preceding these two. As predicted in our pre-registered analysis plan, we also analyzed raw reading times using generalized linear mixed-effects models fit for a Gamma distribution. These models followed the same syntax as the ones adjusted for the log-transformed data. Once again, clear differences in the expected direction were observed for the critical noun ($p < 0.0001$, $b = -17.44$) and its following word ($p < 0.0001$, $b = -21.95$). Crucially, in this second analysis, a small but significant difference was registered at the conjunction region ("conj" in Figure 1), with slightly faster reading times for the predictable condition ($p = 0.03$, $b = -9.67$). The word in this region, namely the BP conjunction "e" ("and"), was constant across all items and conditions.

Although our results do corroborate our initial hypothesis that predictable nouns and features would be pre-activated, they most importantly seem to contribute to a growing body of evidence suggesting that pre-activation effects in language processing are likely very small even in EEG studies (e.g., Nieuwland et al., 2018; Nicenboim et al., 2020), and that the type of data analysis employed in studies on the subject must be thoughtfully considered. Additionally, when taken in context with previous self-paced reading experiments on the topic (e.g., Guerra et al., 2018), our results suggest that this paradigm requires considerably large samples for studies on predictive language processing (a conclusion in line with the one recently reported by Prasad & Linzen, 2020, for studies on syntactic adaptation).

Figure 1



Example stimuli

1a) The couple looked at the restaurant menu until they could make up their minds. They then called the waitress, who wrote down the_m long_m and detailed_m order_m on her pad. (Likely)

1b) The couple looked at the restaurant menu until they could make up their minds. They then called the waitress, who wrote down the_f long_f and detailed_f note_f on her pad. (Unlikely)

Original example stimuli in Brazilian Portuguese (BP)

1a) O casal olhou o cardápio do restaurante até conseguirem se decidir. Eles chamaram a garçonete, que anotou o_m extenso_m e detalhado_m pedido_m na comanda.

1b) O casal olhou o cardápio do restaurante até conseguirem se decidir. Eles chamaram a garçonete, que anotou a_f extensa_f e detalhada_f mensagem_f na comanda.

Critical words were controlled for length (number of syllables) in BP.

References

DeLong et al. 2005. Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*.

Guerra et al. 2018. A crack in the crystal ball: Evidence against pre-activation of gender features in sentence comprehension. AMLaP 2018, Berlin, Germany. <https://amor.cms.hu-berlin.de/~knoeferp/AMLaP2018/Program_files/AMLaP2018_proceedings.pdf>.

Nicenboim et al. 2020. Are words pre-activated probabilistically during sentence comprehension? Evidence from new data and a Bayesian random-effects meta-analysis using publicly available data. *Neuropsychologia* <<https://doi.org/10.1016/j.neuropsychologia.2020.107427>>.

Nieuwland et al. 2018. Large-scale replication study reveals a limit on probabilistic prediction in language comprehension. *eLife* <<https://doi.org/10.7554/eLife.33468>>.

Prasad & Linzen. 2020, February 21. Rapid syntactic adaptation in self-paced reading: detectable, but requires many participants. Preprint <<https://doi.org/10.31234/osf.io/9ptg4>>.

Van Berkum et al. 2005. Anticipating upcoming words in discourse: evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.