

Parts of Speech Shape Reading-Time Variability in Brazilian Portuguese

Diego Alves

Saarland University
Saarbrücken, Germany
diego.alves@uni-saarland.de

Abstract

This study uses regression analysis of Brazilian Portuguese eye-tracking data to examine variability in reading times across grammatical categories. Mixed-effects models reveal distinct patterns: numerals elicit high individual variability in early-stage reading, while function words (e.g., adpositions, determiners) drive differences in late-stage integration. In contrast, nouns show stable effects. These findings demonstrate that individual differences in reading are systematically linked to specific parts of speech, with numerals and function words as key loci of variability.

Keywords: reading time, Portuguese, parts-of-speech

1. Introduction

Theoretical models of reading have largely been built to explain systematic effects of linguistic input on eye movements, emphasising factors such as lexical frequency, predictability, and structural complexity.

Recent work have demonstrated that variation in cognitive resources, reading skill, and linguistic experience systematically modulates reading behaviour (e.g., [Haeuser and Kray, 2024](#); [Staub, 2021](#)). These findings motivate models in which eye movements reflect not only properties of the text, but also enduring characteristics of individual readers.

Empirical investigations of individual variability in reading have typically focused on how readers differ in their sensitivity to specific psycholinguistic phenomena, such as lexical frequency, word predictability, syntactic complexity, ambiguity resolution, and integration costs (e.g., [Kuperman et al., 2018](#); [Nicenboim et al., 2016](#); [Staub, 2021](#)). Variability is usually assessed in relation to experimental manipulations or item-level properties, rather than to the grammatical units over which these effects are realised. As a result, it remains largely unexplored whether individual differences in reading behaviour are distributed uniformly across grammatical categories, or whether certain parts of speech are more prone to variability than others.

This paper investigates the role of part-of-speech (PoS) categories in shaping reading behaviour in Brazilian Portuguese. Using eye-tracking data, we examine how different PoS influence multiple stages of lexical processing, as reflected in first fixation duration, first-run dwell time, and total dwell time. To assess these effects, we analysed inter-participant variability using linear mixed-effects models applied to word-level eye-tracking

measures, allowing us to estimate both average PoS effects and individual differences in sensitivity to grammatical category. By comparing these measures across readers, we aim to identify which PoS categories are most strongly associated with individual variability in reading times, thereby providing insight into the linguistic factors that drive individual differences in sentence processing.

2. Related Work

Individual variability in reading behaviour has become an important topic in psycholinguistic research, particularly in eye-tracking studies of sentence processing.

Previous work has shown that readers differ reliably in their sensitivity to specific perceptual and linguistic factors, including word frequency, predictability, visual contrast, and font difficulty, indicating that individual differences in eye movements reflect stable properties of the reading system rather than measurement noise ([Staub, 2021](#)). Other studies have linked variability in eye movements to reader characteristics, including age, demonstrating systematic differences in the use of contextual information during reading ([Haeuser and Kray, 2024](#)).

In parallel, psycholinguistic research has increasingly drawn on computational models of prediction, including surprisal-based approaches and large language models, to explain reading times across languages ([Wilcox et al., 2023](#); [Xu et al., 2023](#)). Although this work has provided strong evidence for the role of predictability in sentence processing, analyses typically focus on average effects and offer limited insight into how predictive processing varies between individuals.

Beyond reader-level factors, eye-tracking studies have also shown that gaze patterns are sensitive to

grammatical structure. In particular, eye-movement features have been shown to reliably distinguish between major parts of speech, such as nouns and verbs, suggesting that grammatical categories are associated with distinct processing profiles (Barrett and Søgaard, 2015). However, this work has largely focused on classification and representation, rather than on whether different parts of speech differentially contribute to inter-participant variability in reading time.

Finally, most research on individual differences in reading has been conducted in English. Comparatively little psycholinguistic work has examined how inter-participant variability is distributed across grammatical categories in other languages, including Brazilian Portuguese, which differs from English in both syntactic and morphological structure. Recent eye-tracking resources for Brazilian Portuguese make it possible to address this gap (Sardinha, 2010). In addition, recent work has shown that parts of speech differ systematically in their reading-time profiles as a function of information content, as estimated by surprisal from large language models (Alves, 2025). However, it remains unclear whether such PoS-specific processing differences are also associated with differences in inter-participant variability during reading.

3. Methodology

3.1. Eye-tracking Data

The RastrOS corpus was created to support psycholinguistic research in Brazilian Portuguese (BP), with a particular emphasis on lexical predictability and sentence processing. It consists of two primary components: predictability norms obtained through a Cloze task and eye-tracking data collected during reading experiments.

The Cloze task was completed by 393 native speakers of BP recruited from six Brazilian universities, most of whom were undergraduate students. Each participant filled in five randomly selected paragraphs, balanced across three text genres: journalistic (40%), literary (20%), and popular science (40%). The dataset is annotated with PoS tags (generated with the Palavras parser; Bick 2000) and word frequency measures derived from Corpus Brasileiro (Sardinha, 2010) and BrWaC (Wagner Filho et al., 2018). In addition, the corpus includes surprisal and entropy-reduction values computed from the Cloze responses.

Eye-tracking data were collected from 37 undergraduate students using an EyeLink 1000 eye-tracker with a sampling rate of 1000 Hz. Participants read the same 120 sentences included in the Cloze corpus (2,494 words; 2,831 tokens including punctuation). Each sentence is annotated with

36 eye-movement measures, including first fixation duration, first-run dwell time, and total dwell time.

For our analysis, we parsed the RastrOS sentences using the Stanza parser (Qi et al., 2020) and assigned each word a Universal Part-of-Speech (UPOS) tag according to the Universal Dependencies guidelines (De Marneffe et al., 2021). This annotation scheme was chosen to facilitate future cross-linguistic comparisons.

As Portuguese has contractions (e.g., *da*, composed of the adposition *de* combined with the determiner *a*, equivalent to “of the” in English), an alignment phase was necessary. For each contraction in the eye-tracking data, we retained the PoS tag of the head of the contraction.

4. Evaluation Method

To examine inter-participant variability in reading behaviour across parts of speech (PoS), we analysed three standard eye-tracking measures: first fixation duration, first-run dwell time, and total dwell time.

1. First fixation duration - the duration of the first fixation on a word during its first pass. Annotated as `IA_FIRST_FIXATION_DURATION` in RastrOS.
2. First-run dwell time - the sum of all first-pass fixations on a word. `IA_FIRST_RUN_DWELL_TIME` in RastrOS.
3. Total dwell time - the sum of all fixations on a word during the trial. `IA_DWELL_TIME` in RastrOS.

First fixation reflects the initial processing of a word and is associated with early stages of lexical access. Gaze duration captures the time spent on a word during first-pass reading and is sensitive to lexical and syntactic processing. Total fixation time includes any regressions back to the word and reflects later stages of comprehension, such as reanalysis or integration difficulties (Rayner, 1998).

All reading-time measures were log-transformed after adding a constant to reduce skewness. Analyses were conducted at the word level.

To avoid unstable coefficient estimates driven by sparse data, only UPOS categories appearing in three or more distinct sentences were retained for analysis; all other categories were excluded prior to model estimation.

For each reading-time measure, we fitted a linear mixed-effects regression model with part of speech (UPOS) as the predictor of interest as shown in Equation 1. The models included established lexical and contextual control variables: word length,

word frequency (Freq, log-transformed), and word surprisal (Srp) estimated using the LLaMA-3.2-3B transformer language model¹ (Alves, 2025). To account for spillover effects and temporal dependencies in eye movements, we additionally included lagged predictors for reading time, word frequency, and surprisal from the one and two preceding words.

Participant was included as a random effect, with both a random intercept and random slopes for UPOS, allowing the effect of part of speech on reading time to vary across individuals. This hierarchical structure captures individual differences in baseline reading speed as well as in sensitivity to grammatical category.

$$\begin{aligned} \log(RT) \sim & \text{UPOS} + \text{WordLength} + \text{Freq} + \text{Srp} \\ & + \log(RT_{-1}) + \log(RT_{-2}) \\ & + \text{Freq}_{-1} + \text{Freq}_{-2} \\ & + \text{Srp}_{-1} + \text{Srp}_{-2} \\ & + (1 + \text{UPOS} \mid \text{Participant}) \end{aligned} \quad (1)$$

To facilitate comparison of effect magnitudes across predictors, fixed-effect coefficients were fully standardized by refitting the models on variables that had been centred and scaled to unit variance, using the *effectsize* package in R. The resulting coefficients thus represent changes in reading time (in standard deviation units) associated with one standard deviation changes in the predictors.

Inter-individual variability in part-of-speech effects was quantified using the standard deviation of the corresponding random slopes estimated by the mixed-effects model. These random-slope standard deviations provide a model-based estimate of between-participant variability, reflecting individual differences in the strength of UPOS effects after partial pooling and accounting for noise in participant-specific estimates.

This modelling strategy allows us to simultaneously assess (i) the average effect of grammatical category on reading time and (ii) the extent to which these effects systematically vary across readers.

5. Results

Table 1 reports, for each UPOS category, the number of distinct Text–Sentence contexts in which it appears and the number of recording sessions in which it is attested.

As previously described, UPOS categories appearing in fewer than three sentences were not considered in the analysis; consequently, symbols (SYM) and interjections (INTJ) were excluded. All

UPOS	Sentences	Participants
NOUN	104	37
DET	99	37
VERB	97	37
ADP	91	37
ADJ	71	37
ADV	71	37
PRON	61	37
AUX	60	37
CCONJ	57	37
SCONJ	48	37
PROPN	40	37
NUM	27	37
SYM	2	35
INTJ	1	36

Table 1: Number of unique Text–Sentence pairs in which each UPOS appears, and number of unique recording sessions (participants) in which each UPOS is attested.

remaining UPOS categories occur in more than 25 distinct sentences and were read at least once by all participants.

Regarding the predictors that are not UPOS in equation 1, as expected, across all three reading-time measures, word length showed a reliable positive effect and lexical frequency a reliable negative effect, indicating longer reading times for longer and less frequent words.

Surprisal also exhibited a consistent positive effect across measures. Autoregressive effects of previous reading times were robustly positive in all models. In contrast, spillover effects of frequency and surprisal varied across measures, with early measures showing facilitation from previous-word frequency, whereas total dwell time showed positive lagged frequency effects and largely absent surprisal spillover.

Figures 1, 2, and 3 present the effects of UPOS on the three reading-time measures analysed: first fixation duration, first-run dwell time, and total dwell time, respectively. For each measure, bars show standardized fixed-effect estimates from the linear mixed-effects models, with error bars indicating 95% confidence intervals. Point size represents the standard deviation of participant-specific UPOS effects estimated by the model, reflecting the extent to which the influence of each UPOS category varies across readers.

The figures suggest that grammatical category exerts a stronger influence on total dwell time than on earlier reading-time measures. Eight UPOS categories show statistically reliable positive effects on total dwell time, compared to only three categories for first-fixation duration (two negative and one positive) and four for first-run dwell time (three positive and one negative).

¹<https://huggingface.co/meta-llama/Llama-3.2-3B>

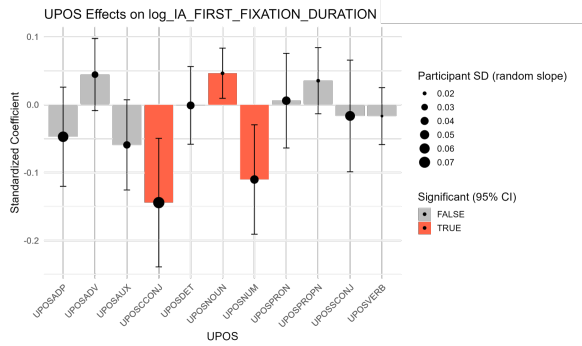


Figure 1: Standardized fixed effects of UPOS on log first fixation duration. Bars show fixed-effect estimates ($\pm 95\%$ CI); point size indicates the standard deviation of participant-level random slopes.

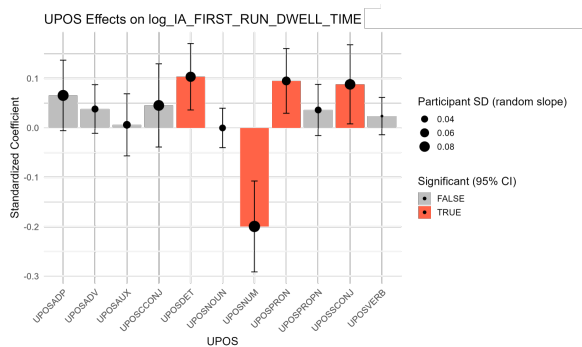


Figure 2: Standardized fixed effects of UPOS on log total dwell time. Bars show fixed-effect estimates ($\pm 95\%$ CI); point size indicates the standard deviation of participant-level random slopes.

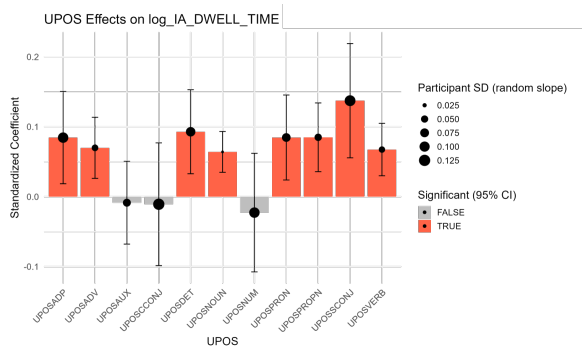


Figure 3: Standardized fixed effects of UPOS on log first-run dwell time. Bars show fixed-effect estimates ($\pm 95\%$ CI); point size indicates the standard deviation of participant-level random slopes.

Across measures, the overall magnitude of inter-participant variability is broadly comparable, although it tends to reach higher levels for total dwell time, indicating greater individual differences at later stages of processing.

For first-fixation duration, reliable effects are lim-

ited to a small number of categories: numerals (NUM) and coordinating conjunctions (CCONJ) show negative effects accompanied by substantial inter-participant variability, whereas nouns (NOUN) exhibit a positive effect with comparatively low variability across readers.

A similar pattern of high variability is observed for numerals in first-run dwell time, where they again show a negative effect, although this effect does not extend to total dwell time. Additional effects on first-run dwell time are observed for subordinate conjunctions (SCONJ), pronouns (PRON), and determiners (DET), all of which display considerable between-participant variability.

In contrast, total dwell time reveals both a larger number of reliable UPOS effects and clearer differentiation in inter-participant variability. Adpositions (ADP), determiners (DET), and subordinate conjunctions (SCONJ), pronouns (PRON), and determiners (DET), all of which display considerable between-participant variability.

With the exception of numerals, the PoS categories showing greater variability are predominantly function words, which are typically short. These words frequently occur at the beginning of phrases or clauses, a position that is also commonly associated with the presence of filler particles. Filled pauses have been shown to occur preferentially at utterance- and phrase-initial positions, suggesting that they are linked to structural planning and boundary marking in speech production and comprehension (Maclay and Osgood, 1959).

Numerals occur in both numerical and written forms and fulfil a range of syntactic functions, including nominal subjects (*nsubj*), nominal modifiers (*nmod*), numeric modifiers (*nummod*), and oblique arguments (*obl*). Their effects appear to be confined to early stages of processing and show substantial variability across participants, suggesting that a more fine-grained analysis incorporating syntactic function may be informative.

6. Conclusion and Future Work

This paper investigated how part-of-speech (PoS) categories influence reading behaviour and its variability across individuals in Brazilian Portuguese.

Analysing eye-tracking data, we found that grammatical category has a stronger effect on late integration measures (total dwell time) than on early lexical access. Crucially, inter-participant variability was not uniform: function words (e.g., adpositions, determiners) and numerals showed the greatest individual differences, while content words such as nouns were more stable.

These results indicate that readers vary most in their processing of syntactic and structural elements, highlighting the need for models of reading

to account for how individual cognitive systems interact with the grammatical architecture of language.

Future work should extend this approach to other languages and incorporate finer-grained syntactic information, as individual UPOS categories can fulfil distinct syntactic functions that may differentially shape reading behaviour.

7. Limitations

The present study has limitations that should be considered when interpreting the results. First, the eye-tracking experiment was conducted exclusively with undergraduate students, resulting in a relatively homogeneous participant sample. This limits the extent to which the observed inter-individual variability can be generalised to broader populations with greater diversity in age, educational background, and reading experience. Future work including more heterogeneous reader groups may reveal different patterns of variability.

Second, the corpus used in the experiment was limited in terms of textual diversity, drawing from a restricted set of sources. Reading behaviour, and its variability, may differ across registers, genres, or communicative contexts, particularly with respect to the processing of function words and syntactic structure. Expanding the range of text types would therefore be important for assessing the robustness and generality of the observed effects.

8. Bibliographical References

- Diego Alves. 2025. Benchmarking language model surprisal for eye-tracking predictions in Brazilian Portuguese. In *Proceedings of the First International Workshop on Gaze Data and Natural Language Processing*, pages 7–17.
- Maria Barrett and Anders Søgaard. 2015. Reading behavior predicts syntactic categories. In *Proceedings of the nineteenth conference on computational natural language learning*, pages 345–349.
- Eckhard Bick. 2000. *The parsing system palavras: Automatic grammatical analysis of Portuguese in a constraint grammar framework*. Aarhus Universitetsforlag.
- Marie-Catherine De Marneffe, Christopher D Manning, Joakim Nivre, and Daniel Zeman. 2021. Universal dependencies. *Computational linguistics*, 47(2):255–308.
- Katja I Haeuser and Jutta Kray. 2024. Age differences in context use during reading and downstream effects on recognition memory. *Psychology and Aging*.
- Victor Kuperman, Kazunaga Matsuki, and Julie A Van Dyke. 2018. Contributions of reader- and text-level characteristics to eye-movement patterns during passage reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 44(11):1687.
- Howard Maclay and Charles E Osgood. 1959. Hesitation phenomena in spontaneous English speech. *Word*, 15(1):19–44.
- Bruno Nicenboim, Pavel Logačev, Carolina Gattei, and Shravan Vasishth. 2016. When high-capacity readers slow down and low-capacity readers speed up: Working memory and locality effects. *Frontiers in psychology*, 7:280.
- Peng Qi, Yuhao Zhang, Yuhui Zhang, Jason Bolton, and Christopher D Manning. 2020. Stanza: A Python natural language processing toolkit for many human languages. *arXiv preprint arXiv:2003.07082*.
- Keith Rayner. 1998. Eye movements in reading and information processing: 20 years of research. *Psychological bulletin*, 124(3):372.
- Tony Berber Sardinha. 2010. Corpus brasileiro. *Informática*, 708:0–1.
- Adrian Staub. 2021. How reliable are individual differences in eye movements in reading? *Journal of Memory and Language*, 116:104190.
- Jorge A Wagner Filho, Rodrigo Wilkens, Marco Idiart, and Aline Villavicencio. 2018. The brWaC corpus: A new open resource for Brazilian Portuguese. In *Proceedings of the eleventh international conference on language resources and evaluation (LREC 2018)*.
- Ethan G Wilcox, Tiago Pimentel, Clara Meister, Ryan Cotterell, and Roger P Levy. 2023. Testing the predictions of surprisal theory in 11 languages. *Transactions of the Association for Computational Linguistics*, 11:1451–1470.
- Weijie Xu, Jason Chon, Tianran Liu, and Richard Futrell. 2023. The linearity of the effect of surprisal on reading times across languages. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 15711–15721.