EYE-MOVEMENT PATTERNS IN ROMANIAN TEXT READING: A LINGUISTIC PERSPECTIVE

Horga Daria[1] **and Popeanga Antonia Maria**[2]

[1]Faculty of Mathematics and Computer Science, University of Bucharest

*contact: daria.horga@s.unibuc.ro, antonia-maria.popeanga@s.unibuc.ro

Eye-tracking research has consistently shown that reading behavior is shaped by linguistic factors such as word length and grammatical category (Luke and Christianson, 2017). However, most studies have focused on languages with relatively simple morphology, leaving morphologically rich languages comparatively underexplored. This study investigates reading behavior in Romanian, a language characterized by complex inflectional morphology and clitic constructions, using eye-tracking data from the MultiplEYE corpus (Jakobi, 2025).

To support this analysis, we first detected fixations using the pymovements library (Krakowczyk *et al.*, 2023), and subsequently implemented a velocity-based procedure for saccade detection, together with a set of metrics capturing the spatial and temporal characteristics of fixations. Building on this preprocessing pipeline, we then developed predictive models that integrate standard eye-tracking metrics with variables tailored to Romanian morphology, including enclitic definite articles (-ul, -ului, -lor), clitic pronoun constructions (s-au, și-a), noun case and number marking, diacritics (ă, â, î, ș, ț), hyphenated forms (după-amiază). In this pilot study, we evaluated the predictability of eye movements using classifiers for word skipping, fixation duration, and regression events. To ensure generalizability across reading patterns, all models were trained using a subject-independent leave-one-subject-out validation framework.

Our results show that word skipping in Romanian is primarily driven by word length and frequency, with additional linguistic features—including function word status and clitic constructions—providing further predictive signal beyond these factors (ROC–AUC ≈ 0.78).

Fixation duration, modeled as a classification task (short, medium, long; ROC–AUC 0.75), is primarily influenced by word length and word position. Romanian-specific features, such as enclitic articles, clitic constructions, and case marking, provide additional predictive value, while diacritics also contribute, likely reflecting increased visual and orthographic processing demands.

For regression events, the models achieved a mean ROC-AUC ≈ 0.68. These regressions are influenced by both general and language-specific features, including fixation duration and surprisal (Jakobi *et al.*, 2023). Shorter regressions tend to land on oblique cases and function words, while longer ones target content words, suggesting a greater loss of context.

In conclusion, this study takes a first step toward addressing the scarcity of Romanian eye-tracking research, emphasizing the value of incorporating language-specific cues for more linguistically informed reading models.

References

Jakobi D. N., Kern T., Reich D. R., Haller P., Jäger L. A. (2023). Potec: A german naturalistic eye-tracking-while-reading corpus. *Scientific Data 10*(1), 640. doi:10.1038/s41597-023-02534-1.

Jakobi D. N. e. a. (2025). Multipleye: Creating a multilingual eye-tracking-while-reading corpus. In *Proceedings of the 2025 Symposium on Eye Tracking Research and Applications (ETRA '25)*, New York, NY, USA. Association for Computing Machinery.

Krakowczyk D. G., Reich D. R., Chwastek J., Jakobi D. N., Prasse P., Süss A., Turuta O., Kasprowski P., Jäger L. A. (2023). pymovements: A python package for processing eye movement data. In *Proceedings of the 2023 Symposium on Eye Tracking Research and Applications*, New York, NY, USA. Association for Computing Machinery. URL: https://link.springer.com/article/10.3758/s13428-017-0908-4.

Luke S. G., Christianson K. (2017). Predictors of eye movements during reading: A replication of rayner, slattery, drieghe, and liversedge (2011) with english readers. *Behavior Research Methods 49*(2), 623–631. URL: https://pmc.ncbi.nlm.nih.gov/articles/PMC12208991/#CR55.