The effect of spelling entropy on reading: an eye tracking study with reference to Russian

Polina Maksimova[1], *Daria Chernova*[1]

[1] *Institute of Cognitive Studies, St. Petersburg State University, Russia*

During word recognition in reading or during written production, the mental lexicon is accessed, where a word's lexical representation—namely its orthographic, phonological, and semantic information—is stored. A high-quality lexical representation is formed through the simultaneous activation of these connections, whereas instability in the graphic or phonological form of a word impedes its development. As the volume of unedited content increases, the frequency of misspellings rises, and such erroneous forms may become entrenched in the mental lexicon. Orthographic errors are considered a consequence of weak orthographic representations, while alternative spellings increase entropy, thereby slowing the recognition even of correctly spelled words [1].

The present paper reports the results of an experiment designed to test the effects of orthographic entropy and the stability of a word's orthographic form in sentence context, which models conditions of natural reading. We hypothesize that higher entropy values for a word will be associated with greater difficulty in recognizing that word. For the experiment, we constructed sentences containing three variants of a target word: correct spelling, spelled with frequent error (according to The General Internet-Corpus of Russian (GICR)[2]), spelled with infrequent error. In this way, we compare frequent misspellings with other distortions of the orthographic form to test whether such spellings become entrenched in the mental lexicon. These words are presented in context, so, in this experiment, our participants do a reading for comprehension task while we record their eye movements. Using a range of eye tracking measures, we assess the difficulty of word processing in context as a function of the orthographic entropy coefficient and the frequency of incorrect spelling. For example, first-pass reading time is an early measure reflecting lexical access, whereas regressions are a late measure reflecting integration with context. To assess individual differences in reading experience, an Author Recognition Test [3], adapted for Russian [4], is used. Data collection is currently ongoing.

The results obtained will allow us to draw conclusions about the extent to which alternative spellings can become entrenched in a speaker's mental lexicon, how they may affect lexical access during reading, and how these processes depend on readers'experience and on the frequency range of words. The findings will be compared with existing data for English, Greek, Finnish, and Hebrew [1] to evaluate how the strength of the orthographic entropy effect across different frequency ranges depends on the type of orthographic system adopted in a given language.

References

[1] Kuperman, V., Bar-On, A., Bertram, R., Boshra, R., Deutsch, A., Kyröläinen, A.-J., Mathiopoulou, B., Oralova, G., & Protopapas, A. (2021). Prevalence of spelling errors affects reading behavior across languages. *Journal of Experimental Psychology: General*, *150*(10), 1974–1993. https://doi.org/10.1037/xge0001038

[2] Belikov, V., Kopylov, N., Piperski, A., Selegey, V., & Sharoff, S. (2013). Big and diverse is beautiful: A large corpus of Russian to study linguistic variation.

[3] Stanovich, K., & West, R. (1989). Exposure to Print and Orthographic Processing. *Reading Research Quarterly*, *24*(4), 402. https://doi.org/10.2307/747605

[4] Chernova, D., & Bakhturina, P. (2023). Method of print exposure assessment: Application in psycholinguistics and adaptation for the Russian language. *Vestnik of Saint Petersburg University. Language and Literature*, *20*(4), 872–887. https://doi.org/10.21638/spbu09.2023.412

[5] Perfetti, C. (2007). Reading Ability: Lexical Quality to Comprehension. *Scientific Studies of Reading*, *11*(4), 357–383. https://doi.org/10.1080/10888430701530730

[6] Rahmanian, S., & Kuperman, V. (2019). Spelling Errors Impede Recognition of Correctly Spelled Word Forms. *Scientific Studies of Reading*, *23*(1), 24–36. https://doi.org/10.1080/10888438.2017.1359274