

"Reading Polish with Czech Eyes" or "How Russian Can a Bulgarian Text Be?": Orthographic Differences as an Experimental Variable in Reading Comprehension

The human language processing mechanism shows a remarkable robustness to different kinds of imperfect linguistic signals. However, it is unclear how exactly a message encoded in one system is decoded by persons used to a different system. We are interested in gaining insights about human performance at retrieving information encoded in an unfamiliar encoding system. Our focus lies on reading intercomprehension, i.e. the task of retrieving information from texts in genetically related languages that have not been acquired through language learning. Intercomprehension is particularly apparent in the Slavic language family, where similarities between languages often enable native speakers of one Slavic language to effectively understand another Slavic language without having learned it (Meľničuk 1986). The observed mutual intelligibility is supported by the fact that individual languages are descended from a common ancestor, traditionally referred to as Proto-Slavic or Common Slavic (Carlton 1991). The degree of intelligibility of an unknown but closely related language may depend on various linguistic as well as extra-linguistic factors (Gooskens 2013). As a mode of languages use, processing a written text systematically differs from understanding a spoken language variety in a dialogue communication or in listening comprehension. In this paper, we investigate mutual intelligibility of two selected pairs of Slavic languages (Czech and Polish, both using the Latin alphabet; Bulgarian and Russian, both using the Cyrillic alphabet) by establishing orthographic correlates, taking into consideration historically conditioned cross-linguistic variations between sound sequences.

In a preceding research effort, we carried out large-scale computational transformation experiments on parallel word sets, with orthographic correspondences based on traditional approaches and comparative historical linguistics. Our aim was to investigate to what degree the selected Slavic languages are mutually intelligible at the orthographic level and to analyze the most frequent orthographic correspondences and differences between the respective Slavic language pairs. Based on the insights we gained from our previous experiments, we now investigate the effect of orthographic differences in parallel word sets on readers' ability to comprehend a text in a closely related language, e.g. native speakers of Czech reading Polish words or native speakers of Bulgarian reading Russian words, and vice versa. To this end, we design a number of web-based experiments, including word translation tasks, multiple choice tasks of isolated words and recognition of cognate words in context. The fact that a divergent orthography influences intelligibility is generally accepted, but what kind of differences cause larger problems for readers and might worsen the ability to comprehend a closely related language?

References

Carlton, T.R. (ed) *Introduction to the Phonological History of the Slavic Languages*. Slavica Publishers, INC. Columbus, Ohio, 1991

Gooskens, C. Methods for measuring intelligibility of closely related language varieties. In: Bayley, R., Cameron, R., and Lucas C. (eds.) *Handbook of Sociolinguistics*. Oxford University Press, 2013

Meľničuk, O.S. (ed) *Istoričeskaja tipologija slavjanskich jazykov*. Naukova dumka, Kiev, 1986