
Tracing patterns of meaning in Japanese writing systems

Motoki Saito¹ & Ruben van de Vijver²

¹Carl von Ossietzky Universität Oldenburg, ²Heinrich-Heine-Universität
Düsseldorf

motoki.saito@uni-oldenburg.de, ruben.vijver@hhu.de

Is meaning expressed in the details of the realization of words? This is some evidence that this is true (Schmitz & Baer-Henney 2024, Berg et al. 2023). This leads one to expect that this should also hold for writing systems.

We focused on Japanese, which mixes three different writing systems. We hypothesized that if the choice of a writing system is affected by the meaning, we can predict a writing system of a word given its meaning.

We collected 383392 words from the Japanese CHILDES corpora (Miyata 2012). For each word a contextualized embedding was calculated (Devlin et al. 2018). We tested our hypothesis by training a DLM model (Baayen et al. 2019) to predict the writing system of a word on the basis of its meaning.

The prediction accuracy of the model was 94.6%, and an LDA was performed to predict the writing system of each word given its meaning (Figure 1). Our analyses show that semantics has information affecting a word's writing system.

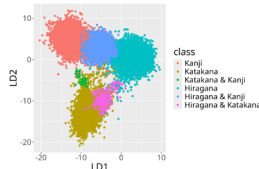


Figure 1: Coefficients of the LDA analysis. The red data points represent words written in kanji. The turquoise data points are words written in hiragana. The yellow-brown data points are words written in katakana.

References: • Baayen, R. H., Chuang, Y.-Y., Shafaei-Bajestan, E. & J. P. Blevins (2019). The discriminative lexicon: A unified computational model for the lexicon and lexical processing in comprehension and production grounded not in (de)composition but in linear discriminative learning. *Complexity*, 2019, 1–39. • Berg, K., Hartmann, S., & D. Claeser (2023). Are some morphological units more prone to spelling variation than others? A case study using spontaneous handwritten data. *Morphology* 34, 173–188. • Devlin, J., Chang, M.-W., Lee, K. & K. Toutanova (2018). BERT: pre-training of deep bidirectional transformers for language understanding. *Computing Research Repository (CoRR)*, abs/1810.04805. • Miyata, S. (2012). *CHILDES nihongoban: Nihongoyoo CHILDES manyuaru 2012*. [Japanese CHILDES: The 2012 CHILDES manual for Japanese]. <http://www2.aasa.ac.jp/people/smiyata/CHILDESmanual/chapter01.html> • Schmitz, D., & D. Baer-Henney (2024). Morphology renders homophonous segments phonetically different: Word-final /s/ in German. *Proc. Speech Prosody 2024 (SP2024)*, 587–591.