From pixels to meaning:
Pixel-based versus trigram-based discriminative models of visual word comprehension

Computational models of visual word comprehension often represent orthographic form through predefined units such as letters or letter n-grams. While effective, such representations presuppose substantial preprocessing of the visual input. In Linear Discriminative Learning (LDL; Baayen et al., 2019), word forms are commonly represented by letter trigrams, which treat visually distinctive elements such as the German gender star simply as discrete symbols. The present study investigates whether lexical comprehension can instead be modelled from pixel-based form representations that remain closer to the perceptual information available during reading.

German gender-star forms such as *Lehrer*in* 'teacher of any gender' provide a particularly informative test case. They combine familiar lexical material with a visually salient and comparatively recent orthographic element while differing semantically from related masculine and feminine forms. I modelled 100 person-denoting nouns in four conditions: generic masculine, specific masculine, specific feminine, and gender-star forms. Semantic representations were derived from a one-million-sentence corpus of contemporary German using Naive Discriminative Learning (Baayen et al., 2011). Two LDL comprehension models were then trained to map written forms onto these semantic vectors: one using conventional letter trigrams, the other using pixel intensities from rendered word forms.

The pixel-based models performed at least as well as the trigram-based models and outperformed them on several measures, despite receiving substantially less linguistically pre-structured input. Both representations captured systematic semantic differences between gender-star and masculine forms, and model-derived measures of lexical processing predicted reading times from an independent self-paced reading study (Haan 2025).

These findings show that successful discriminative modelling of lexical comprehension need not rely on predefined orthographic units. Pixel-based representations provide a viable, and in some cases superior, alternative to letter trigrams while incorporating visual form directly into the mapping from form to meaning. More generally, they offer a step towards less pre-structured and more perceptually grounded models of visual word comprehension.

References

Baayen, R. H., Chuang, Y.-Y., Shafaei-Bajestan, E., & Blevins, J. P. (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, 4895891. https://doi.org/10.1155/2019/4895891

Baayen, R. H., Milin, P., Đurđević, D. F., Hendrix, P., & Marelli, M. (2011). An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychological Review, 118*(3), 438–481. https://doi.org/10.1037/a0023851

Haan, Ann-Sophie. 2025. Gender-fair language processing: A self-paced reading study on the German gender star. MA thesis, Heinrich Heine University Düsseldorf.