Revealing patterns of visual salience in the embedding space of Kenyan Sign Language

Lisa Loy & Hope E. Morgan
Universität Hamburg
lisa.loy@uni-hamburg.de, hope.morgan@uni-hamburg.de

Phonetics in the visual modality are fundamentally different to auditory phonetics, using different articulators and being produced both sequentially and simultaneously (e.g., Meier et al. 2002). Hence, it is often unclear whether research tools created for spoken languages are suitable for sign language research. One such tool are phonetic word embeddings. The intuition behind phonetic word embeddings is that words with similar pronunciation are closer together in the embedding space than words that sound completely different (Fang et al. 2020). In this study, we explore what insights can be gained from phonological sign embeddings. More specifically, we investigate what three different embedding methods – word2vec, dimensionality reduction using truncated SVD on one-hot encoded vectors (OHE-SVD), and a feedforward neural network (FNN) – can reveal about the saliency of individual phonological features. To do so, we examine a dataset of 1,880 signs in Kenyan Sign Language (KSL), annotated as SL CatForm strings consisting of 41 categorical units of form (based on Morgan 2022). We use the concept of minimal pairs to examine the functional load and contributions of individual features to distinguish lexical meanings. We train our embeddings on the 1,880 feature strings, then use cosine similarity to calculate phonological similarity between sign pairs for each approach. Finally, we compare and contrast them to an existing dataset of minimal pairs in KSL (Morgan 2022). We find word2vec is not fruitful for sign language data. However, the OHE-SVD and FNN approaches both successfully assign plausible similarity scores to both minimal pairs and homophones in the KSL data. Furthermore, significant differences in average similarity scores according to phonological feature unveil valuable insights into the functional load of the parameters, specifically *number of hands*, *path movement* and *handshape*. These findings illustrate how embeddings can serve as a tool, contributing to our understanding of which visual features are most salient and potentially relevant for the constitution of meaning in the visual modality.

References: • Fang, A., Filice, S., Limsopatham, N. & O. Rokhlenko (2020). Using phoneme representations to build predictive models robust to asr errors, *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval,* 699–708. • Meier, R. P., Cormier, K. & D. Quinto-Pozos (2002). *Modality and structure in signed and spoken languages*, Cambridge University Press. • Morgan, H. E. (2022). *A Phonological Grammar of Kenyan Sign Language*, Vol. 11, Walter de Gruyter GmbH & Co KG.