Deep neural networks can identify historical phonetic features

Universität Konstanz

Frederik Hartmann

Department of Linguistics, University of Konstanz

Application to New High German

To test the validity of the approach, it is necessary to apply the method to a modern language where all phonetic features are known.

- **Approach**: Analogous to the investigation below, but applied to modern High German. This time, the sample sounds [p], [r], [ɛː], and [aː] were withheld for each network training, respectively.
- **Source**: CELEX phonetic transcriptions (Baayen et al., 1995)
- After the networks were trained, they were tasked with predicting the phonetic features of the sounds withheld.

Results

The results show that

- all 13 tested features of [p] are predicted correctly,
- [r] is correctly predicted to be a voiced liquid, yet regarding place of articulation, only dental/alveolar is predicted which makes a total of 11 out of 13 features.
- The German vowels were less well detected, with a total of 8 out of 10 for [ɛː], and 6 out of 10 for [aː].

Local predictability

- Predictability of sounds based on features in their environment (e.g. Van Son et al., 2005)
- This is due to the phonotactic structure of a language including absolute and statistical constraints

Predicting the phonetics of Proto-Indo-European ‘laryngeals’

The Proto-Indo-European (PIE) ‘laryngeals’ were three sounds of debated quality that are reconstructed by means of the comparative method. Only the position of the laryngeals in PIE words is clearly reconstructible (see e.g. Kümmel, 2007)

- **Approach**: Utilizing local predictability and deep neural networks to investigate the most likely phonetic realization given their phonetic environment. Although previous studies have used machine learning approaches in historical linguistics and cladistics (see Jäger et al., 2016), phonetic feature prediction has not been studied so far.
- **Source**: Wiktionary – large digital database for reconstructed words from proto-languages
- **Method:** After extraction, PIE lemmas were split into trigrams, and each sound was annotated according to its phonetic features
  - Subsequently, binary dense neural networks were optimized and trained on predicting the presence or absence of a phonetic feature in the target sound based on the feature vectors provided by all trigrams.
  - After training and evaluating each network 100 times, the networks were tasked with predicting the phonetic properties of the laryngeals by feeding the networks with the laryngeal environments.

Results

The output of this task was a prediction matrix for each one of the three laryngeals that can be statistically evaluated using Whitney’s U. E.g. for the feature [+asp] in the first laryngeal:

- A median of 115 samples were classified as [+asp] and 85 samples as [-asp] ($p < 0.001$).

➢ The first laryngeal is predominantly found in positions occupied by aspirated sounds in PIE and therefore likely to be [+asp] itself.

References

- Harald Baayen, Richard Piepenbrock, and Léon Gulikers. 1995. Celex2 ldc96l14. Web Download. Philadelphia: Linguistic Data Consortium.
- Martin Kümmel. 2007. Konsonantenwandel. Reichert, Wiesbaden.
- Rob JJH Van Son and Jan PH Van Santen. 2005. Duration and spectral balance of intervocalic consonants: Speech Communication, 47(1-2):100–123.
- Gerhard Jäger and Pavel Sofroniev. 2016. Automatic cognate classification with a support vector machine. In Proceedings of the 13th Conference on Natural Language Processing, Bochumer Linguistische Arbeitsberichte, 128–134.