

Representativeness of Cross-language Word Embeddings on a Lexical Level: Could it be General, or is it Always Task-specific?

Amir Bakarov

Higher School of Economics, Russia
Huawei R&D Russia

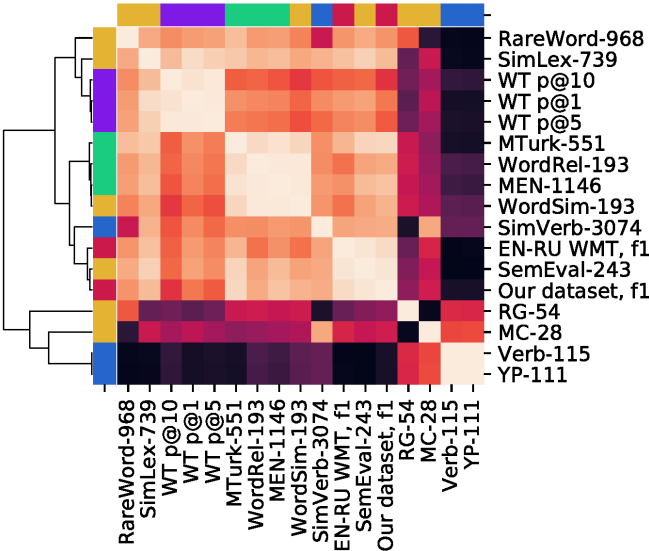
Found in Translation 2018

What is representativeness on a lexical level?

- ▶ Meaning is only one part of information embedded in a word (lexical information).
- ▶ One cannot separate meaning from other types of word information in distributional semantic models (DSM).
- ▶ But one could try to formalize all properties of a word (or, more precisely, a token) and evaluate them separately.
- ▶ **Would the model be really task-independent in this case, or is it always task-specific?**

Results

Correlation of normalized benchmarks scores of several cross-lingual DSMs (the lighter the colors, the higher the correlation).



The poster: <https://bit.ly/2OffEZH>

Thank you!

amirbakarov@gmail.com

<https://github.com/bakarov/cross-lang-embeddings>

