

# Exploring semantic organization across mental lexicons: Perception verbs in Mandarin and English

Yi Yang<sup>1</sup> & Harald Baayen<sup>2</sup>

<sup>1</sup> Beihang University/ University of Tuebingen, heyangyihe@163.com / yi.yang@uni-tuebingen.de

<sup>2</sup> University of Tuebingen, harald.baayen@uni-tuebingen.de

**Keywords:** perception verbs, distributional semantics, mental lexicon, semantic vectors, category

Verbs of perception tend to have many shades of meaning (Viberg, 1983; Marco & Jansegers, 2019). This study addresses perception verbs from a cross-linguistic perspective, comparing Mandarin Chinese and English. We made use of distributional semantics (Firth, 1968; Landauer & Dumais, 1997), taking pre-compiled word embeddings obtained with FastText (taken from <https://fasttext.cc/docs/en/crawl-vectors.html>) as starting point for our analyses.

As a first step, we defined a series of categories comprising concrete nouns (for, e.g., food, animals, and persons), as well as evaluative adjectives, modal verbs, cognitive verbs, time expressions, and onomatopoeia. Furthermore, we defined a highly culture-specific category comprising words for supernatural beings. For Mandarin, we also included the philosophical terms *yīn*, *yáng*, and the 5 elements (*wǔ-xíng*).

As a second step, we then used unsupervised clustering methods (PCA, MDS, t-SNE) to investigate how the perception verbs are positioned with respect to the clusters comprising the words of the above categories. The maps for the categories in Mandarin and English show both similarities and differences. Perception verbs in both languages were closer to cognitive verbs, modal expressions, and adjectives, and positioned further away from concrete categories (e.g. food and plants). Cultural differences also clearly emerged. Mandarin perception verbs are generally closer to words for supernatural beings, whereas English perception verbs do not show this tendency.

For the contrastive analysis, we made use of pairs of translation equivalents, and calculated, for each pair, the Jaccard index of the 20 nearest neighbor translations. That is, we calculated for each pair the ratio of the number of shared translation neighbors to the total set of neighbors in both languages. Of all semantic categories, perception verbs shared the fewest neighbors.

A second contrastive analysis was based on the correlations of the embeddings of the perception verbs with the embeddings of the words in the different categories. For each language, and for a given category, we calculated the mean of the correlations of all the word embeddings of perception verbs with the word embeddings of all the words in that category, henceforth *mean by-category correlation*. Mandarin revealed a preference for more abstract categories. Perception verbs in Mandarin have higher correlations compared to English for communication nouns, cognition nouns, phenomenon nouns, persons, and sports. By contrast, words for plants, objects, animals, body parts, and colors tend to be somewhat more correlated with English perception verbs.

Finally, correlational profiles for individual perception verbs proved useful for clarifying differences in the meanings of near-synonyms. For example, *guān-kàn* is typically used in the context of sports. In English, *sound* and *smell* show marked preferences for evaluative adjectives. The methods proposed in this study makes it possible to trace subtle differences in lexical semantics across languages and are offered in the hope that they will prove a useful complement to existing methods such as behavioral profiling.

In conclusion, the method proposed in this study makes it possible to trace subtle differences in verbal semantics across languages.

## References

- Firth, John Rupert. 1968. *Selected papers of j.r.firth, 1952-59*. Indiana University Press.
- Landauer, Thomas K. & Susan T. Dumais. 1997. A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction and representation of knowledge. *Psychological Review* 104(2). 211–240.
- Marco, Marta Albelda & Marlies Jansegers. 2019. From visual perception to evidentiality: A functional empirical approach to se ve que in spanish. *Lingua* 220. 76–97.
- Viberg, Åke. 1983. The verbs of perception: A typological study. In *Explanations for language universals*, Mouton de Gruyter, Berlin/New York Berlin, New York.