

Sound Symbolism Revisited – a Reconstruction Approach

Kevin Tang*

Department of Linguistics, University College London, United Kingdom

Paweł Mandera, Emmanuel Keuleers†

Department of Experimental Psychology, Ghent University, Belgium

November 11, 2013

A well-known linguistic fact about the linkage between sound and meaning is that the relationship is almost or completely arbitrary (Hinton et al., 2006). Any evidence to the contrary is seen to conflict with one of the axioms of linguistics – Saussure (1916/1995)’s “l’arbitraire du signe”.

Native speakers’ intuition has served to be a powerful linguistic tool (Fodor, 1981; Garrett and Fodor, 1968; Chomsky, 1986). Jespersen (1922) suggested that native speakers have intuitions of sound symbolism, and as he put it, “There is no denying, however, that there are words which we feel instinctively to be adequate to express the ideas they stand for...” (Jespersen, 1922, p.397-398)

Many approaches have been developed to tackle this long-debated phenomenon with varying degrees of success – from testing the perception of phonetic properties (Sapir, 1929; Newman, 1933; Kohler, 1947) e.g. [a](“large”) versus [i](“small”), comparing basic vocabulary cross-linguistically (Wichmann et al., 2010), to validating phonesthemes (Householder, 1946; Bergen, 2004; Drellishak, 2006).

Given the varying degrees of success with previous findings, it is reasonable to assume that the sound symbolic phenomenon is likely to be elusive. However, with an access to large text corpora and new methods of modelling semantic space, we attempted to capture these subtle effects. In this study, we examined sound symbolism from a new perspective, using a novel naturalistic speech corpus of 400 million words, SUBTLEX English, and a machine learning technique, Topic modelling (Latent Dirichlet Allocation, LDA, (Blei et al., 2003)). Using English as a test case, we demonstrated the possibility of reconstructing meaning from sound, and sound from meaning.

To establish semantics, we trained an LDA model on a large subtitle corpus which defines a semantic “space” by representing the semantics of each word as a probability distribution over topics/contexts. We then reconstructed the semantics of each word based on their orthographical/phonological neighbourhood. Similarly, we approached the phenomenon in the opposite direction by reconstructing the form of a word based on its neighbours in the semantic space.

Our preliminary results suggested that the connection between sound and meaning is not entirely arbitrary but consistently above random, and furthermore the results agreed with native speakers/researchers’ intuitions of sound symbolism, supporting Jespersen (1922)’s view on intuitions, that is, allegedly sound-symbolic words (such as “gleam”, “pop” and “slash”) are more reconstructable than just any words. Our methodology was shown to be a promising way of exploring this topic further, to see if this effect still holds, and to what extent, cross-linguistically.

References

- B.K. Bergen. The psychological reality of phonaesthemes. *Language*, pages 290–311, 2004.
- D.M Blei, A.Y. Ng, and M.I. Jordan. Latent dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022, 2003.
- N. Chomsky. *Knowledge of language: Its nature, origins, and use*. Greenwood Publishing Group, 1986.
- S. Drellishak. Statistical techniques for detecting and validating phonesthemes. *Unpublished masters thesis, University of Washington*, 2006.
- J. Fodor. Some notes on what linguistics is about. *Readings in the Philosophy of Psychology*, 2:197–207, 1981.
- M/ Garrett and J. Fodor. Psychological theories and linguistic constructs. *Verbal behavior and general behavior theory*, pages 451–477, 1968.
- L. Hinton, J. Nichols, and J.J. Ohala. *Sound symbolism*. Cambridge University Press, 2006.
- F.W. Householder. On the problem of sound and meaning, an English phonestheme. *Word*, 2:83, 1946.
- O. Jespersen. *Language: Its Nature, Development and Origin*. Allen & Unwin, 1922. ISBN 0044000073.
- Wolfgang Kohler. Gestalt psychology: an introduction to new concepts in psychology. *New York: Liveright*, pages 3–4, 1947.
- S.S. Newman. Further experiments in phonetic symbolism. *The American Journal of Psychology*, 45(1):53–75, 1933.
- E. Sapir. A study in phonetic symbolism. *Journal of Experimental Psychology*, 12(3):225, 1929.
- F de Saussure. *Cours de linguistique générale. Paris: Payot.(Primera edición, 1916)*, 1916/1995.
- S. Wichmann, E.W. Holman, and C.H. Brown. Sound symbolism in basic vocabulary. *Entropy*, 12(4):844–858, 2010.

*Electronic address: kevin.tang.10@ucl.ac.uk

†Electronic address: pawel.mandera@ugent.be; emmanuel.keuleers@ugent.be