Plotting our way through the data: The *way*-construction revisited

Florent Perek, Universität Basel

The *way*-construction (e.g., *She typed her way to a promotion*) is a classic example of a syntactic pattern that is better described in terms of a direct pairing of form with function, in line with the principles of construction grammar (Goldberg, 1995). The history of the construction was famously investigated by Israel (1996) on the basis of a database of citations from the Oxford English Dictionary. Through detailed qualitative analysis based on manual semantic groupings, Israel observes that the distribution of the construction gradually expanded over time through local relations of semantic similarity and analogy. This paper re-examines the diachrony of the *way*-construction in Late Modern American English in a more quantitative and data-driven way, borrowing methods from computational linguistics (Perek, 2014). Some interesting comparisons with Israel's British English data are offered.

Every token of the *way*-construction occurring in the Corpus of Historical American English (Davies, 2010) between 1830 and 2009 was manually extracted. This resulted in a list of 353 verbs attested in the construction at some point of its recent history. This diachronic distribution was analyzed by means of distributional semantic representations as a proxy to the meaning of words. Drawing on the observation that words occurring in similar contexts tend to have similar meanings, distributional semantic representations approximate the meaning of a word by recording its co-occurrence with other words in a vast text corpus (Turney & Pantel, 2010). By computing pairwise distances between the semantic representations of the verbs occurring in a construction and feeding them to a *t*-Distributed Stochastic Neighbor Embedding algorithm (Van der Maaten & Hinton, 2008) that positions the words in a 2-dimensional space according to their distributional semantic similarity, it is possible to reliably visualize the semantic domain of the construction and observe how words in it are related to each other.

Distributional semantic maps of the *way*-construction were plotted for six successive 30-year periods from the 1830s to the 2000s. Comparing these plots allows us to readily identify the semantic development of the construction and track its time course with precision. It is observed that the productivity of the construction is fairly constant but mostly concerns already established semantic classes. More peripheral regions of the semantic domain are slower to attract new members, which lines up with Israel's (1996) observation that the development of the construction is gradual and based on local relations of analogy. Notable recent developments includes the class of verbs of sound emission (*whistle*, *grunt*, *shriek*), which are attested from the 1860s but only become a fully productive class in the first half of the 20th century. In this and other respects, the American *way*-construction appears to be ahead of its British counterpart.

References

Davies, M. (2010). *The Corpus of Historical American English: 400 million words, 1810-2009*. Available online at http://corpus.byu.edu/coha/

Goldberg, A. (1995). *Constructions: A construction grammar approach to argument structure*. Chicago: University of Chicago Press.

Israel, M. (1996). The way constructions grow. In A. Goldberg (ed.), *Conceptual structure, discourse and language*. Stanford, CA: CSLI Publications, 217-230.

Perek, F. (2014). Vector spaces for historical linguistics: Using distributional semantics to study syntactic productivity in diachrony. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, Baltimore, Maryland USA, June 23-25 2014*.

Turney, P. and P. Pantel (2010). From Frequency to Meaning: Vector Space Models of Semantics. Journal of Artificial Intelligence Research 37, 141-188.

Van der Maaten, L. & Hinton, G. (2008). Visualizing Data using *t*-SNE. *Journal of Machine Learning Research*, *9*, 2579-2605