Tracking change in word meaning.
A dynamic visualization of diachronic distributional semantic models.

Kris Heylen, Thomas Wielfaert and Dirk Speelman

Within Computational Linguistics, distributional models of semantics have become the mainstay of large-scale modeling of lexical semantics (see Turney and Pantel 2010 for an overview). Distributional modeling also hold a large potential for research in Linguistics proper: It allow linguists to base their analysis on large amounts of usage data, thus vastly extending their empirical basis, AND they make it possible to detect potentially interesting semantic patterns. However, so far, there have been relatively few applications, mainly because of the technical complexity and the lack of a linguist-friendly interface to explore the output. To address this issue, Heylen et al. (2012) proposed an interactive visualization of a distributional similarity matrix based on Multi-Dimensional Scaling for synchronic data. In this paper, we extend this approach to diachronic data and propose a dynamic visualization of distributional semantic change through motion charts. As a case study, we look at the meaning changes that 17 positive evaluative adjectives have undergone in the Corpus of Historical American English (COHA, Davies 2012) between 1860 and 2000.

Visualization of diachronic distributional data has been proposed previously by a.o. Rohrdantz et al. (2011) but these representations were static. In this paper, we use a dynamic visualization of linguistic change, first proposed by Hilpert (2011) for manually coded data sets, and extend here to the large-scale, unsupervised distributional models. For a set of adjectives that express positive evaluation (a.o. brilliant, magnificent, fantastic, terrific, superb). We investigate how they carve up this semantic space and how this changes over time. From COHA, we extracted a word-by-context co-occurrence vector using a window of 4 left and right for each adjective in each of the 14 decades between 1860 and 2000. Next, we calculated the cosine similarity between all adjective/decade vectors and used non-parametric MDS to represent these similarities in 2 dimensions. The MDS solution with adjective and decade information was then visualized with the R-package googleVis, an interface between R and the Google Visualization API. The resulting, dynamic motion chart is available online under https://perswww.kuleuven.be/~u0038536/magnificent/Magnificent3D.html.

The chart shows adjectives as clickable bubbles with a time slider to move between decades. ‘Playing’ the chart shows dynamically how the semantic distances between the adjectives changes over time. In the center, adjectives like splendid, magnificent or great represent the core of the concept and they remain relatively stable over time. However, figure 1 shows that terrific was in 1860 still quite far removed from the center, probably because it was still predominantly used in its literal sense of frightening. Only around 1950, terrific starts to move to the core and acquires its positive evaluative meaning. Since, distributional models are a completely automatic technique with a multitude of possible parameter settings, this particular solution is probably not yet optimal. An important next step is therefore the evaluation of the automatically induced patterns against a manually coded and interpreted dataset.

References


Figure 1: Movement of *terrific* and *great* through distributional space from 1860 to 2000