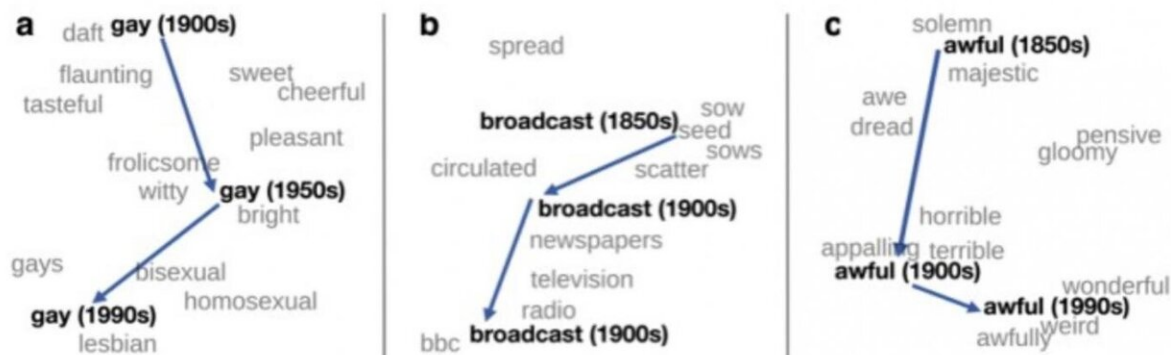


How computational linguistics helps to understand how language works

March 3 2020



Two-dimensional view of the change in meaning of three English words, taken from Hamilton et al. (2016). Credit: upf

Distributional semantics obtains representations of the meaning of words by processing thousands of texts and extracting generalizations using computational algorithms. Despite the popularity of distributional semantics in such fields as computational linguistics and cognitive science, its impact on theoretical linguistics has so far been very limited.

Research by Gemma Boleda, head of the Computational Linguistics and Language Theory (COLT) research group and ICREA research professor with the Department of Translation and Language Sciences at UPF, published in the journal *Annual Review of Linguistics*, provides a

critical review of the abundant studies available on distributional semantics, putting special emphasis on the results that are relevant for theoretical linguistics. Specifically there are three areas: semantic change, polysemy and composition, and the grammar-semantics interface.

The research by Gemma Boleda seeks to connect theoretical and computational approaches to advance in the collective knowledge about how [language](#) works. One of the methods she has extensively researched is distributional semantics, which allows obtaining representations of words automatically. These representations have been shown to reflect significant linguistic properties, such as how two words are similar: a person will tell you that "dog" and "puppy" are very similar, and yet "dog" and "democracy" are hardly similar at all; distributional semantics will say the same, thanks to the fact that it induces linguistic properties based on texts written by people. Therefore, distributional semantics provides radically empirical representations.

Distributional semantics allows analysing the use of words and the evolution of their meaning

Distributional semantics provides an attractive, complementary framework to other, more traditional methods, not only because it is radically empirical but also because it provides multidimensional representations: two words can be likened on one dimension of meaning ("pizza" and "pasta" are types of food), or on another ("pizza" and "wheel" are round). To represent all aspects of meaning, multidimensional representations are needed. Distributional semantics can capture the common uses of two words, as well as their differentiating factors.

One of the important applications of distributional semantics in

theoretical linguistics is the detection of changes in meaning. If language data from different periods are processed, such as books in English from 1900, 1950 and 1990, distributional semantics can be used to automatically detect some words' change in [meaning](#). For example, the word "gay" in English at the beginning of the last century meant "happy" and has been used increasingly to mean "homosexual."

Aspects of research into distributional semantics that contribute to language theory

From the analysis of the works studied, Boleda concludes that there is sufficient evidence for the solid results of distributional semantics to be imported directly to research in theoretical linguistics.

"There are at least four aspects of research in distributional semantics that can contribute to language theory. The first aspect is exploratory: distributional representations can be used to explore large-scale data, for example by examining the similarity of words. The second is as a tool to identify specific cases of linguistic phenomena. For example, words can be identified whose meanings have changed when comparing the representations obtained from texts from different periods. The third is as a test bench: evaluating different linguistic hypotheses in distributional terms. The fourth and most difficult is the discovery of new linguistic phenomena or relevant theoretical trends in the data," the author explains in her work.

More information: Gemma Boleda, Distributional Semantics and Linguistic Theory, *Annual Review of Linguistics* (2019). [DOI: 10.1146/annurev-linguistics-011619-030303](https://doi.org/10.1146/annurev-linguistics-011619-030303)

Provided by Universitat Pompeu Fabra - Barcelona

Citation: How computational linguistics helps to understand how language works (2020, March 3) retrieved 26 June 2024 from <https://phys.org/news/2020-03-linguistics-language.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.