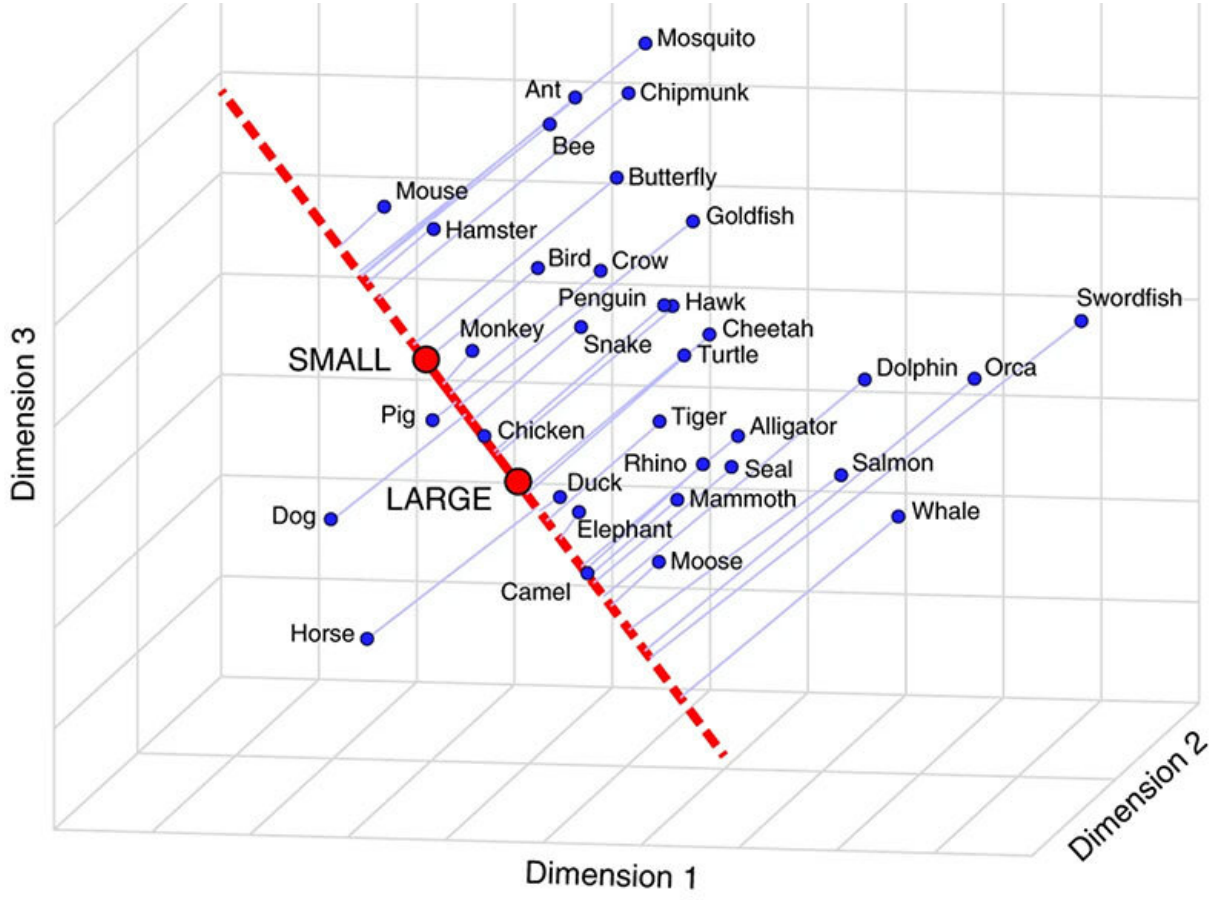
# Language processing programs can assign many kinds of information to a single word, like the human brain

May 4 2022, by Jennifer Michalowsk




Word vectors in the category "animals" (blue circles) are orthogonally projected (light-blue lines) onto the feature subspace for "size" (red line), defined as the vector difference between large and small (red circles). The three dimensions in this figure are arbitrary and were chosen via principal component analysis to enhance visualization; the original GloVe word embedding has 300 dimensions,

and projection happens in that space. Credit: Fedorenko lab

From search engines to voice assistants, computers are getting better at understanding what we mean. That's thanks to language-processing programs that make sense of a staggering number of words, without ever being told explicitly what those words mean. Such programs infer meaning instead through statistics—and a new study reveals that this computational approach can assign many kinds of information to a single word, just like the human brain.

The study, published April 14 in the journal *Nature Human Behavior*, was co-led by Gabriel Grand, a graduate student in electrical engineering and computer science who is affiliated with MIT's Computer Science and Artificial Intelligence Laboratory, and Idan Blank Ph.D. '16, an assistant professor at the University of California at Los Angeles. The work was supervised by McGovern Institute for Brain Research investigator Ev Fedorenko, a cognitive neuroscientist who studies how the human brain uses and understands language, and Francisco Pereira at the National Institute of Mental Health. Fedorenko says the rich knowledge her team was able to find within computational language models demonstrates just how much can be learned about the world through language alone.

The research team began its analysis of statistics-based language processing models in 2015, when the approach was new. Such models derive meaning by analyzing how often pairs of words co-occur in texts and using those relationships to assess the similarities of words' meanings. For example, such a program might conclude that "bread" and "apple" are more similar to one another than they are to "notebook," because "bread" and "apple" are often found in proximity to words like "eat" or "snack," whereas "notebook" is not.

The models were clearly good at measuring words' overall similarity to one another. But most words carry many kinds of information, and their similarities depend on which qualities are being evaluated. "Humans can come up with all these different mental scales to help organize their understanding of words," explains Grand, a former undergraduate researcher in the Fedorenko lab. For example, he says, "dolphins and alligators might be similar in size, but one is much more dangerous than the other."

Grand and Blank, who was then a graduate student at the McGovern Institute, wanted to know whether the models captured that same nuance. And if they did, how was the information organized?

To learn how the information in such a model stacked up to humans' understanding of words, the team first asked human volunteers to score words along many different scales: Were the concepts those words conveyed big or small, safe or dangerous, wet or dry? Then, having mapped where people position different words along these scales, they looked to see whether language processing models did the same.

Grand explains that distributional semantic models use co-occurrence statistics to organize words into a huge, multidimensional matrix. The more similar words are to one another, the closer they are within that space. The dimensions of the space are vast, and there is no inherent meaning built into its structure. "In these word embeddings, there are hundreds of dimensions, and we have no idea what any dimension means," he says. "We're really trying to peer into this black box and say, 'is there structure in here?'"

Specifically, they asked whether the semantic scales they had asked their volunteers use were represented in the model. So they looked to see where words in the space lined up along vectors defined by the extremes of those scales. Where did dolphins and tigers fall on line from "big" to

"small," for example? And were they closer together along that line than they were on a line representing danger ("safe" to "dangerous")?

Across more than 50 sets of world categories and semantic scales, they found that the model had organized words very much like the human volunteers. Dolphins and tigers were judged to be similar in terms of size, but far apart on scales measuring danger or wetness. The model had organized the words in a way that represented many kinds of meaning—and it had done so based entirely on the words' co-occurrences.

That, Fedorenko says, tells us something about the power of language. "The fact that we can recover so much of this rich semantic information from just these simple word co-occurrence statistics suggests that this is one very powerful source of learning about things that you may not even have direct perceptual experience with."

**More information:** Gabriel Grand et al, Semantic projection recovers rich human knowledge of multiple object features from word embeddings, *Nature Human Behaviour* (2022). DOI: 10.1038/s41562-022-01316-8

Provided by Massachusetts Institute of Technology

Citation: Language processing programs can assign many kinds of information to a single word, like the human brain (2022, May 4) retrieved 18 April 2024 from https://techxplore.com/news/2022-05-language-assign-kinds-word-human.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.