Researchers use machine learning to examine meanings across languages

January 9 2024, by Sarah Steimer

Credit: Pixabay/CC0 Public Domain

How meanings align across languages suggests that translation and cross-cultural communication may be challenging, but researchers haven't been clear as to what characterizes this alignment, and what factors might

affect it.

[A recent study published in *PNAS*](#), however, used machine learning to find that meanings across languages are similar within domains of [meaning](#), but they vary across domains. The research also showed that concrete meanings (e.g., hand, tree, pot) varied less than abstract meanings across languages (democracy, truth, happiness), yet languages spoken in closer proximity tended to have much more similar meanings.

The research team was led by Prof. James Evans in the Department of Sociology and Molly Lewis, who began the research as a postdoc at UChicago. Their study used large-scale data to consider core linguistic anthropology questions.

Evans points to the Sapir-Whorf hypothesis, which suggests that the structure of a language influences its speakers' worldviews and cognition. This would mean that language guides thinking. Research since has argued against a strong version of this hypothesis, and Evans's team used machine learning to look at many languages across all domains: Do they line up in particular ways? How can researchers characterize the difference in their meanings?

The researchers used two large data sources: Wikipedia articles, which cover the same topics in [different languages](#), and Test of English as a Foreign Language (TOEFL) essays in which people whose first language isn't English write on a topic in English. What they found was strong evidence for relativism in meaning, demonstrating how the particular language a person speaks influences the way they assemble ideas and think about reality.

"We show that concrete objects are more conserved across languages, but they still vary," said Evans. "So, if you're farther away in culture space, or how you structure kinship, or your experience with

environment and climate, or physical distance; each contributes to the difference in how you structure meanings across words."

Evans explains that the study showed a much more powerful way of thinking about the difference between languages. Not in terms of abstractness and concreteness of words, but their local and global associations with one another.

"Domains like health care—doctors, nurses, disease; or religion—priests, shaman, blessings, curses—when you use one of the words you use other words as well," said Evans. "These domains tend to be highly conserved across languages. But what's different is the distance between those domains."

He gives an example in Mandarin Chinese. In this language, discussing family will often include the use of physical space metaphors, such as mountains and oceans. In another language, however, they use metaphors related to health and healing.

"Within domains, languages are largely the same; but across domains, these differences create more and less available metaphors, turns of phrase, shifts in a narrative," he said. "Those cognitive proximities—available to you through your native language—dramatically shape the way you'll write and read a narrative and the degree to which an explanation will feel familiar and convincing, or surprising and suspect to you."

If we can learn to anticipate these associations, Evans said, we can better tune and improve translations. For example, rather than Google offering a word-for-word translation that may lose the meaning of a metaphor, it could shape it to an association that makes sense in the reader's native language. It can shape the way a second language is taught as well; we could teach metaphor and association in the same way we teach syntax

and grammar. And it could also help in writing edicts and laws that affect people speaking various languages.

"There is measurable culture embedded in language," said Evans. "And it deeply shapes the way in which people experience the world, construct metaphors and communicate ideas."

More information: Molly Lewis et al, Local similarity and global variability characterize the semantic space of human languages, *Proceedings of the National Academy of Sciences* (2023). DOI: 10.1073/pnas.2300986120

Provided by University of Chicago