

Cognitive scientists develop new take on old problem: why human language has so many words with multiple meanings

January 19 2012, by Emily Finn



Graphic: Christine Daniloff

Why did language evolve? While the answer might seem obvious -- as a way for individuals to exchange information -- linguists and other students of communication have debated this question for years. Many prominent linguists, including MIT's Noam Chomsky, have argued that language is, in fact, poorly designed for communication. Such a use, they say, is merely a byproduct of a system that probably evolved for other reasons -- perhaps for structuring our own private thoughts.

As evidence, these linguists point to the existence of ambiguity: In a system optimized for conveying information between a speaker and a

listener, they argue, each word would have just one meaning, eliminating any chance of confusion or misunderstanding. Now, a group of MIT cognitive scientists has turned this idea on its head. In a new theory, they claim that ambiguity actually makes language more efficient, by allowing for the reuse of short, efficient sounds that listeners can easily disambiguate with the help of context.

“Various people have said that ambiguity is a problem for communication,” says Ted Gibson, an MIT professor of cognitive science and senior author of a paper describing the research to appear in the journal *Cognition*. “But once we understand that context disambiguates, then ambiguity is not a problem — it’s something you can take advantage of, because you can reuse easy [words] in different contexts over and over again.”

Lead author of the paper is Steven Piantadosi PhD ’11; Harry Tily, a postdoc in the Department of Brain and Cognitive Sciences, is another co-author.

What do you ‘mean’?

For a somewhat ironic example of ambiguity, consider the word “mean.” It can mean, of course, to indicate or signify, but it can also refer to an intention or purpose (“I meant to go to the store”); something offensive or nasty; or the mathematical average of a set of numbers. Adding an ‘s’ introduces even more potential definitions: an instrument or method (“a means to an end”), or financial resources (“to live within one’s means”).

But virtually no speaker of English gets confused when he or she hears the word “mean.” That’s because the different senses of the word occur in such different contexts as to allow listeners to infer its meaning nearly automatically.

Given the disambiguating power of context, the researchers hypothesized that languages might harness ambiguity to reuse words — most likely, the easiest words for language processing systems. Building on observation and previous studies, they posited that words with fewer syllables, high frequency and the simplest pronunciations should have the most meanings.

To test this prediction, Piantadosi, Tily and Gibson carried out corpus studies of English, Dutch and German. (In linguistics, a corpus is a large body of samples of language as it is used naturally, which can be used to search for word frequencies or patterns.) By comparing certain properties of words to their numbers of meanings, the researchers confirmed their suspicion that shorter, more frequent words, as well as those that conform to the language's typical sound patterns, are most likely to be ambiguous — trends that were statistically significant in all three languages.

To understand why ambiguity makes a language more efficient rather than less so, think about the competing desires of the speaker and the listener. The speaker is interested in conveying as much as possible with the fewest possible words, while the listener is aiming to get a complete and specific understanding of what the speaker is trying to say. But as the researchers write, it is “cognitively cheaper” to have the listener infer certain things from the context than to have the speaker spend time on longer and more complicated utterances. The result is a system that skews toward ambiguity, reusing the “easiest” words. Once context is considered, it's clear that “ambiguity is actually something you would want in the communication system,” Piantadosi says.

Tom Wasow, a professor of linguistics and philosophy at Stanford University, calls the paper “important and insightful.”

“You would expect that since languages are constantly changing, they

would evolve to get rid of ambiguity,” Wasow says. “But if you look at natural languages, they are massively ambiguous: Words have multiple meanings, there are multiple ways to parse strings of [words](#). ... This paper presents a really rigorous argument as to why that kind of ambiguity is actually functional for communicative purposes, rather than dysfunctional.”

Implications for computer science

The researchers say the statistical nature of their paper reflects a trend in the field of linguistics, which is coming to rely more heavily on information theory and quantitative methods.

“The influence of computer science in linguistics right now is very high,” Gibson says, adding that natural language processing (NLP) is a major goal of those operating at the intersection of the two fields.

Piantadosi points out that ambiguity in natural language poses immense challenges for NLP developers. “Ambiguity is only good for us [as humans] because we have these really sophisticated cognitive mechanisms for disambiguating,” he says. “It’s really difficult to work out the details of what those are, or even some sort of approximation that you could get a computer to use.”

But, as Gibson says, computer scientists have long been aware of this problem. The new study provides a better theoretical and evolutionary explanation of why ambiguity exists, but the same message holds: “Basically, if you have any human [language](#) in your input or output, you are stuck with needing context to disambiguate,” he says.

This story is republished courtesy of MIT News (web.mit.edu/newsoffice/), a popular site that covers news about MIT research, innovation and teaching.

Provided by Massachusetts Institute of Technology

Citation: Cognitive scientists develop new take on old problem: why human language has so many words with multiple meanings (2012, January 19) retrieved 3 April 2024 from <https://phys.org/news/2012-01-cognitive-scientists-problem-human-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.