

Applying information theory to linguistics suggests 'functional design' in cross-language variations

October 10 2012, by Larry Hardesty



The majority of languages—roughly 85 percent of them—can be sorted into two categories: those, like English, in which the basic sentence form is subject-verb-object ("the girl kicks the ball"), and those, like Japanese, in which the basic sentence form is subject-object-verb ("the girl the ball kicks").

The reason for the difference has remained somewhat mysterious, but researchers from MIT's Department of Brain and Cognitive Sciences now believe that they can account for it using concepts borrowed from [information theory](#), the discipline, invented almost singlehandedly by longtime MIT professor Claude Shannon, that led to the [digital revolution](#) in communications. The researchers will present their hypothesis in an upcoming issue of the journal [Psychological Science](#).

Shannon was largely concerned with faithful communication in the presence of "noise"—any external influence that can corrupt a message on its way from sender to receiver. Ted Gibson, a professor of cognitive sciences at MIT and corresponding author on the new paper, argues that [human speech](#) is an example of what Shannon called a "noisy channel."

"If I'm getting an idea across to you, there's noise in what I'm saying," Gibson says. "I may not say what I mean—I pick up the wrong word, or whatever. Even if I say something right, you may hear the wrong thing. And then there's ambient stuff in between on the signal, which can screw us up. It's a real problem." In their paper, the MIT researchers argue that languages develop the word order rules they do in order to minimize the risk of [miscommunication](#) across a noisy channel.

Gibson is joined on the paper by Rebecca Saxe, an associate professor of [cognitive neuroscience](#); Steven Piantadosi, a postdoc at the University of Rochester who did his doctoral work with Gibson; Leon Bergen, a graduate student in Gibson's group; research affiliate Eunice Lim; and Kimberly Brink, who graduated from MIT in 2010.

Mixed signals

The researchers' hypothesis was born of an attempt to explain the peculiar results of an experiment reported in the *Proceedings of the National Academy of Sciences* in 2008; Brink reproduced the experiment as a class project for a course taught by Saxe. In the experiment, native English speakers were shown crude digital animations of simple events and asked to describe them using only gestures. Oddly, when presented with events in which a human acts on an inanimate object, such as a girl kicking a ball, volunteers usually attempted to convey the object of the sentence before trying to convey the verb—even though, in English, verbs generally precede objects. With events in which a human acts on another human, such as a girl kicking a boy, however, the volunteers would generally mime the verb before the object.

"It's not subtle at all," Gibson says. "It's about 70 percent each way, so it's a shift of about 40 percent."

The tendency even of speakers of a subject-verb-object (SVO) language like English to gesture subject-object-verb (SOV), Gibson says, may be an example of an innate human preference for linguistically recapitulating old information before introducing new information. The "old before new" theory—which, according to the University of Pennsylvania linguist Ellen Price, is also known as the given-new, known-new, and presupposition-focus theory—has a rich history in the linguistic literature, dating back to at least the work of the German philosopher Hermann Paul, in 1880.

Imagine, for instance, the circumstances in which someone would actually say, in ordinary conversation, "the girl kicked the ball." Chances are, the speaker would already have introduced both the girl and the ball—say, in telling a story about a soccer game. The sole new piece of information would be the fact of the kick.

Assuming a natural preference for the SOV word order, then—at least in cases where the verb is the new piece of information—why would the volunteers in the PNAS experiments mime SVO when both the subject and the object were people? The MIT researchers' explanation is that the SVO ordering has a better chance of preserving information if the communications channel is noisy.

Suppose that the sentence is "the girl kicked the boy," and that one of the nouns in the sentence—either the subject or the object—will be lost in transmission. If the word order is SOV, then the listener will receive one of two messages: either "the girl kicked" or "the boy kicked." If the word order is SVO, however, the two possible messages on the receiving end are "the girl kicked" and "kicked the boy": More information will have made it through the noisy channel.

Down to cases

That is the MIT researchers' explanation for the experimental findings reported in the 2008 PNAS paper. But how about the differences in word order across languages? A preliminary investigation, Gibson says, suggests that there is a very strong correlation between word order and the strength of a language's "case markings." Case marking means that words change depending on their syntactic function: In English, for instance, the pronoun "she" changes to "her" if the kicker becomes the kicked. But case marking is rare in English, and English is an SVO language. Japanese, a strongly case-marked language, is SOV. That is, in Japanese, there are other cues as to which noun is subject and which is object, so Japanese speakers can default to their natural preference for old before new.

Gibson adds that, in fact, some languages have case markings only for animate objects—an observation that accords particularly well with the MIT researchers' theory.

"It's an extremely valuable study," says Steven Pinker, the Johnstone Family Professor in the Department of Psychology at Harvard University. "The design of any language reflects a compromise between properties that make it more useful—clarity, expressiveness, ease of articulation—and properties that are standardized across a community of speakers so that everyone is using the same code. Most grammatical theorists have focused on the arbitrary nature of the community-wide grammar. Gibson has now shed light on how each of these grammars has evolved, in a few predictable ways, to maximize clarity in communicating who did what to whom. That is, much more can be said than just 'That's the way English is; that's the way Turkish is,' and so on. Gibson's study shows that there is a great deal of functional design in seemingly arbitrary patterns of variation across languages."

In order to make their information-theoretical model of [word order](#) more rigorous, Gibson says, he and his colleagues need to better characterize the "noise characteristics" of spoken conversation—what types of errors typically arise, and how frequent they are. That's the topic of ongoing experiments, in which the researchers gauge people's interpretations of sentences in which words have been deleted or inserted.

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: Applying information theory to linguistics suggests 'functional design' in cross-language variations (2012, October 10) retrieved 2 May 2024 from <https://medicalxpress.com/news/2012-10-theory-linguistics-functional-cross-language-variations.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.