Entropy is Universal Rule of Language

By [Lisa Grossman](#) ✉ May 17, 2011 | 5:15 pm | Categories: [Anthropology](#), [Brains and Behavior](#)

The amount of information carried in the arrangement of words is the same across all languages, even languages that aren't related to each other. This consistency could hint at a single common ancestral language, or universal features of how human brains process speech.

"It doesn't matter what language or style you take," said systems biologist [Marcelo Montemurro](#) of England's University of Manchester, lead author of a study May 13 in [PLoS ONE](#). "In languages as diverse as Chinese, English and Sumerian, a measure of the linguistic order, in the way words are arranged, is something that seems to be a universal of languages."

Language carries meaning both in the words we choose, and the order we put them in. Some languages, like Finnish, carry most of their meaning in tags on the words themselves, and are fairly free-form in how words are arranged. Others, like English, are more strict — "John loves Mary" means something different from "Mary loves John."

Montemurro realized that he could quantify the amount of information encoded in word order by computing a text's "entropy," or a measure of how evenly distributed the words are. Drawing on methods from [information theory](#), Montemurro co-author Damián Zanette of the National Atomic Energy Commission in Argentina calculated the entropy of thousands of texts in eight different languages:

English, French, German, Finnish, Tagalog, Sumerian, Old Egyptian and Chinese.

Then the researchers randomly rearranged all the words in the texts, which ranged from the complete works of Shakespeare to *The Origin of Species* to prayers written on Sumerian tablets.

"If we destroy the original text by scrambling all the words, we are preserving the vocabulary," Montemurro said. "What we are destroying is the linguistic order, the patterns that we use to encode information."

The researchers found that the original texts spanned a variety of entropy values in different languages, reflecting differences in grammar and structure.

But strangely, the difference in entropy between the original, ordered text and the randomly scrambled text was constant across languages. This difference is a way to measure the amount of information encoded in word order, Montemurro says. The amount of information lost when they scrambled the text was about 3.5 bits per word.

"We found, very interestingly, that for all languages we got almost exactly the same value," he said. "For some reason these languages evolved to be constrained in this framework, in these patterns of word ordering."

This consistency could reflect some cognitive constraints that all human brains run up against, or give insight into the evolution of language, Montemurro suggests.

Cognitive scientists are still debating whether languages have universal features. Some pioneering linguists suggested that languages should evolve according to a limited set of rules, which would produce similar features of grammar and structure. But a study published last month that looked at the structure and syntax of thousands of languages found no such rules.

It may be that universal properties of language show up only at a higher level of organization, suggests linguist Kenny Smith of the University of Edinburgh.

"Maybe these broad-brushed features get down to what's really essential" about language, he said. "Having words, and having rules for how the words are ordered, maybe those are the things that help you do the really basic functions of language. And the places where linguists traditionally look to see universals are not where the fundamentals of language are."

Image: James Morrison/Flickr.

Citation:
"Universal Entropy of Word Ordering Across Linguistic Families." Marcelo A. Montemurro and Damián H. Zanette. PLoS ONE, Vol. 6, Issue 5, May 13, 2011. DOI: 10.1371/journal.pone.0019875.

See Also:

- Evolution of Language Takes Unexpected Turn
- Evolution of Language Parallels Evolution of Species
- Study: Math Skills Rely on Language, Not Just Logic
- Cultural Evolution Could Be Studied in Google Books Database
- Roots of Language Run Deeper Than Speech