

Linguists to re-think reason for short words

January 25 2011, by Lin Edwards

(PhysOrg.com) -- Linguists have thought for many years the length of words is related to the frequency of use, with short words used more often than long ones. Now researchers in the US have shown the length is more closely related to the amount of information the words carry than their frequency of use.

A link between the length of words and how frequently they are used was first proposed in 1935 by George Kingsley Zipf, a Harvard University [linguist](#) and philologist. [Zipf's idea](#) was that people would tend to shorten words they used often, to save time in writing and speaking. The relationship seems intuitive and it seems to apply to many languages with short words such as “the”, “a”, “to”, “and”, “so” (and equivalents in other languages) being frequently used.

Researchers at the Massachusetts Institute of Technology (MIT), led by Steven Piantadosi, tested the Zipf relationship by analysing word use in 11 European languages. They analyzed digitized texts for correlations between words by counting how often all pairs of words occurred in sequence. This information was then used to estimate the probability of words occurring after given previous words or sequences of words. They made the assumption that the more predictable a word is, the less information it conveys, and estimated the information content from information theory, which says the information content is proportional to the negative logarithm of the probability of a word occurring.

Piantadosi said if the word length is directly related to information content this would make the transmission of information through

[language](#) more efficient and also make speech and written texts easier to understand. This is because shorter words, carrying less information, would be scattered through the speech, essentially “smoothing out” the information density and delivering the important information at a steady rate.

The studies suggest that the short words are in fact the least informative and most predictable words rather than the most often used, and that word length is more closely related to the information the [words](#) contain.

The paper is soon to be published in the *Proceedings of the National Academy of Sciences (PNAS)*. Steven Piantadosi belongs to the PhD program with MIT’s Department of Brain and Cognitive Sciences.

More information: Piantadosi, S. T., et al. *Proceedings of the National Academy of Sciences* (2011). *PNAS* paper will appear online at [dx.doi.org/10.1073/pnas.1012551108](https://doi.org/10.1073/pnas.1012551108)

© 2010 PhysOrg.com

Citation: Linguists to re-think reason for short words (2011, January 25) retrieved 19 April 2024 from <https://phys.org/news/2011-01-linguists-re-think-short-words.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.