

Decoding the dictionary: Study suggests lexicon evolved to fit in the brain

The latest edition of the Oxford English Dictionary boasts 22,000 pages of definitions. While that may seem far from succinct, new research suggests the reference manual is meticulously organized to be as concise as possible — a format that mirrors the way our brains make sense of and categorize the countless words in our vast vocabulary.

“Dictionaries have often been thought of as a frustratingly tangled web of words where the definition of word A refers users to word B, which is defined using word C, which ends up referring users back to word A,” said Mark Changizi, assistant professor of cognitive science at Rensselaer Polytechnic Institute. “But this research suggests that all words are grounded in a small set of atomic words — and it’s likely that the dictionary’s large-scale organization has been driven over time by the way humans mentally systematize words and their meanings.”

Dictionaries are built like an inverted pyramid. The most complex words (e.g., “albacore” and “antelope”) sit at the top and are defined by words that are more basic, and thus lower on the pyramid. Eventually all words are linked to a small number of words — called “atomic words,” (such as “act” and “group”) — that are so fundamental they cannot be defined by simpler terms. The number of levels of definition it takes to get from a word to an atomic word is called the “hierarchical level” of the word.

Changizi’s research, which was published online this week and will appear in the June print edition of the *Journal of Cognitive Systems Research*, indicates that the dictionaries we use every day utilize approximately the optimal number of hierarchical levels — and provide a visual roadmap of how the lexicon itself has culturally evolved over tens of thousands of years to help lower the overall “brain space” required to encode it, according to Changizi.

Many other human inventions — such as writing and other human visual signs — have been designed either explicitly or via cultural selection over time so as to minimize their demands on the brain, Changizi said.

By conducting a series of calculations based on the estimation that the most complex words in the dictionary total around 100,000 different terms, and that the number of atomic words range from 10 to 60, Changizi was able to devise three signature features present in the most efficient dictionaries — as well as in their human counterpart, the brain.

Most importantly, he discovered that the total number of words across all the definitions in the dictionary (and thus the size of the dictionary) changes in relation to the total number of hierarchical levels present. Optimal dictionaries should have approximately seven hierarchical levels, according to Changizi.

“The presence of around seven levels of definition will reduce the overall size of the dictionary, so that it is about 30 percent of the size it would be if there were only two hierarchical levels,” Changizi said.

Additionally, users will find that there are progressively more words at each successive hierarchical level, and that each hierarchical level contributes mostly to the definitions of the words just one level above their own, according to Changizi, who put his three predictions to the test by studying actual dictionaries.

The Oxford English Dictionary and WordNet — a large, online lexical database of English, developed at Princeton University — were found to possess all three signatures of an economically organized dictionary,

and thus were organized in such a way as to economize the amount of dictionary space required to define the lexicon, according to Changizi.

“Somehow, over centuries, these revered reference books have achieved near-optimal organization,” Changizi said. “That optimality can likely be attributed to the fact that cultural selection pressures over time have shaped the organization of our lexicon so as to require as little mental space and energy as possible.”

Changizi believes his research has potential applications in the study of childhood learning, where scientists could analyze how students learn vocabulary words and possibly develop ways to optimize that learning process.

Source: Rensselaer Polytechnic Institute

This document is subject to copyright. Apart from any fair dealing for the purpose of private study, research, no part may be reproduced without the written permission. The content is provided for information purposes only.