



A pair of researchers, one with the University of Rochester the other the University of California has found that combining all the data necessary to store and use the English language in the brain adds up to approximately 1.5 megabytes. In their paper published in the journal *Royal Society Open Science*, Francis Mollica and Steven Piantadosi describe applying information theory to add up the amount of data needed to store the various parts of the English language.

As infants, humans begin acquiring and speaking the [language](#) of those around them—how it happens is still a mystery, but scientists know that it entails much more than storing words alongside definitions like a dictionary. There are associative clues with words, for example, such as the concept of flight with the word "bird," or even "wing," or "robin." There is also information that tells the brain how to pronounce a word and how it can and cannot be used with other words, and the sounds that make up a word when spoken. In the new effort, Mollica and Piantadosi undertook the task of converting all of the ways our brain might store a language into data amounts. To do so, they used [information theory](#), a branch of mathematics that focuses on how information is coded via sequences of symbols.

To make their calculations, the researchers assigned quantifiable size estimates to the various aspects of the English language. They began by assigning phonemes, the sounds that stack into spoken words. They noted that humans use approximately 50 phonemes and suggested each would require approximately 15 bits to store. They next moved on to vocabulary, estimating that the [average person](#) knows approximately 40,000 words—taken together, they estimated it would add up to approximately 400,000 bits. Next on the list was semantics for those 40,000 [words](#)—that added up to approximately 12 million bits. They also noted that word frequency is important—they added in another

80,000 bits to account for that. They tossed in another 700 bits to store syntax rules. Adding it all up came to approximately 1.56 megabytes—close to the amount needed to [store](#) a single digital picture.

**More information:** Francis Mollica et al. Humans store about 1.5 megabytes of information during language acquisition, *Royal Society Open Science* (2019). [DOI: 10.1098/rsos.181393](https://doi.org/10.1098/rsos.181393)

[Press release](#)

© 2019 Science X Network

Citation: Researchers estimate it takes approximately 1.5 megabytes of data to store language information in the brain (2019, March 27) retrieved 5 May 2024 from <https://medicalxpress.com/news/2019-03-approximately-megabytes-language-brain.html>

<p>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.</p>
--