

"Kernel" lexicon of languages remains stable in the long run

October 7 2014, by Sandra Jacob

The frequency with which we use different words changes all the time, new words are invented or fall out of use. Yet little is known about the dynamics of lexical change across languages. Researchers of Kazan Federal University in Russia and the Max Planck Institute for Evolutionary Anthropology in Leipzig, Germany, have now studied the lexical evolution of English in comparison to Russian, German, French, Spanish and Italian using the Google Books N-Gram Corpus. They found that major societal transformations such as wars cause faster changes in word frequency distributions, whereas lexical evolution is dampened during times of stability, such as the Victorian Era. Furthermore, the researchers found British and American English to drift apart during the first half of the 20th century, but then begin to re-converge, likely due to the mass media. Apart from these peculiarities, however, the researchers also find similar rates of change across languages at larger time scales, revealing universal trends governing lexical evolution.

The lexicon of a language reflects the world of its speakers. Accordingly, changes in the lexicon of a language reflect changes in the environment. In their current study Søren Wichmann of the Max Planck Institute for Evolutionary Anthropology and his collaborators of Kazan Federal University studied the dynamics of lexical evolution over time and across languages. To this aim the researchers used the Google Books N-Gram Corpus to monitor word usage during the past five centuries. Wichmann and colleagues focused on single words, so-called 1-grams, from six different languages and looked specifically at how frequently these words were used year by year.

Wichmann and colleagues found that major frequency changes are likely triggered by historical events, such as World War I and World War II or the October Revolution in Russia. "Any transformation in society will change the frequency of words", says Wichmann. "With the start of a war or during a revolution new words enter a language reflecting changes in the way people are affected by the world around them." During stable times, as for instance the Victorian Era in Britain, the language was relatively constant with lexical changes occurring less frequently. "Interestingly enough, in this process, British English lags behind American English by 20 years."

The researchers furthermore looked at lexical variation between American English and British English and found that in the 1850s these two dialects of the English language were still diverging more and more until, in the 1950s, they started to re-converge. "By the mid-twentieth century the two variants begin to converge, quite likely owing to the advent of the [mass media](#), including radio and television, and in this process of convergence it is British English that tries to catch up with the evolution of American English rather than the other way around", says Wichmann.

In general, the researchers found that words which are more frequently used in a language, like articles, prepositions and conjunctions, remain more stable over time, while less frequently used words are more prone to change. Across languages the rate of change is also smaller for words that are part of the "kernel lexicon", the [words](#) that constitute 75 percent of texts in the written [language](#). "These results suggest that once we zoom out and look at wider time intervals, the effects of historically contingent changes cancel out, and languages begin to behave quite similarly", says Wichmann.

More information: V. Bochkarev, V. Solovyev, S. Wichmann, "Universal versus historical contingencies in lexical evolution," *Journal*

of the Royal Society Interface, October 2014

Provided by Max Planck Society

Citation: "Kernel" lexicon of languages remains stable in the long run (2014, October 7) retrieved 26 April 2024 from <https://phys.org/news/2014-10-kernel-lexicon-languages-stable.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.