

Machine translation tools find word meanings vary based on news viewership

October 23 2020, by Byron Spice



Credit: Pixabay/CC0 Public Domain

It's not news that U.S. politics are highly polarized or that polarization affects cable news channels. But researchers at Carnegie Mellon

University, using computer translation tools in an unprecedented way, have found that even the meanings of some words are now polarized.

Everyone is speaking English, they said, yet the computer analysis of social media discussions shows viewers of different news channels are, in a sense, speaking different languages.

Based on millions of user comments on the YouTube channels for four leading cable news outlets, it seems that viewers of right-wing outlets think of "Burisma," in the same way that their left-wing counterparts think of "Kushner." A "protest" to one set of viewers is a "riot" to another. For one, it's a "mask," to another, a "muzzle."

"Black Lives Matter" (BLM) in CNN English is equivalent to "All Lives Matter" in Fox News English. Even more extreme, some right-wing news viewers use "BLM" in the same context as left-wing news viewers use "KKK" (Ku Klux Klan).

"Some of these so-called misaligned pairs seem pretty obvious," said Mark S. Kamlet, University Professor of Economics and Public Policy. "But it's surprising how different some of them are. It gives you a sense of the really tragic polarization that exists today."



Mark S. Kamlet holds joint appointments to the Heinz College of Information Systems and Public Policy and the Dietrich College of Humanities and Social Sciences. He's leading innovative research using computer translation tools. Credit: Carnegie Mellon University

Modern [machine translation](#) methods determine the meaning of a word based in large part on context—the other words that it usually appears closest to in texts. "Hello" in English and "hola" in Spanish are identical greetings and, thus, appear in the same context in different languages.

Ashiqur KhudaBukhsh, a project scientist in the School of Computer Science's Language Technologies Institute, said the idea behind the new research was to use the same method to analyze the polarization of social media. The goal was to find different English words that are used in the

same context by people speaking different news languages.

For instance, a conservative might say "Democrats are the greatest threat to America today," while liberals might say "Republicans are the greatest threat to America today." Democrats and Republicans are used in the same context, making them misaligned pairs and an indication of political polarization.

To perform their analysis, the researchers used a data set of 86.6 million comments by 6.5 million users to more than 200,000 [news](#) videos from CNN, Fox News, MSNBC and One America News Network (OANN). The software completes the analysis automatically, without human intervention.

"We think our method is powerful because it's efficient," KhudaBukhsh said. "You don't have to read millions of comments. But if you know that 'mask' translates into 'muzzle,' you immediately know a debate is going on surrounding freedom of speech and mask use."

In addition to detecting these misaligned pairs, the method also calculates the degree of similarity between the "languages." In a four-way analysis of CNN, MSNBC, Fox News and OANN, words translated from MSNBC English to CNN English had a 63% similarity, while words translated from MSNBC English to OANN English had just a 42% similarity.

The researchers also compared the comments of viewers of CNN, Fox News and MSNBC with more than 4 million comments by viewers of late night comedians Trevor Noah, Seth Meyers, Stephen Colbert, Jimmy Kimmel and John Oliver. They found words translated from Fox News English to comedian English were 75% similar, while words translated from CNN English to comedian English were 83% similar.

Doing the same analysis by hand would be impossible, said Kamlet, who holds joint appointments to the Heinz College of Information Systems and Public Policy and the Dietrich College of Humanities and Social Sciences.

"We use a standard statistical package that takes each word and maps it into a 100 dimension space," he explained. "Obviously, you might be able to do cross tabs by hand. But even with cross tabs, you're talking about millions of comments."

The research team includes Tom Mitchell, Founders University Professor; and Rupak Sarkar, research engineer for a fall 2020 seminar course on tracking political sentiments using machine learning taught by KhudaBukhsh, Kamlet and Mitchell. Their paper has been submitted to a computer science conference and is available online at arXiv.

More information: KhudaBukhsh et al., We Don't Speak the Same Language: Interpreting Polarization Through Machine Translation. arXiv:2010.02339 [cs.CL]. arxiv.org/pdf/2010.02339.pdf

Provided by Carnegie Mellon University

Citation: Machine translation tools find word meanings vary based on news viewership (2020, October 23) retrieved 19 April 2024 from <https://techxplore.com/news/2020-10-machine-tools-word-vary-based.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.
