

Word detectives: Science may help finger opinion columnist

September 7 2018, by Seth Borenstein



Credit: CC0 Public Domain

Language detectives say the key clues to who wrote the anonymous New York Times opinion piece slamming President Donald Trump may not be the odd and glimmering "lodestar," but the itty-bitty words that

people usually read right over: "I," "of" and "but."

And lodestar? That could be a red herring meant to throw sleuths off track, some experts say.

Experts use a combination of language use, statistics and computer science to help figure out who wrote documents that are anonymous or possibly plagiarized. They've even solved crimes and historical mysteries that way. Some call the field forensic linguistics, others call it stylometry or simply doing "author attribution."

The field is suddenly at center stage after an unidentified "senior administration official" wrote in the Times that he or she was part of a "resistance" movement working from within the administration to curb Trump's most dangerous impulses.

"My phone has been ringing off the hook with requests to do that analysis and I just don't have the time," says Duquesne University computer and language scientist Patrick Juola.

Robert Leonard, a Hofstra University linguistics professor who has helped solve murders by examining language, says if experts could get the right number of writing samples from officials whose identities are known, "an analysis could certainly be done."

One political scientist figures there are about 50 people in the Trump administration who fit the Times' description as a senior administration official and could be the author. The key would be to look at how they write, the words they use, what words they put next to each other, spelling, punctuation and even tenses, experts say.

"Language is a set of choices. What to say, how to say and when to say it," Juola says. "And there's a lot of different options."

One of the favorite techniques of Juola and other experts is to look at what's called "function words." These are words people use all the time but that are hard to define because they more provide function than meaning. Some examples are "of," "with," "the," "a," "over" and "and."

"We all use them but we don't use them in the same way," Juola says. "We don't use them in the same frequency." Same goes with apostrophes and other punctuation.

For example, do you say "different from" or "different than?" asks computer science and data expert Shlomo Argamon of the Illinois Institute of Technology.

Women tend to use first- and second-person pronouns more—"I," "me" and "you"—and more present tense, Argamon says.

Men use "the," "of," "this" and "that" more often, he says.

"You look for clues and you try to assess the usefulness of those clues," Argamon says. But he is less optimistic that the Trump opinion piece case will be cracked for various reasons, including the New York Times' editing for style and possible efforts to fool language detectives with words that someone else likes to use such as "lodestar." Mostly, he's pessimistic because to do a proper comparison, samples from all suspects have to be gathered and have to be similar, such as all opinion columns as opposed to novels, speeches or magazine stories.

Rachel Greenstadt at Drexel University studies when people try to throw off investigators with words they don't normally use or purposeful bad spellings. She says her first instinct is that the word "lodestar"—one Vice President Mike Pence has used several times—is "a red herring." It seems too deliberate.

"Most people are still looking for sound bite-sized features like lodestar instead of trying to get a handle on the whole picture," says Hofstra's Leonard.

Greenstadt says language analysis "could kind of contribute to the picture" of who wrote the Times' opinion pieces, but she adds "by itself, I'd be concerned to use it."

Still, with the right conditions words matter.

Juola testified in about 15 trials and handled even more cases that never made it to court. His biggest case was in 2013, when a British newspaper got a tip that the book "The Cuckoo's Calling" by Robert Galbraith was really written by Harry Potter author J.K. Rowling. In about an hour, Juola fed two Rowling books, "The Cuckoo's Calling" and six other novels into his computer, analyzed the language patterns with four different systems and concluded that Rowling did it.

A couple of days later, Rowling confessed.

It was far from the first time that language use fingered the real culprit. The Unabomber's brother identified him because of his distinctive writing style. Field pioneers helped find a kidnapper who used the unique term "devil strip" for the grassy area between the sidewalk and road. The phrase is only used in parts of Ohio.

Even in politics, words are poker tells. In 1996, the novel "Primary Colors" about a Clintonesque presidential candidate set Washington abuzz trying to figure out who was the anonymous author. An analysis by a Vassar professor and other work pointed to Newsweek's Joe Klein and he finally admitted it.

But the literary sleuthing goes back to the founding of the republic.

Historians had a hard time figuring out which specific Federalist Papers were written by Alexander Hamilton and which were by James Madison. A 1963 statistical analysis figured it out: One of the many clues came down to usage of the words "while" and "whilst." Madison used "whilst"; Hamilton preferred "while."

Juola says experts in the field can generally tell introverts from extroverts, men from women, education level, age, location, almost everything but astrological sign.

"The science is very good," Juola said. "It's not quite DNA. It's actually considered by some scientists to be considered the second-most accurate form of forensic identification we have because it is so good."

© 2018 The Associated Press. All rights reserved.

Citation: Word detectives: Science may help finger opinion columnist (2018, September 7) retrieved 1 July 2024 from <https://phys.org/news/2018-09-word-science-finger-opinion-columnist.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.