An Israeli algorithm sheds light on the Bible

June 30 2011, By MATTI FRIEDMAN, Associated Press

This undated file photo made available by the Yad Ben Zvi Institute on Nov. 8, 2007, shows a piece of an ancient parchment believed to be part of the most authoritative manuscript of the Hebrew Bible, the Aleppo Codex. Software developed by an Israeli team of scholars led by Moshe Koppel, of Bar Ilan University near Tel Aviv, is giving intriguing new hints about what researchers believe to be the multiple hands that wrote the Bible. (AP Photo/Yad Ben Zvi Institute, File)

Software developed by an Israeli team is giving intriguing new hints about what researchers believe to be the multiple hands that wrote the Bible. The new software analyzes style and word choices to distinguish parts of a single text written by different authors, and when applied to the Bible its algorithm teased out distinct writerly voices in the holy book.

The program, part of a sub-field of <u>artificial intelligence</u> studies known

as authorship attribution, has a range of potential applications - from helping law enforcement to developing new computer programs for writers. But the Bible provided a tempting test case for the algorithm's creators.

For millions of Jews and Christians, it's a tenet of their faith that God is the author of the core text of the Hebrew Bible - the Torah, also known as the Pentateuch or the Five Books of Moses. But since the advent of modern biblical scholarship, academic researchers have believed the text was written by a number of different authors whose work could be identified by seemingly different ideological agendas and linguistic styles and the different names they used for God.

Today, scholars generally split the text into two main strands. One is believed to have been written by a figure or group known as the "priestly" author, because of apparent connections to the temple priests in Jerusalem. The rest is "non-priestly." Scholars have meticulously gone over the text to ascertain which parts belong to which strand.

When the new software was run on the Pentateuch, it found the same division, separating the "priestly" and "non-priestly." It matched up with the traditional academic division at a rate of 90 percent - effectively recreating years of work by multiple scholars in minutes, said Moshe Koppel of Bar Ilan University near Tel Aviv, the computer science professor who headed the research team.

"We have thus been able to largely recapitulate several centuries of painstaking manual labor with our automated method," the Israeli team announced in a paper presented last week in Portland, Oregon, at the annual conference of the Association for Computational Linguistics. The team includes a computer science doctoral student, Navot Akiva, and a father-son duo: Nachum Dershowitz, a Tel Aviv University computer scientist, and his son, Idan Dershowitz, a Bible scholar at Hebrew

University in Jerusalem.

The places in which the program disagreed with accepted scholarship might prove interesting leads for scholars. The first chapter of Genesis, for example, is usually thought to have been written by the "priestly" author, but the software indicated it was not.

Similarly, the book of Isaiah is largely thought to have been written by two distinct authors, with the second author taking over after Chapter 39. The software's results agreed that the book might have two authors, but suggested the second author's section actually began six chapters earlier, in Chapter 33.

The differences "have the potential to generate fruitful discussion among scholars," said Michael Segal of Hebrew University's Bible Department, who was not involved in the project.

Over the past decade, computer programs have increasingly been assisting Bible scholars in searching and comparing texts, but the novelty of the new software seems to be in its ability to take criteria developed by scholars and apply them through a technological tool more powerful in many respects than the human mind, Segal said.

Before applying the software to the Pentateuch and other books of the Bible, the researchers first needed a more objective test to prove the algorithm could correctly distinguish one author from another.

So they randomly jumbled the Hebrew Bible's books of Ezekiel and Jeremiah into one text and ran the software. It sorted the mixed-up text into its component parts "almost perfectly," the researchers announced.

The program recognizes repeated word selections, like uses of the Hebrew equivalents of "if," "and" and "but," and notices synonyms: In

some places, for example, the Bible gives the word for "staff" as "makel," while in others it uses "mateh" for the same object. The program then separates the text into strands it believes to be the work of different people.

Other researchers have looked at linguistic fingerprints in less sacred texts as a way of identifying unknown writers. In the 1990s, the Vassar English professor Donald Foster famously identified the journalist Joe Klein as the anonymous author of the book "Primary Colors" by looking at minor details like punctuation.

In 2003, Koppel was part of a research team that developed software that could successfully tell, four times out of five, if the author of a text was male or female. Women, the researchers found, are far more likely to use personal pronouns like "she" and "he," while men prefer determiners like "that" and "this" - women, in other words, talk about people, while men prefer to talk about things. That success sparked debate about how gender shapes the way we think and communicate.

Research of this kind has potential applications for law enforcement, allowing authorities to catch imposters or to match anonymous texts with possible authors by identifying linguistic tics. Because the analysis can also help identify gender and age, it might also allow advertisers to better target customers.

The new software might be used to investigate Shakespeare's plays and settle lingering questions of authorship or co-authorship, mused Graeme Hirst, a professor of computational linguistics at the University of Toronto. Or it could be applied to modern texts: "It would be interesting to see if in more cases we can tease apart who wrote what," Hirst said.

The algorithm might also lead to the creation of a style checker for documents prepared by multiple authors or committees, helping iron out

awkward style variations and creating a uniform text, Hirst suggested.

What the algorithm won't answer, say the researchers who created it, is the question of whether the Bible is human or divine. Three of the four scholars, including Koppel, are religious Jews who subscribe in some form to the belief that the Torah was dictated to Moses in its entirety by a single author: God.

For academic scholars, the existence of different stylistic threads in the Bible indicates human authorship.

But the research team says in their paper they aren't addressing "how or why such distinct threads exist."

"Those for whom it is a matter of faith that the Pentateuch is not a composition of multiple writers can view the distinction investigated here as that of multiple styles," they said.

In other words, there's no reason why God could not write a book in different voices.

"No amount of research is going to resolve that issue," said Koppel.

Citation: An Israeli algorithm sheds light on the Bible (2011, June 30) retrieved 2 May 2024 from https://phys.org/news/2011-06-israeli-algorithm-bible.html