'Ghostwriting' the Torah? New algorithm distinguishes contributors to the Old Testament with high accuracy

October 11 2011

In both Jewish and Christian traditions, Moses is considered the author of the Torah, the first five books of the Bible. Scholars have furnished evidence that multiple writers had a hand in composing the text of the Torah. Other books of the Hebrew Bible and of the New Testament are also thought to be composites. However, delineating these multiple sources has been a laborious task.

Now researchers have developed an algorithm that could help to unravel the different sources that contributed to individual books of the Bible. Prof. Nachum Dershowitz of Tel Aviv University's Blavatnik School of Computer Science, who worked in collaboration with his son, Bible scholar Idan Dershowitz of Hebrew University, and Prof. Moshe Koppel and Ph.D. student Navot Akiva of Bar-Ilan University, says that their computer algorithm recognizes linguistic cues, such as word preference, to divide texts into probable author groupings.

By focusing exclusively on writing style instead of subject or genre, Prof. Dershowitz and his colleagues sidestepped several methodological hurdles that hamper conventional Bible scholarship. These issues include a potential lack of objectivity in content-based analysis and complications caused by the multiple genres and literary forms found in the Bible — including poetry, narrative, law, and parable. Their research was presented at the 49th Annual Conference of the Association for Computational Linguistics in Portland.

A keen eye for detail

According to Prof. Dershowitz, the software searches for and compares details that human scholars might have difficulty detecting, such as the frequency of the use of "function" words and synonyms. Such details have little bearing on the meaning of the text itself, but each author or source often has his own style. This could be as innocuous as an author's preference for using the word "said" versus "spoke."

To test the validity of their method, the researchers randomly mixed passages from the two Hebrew books of Jeremiah and Ezekiel, and asked the computer to separate them. By searching for and categorizing chapters by synonym preference, and then looking at usage of common words, the computer program was able to separate the passages with 99 percent accuracy. The software was also able to distinguish between "priestly" materials — those dealing with issues such as religious ritual — and "non-priestly" material in the Torah, a categorization that is widely used by Bible scholars.

While the algorithm is not yet advanced enough to give the researchers a precise number of probable authors involved in the writing of the individual books of the Bible, Prof. Dershowitz says that it can help to identify transition points within the text where a source changes, potentially shedding new light on age-old debates.

Categorizing the unknown

Part of a new field called "digital humanities," computer software like Prof. Dershowitz's is being developed to give more insight into historical sources than ever before. Programs already exist to help attribute previously anonymous texts to well-known authors by writing style, or uncover the gender of a text's author. But the Bible presents a new

challenge, says Prof. Dershowitz, as there are no independently attributed works to which to compare the Biblical books.

The Torah algorithm may also provide new information about other enigmatic source material, such as the many pamphlets and treatises of unknown composition that are scattered throughout history. And because the software can identify subtle linguistic cues, it is able to uncover differences within mere percentage points, a feat that has never before been possible. "If the computer can find features that Bible scholars haven't noticed before, it adds new dimensions to their scholarship. That would be gratifying in and of itself," says Prof. Dershowitz.

Provided by Tel Aviv University