Orthography and Biblical Criticism

Idan Dershowitz, Nachum Dershowitz, Tomer Hasid, Amnon Ta-Shma
Department of Bible, Hebrew University, Israel
School of Computer Science, Tel Aviv University, Israel

Who Wrote the Bible?

Hypothesis: Differences in spelling reflect different authorship and/or different genres.

Computers know how to count. How do we measure the statistical significance of differences?

Previous work used ad-hoc filters

<table>
<thead>
<tr>
<th>Words that do not appear in both components</th>
</tr>
</thead>
<tbody>
<tr>
<td>Words that are monochromatic</td>
</tr>
<tr>
<td>Words that are almost entirely in one component</td>
</tr>
<tr>
<td>Words that are almost always the same</td>
</tr>
</tbody>
</table>

Bible scholars often speak of several major components in the Pentateuch (Torah), some of which are disputed.

- Jahwistic (much of Genesis and Exodus)
- Elohistic (much of Exodus and Numbers)
- Priestly (first part of Leviticus)
- Holiness (latter part of Leviticus)
- Deuteronomic (most of Deuteronomy)

Problem: Polysynyny — Words may have multiple meanings, and senses may have preferred spellings.

Solution: Use sense-tagged Bible (Strong’s Concordance).

Problem: Aggregating all words and using the standard \( \chi^2 \) test biases the outcome by word frequencies (Simpson’s Paradox).

Solution: Use the Cochran-Mantel-Haenszel (CMH) test of significance.

<table>
<thead>
<tr>
<th>Source A</th>
<th>Source B</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plene</td>
<td>m</td>
<td>m</td>
</tr>
<tr>
<td>Defective</td>
<td>n</td>
<td>n</td>
</tr>
</tbody>
</table>

Results

- Priestly law has significantly different orthography than Deuteronomic narrative.
- Deuteronomic law is significantly different than Elohistic narrative.
- Which is the more paleological? Look at the odds ratios.
- Priestly law is more paleological than Deuteronomic narrative.
- The putative Elohistic document is more paleological than the Deuteronomic source.

Conclusions

- New tool for textual studies.
- Ideal for many sparse features (e.g. word choice).

Ramifications

- Pentateuch is a combination of written, rather than oral, sources.
- The 10th c. Masoretic Text retains subtle differences in spelling that predate reduction more than a millennium earlier.
- Potential ramifications for debates regarding the relative dating of the Priestly and Deuteronomic sources.
- Linguistic suggestions: זכרון is a late form.

Future Work

- Consider the (relatively short) poetic sections, as well.
- Fill in the blanks. (Too few words for CMH.)
- Consider orthographic variations even if not consistently plene or defective.
- Consider additional proposed divisions of the text.
- Conduct experiment on Samaritan Pentateuch.