



# Orthography and Biblical Criticism



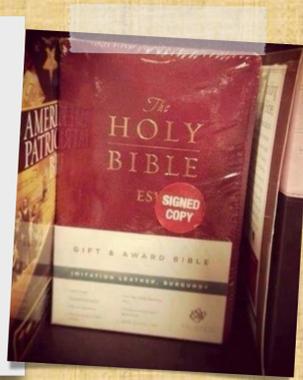
Idan Dershowitz,\* Nachum Dershowitz,\* Tomer Hasid,\* Amnon Ta-Shma\*  
 \* Department of Bible, Hebrew University, Israel      \* School of Computer Science, Tel Aviv University, Israel

## 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 the differences?



## Orthographic Variants of Genesis 1:14

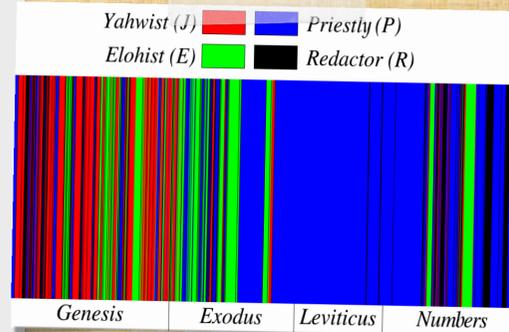
:MT יהי מארת ברקיע השמים להבדיל בין היום ובין הלילה והיו לאתת ולמועדים ולימים ושנים  
 :4Q7 יהי מארות ברקי[ע] השמים להבדיל בין היום ובין הלילה ויהיו לאתות ולמעדים ל[ימים ושנים]  
 :Sam יהי מאורות ברקיע השמים להאיר על הארץ ולהבדיל בין היום ובין הלילה והיו לאתות ולמועדים ולימים ושנים

"They were experts regarding defective and plene spelling; we are not." — *Talmud*

	D	D	E	E	P	P	H	H	J
D		0.90	0.07	—	0.32	0.00	—	0.30	0.11
D			0.00	0.20	0.85	0.09	—	0.80	0.03
E				—	0.78	0.56	—	0.48	0.67
E					—	—	—	—	0.79
P						—	—	—	0.79
P						0.37	0.45	0.24	0.85
P								—	0.07
H								—	—
H									0.18

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

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



## Hebrew Spelling — Plene and Defective

Several letters that originally represented consonants evolved to reflect vowels, as well. When used in such a way, they are known as *matres lectionis* — “mothers of reading” — as they facilitate comprehension.

Conventions have changed over the centuries.

Four ways to write a single biblical word (attested in Masoretic Text and/or in Dead Sea Scrolls):

לוחות  
 לוחת  
 להות  
 להת

**Problem:** Plene vs. defective does not necessarily reflect modern vs. archaic spelling. Spelling generally became “fuller” over time, but the reverse direction is documented, as well.

Example: \*יוצת > יוצאת

**Solution:** Use neological (new) vs. paleological (old) categories.

Consider each spelling of each syllable of each word sense as a separate feature.

## Results

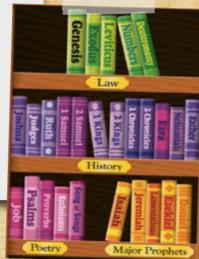
- Priestly law has significantly different orthography than Deuteronomistic narrative.
- Deuteronomistic law is significantly different than Elohist narrative.
- Which is the more paleological? Look at the odds ratios.
  - Priestly law is more paleological than Deuteronomistic narrative.
  - The putative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

## Ramifications

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

## Genres

Narrative  
 Law  
 Poetry



**Problem:** Polysemy — 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.

- Works with multiple features.
- A 2x2 frequency table for each feature.
- Assigns the correct weight to each feature.

## Conclusions

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

	Source A	Source B	Total
Plene	X	m	m
Defective	Y	n <sub>i</sub>	n
Total	X	(m	m

## 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.

## Previous work used ad-hoc filters

- Words that do not appear in both components
- Words that are monochromatic
- Words that are almost entirely in one component
- Words that are almost always the same