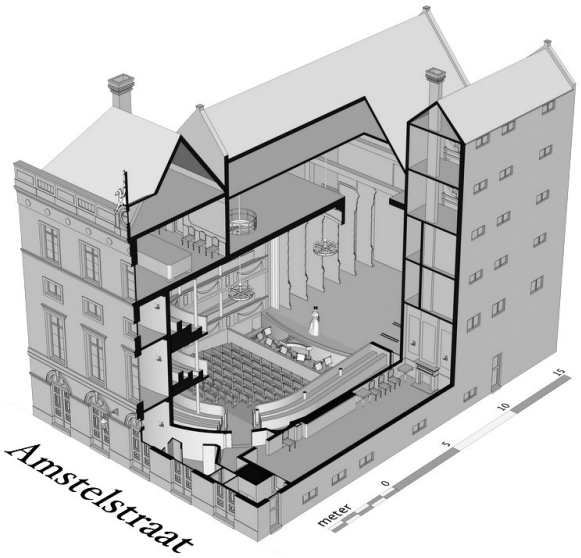

Fig. 2: German Theatre, Amsterdam, ca.1795

References

1. **C.M. Fogarty and Tom Lawrenson** (1981), *The lessons of the reconstructed performance*, Theatre Survey 22, 141-159. Also see: • **Robert K. Sarlós** (1989), *Performance reconstruction: the virtual link between the past and the future*, in Thomas Postlewait & Bruce A. McConachie (eds.), Interpreting the Theatrical Past. Essays in the Historiography of Performance. Iowa City: University of Iowa Press), pp. 198-247.

2. **John Golder.** (1984) *The Théâtre du Marais in 1644: a new look at the old evidence concerning France's second public theatre*, Theatre Survey 25, pp.127-152.

3. **David Thomas** (1999), *The design of the Théâtre du Marais and Wren's Theatre Royal, Drury Lane: A computer-based investigation*, Theatre Notebook 53:3, pp. 127-145; **Christa Williford** (2001), *Computer Modelling Classical French Theatre Spaces: Three Reconstructions*, in Philip Tomlinson (ed.), French 'Classical' Theatre Today. Teaching, Research, Performance (Amsterdam: Rodopi), pp. 155-164; **Frank Mohler** (1999a), *The Survival of the Mechanized Flat Wing Change: The Court Theatres of Gripsholm, Cesky Krumlov, and Drottningholm*, in Theatre Design & Technology (1999a): 6-56; **Frank Mohler** (2004), *The Chateau Theatre in Litomysl and the Scenery of Josef Platzer*, in Theatre Design & Technology (2004): 24-31.

4. **David Thomas** (1999), *The design of the Théâtre du Marais and Wren's Theatre Royal, Drury Lane: A computer-based investigation*, Theatre Notebook 53:3, pp. 127-145; **Christa Williford** (2001), *Computer Modelling Classical French Theatre Spaces: Three Reconstructions*, in Philip Tomlinson (ed.), French 'Classical' Theatre Today. Teaching, Research, Performance (Amsterdam: Rodopi), pp. 155-164; **Frank Mohler** (1999a), *The Survival of the Mechanized Flat Wing Change: The Court Theatres of Gripsholm, Cesky Krumlov, and Drottninghol*, in Theatre Design & Technology: 6-56; **Frank Mohler** (2004), *The Chateau Theatre in Litomysl and the Scenery of Josef Platzer*, in Theatre Design & Technology: 24-31.

5. **Richard Beacham** (1999, *'Eke out our performance with your mind': reconstructing the theatrical past with the aid of computer simulation*, in Terry Coppock (ed.), Information Technology and Scholarship. Applications in the Humanities (London: The British Library), pp. 131-154; **Frank Mohler** (1999), *Computer Modelling as a Tool for the Reconstruction of Historic Theatrical Production Techniques*, Theatre Journal 51:4, pp. 417-431.

Orthography and Biblical Criticism

Dershowitz, Idan
Tel Aviv University, IL

Dershowitz, Nachum
Tel Aviv University, IL

Hasid, Tomer
Tel Aviv University, IL

Ta-Shma, Amnon
Tel Aviv University, IL

Abstract

Biblical Hebrew exhibits considerable orthographic variability. A single word may be spelled in multiple ways—often within the same book. In this study, we set out to determine if these differences in spelling correspond in any way to scholarly theories regarding the authorship of the Pentateuch. We use a statistical test that is designed for use when there are many features to take into account, each of which occurs only sparsely. Our results indicate that despite the tortuous editing processes and countless generations of hand-copied manuscripts, certain statistically significant correlations between orthography and the hypothesized sources remain.

1 Introduction

The Pentateuch has been attributed to several major sources. We investigate here whether there exists a statistically significant correlation between these postulated sources and variations in spelling in the received Masoretic text.

We consider the source units about which there is broad agreement among Bible scholars, namely, the classic four-source division of the text into J, E, P, and D[1], plus the consensual source H. We only considered words occurring in paragraphs for which there is relative agreement among scholars. We also compared genres—narrative and legal—since different genres might employ different conventions. (We ignored poetry with its distinctive language register.)

Regarding orthography, we examined the use of consonants to represent vowels, a practice that has changed over time. The Canaanite languages—Hebrew among them—were generally recorded in alphabetic writing sans vowels. With time, certain characters began to serve double duty, representing vowels, as well as consonants. These letters are known as matres lectionis ("mothers of reading"). The written representation of vowels increased from one century to the next, but it appears there was variation even within a single period.

While matres lectionis proliferated, another process complicated matters. When a word's pronunciation evolved so that a particular consonant stopped being pronounced, the letter representing that consonant was not always written. For this reason, we classify spellings as either "neological" (reflecting innovative orthography) or "paleological" (conforming to earlier norms).

We have, then, two labelings to work with:

– By source and genre: J, E, E-law, P, P-law, D, D-law, H, or H-law. For simplicity, we will refer to these nine categories as "sources."
– By orthography: paleological or neological.

We apply a statistical test, Cochran-Mantel-Haenszel (CMH)[2], to check whether there is a correlation between the two labelings, that is, whether any particular source is more paleological than others.

2 Possible Approaches to the Statistical Problem

Assume we have two sources, A and B, plus an orthographic classification, and would like to check whether the classifications are correlated.

2.1 The Naïve Approach

A naïve approach is to count the total frequencies of neological and paleological syllables for each one of the sources and then run a χ^2 test for the resulting 2×2 table. We believe this approach is not good. If A has a different word distribution than B, it is possible that even when the sources have identical spelling the naïve test would declare the two classifications strongly correlated, simply because one tends to use certain words that are spelled as neological more often. Working with aggregated data is therefore most likely to catch word distribution differences between sources, rather than spelling differences.

2.2 Filtering

Andersen and Forbes[3] conducted an extensive automated study of spelling in the Bible. They created 65 classes, based on grammatical form, vocalization and stress. Within each, they used the naïve approach, aggregating all words in a class and checking the χ^2 score of the A/B and plene/defective classifications. As they use aggregated data (within each class), they still face the word distribution problem, and in particular their method is vulnerable to words like ðð (lo ̄) that appear frequently and mostly in one form. They tackled this problem with several ad-hoc filters and rules, such as filtering out words that almost always appear only in one form (see, e.g., [1, Chap. 10]). However, on a conceptual level, it seems that whatever set of filters is used, there is still the problem that differences in word distribution between different sources is interpreted as spelling differences. It appears they could not exhibit a conclusive relationship between stress and spelling (see [1, Epilog]), which seems to undermine the rationale behind dividing words into those classes.

2.3 New Approach

Our goal is to identify spelling differences even when each source may have a different distribution of words (e.g., legal texts tend to use legal terminology). The appropriate statistical technique is CMH. The idea is to bypass the language distribution problem by having a 2 × 2 contingency table for each word in the language, describing the number of neological/ paleological occurrences of the word in each source. The test combines the data from all the 2 × 2 tables in a way that gives weight to the statistical significance of the data in each table, ignoring the frequencies of the word in each source.

We enumerate events at the finest possible granularity, classifying each syllable of each occurrence of a particular word in the text. For each syllable, we have one stratum (in the statistical sense of stratified data) containing a 2×2 contingency table describing the number of neological/ paleological occurrences of that syllable of the word in each source.

This observation, as simple as it is, is conceptually important and is crucial for getting sound statistical data on the problem. As a side effect of using CMH, we avoid ad-hoc filters and rules.

3 Experimental Design and Results

We have a stratum for each syllable and use the tagging of the biblical text into word senses (Strong number). We consider two sources at a time, computing the following statistics:

1. the χ^2 and p-value of the CMH test;
2. the validity of the χ^2 test with the Rule of 5;
3. the common odds ratio;
4. the $p=1-\alpha$ confidence intervals for the logarithm of the common odds ratio, taking $\alpha = 0.05$.

The *p*-values and the ln(odds) values for the pairs of sources are tabulated as follows:

	D	D-law	E	E-law	P	P-law	H	H-law
D-law	0.900							
E	0.073	**0.000**						
E-law	∼	0.198	∼					
P	0.323	0.848	0.777	∼				
P-law	**0.000**	0.087	0.588	∼	0.327			
H	∼	∼	∼	∼	0.445	∼		
H-law	0.296	0.804	0.482	∼	0.240	0.067	∼	
J	0.108	0.033	0.671	0.790	0.852	0.276	∼	0.184

The cells with tildes are those that failed to pass the Rule of 5 [3].

In the following table, the number in cell (*i, j*) tells us how much source *i* is more likely to be paleological than source *j*. Roughly speaking, if the cell (*i, j*) = 0.44, then *i* uses the paleological form $20.44 \geq 1$ more often then *j*. If it is zero they have the same frequencies, $20 = 1$; if it is negative, source *i* is less paleological, $2^{-0.44} \leq 1$.

	D	D-law	E	E-law	P	P-law	H	H-law
D-law	−0.080							
E	0.460	1.192						
E-law	∼	0.776	∼					
P	0.263	0.118	−0.087	∼				
P-law	0.818	0.503	0.210	∼	0.181			
H	∼	∼	∼	∼	0.546	∼		
H-law	0.516	0.161	−0.388	∼	−0.389	−0.538	∼	
J	0.351	0.521	0.107	−0.237	0.054	0.267	∼	0.653

Thus, D-law appears to be the most neological.

4 Discussion

Our results appear to be of potential interest for several reasons.

– For Bible scholars, they suggest that the countless scribes who edited, expanded, and copied the text(s) that eventually crystallized into the Masoretic text did not change enough to obscure the characteristic spelling of individual units. Our findings open the door to new approaches in the critical analysis of biblical texts, as the value of orthography in such contexts has thus far been underestimated.
– The simple statistical test we use cannot disentangle the many authors of the Bible. However, it does produce some interesting results, that we hope would be combined with other data to shed light on the fascinating question of how the Bible, as we know it today, evolved. In particular, the observation that Deuteronomic narrative is more neological in spelling than Priestly law may be of some value in the ongoing debate regarding the relative dating of P and D. It is possible that that there is some hidden random variable that strongly affects spelling, which might better explain the results. However, in spite of much effort (e.g., see [1]), such a hidden random variable has not been identified.
– Outliers in the analysis may suggest alternate classifications for some linguistic phenomena. The most prominent example for that is the holam syllable פֹה (po) in the word ציפור (tzi-por). The experts labeled ציפור as paleological and ציפר as neological, while our data seems to indicate the opposite (in D-law there are 2 occurrences of ציפור and none of ציפר, in P-law 2 occurrences of ציפור and 9 of ציפר, though D-law is overall more neological than P-law).
– For textual analysis,more generally, this work suggests that the Cochran-Mantel-Haenszel is an appropriate statistical measure, when features are sparse.

References

1. **S. R. Driver** (1913). *An Introduction to the Literature of the Old Testament*. Edin- burgh.

2. **N. Mantel and J. L. Fleiss** (1980). *Minimum expected cell size requirements for the Mantel-Haenszel one-degree-of-freedom chi-square test and a related rapid procedure*. American Journal of Epidemiology, 112(1):129–134.

3. **F. I. Andersen and A. D. Forbes** (1986). *Spelling in the Hebrew Bible*. Biblical Institute Press.

L'Innommable / The Unnamable: The Second Module of the Samuel Beckett Digital Manuscript Project's Hybrid Genetic Edition.

Dillen, Wout
wout.dillen@gmail.com
University of Antwerp

This poster will offer an interactive demonstration of the second module of the *Samuel Beckett Digital Manuscript Project* (BDMP) – an international collaboration between the Centre for Manuscript Genetics at the University of Antwerp, the Beckett International Foundation at the University of Reading, and the Harry Ransom Humanities Research Center at the University of Austin, Texas, with the kind permission of the Estate of Samuel Beckett. As its name implies, the BDMP aims to reunite and make publicly accessible all manuscripts of Samuel Beckett's works, the physical documents of which are located in different holding libraries around the world. This goal will be realized in the form of a hybrid genetic edition that combines a digital archive of the manuscripts organized in twenty-six research modules (one published each year) with a series of twenty-six accompanying volumes analysing the geneses of the texts contained in the corresponding modules. Each of the modules comprises digital facsimiles and transcriptions of all extant manuscripts pertaining to an individual text – or to a collection of shorter texts. The digital archive can be accessed online at www.beckettarchive.org , where you can currently find the project's first module, which combines the manuscripts of the late short prose text *Stirrings Still / Soubresauts*with those of Beckett's last poem *Comment dire / what is the word*[1] . The project's second module, which will present the bilingual genetic dossier of Beckett's novel *L'Innommable / The Unnamable* edited by Dirk Van Hulle, Shane Weller and Vincent Neyt, will be made available online towards the end of 2013[2].

As a hybrid genetic edition, the BDMP combines digital scholarly editing with genetic criticism – a form of literary criticism that studies the dynamics of the writing process. As such, the edition does not aim to support a new reading text of Beckett's works, but rather to highlight the creative process that brought those works about. Traces of this process can be found in the extant manuscripts, more specifically in their many deletions, additions, paralipomena, doodles, etc. According to Pierre-Marc de Biasi, the objective of genetic criticism is twofold (42)[3]: first (1) to locate, collate, and transcribe all of the work's extant versions in order to make them analysable, and then (2) to reconstruct the logic of the work's genesis (also called the 'avant texte') from a chosen perspective. Therefore, rather than focussing solely on an *analysis* of the works' geneses – which can be found in the interpretative, printed component of our edition – we also want to make these geneses *analysable*, by offering fully transcribed (in TEI-compliant XML) and searchable facsimiles in the edition's digital component. Hence, the BDMP complies with Patrick Sahle's recent definition of scholarly hybrid editions, stipulating that such editions should not only be published in different media, but that their different components should complement one another, and that each component should take the possibilities and limitations of its medium into account (64)[4].

Because the modern manuscripts we exhibit are still under copyright and therefore do not yet belong to the public domain, the BDMP still requires its users (either individuals or institutions) to pay a subscription fee to gain access to its materials. Therefore, this poster will be a great opportunity for potential users to receive a personal, hands-on introduction to the project. The poster's main focus will be on the project's newest module, and on the differences between both modules. Because the genesis of every work is different, even within the oeuvre of a single author, we try to determine what the specific needs of the texts in each subsequent module are, and to re-evaluate the tools and functionalities the module will provide to satisfy those needs accordingly. For the BDMP's second module, for example, this resulted in changes to the Synoptic Sentence View (which allows the user to grab any sentence in our corpus, and generate a chronological list of its different versions), to the image-text linking tool, etc. Furthermore, this poster will also demonstrate the BDMP's improved integration of CollateX[5] – the interoperable collation tool that is part of the Interedition project[6].

References

1. **Beckett, Samuel** (2011). *Stirrings Still / Soubresauts and Comment Dire / what is the word: an electronic genetic edition* (Series 'The Beckett Digital Manuscript Project' module 1), edited by Dirk Van Hulle and Vincent Neyt. Brussels: University Press Antwerp (ASP/UPA). www.beckettarchive.org (accessed 30 October 2013).
2. **Beckett, Samuel** (2011). *Stirrings Still / Soubresauts and Comment Dire / what is the word: an electronic genetic edition* (Series 'The Beckett Digital Manuscript Project' module 1), edited by Dirk Van Hulle and Vincent Neyt. Brussels: University Press Antwerp (ASP/UPA). www.beckettarchive.org (accessed 30 October 2013).
3. **de Biasi, Pierre-Marc** (2004) *Toward a Science of Literature: Manuscript Analysis and the Genesis of the Work.* Genetic Criticism. Texts and Avant-textes. Philadelphia: University of Pennsylvania Press: 36-68.
4. **Sahle, Patrick** (2013) *Digitale Editionsformen, Zum Umgang mit der Überlieferung unter den Bedingungen des Medienwandels*. Teil 2: Befunde, Theorie und Methodik. Nordstedt: Books on Demand.
5. **CollateX** (2013) *Home.* CollateX. www.collatex.net (accessed 30 October 2013).
6. **Interedition** (2013) *Home.* Interedition. www.interedition.eu (accessed 30 October 2013).

Spreading DiRT: extending the Digital Research Tools directory

Dombrowski, Quinn
quinnd@berkeley.edu
UC Berkeley

Gold, Matthew
mattgold@gmail.com
City University of New York

1. Background

The DiRT (Digital Research Tools, dirt.projectbamboo.org) directory is a longstanding resource for scholars interested in digital tools and methodologies, providing basic information about software that can facilitate different stages of the research process. DiRT was originally designed as a wiki, where a single wiki page contained information about all tools in a given category. In 2011, under the auspices of Project Bamboo, DiRT was completely rebuilt using the Drupal content management system, which allowed for data to be stored in a structured manner. This enabled more complex searching and browsing options (such as allowing the user to limit results based on criteria like platform or cost), and provided individual profile pages for each tool, which could then serve as a locus for specific comments, or be referenced in other tool profiles. For instance, if a profile page indicates that Neatline is a suite of add-on tools for Omeka, a link to Omeka appears on the Neatline tool profile page, and vice versa.