

White Paper for:

NEH HG-229342-15

A Digital Synopsis of Mishnah and Tosefta

Hayim Lapin

University of Maryland 9/25/19

(with Professor Tal Ilan, FSU, Berlin, as German PI)

Description of the Project

A Digital Synopsis of Mishnah and Tosefta is a joint project between Tal Ilan (FUB, Berlin, Germany) and Hayim Lapin (University of Maryland, College Park, MD, USA), funded under a joint NEH/DFG funding opportunity. We are grateful to both our home institutions and to the funding agencies that made our work possible. The fundamental planned activities of the project were to prepare (1) aligned transcriptions of several witnesses each of (a) the Mishnah and (b) the Tosefta, respectively as well as (2) an alignment of the Mishnah to the Tosefta based on (a) string-matching algorithms to identify shared text as well as (b) manually identified parallels based a review of scholarly editions and scholarly studies. The basic division of labor was that the actual transcription and other direct textual work would be carried out in Berlin; the web application would be developed at Maryland, and the string-matching work would be coordinated from Maryland as well.

Along with Ilan and Lapin, Professor Shamma Friedman (Bar Ilan University and Jewish Theological Seminary, Ramat Gan and Jerusaelm, Israel) was an important contributor in the early stages of the project. Daniel Stoekl Ben Ezra (EPHE, Paris, France) was instrumental in providing OCR's materials for our use (the Mishnah extracted from the Munich MS of the Babylonian Talmud and the Cambridge MS of the Mishnah), and, increasingly Stoekl became a core member of the leadership for the project. Although not formally part of the team, Avi Shmidman and Moshe Koppel (both of Bar Ilan University and of Dicta) developed the algorithm used for string-matching, modified parameters and re-ran the data, and, since they have familiarity with the texts themselves were able to contribute some first-hand knowledge as well. Finally, Raffaele Vigiante (University of Maryland) was the lead developer for the project. *Mishnah editions.*

Our Mishnah edition exists in two forms: A feature-rich form implemented in TEI-Publisher that includes only a select portion of the text (editions.erabbinica.org), and the original "Maryland Digital Mishnah" site (mishnah.mith.us), which, alongside several partial transcriptions has the full text of several witnesses to the Mishnah:

- Kaufmann MS (MTAK Kaufmann A 60)
- Parma Biblioteca Palatina 3173 (formerly De Rossi 138-139)
- Cambridge Add. 470.1
- BSB Cod. Hebr. 95 (Babylonian Talmud)
- Vilna/Vilnius 1913-1923 (the modern text reproduced in most current printed editions)
- Leiden University Libraries Or. 4720 (Palestinian Talmud)¹
- Venice 1523 Printed Edition of the Palestinian Talmud

Tosefta edition.

¹ Because the Palestinian Talmud does not cover the whole of the Talmud, this and the printed edition of the Palestinian Talmud only include about two thirds of the Mishnah.

Our Tosefta edition makes use of Österreichische Nationalbibliothek Cod hebr. 20 (Vienna), supplemented, where that manuscript has missing leaves, from British Library Add. 27296 (London). The text is currently queryable via the Mishnah-Tosefta alignment and through our experiments using the DTS standard (see below). When the whole application is ported to TEI publisher, we plan to implement direct edition of that work as well. *Mishnah-Tosefta Alignments*.

The full set of Mishnah-Tosefta alignments is available at mishnah.mith.us. The data is based on the string-matching algorithm. Users can choose a mishnah chapter as base or a Tosefta chapter as base and navigate through the work to find overlapping passages. Still to be added is navigation based on the Tosefta.

As promised we have also collected parallels reported in the scholarly literature. When the application is ported to TEI-Publisher, we will incorporate these into the application as well.

CTS/DTS Experiments.

We discuss the significance of DTS as a standard for projects above. We have published a few texts on the DTS standard on the newest Perseus platform. Access to the Mishnah, for instance, is available here:

- <https://scaife.perseus.org/library/urn:cts:ancJewLit:mishnah/>

For the Mishnah and Tosefta in particular, the results of a very preliminary “GetPassage” request can be viewed by directly within our publication. Some sample queries are:

- Mishnah Yevamot 4:1 word three to 4:2 word 5 from a reference text:
<http://mishnah.mith.us/exist/apps/digitalmishnah/modules/pre-cts.xql?GetPassage=ref.3.1.4.1.3-ref.3.1.4.2.5>
- Mishnah Yevamot 4:1 from specifically the Parma MS:
<http://mishnah.mith.us/exist/apps/digitalmishnah/modules/pre-cts.xql?GetPassage=S00483.3.1.4.1>
- Tosefta Yevamot 4:1 word three to 4:2 word 5:
<http://mishnah.mith.us/exist/apps/digitalmishnah/modules/pre-cts.xql?GetPassage=ref-t.3.1.4.1.3-ref-t.3.1.4.2.5>

Continuation of the Project

Integration of the Multiple Project Sites.

As we have noted several times in this report, the way, the promised work products from this project exist as separate and only partially integrated applications. Over the next months, we look forward to porting the application at mishnah.mith.us to TEI-Publisher and integrating all the ongoing projects into one application.

As we move forward we see two immediate additional areas for future work.

Linked Open Data.

The idea of a Distributed Text Service has been raised several times in this report as a way of sharing data within and between projects. Another way of thinking of this is bringing the texts that we study into the universe of Linked Open Data: our materials, transcriptions, annotations, commentaries are all describable in ways that other humans and computers can access them in predictable ways. Already, Lapin and Stoekl are collaborating on several large scale digitization projects in which an emerging DTS for rabbinic and classical Jewish materials forms part of the infrastructure of the project.

Also central to the Linked Open Data enterprise is the integration of named entities. Ilan is almost single-handedly the author of the authoritative *Lexicon of Jewish Names* in Antiquity and has discussed with Lapin creating a digital edition which would then allow us to link all the proper names in the Mishnah to this Dictionary. Our project has also contributed lists of place names to ongoing geographical and toponymic projects.³ As these projects mature our editions can link to these authoritative lists; at the same time, those gazetteers and onomastica will be able to link to our editions through DTS citations.

Text Reuse.

Directly inspired by the NEH/DFG funded project on Mishnah-Tosefta overlaps, Lapin proposed (and was awarded startup funding from the University of Maryland for) a small pilot using the Shmidman/Koppel algorithm applied to a broad subset of rabbinic literature (the literature known as Tannaitic, thought to have been composed between ca. 200 and 300 AD/CE, including the Mishnah and Tosefta) and the two Talmuds that cite this material extensively. We now have an extensive preliminary set of matches among all of these texts, numbering in the hundreds of thousands.

Perhaps the most promising model for this overlapping material is a graph database, in which any given word in any given text can be connected to any other based on a variety of criteria. In addition to fast data retrieval, the graph model has the potential to facilitate data-mining queries such as social connectivity (how are people associated with one another or with particular locations, phrases, or topics), textual “archaeology” (which passages, keywords, or topics cluster together across literary works), and more.