An Interactive Visualization of Recurring Microforms Across Various Hebrew Works and their Witnesses

Marcus Pöckelmann
marcus.poeckelmann@informatik.uni-halle.de
Martin Luther University Halle-Wittenberg
(ORCID iD: 0000-0001-8718-1198)

Alexandra Damm
alexandra.damm@student.uni-halle.de
Martin Luther University Halle-Wittenberg

Bill Rebiger
bill.rebiger@judaistik.uni-halle.de
Martin Luther University Halle-Wittenberg

Introduction

Keter Shem Ṭov (KST) – or the "Crown of the Good Name" – is a medieval kabbalistic treatise written in Hebrew probably in Spain in the middle of the 13th century. KST is preserved in about one hundred manuscripts stemming from various European regions. It can be seen as a set of microforms[1] or textual modules, which typically consist of some consecutive text segments. Within the manuscript evidence of KST, microforms are configured differently (Necker and Rebiger 2024) thus creating different versions of KST. These microforms can also be found as components of various other works (macroforms). In this manner, KST is closely linked to the so-called *Divre Menaḥem* (DM or "Words of Menaḥem"). Furthermore, a rather long textual module from KST consisting of a commentary on the biblical book of Job, chapter 28, is also known from other works from the 13th century like the "Commentary on Song of Songs" by Ezra of Gerona or as an abridged version in the "Commentary on Job" by Naḥmanides or partly in the "Commentary on the Torah" by Menaḥem Recanati. There are many other examples attesting the same microforms or textual modules known from KST in different other works (macroforms).

In order to find microforms (originating from KST) in texts other than KST and to investigate the relationship of these text components across different macroforms, a reasonable distant-reading visualization is needed. Developing and providing such a visualization option is part of the ongoing research project *Digital Multi-layer Synopsis and Mystical Traditions of the Hebrew Alphabet according to the Kabbalistic Treatise Keter Shem Ṭov[2]*.

Related Work

The general problem described in the introduction touches two fields of research in digital humanities. First, in digital scholarly editions it is common to collate different witnesses of the edited work with specialized software (Nury 2018). On the other hand, intertextual relationships among different works are studied under the label of text reuse (detection), especially in classicizes where the used sources in a work are seldom cited directly (Büchler et al. 2014, Pöckelmann et al. 2020). While both fields are related to each other, they differ as collation often aims for mostly exact matching of the witnesses (besides some grade of normalization), while text reuse detection aims to identify allusions and paraphrases. There is a wide range of distant-reading visualization techniques applicable for both (Yousef and Jänicke 2021). However, none of these tools is directly applicable to the problem considered here, as they are not designed for visualizing both, different witnesses of a work and recurring microforms among different works, at the same time.

Approach and Implementation

Our approach of visually representing microforms in macroforms is based on CATview (Pöckelmann et al. 2015)[3]. CATview is a JavaScript library, originally developed for the visualization of collation results in the context of scholarly editions, and integrated in extensive collation tools such as LAKomp (Leipold et al. 2015) and LERA (Pöckelmann et al. 2023). For each Hebrew work under scrutiny, we generate one CATview, which allows to show information for different witnesses of that text. The CATviews will then be linked to interactively show the recurring microforms across works. Additionally, a full text synopsis can be displayed to show details on a word level for a selected microform.

In our experiments we used four Hebrew works. The text witnesses of KST and DM have been manually transcribed earlier and are considered as a single work. The three commentaries by Ezra, Naḥmanides and Recanati have been

---

1   The terms micro- and macroforms have been established by Peter Schäfer (1981) in order to describe the phenomenon of fluctuating textual transmission and redaction in early Jewish mysticism.

2   The project is funded by Deutsche Forschungsgemeinschaft (DFG) – Grant number 414786977.

3   We used version 1.6.3, available under https://catview.uzi.uni-halle.de

transcribed for this experiment via OCR with the aid of eScriptorium (Kiessling et al. 2019) and a model trained for Hebrew (Stoekl et al. 2021). The results have been manually corrected.

All text witnesses have been segmented manually according to sense units, which typically range from some words up to some sentences. The segment alignment of the witnesses of *Keter Shem Ṭov* was automatically done with LERA based on their similarity via the Jaccard-coefficient (Jaccard 1901) and manually corrected afterwards. As a result, we have a CATview showing the variations among its witnesses as illustrated in Figure 1.
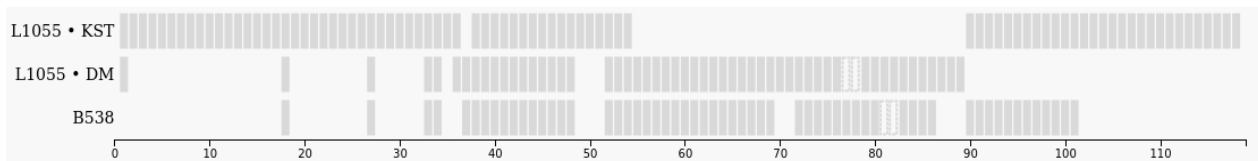

Fig. 1: CATview for three witnesses of *Keter Shem Ṭov* (*Divre Menaḥem* respectively). Segments are represented as rectangles that share the same column when aligned. Note, there is also a transposition of segments which is indicated by dotted rectangles.

The identification of recurring microforms among the four works on the segment level has been done manually. This will later be supported using appropriate semi-automatic algorithms. To visualize the relationships across the four CATviews, we had to extend the CATview software. First, a feature to highlight rectangles in red hues was added as well as callback functions that react on the mouse cursor entering or leaving rectangles and allow to run custom JavaScript code on those events. On the other hand, we implemented an additional JavaScript-component, which manages all necessary data for linking the CATviews, receives the callbacks from mouse hovering and triggers the red coloring of rectangles according to the data. Thereby, the red hues are used to indicate the similarity of segments (which will be calculated via the Jaccard-coefficient). If there are sequences of segments that together form larger microforms, all these segments are highlighted in red, with the currently considered segment and its directly connected counterparts marked by an additional circle. Please see Figure 2 for an illustration.
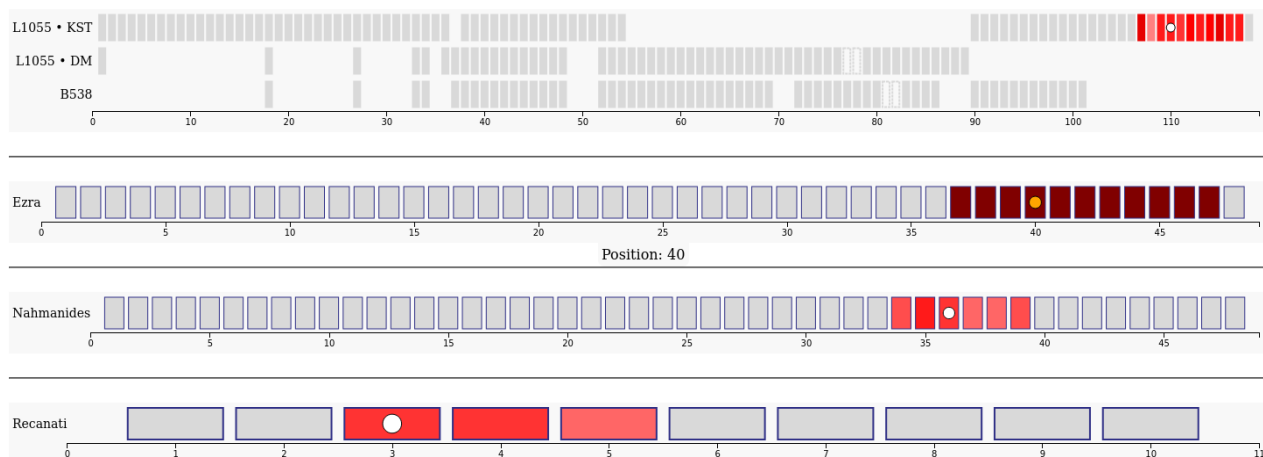

Fig. 2: Screenshot of four CATviews. While the first one for *Keter Shem Ṭov* visualizes three witnesses of the work, the others represent only one witness each. One of the recurring microforms found in all four works spanning multiple consecutive segments is highlighted in red hues, as the mouse was hovered over segment 40 in the commentary of Ezra (indicated by an orange circle). Thereby the darker the red hues, the more similar the segments are compared to the segments of the base text (Ezra).

This combination of multiple CATviews is useful for distant-reading in order to get an interactive overview of the relationships among the works and their witnesses on a segment level. It also allows to spot recurring microforms. This presentation is accompanied by a full-text synopsis of the segments currently in focus. It allows the corresponding texts to be studied in close-reading on a word level, see Figure 3. Again, LERA was used to generate the synoptic view.

Recanati	Nahmanides	Ezra	KST · L1055

ואמר יתפשטו ויתגלו שנאמר[1] בחלמיש שלח ידו המשיל[2] החכמה לחלמיש על הנתינות שהיו סתומים בה[4]

בחלמיש שלח ידו, המשיל[2] החכמה לחלמיש צור[3] על הנתיבות שהיו סתומים בה. הפך משרש הרים, וגלה הנסתר תחתם. ויקרא יאורים ההוויות הנכללות בשם מים במעשה בראשית, ואחר שיצאו היאורים[4],

בחלמיש שלח ידו, המשיל[2] החכמה לחלמיש על הנתיבות שהיה בם סתומים, והמשיל העניין ההוא לדבר הנסתר במקום קשה וחזק, והפך שרשה[4].

ואז[1] בחלמיש שלח ??? ??? [126א] החכמה לחלמיש על שם ההוויות שהיו בה סתומים והמשיל העמק לדבר נסתר במקום קשה וחזק והפך שרשם[4]

Fig. 3: Screenshot of a full-text synopsis for the segment hovered in Fig. 2.

Conclusion and Future Work

We have presented a visualization that simultaneously depicts both, intertextual relationships between several works in the form of recurring segments, as well as the alignment of several text witnesses of these works.

The next steps include the embedding of this visualization into a comprehensive platform for studying these relationships across different works and the processes of their creation alongside the usual collation of witnesses of a single work. To recognize more recurring microforms, we are working on a semi-automatic solution that provides possible candidates that then can be checked by a specialist scholar and added to the database. This goes hand in hand with the question of a meaningful semi-automatic segmentation of the texts. The aim is to gradually open up further works for the corpus and enable collaborative work. Building on this, a three-dimensional visualization is planned to better link the two existing visualizations in order to create a novel multi-layer synopsis of different versions of *Keter Shem Ṭov* and other treatises related to it and attesting the same microforms or textual modules.

References

Büchler, M. / Burns, P.R. / Müller, M. / Franzini, E. / Franzini, G. (2014): *Towards a Historical Text Re-use Detection*. In: Biemann, C. / Mehler, A. (eds) Text Mining. Theory and Applications of Natural Language Processing. Springer, Cham. https://doi.org/10.1007/978-3-319-12655-5_11

Jaccard, P. (1901): *Étude comparative de la distribution florale dans une portion des Alpes et des Jura*. In: Bulletin de la Société Vaudoise des Sciences Naturelles, 37, pp. 547-579. https://doi.org/10.5169/seals-266450

Nury, E. (2018): *Automated collation and digital editions: from theory to practice*. Dissertation. King's College London.

Kiessling, B. / Tissot, R. / Stokes P. / Stökl Ben Ezra / D. (2019): *eScriptorium: An Open Source Platform for Historical Document Analysis.* In: Proceedings of the International Conference on Document Analysis and Recognition Workshops (ICDARW), Sydney, NSW, Australia, 2019, pp. 19-19. https://doi.org/10.1109/ICDARW.2019.10032

Leipold, A. / Kösser, S. / Medek (*Gießler), A. / Solms, H.-J. (2015): *Zwischen Online-Korpus und Buch – Die Hybridedition der Wundarznei des Heinrich von Pfalzpaint.* In: Bein, T. (eds) Vom Nutzen der Editionen, Beihefte zu Editio, Band 39, pp. 167-184

Pöckelmann, M. / Medek, A. / Molitor, P. / Ritter, J. (2015): *CATview - Supporting The Investigation Of Text Genesis Of Large Manuscripts By An Overall Interactive Visualization Tool*. In: 26. international annual conference of Digital Humanities (DH2015), Sydney, 29.06.-03.07.2015.

Pöckelmann, M. / Dähne, J. / Ritter, J. / Molitor, P. (2020): *Fast paraphrase extraction in Ancient Greek literature*. In: it - Information Technology, Special Issue "Digital Tools form Intertextuality Studies", 62(2), pp. 75-89, de Gruyter. https://doi.org/10.1515/itit-2019-0042

Pöckelmann, M. / Medek, A. / Ritter, J. / Molitor, P. (2023): *LERA - An interactive platform for synoptical representations of multiple text witnesses*. In: Digital Scholarship in the Humanities, 38(1), pp. 330–346. https://doi.org/10.1093/llc/fqac021

Necker, G. / Rebiger. B. (2024): *The One and the Many: The Structure and Versions of Keter Shem Tov According to the New Synoptic Edition with a Special Focus on MS Jerusalem, NLI, 8° 541*. In: Rebiger, B. / Necker, G. (eds) Editing Kabbalistic Texts. Wiesbaden: Harrassowitz.

Schäfer, P. (1981): *Hekhalot-Studien*, Tübingen (pp. 15f. and 199-201).

Stoekl Ben Ezra, D. / Brown-DeVost, B. / Jablonski, P. / Lapin, H. / Kiessling, B. / Lolli, E. (2021): *BiblIA-a General Model for Medieval Hebrew Manuscripts and an Open Annotated Dataset*. In: The 6th International Workshop on Historical Document Imaging and Processing, pp. 61-66.

Yousef, T. / Janicke, S. (2021): *A Survey of Text Alignment Visualization.* In: IEEE Transactions on Visualization and Computer Graphics, vol. 27, no. 2, pp. 1149-1159. https://doi.org/10.1109/TVCG.2020.3028975.