# JEWISH STUDIES IN THE DIGITAL AGE

Edited by Gerben Zaagsma, Daniel Stökl Ben Ezra,
Miriam Rürup, Michelle Margolis, and Amalia S. Levi

## STUDIES IN DIGITAL HISTORY AND HERMENEUTICS

**Jewish Studies in the Digital Age**

# Studies in Digital History and Hermeneutics

## Volume 5

# Jewish Studies in the Digital Age

Edited by
Gerben Zaagsma, Daniel Stökl Ben Ezra,
Miriam Rürup, Michelle Margolis, and Amalia S. Levi

UNIVERSITÉ DU LUXEMBOURG

C²DH
LUXEMBOURG CENTRE FOR
CONTEMPORARY AND DIGITAL HISTORY

# Contents

Gerben Zaagsma, Daniel Stökl Ben Ezra, Miriam Rürup,
Michelle Margolis, and Amalia S. Levi

# Jewish Studies in the Digital Age: Introduction

In the past two decades we have witnessed the rapid increase of events, work-shops, sessions, and projects that, in one way or another, are positioned at the intersection of Jewish Studies and Digital Humanities (henceforth: DHJewish). This should hardly come as a surprise given the institutionalization of the digital humanities as a field and the ongoing proliferation of cultural heritage online. These developments have inevitably led to the need to confront the consequences of the so-called digital turn in all fields and disciplines of the humanities and social sciences. The international conference *#DHJewish – Jewish Studies in the Digital Age*, organized by the Centre for Contemporary and Digital History (C²DH) of the University of Luxembourg in January 2021, aimed to do so with regard to the field known as Jewish Studies.[1] As in so many comparable events, the confer-ence highlighted the common methodological and epistemological challenges of much work in the digital humanities while, at the same time, seeking to probe the particular. After all, every field and (sub-)discipline has its own characteris-tics, preferred methodologies, and specific research questions for which certain digital approaches might be particularly suitable. This volume gathers a selection of revised and extended papers from the conference that, taken together, reflect the current state of what we call *DHJewish*.

The question of how technology affects or might affect Jewish Studies in the present and future is not new.[2] In fact, attempts to chart how new technologies could alter research practices in Jewish Studies date back to at least 1969 when Henry Ekstein, during the Fifth World Congress of Jewish Studies in Jerusalem, presented a paper outlining a vision for a database that, in his view, would solve

---

**1** For more information and a link to the online conference archive with all session recordings, see: https://www.c2dh.uni.lu/events/dhjewish-jewish-studies-digital-age (all URLs in this intro-duction were last accessed on January 18, 2022, unless noted otherwise). All authors of this intro-duction were part of the conference program committee.

**2** This introduction is partly based upon: Gerben Zaagsma, "#DHJewish – Jewish Studies in the Digital Age," *Medaon* 12 (2018): 1–11; Gerben Zaagsma, Keynote lecture, "Exploring Jewish His-tory in the Digital Age," *International Conference What's New, What's Next? Innovative Methods, New Sources, and Paradigm Shifts in Jewish Studies*, POLIN Museum of the History of Polish Jews, Warsaw, October 3–6, 2021. See: https://doi.org/10.5281/zenodo.6631756.

many of the information storage and retrieval challenges facing the field of Jewish Studies.[3] Ekstein argued that, given the state of technology, it was possible to "establish a large scale data bank which would include information in the field of Jewish Studies, and to connect this bank by means of a telecommunication network to the most important research centers in this field all over the world."[4] Ekstein's vision is reminiscent of early information organization systems, as developed and realized by the Belgian lawyer and proto information scientist Paul Otlet and his companion Henri Lafontaine.[5] It also echoes developments in data communications networks that got underway in the 1960s and would eventually evolve into the Internet in the 1980s and World Wide Web in the 1990s. Ekstein, in sum, was effectively outlining what an online library for Jewish Studies could look like in an age when digital computing began to replace its analog siblings.

Ekstein's remarks raise a number of broader questions, however. To begin with, they hint at the *longue durée* of the encounter between technology and the humanities, in techno-material terms as well as in thinking through its broader consequences for humanities research practices. Secondly, they highlight challenges in the realm of information management that were by no means exclusive to the field of Jewish Studies alone and had been discussed by early information scientists for some decades already. And finally, Ekstein's remarks notwithstanding, the mechanical tools that he spoke of had in fact undergone significant change since the late 19th century and Jewish Studies scholars had keenly taken advantage of the possibilities they afforded ever since then. Indeed, while Ekstein might have been one of the first to reflect upon the overall state of Jewish Studies at the dawn of humanities computing, he clearly drew upon important developments preceding him. In fact, historians such as Samuel Oppenheim had been using the photostat since the 1920s to gather dispersed materials about their research topics. In the late 1930s and early 1940s geographers and demographers began to use punched card technology in the realm of Jewish Population Studies, following its entry in academic research in general, not only in the sciences, but also anthropology, literature, the social sciences, and (economic) history.[6] Not

---

**3** Henry C. Ekstein, "How to Increase Effectiveness of Research in Jewish Studies," *Proceedings of the World Congress of Jewish Studies* (1969): 3–7.
**4** Ekstein, "How to Increase Effectiveness of Research in Jewish Studies," 5.
**5** W. Boyd Rayward, ed., *International Organisation and Dissemination of Knowledge: Selected Essays of Paul Otlet* (Amsterdam: Elsevier, 1990).
**6** Sophia Robison and Joshua Starr, *Jewish Population Studies: Conference on Jewish Relations, New York, 1943* (New York: Conference on Jewish Relations, 1943). For a general overview of the emerging application of punched card technology in academia in the 1930s, see: G.W. Baehne, *Practical Applications of the Punched Card Method in Colleges and Universities* (New York: Columbia University Press, 1935).

much later, in the late 1940s, microfilming Jewish archives got underway, while work also began on automated machine translation, concordances, and word indexes.[7]

The latter's impact in Jewish Studies would soon be felt. On March 26, 1958, the Jesuit priest and scholar Roberto Busa and engineer Paul Tasman held a press conference at IBM's headquarters in New York to describe their literary data processing work on the Dead Sea Scrolls using punched card technology. Their presentation made headlines all over the world and Busa subsequently also presented the research in July 1961 at the third World Congress of Jewish Studies in Jerusalem.[8] This work was an offshoot of Busa's work on the Index Thomisticus, begun in 1949 with the help of IBM, which would eventually earn him the reputation of "founding father" of the Digital Humanities, though this foundational myth has meanwhile been thoroughly questioned and scholars have recently also emphasized the "female operatives" of his work and interrogated the "cultural, intellectual, and social conditions that shaped the earliest work in digital humanities."[9]

In 1959, inspired by Busa, the recently founded Israel Academy of the Hebrew language decided to create a database for its *Historical Dictionary*, today's *ma'agarim*, one of the first of its kind for a major historical language.[10] Around the same time, great strides were also made in the realm of Yiddish Studies and

**7** Dolores M. Burton, "Automated Concordances and Word Indexes: The Fifties," *Computers and the Humanities* 15, no. 1 (1981), https://doi.org/10.1007/BF02404370.

**8** Paul Tasman, *Indexing the Dead Sea Scrolls by Electronic Literary Data Processing Methods* (New York: IBM, 1958). Roberto Busa, "The Index of All Non Biblical Dead Sea Scrolls Published up to December 1957," *Revue de Qumrân* 2 (October 1958): 187–98. For the global resonance see, for example, this set of articles on the Dutch Delpher website: accessed May 2, 2018, http://bit.ly/2mu7htf. Busa himself reminisced about the occasion in: Roberto Busa, "The Annals of Humanities Computing: The Index Thomisticus," *Computers and the Humanities* 14, no. 2 (1980): 83–90, 85, https://doi.org/10.1007/bf02403798. For an elaborate analysis of this work see also: Steven E. Jones, *Roberto Busa, S.J., and the Emergence of Humanities Computing: The Priest and the Punched Cards* (Routledge: London, 2016), ch. 5.

**9** Steven Jones has complicated and contextualized this founding myth in his important work on Busa, see: Jones, *Roberto Busa*. See also, Melissa Terras and Julianne Nyhan, "Father Busa's Female Card Operatives," in *Debates in the Digital Humanities 2016*, https://dhdebates.gc.cuny.edu/read/untitled/section/1e57217b-f262-4f25-806b-4fcf1548beb5; Geoffrey Rockwell and Stefan Sinclair, "Tremendous Mechanical Labor: Father Busa's Algorithm," *DHQ* 14, no. 3 (2020), http://www.digitalhumanities.org/dhq/vol/14/3/000456/000456.html.

**10** See: https://maagarim.hebrew-academy.org.il/. On the history of the project see: Zeev Ben-Haim, "On the Making of the Historical Dictionary of the Hebrew Language of the Academy of the Hebrew Language," *Leshonenu* 23 (1959): 102–23 [Hebrew]; R. Merkin, Z. Busharia, and E. Meir, "The Historical Dictionary of the Hebrew Language," *Literary and Linguistic Computing* 4, no. 4 (January 1, 1989): 271–73, https://doi.org/10.1093/llc/4.4.271; Israel Yeivin, "Le dictionnaire historique de la langue hébraïque," *Meta* 43, no. 1 (1998): 19–26.

Yiddish linguistics to expand the use of punched cards to enable the processing of linguistic data. In 1961 the linguist Uriel Weinreich managed to acquire funding to use "machine aids" for the creation of the Language and Culture Atlas of Ashkenazic Jewry (LCAAJ). A proposal from around 1960 in the LCAAJ archives shows a request for 200,000 cards and 500 hours of machine use at the newly created IBM Watson Scientific Laboratory at Columbia University.[11] As Weinreich noted: "When one tries to visualize the editing of an atlas of many hundreds of maps, with up to 500 locations on each, it becomes clear what advantages are gained by the electronic filing and sorting of the data."[12] A few years later, in 1967, the Responsa Project launched, a computerized full-text retrieval system providing access to, as Yaacov Choueka described it, "Rabbinical case-law documents spanning more than ten centuries."[13]

# 1 From Humanities Computing to Digital Humanities

As will be clear from this very sketchy overview, when Ekstein told his audience at the World Congress of Jewish Studies in 1969 that "the mechanical tools with which the researcher works have remained almost unchanged in the last five hundred years," he missed some very important developments. At that point in time, the late 1960s, humanities computing had firmly taken hold in the academy with its own conferences and events being organized as well as its own publications. It would continue to develop as the era of mainframe computing began to give way to micro- and personal computers which were introduced at universities from the late 1970s onwards. From 1973 onwards, a long series of 90 colloquia on humanities computing at the University of Tübingen demonstrates well the growing use of computers, especially the TUebinger System von TExtverarbeitungs-Programmen (TUSTEP),[14] with a considerable impact on Jewish Studies in (printed) editions or

---

**11** "Description of Project," Box 236, Folder 13, Archives of the Language and Culture Atlas of Ashkenazic Jewry, Rare Book and Manuscript Library, Columbia University.

**12** Uriel Weinreich, "Machine Aids in the Compilation of Linguistic Atlases," *American Philosophical Society Yearbook* 1963 (1964): 622–25.

**13** Yaacov Choueka, "Computerized Full-Text Retrieval Systems and Research in the Humanities: The Responsa Project," *Computers and the Humanities* 14, no. 3 (November 1, 1980): 153–69, https://doi.org/10.1007/BF02403764.

**14** This powerful computer program for philological work, editions, synopses, concordances, and sophisticated page-layout is still used in many large projects more than 40 years after its first version.

synopses of the Mishnah, the Palestinian Talmud, and Hekhalot Literature and a concordance of the latter.[15] For the study of the biblical text, morphologically tagged versions of Hebrew, Greek, Syriac, Latin, and other texts were developed and aligned from the early 1980s and led to breakthroughs in the study of translation and a series of five conferences on computing and the Bible.[16] When the Internet and World Wide Web arrived on the scene in the early 1990s, as before, Jewish Studies scholars were quick to take advantage. Beginning in 1988, for instance, Martin Abegg famously used a personal computer to reconstruct the text of the unpublished Dead Sea Scrolls from the unpublished Preliminary Concordance, using a Macintosh SE that some nicknamed "Rabbi Computer."[17] His work led to their groundbreaking but controversial publication in 1991 and was quickly integrated into the first commercial program packages for Biblical Studies.[18]

During all this time, librarians and archivists eagerly discussed and explored the possible applications of new technologies too, including for instance on the pages of the American journal *Judaica Librarianship*.[19] They also actively sought to shape the future form of the Jewish documentary record. In October 1991, at a

---

**15** Michael Krupp, "Das Mischna-Editionsprojekt" (February 2, 1974), Michael Krupp, "Computer-unterstützte Zusammenstellung von textkritischen Apparaten. Erfahrungen bei der Vorbereitung der Mischna-Edition" (July 2, 1977), Gottfried Reeg, "Spaltensynopse und Zeilensynopse als Darstellungsformen für kritische Texteditionen" and "Konkordanz zur Hekalot-Literatur" (both June 30, 1984). Summaries accessible at http://www.tustep.uni-tuebingen.de/kolloq.html. The projects led to Michael Krupp et al., *Mischna*, 7 vols. (Jerusalem: Lee Achim, 2018); Peter Schäfer with Margarete Schlüter and Hans-Georg von Mutius, *Synopse zur Hekhalot-Literatur*. Texts and Studies in Ancient Judaism 2. (Tübingen: Mohr-Siebeck, 1981); Peter Schäfer with Gottfried Reeg et al., *Konkordanz zur Hekhalot-Literatur*, 2 vols. TSAJ 12–13 (Tübingen: Mohr-Siebeck, 1986 and 1988); Peter Schäfer, Hans-Jürgen Becker with Gottfried Reeg et al., *Synopse zum Talmud Yerushalmi*, 7 vols.) (TSAJ 31, 33, 35, 82, 83, 67, 47) (Tübingen: Mohr-Siebeck, 1991–2001).

**16** Robert Kraft and Emmanuel Tov, "Computer-Assisted Tools for Septuagint Studies," *Bulletin of the International Organization of Septuagint and Cognate Studies* 14 (1981): 22–40. Robert A. Kraft, Emanuel Tov, and John R. Abercrombie, *Computer Assisted Tools for Septuagint Studies* (Atlanta, GA: Scholars Press, 1986). The first conference was published as *Proceedings of the First International Colloquium Bible and Computer, The Text, Louvain-la-Neuve (Belgique) 2–3–4 September 1985* (Paris: Champion; Geneva: Sladkine, 1986).

**17** Jones, *Roberto Busa*, 144–47, 145.

**18** In particular, Accordance Software, which started in 1988 and developed many supplementary modules and complex search syntactical feature queries has a great impact in the study of ancient Judaism. See also: Johanna Sprondel, "Toward a Humanities of the Digital? Reading Search Engines as a Concordance," in The Making of the Humanities, vol. 3, The Modern Humanities, ed. Rens Bod, Jaap Maat, and Thijs Weststeijn (Amsterdam: Amsterdam University Press, 2014), 479–93, 483.

**19** *Judaica Librarianship*, last accessed May 2, 2018, https://ajlpublishing.org/. See, e.g., Diane Romm, "Retrieval of Judaica from Electronic Media: An Overview," *Judaica Librarianship* 8, no.

time when the PC had become a common tool for scholars and the Internet had arrived at universities, the Leo Baeck Institute organized a conference on "Problems and Issues in Jewish Archives and Historiography in the Five New States of Germany." The meeting resulted in a plan to create a database of Jewish archival holdings in the states of the former German Democratic Republic (GDR) with the aim of enhancing access to these dispersed collections. In an echo of Ekstein's earlier vision, Robert Jacobs noted: "How very fortunate we are to live in a time when electronic capabilities enable us to provide bibliographic references that will allow subsequent generations of scholars to find the materials they seek with no more than a few keystrokes."[20] The plan was an excellent example of the kind of "database" that Ekstein had in mind. Importantly, the idea was for an online catalog, not a digitization project.

Soon, however, the publication of primary sources in the form of CD-ROMs and online databases and textual editions began, and it would lead to a veritable sea change during the 1990s, enabling far easier access to source materials and fundamentally changing scholarly information management practices. The impact of these developments on the field of Jewish Studies was a major focus of Heidi Lerner's *Perspectives on Technology* column in the Association of Jewish Studies' (AJS) *Perspectives* magazine, published on a regular basis between 2003 and 2011.[21] Lerner's 2002 article "New Technologies and Old Methodologies: Jewish Studies Research in the Digital Age" was probably one of the first to comprehensively address the possibilities that the digital turn offered for Jewish Studies scholars.[22] And then as now, as Lerner's column highlights, librarians and archivists were at the forefront of digital developments and pointed the way for humanities scholars.

The early 2000s was also the period of the transition from humanities computing to what we now call "digital humanities," characterized by mass digitization, "big data," and the proliferation of new tools and new forms of knowledge dissemination. Digital humanities work in Jewish Studies in the past two decades has been made possible by a plethora of projects that bring together hitherto dispersed, inaccessible, or fragile materials into digital archives and collections that

---

1–2 (1993): 61–63 and Bella Hass Weinberg, "*Judaica Librarianship* in the Age of the Internet," *Judaica Librarianship* 9, no. 1–2 (1995): 3–5.

**20**  Robert Jacobs, "Jewish Archival Holdings in the Five New States of Germany: Creating an Inventory," *Judaica Librarianship* 8, no. 1 (1994): 17–22, 22, https://doi.org/10.14263/2330-2976.1222.

**21**  AJS, Perspectives on Technology, last accessed May 2, 2018, https://www.associationforjewishstudies.org/what-is-jewish-studies/digital-jewish-studies/perspectives-on-technology.

**22**  Heidi Lerner, "New Technologies and Old Methodologies: Jewish Studies Research in the Digital Age," *Shofar: An Interdisciplinary Journal of Jewish Studies* 20, no. 4 (2002): 81–95, https://doi.org/10.1353/sho.2002.0073.

facilitate digital scholarship. If dispersion is the common denominator among many primary source collections for Jewish Studies, then the digital offers the possibility to virtually bring together dispersed materials.

The issue of dispersion also relates to one of the key characteristics of the Jewish historical experience, migration and its transnational aspects. It is hardly surprising then that Jewish migration, especially East European, ranks prominently among digital resources. Among several well-known YIVO projects, such as the Encyclopedia of Jews in Eastern Europe, the exceptional Vilna Collections Project should be mentioned, which will offer unprecedented access to the largest archive on East European Jewish History in the world.[23] These projects exist side by side with a huge variety of national and local history projects, which often also exhibit transnational dimensions, such as Memoria Viva and the Jewish Diaspora Collection (part of the University of Florida's Digital Collections) which focus on Latin America, including the Caribbean.[24] National, regional, and local history projects comprise too many to mention. Some examples with clear transnational dimensions include the South African Jewish Museum's digital archive and the South Africa Jewish Rootsbank, Yiddish Melbourne, and DigiBaeck (the Digital Collections of the Leo Baeck Institute).[25] More nationally focused projects include Key Documents of German-Jewish History and the Digital Library of the Italian Foundation Center for Contemporary Jewish Documentation.[26] Other digitized collections, such as the Sephardic Studies Digital Collection at the University of Washington aim to shed light on lesser known histories and languages of Sephardic Jews and counter the loss of history and culture of dispersed and gradually dwindling communities.[27]

Next to migration, the Holocaust is probably the single most important topic, covered by a wide range of resources that include the European Holocaust Research Infrastructure (EHRI), Arolsen Archives, the USC Shoah Foundation Institute's Visual History Archive Online (VHA Online), the German Memorial Book project, the Czech Holocaust Victims and Document Database, and the New York Public Library's Yizkor Book Collection.[28] As so many Holocaust-related sources and tes-

---

**23** See: https://vilnacollections.yivo.org/.

**24** See: https://yivoencyclopedia.org/default.aspx; https://www.yivo.org/vilna-collections-project; https://mviva.org/; https://ufdc.ufl.edu/judaica.

**25** See: https://sajmarchives.com/; http://www.jewishroots.uct.ac.za/; https://www.monash.edu/arts/acjc/yiddish-melbourne; https://www.lbi.org/collections/digibaeck/.

**26** See: https://jewish-history-online.net/; http://digital-library.cdec.it/cdec-web/.

**27** https://content.lib.washington.edu/sephardicweb/index.html.

**28** See: https://ehri-project.eu/; https://arolsen-archives.org/; https://vhaonline.usc.edu/; https://www.bundesarchiv.de/gedenkbuch/; https://www.holocaust.cz/databaze-obeti/; https://digital-collections.nypl.org/collections/yizkor-book-collection#/.

timonies have become available online, scholarly attention has shifted in recent years towards studying the nature of Holocaust memory in the digital age.[29]

In terms of types of sources being digitized, textual sources such as newspapers, manuscripts, and books, dominate. The National Library of Israel's Historical Jewish Press website contains, at the time of writing, 625 newspapers in 20 languages from all over the world, though with clear foci on the regions of East Europe, the Middle East, and North America, Jewish languages in Yiddish and Hebrew, and non-Jewish languages, such as Arabic, Polish, French, and English.[30] The website provides a robust OCR and search function in four alphabets to allow greater use of and access to the materials, which enables the kind of comparative research that would require extensive research trips only a decade ago. Other newspaper databases complement these holdings, a notable resource to mention is Compact Memory which contains 424 periodicals in nine languages, the bulk being in German.[31] Compact Memory is part of the digitized Judaica collections of Frankfurt University Library which offers thousands of digitized books.[32] In this respect one should also mention important resources such as the Yiddish Book Center's Digital Yiddish Library, and the seminal International Collection of Digitized Hebrew Manuscripts project (KTIV).[33]

The example of KTIV also highlights an important point. While many libraries are digitizing their collections, work in digital humanities goes well beyond posting digital facsimiles online. It is important to distinguish between scholarly work that simply uses "digital" tools (which, it could be argued, applies to nearly everything, since something as basic as word processing uses the digital to express words on paper) and research that uses computational tools and digital methods to analyze and interpret digitized materials.

This distinction is abundantly clear in the case of Hebrew manuscript studies. Over the past three decades, Hebrew manuscripts became the object of several studies in the field of computerized document analysis questions such as script classification, writer identification, layout segmentation, handwritten text recogni-

---

**29** See especially: Jeffrey Shandler, *Holocaust Memory in the Digital Age: Survivors' Stories and New Media Practices* (Stanford, CA: Stanford University Press, 2017) and his chapter Digitizing Holocaust Memories in this volume. For Holocaust research in the digital age see the various articles in special issue no. 13 of *Quest. Issues in Contemporary Jewish History* in 2018. For the introduction: Laura Brazzo and Reto Speck, "Holocaust Research and Archives in the Digital Age: Introduction," *Quest. Issues in Contemporary Jewish History* 13 (2018): V–XIII.
**30** See: https://www.nli.org.il/en/discover/newspapers/jpress.
**31** See: https://sammlungen.ub.uni-frankfurt.de/cm/nav/index/title.
**32** See: https://sammlungen.ub.uni-frankfurt.de/judaica/nav/index/all.
**33** See: https://www.yiddishbookcenter.org/collections/digital-yiddish-library; https://web.nli.org.il/sites/nlis/en/manuscript.

tion, text to image alignment, and crowdsourcing.[34] Larger, accessible text corpora, stronger and more high-performance computers, and the (re)discovery of neural networks also caused a huge increase in the application of Natural Language Processing (NLP) on Semitic and Jewish Languages. Only in the last five years has performance reached production level for morphological, lexical and syntactical tagging, named entity recognition, stylometrics, author attribution, text reuse, topic modeling, and sentiment analysis even for historical texts.[35] Similarly, in the field of computer vision applied to manuscript studies, after the pioneering studies of the early years of the new millennium as, for example, exemplified in the join-discovery tool of the Friedberg Genizah project (FGP),[36] the last five years have led to breakthroughs that allow mass applications of previously unthinkable scale.[37] The FGP, created in 2005 by and around Yaacov Choueka, assembles images, transcriptions, and metadata on more than 250,000 fragments with more than 600,000 images dispersed in more than 70 repositories worldwide.[38] It was also one of the first projects worldwide in any discipline to apply cutting edge computer vision, natural language processing, and data-mining on a very large scale to historical documents.[39]

Alongside the study of Hebrew manuscripts, digital humanities methods have been applied to the study of the Jewish book and Jewish languages. One such approach is offered by Marienberg-Milikowsky focusing on the study of reli-

**34** One of the first computer vision projects applied to Hebrew manuscripts was Laurence Likforman-Sulem, Henri Maître, and Colette Sirat, "An Expert Vision System for Analysis of Hebrew Characters and Authentication of Manuscripts," *Pattern Recognition* 24, no. 2 (1991): 121–37. See also Itay Bar Yosef, Klara Kedem, Its'hak Dinstein, Malachi Beit-Arie, and Edna Engel, "Classification of Hebrew Calligraphic Handwriting Styles: Preliminary Results," *DIAL* (2004): 299–305.
**35** E.g. Avi Shmidman, Shaltiel Shmidman, Moshe Koppel, and Yoav Goldberg, "Nakdan: Professional Hebrew Diacritizer," *ACL* (demo) (2020): 197–203. Compare Shuly Wintner, "Hebrew Computational Linguistics: Past and Future," *Artificial Intelligence Review* 21, no. 2 (2004): 113–38 with Amit Seker, Elron Bandel, Dan Bareket, Idan Brusilovsky, Refael Shaked Greenfeld, and Reut Tsarfaty, "AlephBERT: A Hebrew Large Pre-trained Language Model to Start-Off Your Hebrew NLP Application With," CoRR abs/2104.04052 (2021), https://arxiv.org/abs/2104.04052. See the contributions by Shmidman and Waxman in this volume.
**36** See Roni Shweka, Yaacov Choueka, Lior Wolf, and Nachum Dershowitz, "Veqarev otam ehad el ehad: Zihuy ktav yad vetseruf qit'ei hagnizah beemtsa'ut mahshev," *Ginzei Qedem* 7 (2011): 173–209.
**37** See, e.g., Daniel Stökl Ben Ezra, Bronson Brown-DeVost, Pawel Jablonski, Hayim Lapin, Benjamin Kiessling, and Elena Lolli, "BiblIA: A General Model for Medieval Hebrew Manuscripts and an Open Annotated Dataset," *HIP@ICDAR* 2021: 61–66.
**38** Yaacov Choueka, "Computerizing the Cairo Genizah: Aims, Methodologies and Achievements," *Ginzei Qedem* 8 (2012): 9–30.
**39** Shweka et al., "Veqarev otam ehad el ehad."

gious Hebrew and Aramaic manuscripts and languages.[40] The increase in such initiatives in recent years is evidenced by projects such as *Tikkoun Sofrim*, which uses AI to transcribe Hebrew manuscripts, *Footprints*, which traces the movement of Jewish books through time and place, *HaMapah*, which traces rabbinic networks based on printed responsa, and *Geniza Scribes*, which invites "the crowd" to transcribe manuscripts in Hebrew, Arabic, and other languages, from the Cairo Geniza.[41]

Digital humanities methods are also used to facilitate historical research beyond dispersed primary sources and against national borders. Some examples are projects such as *Mapping Modern Jewish Cultures*,[42] which explores space, time, and multilingual Jewish communities through the lens of cafés in urban environments; the Wikidata-based *EHRI Ghettos*[43] authority list of Holocaust-era ghettos; *Mapping Jewish LA*,[44] which uses digital tools to highlight the diversity of Jewish experiences in one single place; and the *Documenting Judeo-Spanish*[45] project, which aims to document and provide access to Sephardic texts written with the Solitreo script.

Surveying all of these projects, a broader picture emerges. As Gerben Zaagsma argued a few years ago, Jewish history presents specific challenges in the realm of information management due to its textual tradition, its diasporic nature, the – forced – migration of people, texts, ideas, and thus its transnational aspects. These characteristics, which can equally be applied to the field of Jewish Studies in general, are, in turn, reflected in both the state of Jewish heritage (dispersal of sources and objects) and its nature (multilingual, multiscriptual, and often textual). As a result, a key technological challenge for Jewish history in the digital age, and Jewish Studies more generally, is to work towards solutions for information retrieval and analysis from dispersed, multilingual, and multiscriptual sources.[46] Much work in this direction has been done in the past three decades and it has allowed us to engage the transnational, interactional, inter-, intra- and cross-cultural dimensions of Jewish history in new ways.

---

**40** Itay Marienberg-Milikowsky, "Digital Research of Jewish Texts: Challenges and Opportunities," in *Textual Transmission in Contemporary Jewish Cultures*, ed. Avriel Bar-Levav and Uzi Rebhun (New York: Oxford University Press, 2020), 15–25, https://doi.org/10.1093/oso/9780197516485.001.0001.
**41** For digital editions of rabbinical texts see e.g. Chaim Milikowsky, "Scholarly Editions of Three Rabbinic Texts – One Critical and Two Digital," in *Advances in Digital Scholarly Editing*, ed. Peter Boot et al. (Leiden: Sidestone Press, 2017), 137–46.
**42** https://richbrew.org/.
**43** https://portal.ehri-project.eu/vocabularies/ehri_ghettos.
**44** http://www.mappingjewishla.org/.
**45** https://documentingjudeospanish.com/.
**46** Zaagsma, "Jewish Studies in the Digital Age."

Fifteen years ago Paula Hyman argued for the importance of comparative approaches to integrate what she called minority history into the history of majority populations.[47] Around the same time Moshe Rosman sought to probe the challenges that postmodernism posed for engaging with Jewish history, or Jewish histories, including questions of Jewish identity, periodization, and intercultural relations.[48] More recently, Zwiep and Wallet have suggested that "big data" might be one answer to some of these questions as its longitudinal character can help us explore the *longue durée* of Jewish history.[49] If indeed some of the key tenets of Jewish historiography nowadays lie in studying inter- and intra-Jewish as well as Jewish/non-Jewish interactions, and in probing their fluid, constantly changing, and evolving nature in space as well as in and over time, then digitization and digital history are well placed to address the challenges involved.

Online resources make comparisons within and between Jewish populations as well as between Jewish and non-Jewish populations easier than ever; they can help answer the crucial question of what was specific for *which* Jewish experiences, and for *whose* Jewish experiences. Computational techniques allow us to trace long-term trends in, for instance, newspapers and thus shifting discourses and concerns; network analysis can help chart the global migration of books and intellectual ideas and explore Jewish interconnectedness across borders; migration and migrant experiences can be traced, explored, and compared through newspapers and other sources. Both the multiplicity and diversity of Jewish histories and experiences and the commonalities that unite and unify them are thus open for new and renewed explorations.[50]

As exciting as many of these digital projects are, though, a note of caution is in order. The continuity of digital humanities work is intimately bound up with sustainable digital preservation, and obsolescence is a serious concern. Just as the original data from the LCAAJ project could not be accessed digitally without re-digitizing the texts and OCRing them to allow searchability, many digital projects are no longer accessible because of obsolete hardware or software. An innovative virtual exhibition allowing a walkthrough of the Braginsky collection, virtually turning pages of manuscripts in exhibition cases, is no longer accessible because the Flash product that it used is no longer supported. Link rot abounds,

**47** Paula Hyman, "Recent Trends in European Jewish Historiography," *Journal of Modern History* 77 (2005): 345–56.

**48** Moshe Rosman, *How Jewish Is Jewish History?* (Oxford: Littman Library of Jewish Civilization, 2007).

**49** Bart Wallet and Irene Zwiep, "Session 0.8.I/II: Humanities in the Mirror: Writing Jewish History in a Digital Key," EAJS Quadrennial Congress, Kraków, July 15–19, 2018.

**50** See also: Zaagsma, "Exploring Jewish History in the Digital Age."

and many "cutting edge projects" of a decade ago are no longer accessible. There is still much work to be done on this front.

## 2 Taking Stock

To take stock of the myriad developments that were outlined above, a range of events have been organized over the past decade. The Center for Jewish History (CJH) ran a workshop in 2011 entitled *From Access to Integration: Digital Technologies and the Study of Jewish History*, which sought to "explore in a systematic way new approaches to coordinating and integrating the digitization of Jewish historical sources around the world."[51] The workshop also aimed to connect Jewish Studies information specialists as a means of addressing the "challenges faced by many institutions in employing emerging technologies for the study of Jewish history." Perhaps ironically, and as a strong case in point that illustrates the abovementioned point about obsolescence, the conference website and other associated materials are no longer online, and the only live content remaining are the tweets using the hashtag #cjh-a2i.[52]

In his keynote lecture during the workshop, entitled "Digitization and Its Discontents for Jewish History," historian Anthony Grafton outlined the various ways in which the digital turn is affecting academia and academic libraries, including the possibilities of digitally reuniting dispersed material. He also noted that much of Jewish scholarship happens outside academic circles, meaning that open access to online Jewish resources had become highly important.[53] Unfortunately, the various blog posts devoted to the conference do not reveal what, if any, answers were formulated as to the new approaches and challenges mentioned above, or indeed provide much detail as to what these were in the first place.[54] A promised "white paper" following the conference unfortunately did not materialize. On Twitter, though, we find some traces of the debate as it took place. Librar-

---

**51** From Arthur Kiron's introduction to: Anthony Grafton, "Digitization and Its Discontents for Jewish History," a talk delivered at the International Conference *From Access to Integration: Digital Technologies and the Study of Jewish History Center for Jewish History*, New York, 2012, 2, last accessed June 12, 2014, http://www.cjh.org/CJHGraftonDigitization/ (no longer available online).
**52** Workshop, *From Access to Integration: Digital Technologies and the Study of Jewish History*. An archived version of the conference website's About page can be accessed here: accessed November 30, 2021, https://web.archive.org/web/20120126155457/http://techconference.cjh.org/about.php.
**53** Grafton, "Digitization and Its Discontents," 17–19.
**54** Last accessed March 20, 2018, https://16thstreet.tumblr.com/search/access+to+integration.

ian Deanna Marcum, for example, noted the international scope, long history, and multilingualism as distinct features of Jewish Studies, while also stressing that all fields share certain fundamental needs for infrastructure, governance, funding, selection, etc.[55]

The CJH workshop was the start of a decade full of DHJewish events. In 2012 Brown University organized a workshop "Ancient Religions, Modern Technology" devoted to the ways in which the digital humanities has or can change the study of religion in antiquity.[56] In 2013, the Institute for the History of the German Jews (Institut für die Geschichte der deutschen Juden) in Hamburg organized the workshop *Jüdische Geschichte Digital* (Digital Jewish History).[57] Taking stock of a wide variety of digital projects pertaining to German-Jewish history, the event led to the creation of the network Jüdische Geschichte Digital within the digital history working group of the Historikerverband, the German Historical Association.[58]

Meanwhile, the annual AJS conference and the European Association of Jewish Studies (EAJS) conference, held every four years, began to include panels and workshops on Jewish Studies and Digital Humanities. Both the 2012 and 2013 AJS conferences featured a THATCamp Jewish Studies and in 2014 a THATCamp was organized in Haifa, Israel.[59] The Association of Jewish Libraries (AJL) held a digital humanities Roundtable at its 2014 meeting, and the keynote that year, by Emile Schrijver, was titled "Between Being-Wise and Not-Knowing-What-To-Ask: Jewish Librarianship and Digital Humanities." AJL's journal, *Judaica Librarianship*, began publishing a regular review column of DHJewish projects in 2017.[60] In 2015, a conference entitled *On the Same Page: Digital Approaches to Hebrew*

---

**55** Last accessed March 20, 2018, https://twitter.com/search?l=&q=%23cjh-a2i&src=typd&lang=eng.

**56** See the call for papers: http://aramaicnt.org/2011/04/06/call-for-papers-ancient-religion-modern-technology-workshop/.

**57** *Jüdische Geschichte Digital* workshop, last accessed March 20, 2018, https://www.hsozkult.de/event/id/termine-22109, organized by Anna Menny and Miriam Ruerup. For a conference report see: Gerben Zaagsma, Tagungsbericht *Jüdische Geschichte Digital*. 13.06.2013–14.06.2013, Hamburg, in *H-Soz-u-Kult*, September 10, 2013, available online at: last accessed May 2, 2018, www.hsozkult.de/conferencereport/id/tagungsberichte-5011.

**58** Netzwerk Jüdische Geschichte Digital, last accessed April 5, 2018, http://www.historikerverband.de/arbeitsgruppen/ag-digitale-gw/netzwerk-juedische-geschichte-digital.html.

**59** With regard to the 2012 AJS THATCamp see: Jeffrey Shandler, "From the President," *AJS Perspectives*, Fall (2012): 3–4. For the individual camps: http://jewishstudies2012.thatcamp.org/; http://jewishstudies2013.thatcamp.org/; http://haifa2014.thatcamp.org/.

**60** See the introduction to the new column: Michelle Margolis, "JS/DH: An Introduction to Jewish Studies/ Digital Humanities Resources," *Judaica Librarianship* 20, no. 1 (2017), https://doi.org/10.14263/2330-2976.1293.

*Manuscripts* took place at King's College London.[61] A follow-up EAJS round table, *Turning the Page: Jewish Print Cultures & Digital Humanities*, at the University of Amsterdam in February 2017, dealt with "early modern print cultures and the specific questions associated with them, e.g. regarding Jewish multilingualism, geographical space, the linking of various disparate library and archive collections, and methods, scales and techniques of textual analysis."[62] *From "Tablet" to "Tablet": A Digital Humanities Workshop* was held at the Institute for the History of the German Jews in Hamburg in September 2017.[63]

Around the time the EAJS established a *Digital Forum* to engage with digital scholarship more comprehensively,[64] the 2018 EAJS conference in Kraków also featured two "digital" sessions; one on *Humanities in the Mirror: Writing Jewish History in a Digital Key*, which, by focusing on big data, aimed to "address the question whether DH corpora and methods will enable us to find a new common ground in the field of Jewish history" and reconsider its *longue durée*; the second on *New Philologies: Hebrew Manuscript and Print Cultures in a Digital Key* focused on editions and the application of machine learning.[65] In January 2021, the online journal *Reviews in Digital Humanities* published a special issue on Jewish Digital Humanities and since 2022 the *European Journal of Jewish Studies* has included a section dedicated to DH and Jewish Studies.[66]

In short, in recent years there have been several efforts to understand and discuss how the digital turn has affected Jewish Studies and what its intersection with digital humanities looks like. The present volume gathers selected papers from the international online conference *#DHJewish – Jewish Studies in the Digital Age*, which took place in January 2021 and was organized by the Luxembourg

---

**61** Conference: On the Same Page: Digital Approaches to Hebrew Manuscripts, last accessed March 20, 2018, https://www.kcl.ac.uk/artshums/depts/trs/research/seminars/jewish/hebrew 2015.aspx.
**62** EAJS Roundtable Report, "Turning the Page: Jewish Print Cultures & Digital Humanities," University of Amsterdam, February 2017, available online at: last accessed September 7, 2017, https://www.eurojewishstudies.org/colloquia/eajs-programme-in-jewish-studies/eajs-roundtable-report-turning-the-page/.
**63** This workshop was initiated by Gerben Zaagsma and funded by the Rothschild Foundation and organized in cooperation with Miriam Ruerup at the Institute for the History of the German Jews.
**64** See: https://www.eurojewishstudies.org/digital-forum/eajs-digital-forum/.
**65** See the abstracts for both sessions here: https://www.eurojewishstudies.org/digital-forum/eajs-conference-2018/.
**66** See: https://reviewsindh.pubpub.org/v2-n1; https://brill.com/view/journals/ejjs/ejjs-overview.xml.

Center for Contemporary and Digital History (C²DH).[67] The aim of the Luxembourg conference was to take stock of how the digitization boom of the last two decades, and the rapid advancement of digital tools to analyze data in myriad ways, have opened up new avenues for Jewish Studies research. It sought to answer the questions of how digital developments can be harnessed to address specific questions and problems in the field, and what the current state of the art looks like. Supported by an international program committee, it brought together more than 60 scholars and heritage practitioners to discuss how the digital turn affects the field of Jewish Studies.[68] Importantly, the conference was not a stand-alone event but part of a bigger project that includes the online portal *#DHJewish – Jewish Studies and Digital Humanities*, launched in June 2022. #DHJewish offers a database of projects, a news and events section, blog posts, and links to various relevant bibliographies as well as an online Zulip community.[69]

## 3  Sections of the Volume

The present volume contains papers based on presentations given during the 2021 #DHJewish conference, and is divided in four sections: Collections, Spatiality, Text, and Computational. This subdivision should be thought of in terms of where the main emphasis of the individual chapters lies, yet there are intersections in many ways. Thus, the first section, *Collections*, features chapters that take digital resources, and reflect upon their use in Jewish Studies, as their starting point. The chapters in the second section, *Spatiality*, all revolve around the affordances of spatial humanities approaches to understanding Jewish history and primary sources. The section *Text* features chapters that discuss various methods to engage with research focusing on and working with textual sources while the final section, *Computational*, includes chapters that all employ analytical methods derived from computational linguistics. Importantly, the various sections provide a mix of project-oriented case studies and research-oriented digital

---

**67**  See: https://www.c2dh.uni.lu/events/dhjewish-jewish-studies-digital-age. The full archive of the conference with session recordings can be found here: https://www.morressier.com/o/event/5fd2237e54bbb7f516f76f1b.

**68**  The program committee consisted of: Michelle Margolis, Rachel Deblinger, Karin Hofmeester, Gabor Kadar, Amalia Levi, Anna Menny, Miriam Ruerup, Sinai Rusinek, Avi Shmidman, Daniel Stoekl Ben Ezra, Dov Winer, Gerben Zaagsma (Chair), and Irene Zwiep. See also: https://www.c2dh.uni.lu/events/dhjewish-jewish-studies-digital-age.

**69**  See: https://dhjewish.org/.

scholarship. As such, the present volume also seeks to provide a glimpse of the various directions that digital humanities work in Jewish Studies can take.

## 3.1 Collections

The first section, *Collections*, features four chapters and starts with the conference opening keynote lecture by **Jeffrey Shandler**, who takes the reader from the first recorded interviews with Holocaust survivors in Europe, conducted by US-based psychologist David Boder in the immediate aftermath of World War II, to today's new approaches in using immersive storytelling through new visualization techniques. Shandler shows us that the use of all material, whether "born analogue" or "digital" depends on the interaction of the user with the interviews. While each technology enables us to experience a different type of encounter with eyewitnesses, these are always embedded in our own knowledge and research questions, which frame our engagements.

Shandler's chapter is followed by that of **Inna Kizhner**, **Melissa Terras**, and their co-authors, who discuss the question of how Jewish culture is represented in museum collections. They provide us with a comparative survey of metadata on "Jewish" collections in the Metropolitan Museum of Art in the United States and the State Catalogue of Museum Collections of the Russian Federation that provide an entrance point into the complexities of international metadata standards and address the importance of universal standards to provide equal access to collections.

Following this analysis of how digital collections are constituted, **Jakub Mlynář**, **Jiří Kocián**, and **Karin Hofmeisterová** address the question of how users engage with them, specifically how search engines and search-query techniques shape our results. They present us with a small-scale study on how users at the Malach Centre for Visual History (CVHM) at the Charles University in Prague search in corpora of audiovisual Holocaust testimonies such as, among others, the Visual History Archive (VHA) and Fortunoff Video Archive for Holocaust Testimonies. Through their case study they show how search engines and search-query techniques shape our findings and, ultimately, "tools" thus produce "data."

Finally, **Anna Bonazzi**'s essay delves into the challenges involved in indexing the content of digital resources. She discusses work to semi-automate content indexing and analysis of Holocaust testimonies based on N-grams. Traditional, manually assigned keyword-based indexing of testimonies emphasizes verbal content terms such as noun-based facts, names, and historical references that we expect survivors to use. Bonazzi argues for an alternative indexing system that is

not limited to the identification of keywords to summarize content and presents a semiautomated DH approach based on N-grams. This approach allows us to identify patterns and narrative structures that go unseen in traditional keyword-based indexing of the survivors' testimonies, including structural, non-verbal categories like uncertainty, reticence, and emotional insistence on time references.

## 3.2 Spatiality

In the second section, *Spatiality*, four chapters illustrate the breadth of spatial approaches that are current in Jewish Studies. The essay by **Sinja Clavadetscher**, **Stefanie Mahrer** and **Stefanie Salvisberg** discusses the forced migration of Jewish academics from Nazi Germany within the context of a broader research project about Switzerland and academic forced migrants from 1933 to 1950. The research uses nodegoat to map the expulsion of academics and resulting transnational academic network in order to shed light on inner-institutional changes in respect to both academic staff and the academic standing of the research institutions involved.

**Maja Hultman** uses a GIS approach to debunk established historiographical narratives of a spatially inscribed, dichotomized Jewish urban experience in early 20th-century Stockholm. The project interrogates the supposed division between integrated, Reform, northern-residing Jews and Eastern European, poor, orthodox, southern-residing Jews. In this project, GIS is used as a tool to facilitate quantitative analysis of primary sources over the city's topography, in tandem with qualitative sources. Hultman explores the role of this unique topography in shaping complex Jewish approaches to social integration, religious practices, and communal relations.

**Piergabriele Mancuso**'s essay highlights the centrality of archival collections and research for developing virtual reality and 3D modeling reconstructions. Using the Ghetto Mapping Project as a case study, Mancuso outlines the various and heterogeneous types of material used to assess not only the architectural features of the built environment, but also demographic and socioeconomic trends of the long presence of the Jewish community in the Florence ghetto, eventually seeking to understand the politics of segregation.

Finally, **Daniel Stein Kokin** presents a project that can simultaneously serve as a tool for research and pedagogy and as a political intervention that raises awareness of the long history of what we often take as a given – the "points" on our maps. Based upon maps and lists of settlements in the territory of Palestine and what today is Israel from the 1840s to the present, Stein Kokin shows how our contemporary topography is ultimately a temporary point – *nekudah* – on

the map which may have replaced another point before and might be followed by yet other points in the future. Ultimately, he problematizes the chronological matter-of-factness of what we see on our maps.

## 3.3 Text

The third section, *Text*, introduces a variety of methodological perspectives that can be brought to bear upon Jewish textual sources. **Benjamin Lee** reports on the use of machine learning techniques for extracting and analyzing visual content of early 20th-century Ladino newspapers. Lee scales up scholarship using the Newspaper Navigator tool to extract photographs, illustrations, maps, comics, and editorial cartoons alongside advertisements, and facilitate transnational analysis of the Sephardic Jewish experience. Beyond reporting on the actual work of creating and analyzing the dataset, Lee offers insights about interdisciplinary collaborations for digital humanities work in Jewish Studies, as well as reflections on ethical considerations when applying machine learning techniques to Jewish cultural heritage collections.

**Abby Gondek**'s essay explores the question of how a network analysis and visualization tool such as nodegoat can help us overcome biased perceptions of how political agendas were shaped. Her case study of Henrietta Stein Klotz, assistant to Henry Morgenthau Jr., Secretary of the US Treasury between 1934 and 1945, shows how Klotz influenced and shaped Morgenthau's positions in response to the Holocaust. Her sophisticated multi-layered network analysis allows her to trace the gendered dimensions of interpersonal networks and the political influence of actors that so often remain invisible.

**Zef Segal** analyzes the contents of the periodical *HaTzfira* over a six-year period, before and after its transition from a weekly to a daily newspaper and shows what content changes in a newspaper can reveal about readers' interests and concerns. Segal uses topic modeling to show how local events and generational change led to a shift from more scientific content to a widening focus on world events, politics, and Jewish-related news.

**Tatsiana Astrouskaya**, finally, uses the well-known programming language R to study and analyze the correspondence of Russian refuseniks in their struggle to leave the Soviet Union. Focusing on the story of Ernst Markovich Levin, Astrouskaya analyzes the quantitative and the qualitative dimensions of his activities petitioning various authorities. Her argument-driven digital historical analysis shows how human and machine readings can complement each other also in relatively "small data" studies.

## 3.4  Computational

The final and most specialized section of the volume, *Computational*, features four chapters that showcase some of the latest developments in computational techniques as applied to Jewish manuscripts. Turning to a print from the 18th century for a highly structured corpus of texts from early manuscripts, **Luigi Bambaci** reminds us that data for 21st-century computational analysis can be found centuries earlier. Bambaci's chapter describes his work digitizing, parsing, and encoding the Kennicott Bible (1774–1776) to better analyze and understand textual elements in manuscripts of the Hebrew Bible up to constructing the stemmatic tree. One of the most difficult aspects of studying post-canonical Jewish texts is identifying biblical citations that are not exact. Historical writers of religious texts often cited the Bible but might play around with the citation to fit a poem or the context at hand.

Avi Shmidman presents an algorithm and tool to automatically detect biblical paraphrastic quotations in a given text, a non-trivial task when they are short and not identical to the Hebrew text. This groundbreaking work has already been shown to be a tremendous tool in textual research in Jewish Studies.

**Daria Vasyutinsky**, **Jihad El-Sana** and colleagues introduce their work in applying deep learning models to classify script types and sub-types in medieval Hebrew manuscripts. Incorporating, *inter alia*, the techniques and databases of Hebrew paleography their research project is part of a broader ongoing effort to develop algorithmic tools for processing historical documents.

Finally, **Joshua Waxman**'s essay addresses the need for modern punctuated versions of classical Jewish texts which were often composed without punctuation. He describes an automatized system that was developed to create a punctuated version of the Hebrew and Aramaic of the Babylonian Talmud. This has the potential to assist greatly in Talmud use and study.

As we hope this introduction has made clear, the field of Jewish Studies has always been shaped by the uptake of new technologies. It is therefore important to acknowledge the groundbreaking and pioneering work that has led to our present moment and can be traced back to at least the 1950s. As Jewish Studies have now firmly moved into the digital age, we hope that the essays in this volume will enhance critical reflection on the methodological and epistemological consequences for our field and encourage the further uptake of digital approaches to bring about its full potential.

# References

Baehne, G.W. *Practical Applications of the Punched Card Method in Colleges and Universities*. New York: Columbia University Press, 1935.

Ben-Haim, Zeev. "On the Making of the Historical Dictionary of the Hebrew Language of the Academy of the Hebrew Language." *Leshonenu* 23 (1959): 102–23.

Brazzo, Laura and Reto Speck. "Holocaust Research and Archives in the Digital Age: Introduction." *Quest. Issues in Contemporary Jewish History* 13 (2018): V–XIII.

Burton, Dolores M. "Automated Concordances and Word Indexes: The Fifties." *Computers and the Humanities* 15, no. 1 (1981). https://doi.org/10.1007/BF02404370.

Busa, Roberto. "The Annals of Humanities Computing: The Index Thomisticus." *Computers and the Humanities* 14, no. 2 (1980): 83–90. https://doi.org/10.1007/bf02403798.

Busa, Roberto. "The Index of All Non Biblical Dead Sea Scrolls Published up to December 1957." *Revue de Qumrân* 2 (October 1958): 187–98.

Chesner, Michelle. "JS/DH: An Introduction to Jewish Studies/ Digital Humanities Resources." *Judaica Librarianship* 20, no. 1 (2017). https://doi.org/10.14263/2330-2976.1293.

Choueka, Yaacov. "Computerized Full-Text Retrieval Systems and Research in the Humanities: The Responsa Project." *Computers and the Humanities* 14, no. 3 (November 1, 1980): 153–69. https://doi.org/10.1007/BF02403764.

Choueka, Yaacov. "Computerizing the Cairo Genizah: Aims, Methodologies and Achievements." *Ginzei Qedem* 8 (2012): 9–30.

Conference: On the Same Page: Digital Approaches to Hebrew Manuscripts. Last accessed March 20, 2018. https://www.kcl.ac.uk/artshums/depts/trs/research/seminars/jewish/hebrew2015.aspx.

EAJS Roundtable Report. "Turning the Page: Jewish Print Cultures & Digital Humanities." University of Amsterdam, February 2017. Last accessed September 7, 2017. https://www.eurojewishstudies.org/colloquia/eajs-programme-in-jewish-studies/eajs-roundtable-report-turning-the-page/.

Ekstein, Henry C. "How to Increase Effectiveness of Research in Jewish Studies." *Proceedings of the World Congress of Jewish Studies* (1969): 3–7.

Grafton, Anthony. "Digitization and Its Discontents for Jewish History." A talk delivered at the International Conference *From Access to Integration: Digital Technologies and the Study of Jewish History Center for Jewish History*, New York, 2012.

Hyman, Paula. "Recent Trends in European Jewish Historiography." *Journal of Modern History* 77 (2005): 345–56.

Jacobs, Robert. "Jewish Archival Holdings in the Five New States of Germany: Creating an Inventory." *Judaica Librarianship* 8, no. 1 (1994): 17–22. https://doi.org/10.14263/2330-2976.1222.

Jones, Steven E. *Roberto Busa, S.J., and the Emergence of Humanities Computing: The Priest and the Punched Cards*. Routledge: London, 2016.

Kraft, Robert and Emmanuel Tov. "Computer-Assisted Tools for Septuagint Studies." *Bulletin of the International Organization of Septuagint and Cognate Studies* 14 (1981): 22–40.

Kraft, Robert A., Emanuel Tov, and John R. Abercrombie. *Computer Assisted Tools for Septuagint Studies*. Atlanta, GA: Scholars Press, 1986.

Krupp, Michael et al., *Mischna*, 7 vols. Jerusalem: Lee Achim, 2018.

Lerner, Heidi. "New Technologies and Old Methodologies: Jewish Studies Research in the Digital Age." *Shofar: An Interdisciplinary Journal of Jewish Studies* 20, no. 4 (2002): 81–95. https://doi.org/10.1353/sho.2002.0073.

Likforman-Sulem, Laurence, Henri Maître, and Colette Sirat. "An Expert Vision System for Analysis of Hebrew Characters and Authentication of Manuscripts." *Pattern Recognition* 24, no. 2 (1991): 121–37.

Marienberg-Milikowsky, Itay. "Digital Research of Jewish Texts: Challenges and Opportunities." In *Textual Transmission in Contemporary Jewish Cultures*, edited by Avriel Bar-Levav and Uzi Rebhun, 15–25. New York: Oxford University Press, 2020. https://doi.org/10.1093/oso/9780197516485.001.0001.

Merkin, R., Z. Busharia, and E. Meir. "The Historical Dictionary of the Hebrew Language." *Literary and Linguistic Computing* 4, no. 4 (January 1, 1989): 271–73. https://doi.org/10.1093/llc/4.4.271.

Milikowsky, Chaim. "Scholarly Editions of Three Rabbinic Texts – One Critical and Two Digital." In *Advances in Digital Scholarly Editing*, edited by Peter Boot et al., 137–46. Leiden: Sidestone Press, 2017.

*Proceedings of the First International Colloquium Bible and Computer, The Text, Louvain-la-Neuve (Belgique) 2–3–4 September 1985*. Paris: Champion; Geneva: Sladkine, 1986.

Rayward, W. Boyd, ed. *International Organisation and Dissemination of Knowledge: Selected Essays of Paul Otlet*. Amsterdam: Elsevier, 1990.

Robison, Sophia and Joshua Starr. *Jewish Population Studies: Conference on Jewish Relations, New York, 1943*. New York: Conference on Jewish Relations, 1943.

Rockwell, Geoffrey and Stefan Sinclair. "Tremendous Mechanical Labor: Father Busa's Algorithm." *DHQ* 14, no. 3 (2020). http://www.digitalhumanities.org/dhq/vol/14/3/000456/000456.html.

Romm, Diane. "Retrieval of Judaica from Electronic Media: An Overview." *Judaica Librarianship* 8, no. 1–2 (1993): 61–63.

Rosman, Moshe. *How Jewish Is Jewish History?* Oxford: Littman Library of Jewish Civilization, 2007.

Schäfer, Peter, Hans-Jürgen Becker with Gottfried Reeg et al. *Synopse zum Talmud Yerushalmi*. 7 vols. TSAJ 31, 33, 35, 82, 83, 67, 47. Tübingen: Mohr-Siebeck, 1991–2001.

Schäfer, Peter with Gottfried Reeg et al. *Konkordanz zur Hekhalot-Literatur*. 2 vols. TSAJ 12–13. Tübingen: Mohr-Siebeck, 1986 and 1988.

Schäfer, Peter with Margarete Schlüter and Hans-Georg von Mutius. *Synopse zur Hekhalot-Literatur*. Texts and Studies in Ancient Judaism 2. Tübingen: Mohr-Siebeck, 1981.

Shandler, Jeffrey. "From the President." *AJS Perspectives*, Fall (2012): 3–4.

Shandler, Jeffrey. *Holocaust Memory in the Digital Age: Survivors' Stories and New Media Practices*. Stanford, CA: Stanford University Press, 2017.

Shmidman, Avi, Shaltiel Shmidman, Moshe Koppel, and Yoav Goldberg. "Nakdan: Professional Hebrew Diacritizer." *ACL* (demo) (2020): 197–203.

Shweka, Roni, Yaacov Choueka, Lior Wolf, and Nachum Dershowitz. "Veqarev otam ehad el ehad: Zihuy ktav yad vetseruf qit'ei hagnizah beemtsa'ut mahshev." *Ginzei Qedem* 7 (2011): 173–209.

Sprondel, Johanna. "Toward a Humanities of the Digital? Reading Search Engines as a Concordance." In *The Making of the Humanities*, vol. 3, *The Modern Humanities*, edited by Rens Bod, Jaap Maat, and Thijs Weststeijn, 479–93. Amsterdam: Amsterdam University Press, 2014.

Stökl Ben Ezra, Daniel, Bronson Brown-DeVost, Pawel Jablonski, Hayim Lapin, Benjamin Kiessling, and Elena Lolli. "BiblIA: A General Model for Medieval Hebrew Manuscripts and an Open Annotated Dataset." HIP@ICDAR 2021: 61–66.

Tasman, Paul. *Indexing the Dead Sea Scrolls by Electronic Literary Data Processing Methods*. New York: IBM, 1958.

Terras, Melissa and Julianne Nyhan. "Father Busa's Female Card Operatives." In *Debates in the Digital Humanities 2016*. https://dhdebates.gc.cuny.edu/read/untitled/section/1e57217b-f262-4f25-806b-4fcf1548beb5.

Wallet, Bart and Irene Zwiep. "Session 0.8.I/II: Humanities in the Mirror: Writing Jewish History in a Digital Key." EAJS Quadrennial Congress, Kraków, July 15–19, 2018.

Weinberg, Bella Hass. "*Judaica Librarianship* in the Age of the Internet." *Judaica Librarianship* 9, no. 1–2 (1995): 3–5.

Weinreich, Uriel. "Machine Aids in the Compilation of Linguistic Atlases." *American Philosophical Society Yearbook 1963* (1964): 622–25.

Wintner, Shuly. "Hebrew Computational Linguistics: Past and Future." *Artificial Intelligence Review* 21, no. 2 (2004): 113–38.

Yeivin, Israel. "Le dictionnaire historique de la langue hébraïque." *Meta* 43, no. 1 (1998): 19–26.

Yosef, Itay Bar, Klara Kedem, Its'hak Dinstein, Malachi Beit-Arie, and Edna Engel. "Classification of Hebrew Calligraphic Handwriting Styles: Preliminary Results." *DIAL* (2004): 299–305.

Zaagsma, Gerben. "#DHJewish – Jewish Studies in the Digital Age." *Medaon* 12 (2018): 1–11.

Zaagsma, Gerben. Keynote lecture, "Exploring Jewish History in the Digital Age." *International Conference What's New, What's Next? Innovative Methods, New Sources, and Paradigm Shifts in Jewish Studies*. POLIN Museum of the History of Polish Jews, Warsaw. October 3–6, 2021. https://polin.pl/en/keynotes-whats-new-whats-next.

Zaagsma, Gerben. Tagungsbericht *Jüdische Geschichte Digital*. 13.06.2013–14.06.2013, Hamburg. In *H-Soz-u-Kult*, September 10, 2013. Last accessed May 2, 2018. www.hsozkult.de/conferencereport/id/tagungsberichte-5011.

# Collections

Jeffrey Shandler
# Digitizing Holocaust Memories

**Abstract:** Digital media have enabled expansive opportunities to engage the Holocaust and its remembrance, especially by facilitating new forms of research. In particular, digitized audio or video recollections of survivors and other eyewitnesses of the Holocaust attract a wide range of users. These resources, whether "born analog" materials that have been digitized or more recent materials that were created in digital mediums, pose particular challenges with regard to their preservation, archiving, cataloging, indexing, and accessing, even as they offer distinctive opportunities for engaging with Holocaust remembrance. Digitization not only facilitates accessing these materials but also enables their remediation in new formats, shaping how these materials can be accessed through different means of distribution. These factors, in turn, inform how these resources are studied. The scope of materials includes documentations of pre-Holocaust Jewish culture, including collections of prewar photographs and Yiddish language and folklore, many of which are in digital form. But the largest corpus of resources are found in institutional archives of Holocaust survivors' memories, housing tens of thousands of life histories recorded in audio or video formats. These collections straddle a temporal boundary marked by both the impending loss of living witnesses to the Holocaust and the transition from the "video age" to the "digital age," making them especially resonant for studying Holocaust memory practices. At the same time, concerns for providing future generations with access to Holocaust survivors' life stories has engendered the use of new digital technologies – interactive holography and immersive virtual reality – to present these narratives.

**Keywords:** archiving, digital media, Holocaust memory, Holocaust survivors, indexing, life history, remediation, video

Digital media have enabled expansive opportunities to engage the Holocaust and its remembrance. These include new forms of research, such as using non-invasive archeological survey techniques to create three-dimensional visualizations of killing sites.[1] Digital media also facilitate new means of imagining Holocaust scenarios, whether through gaming or through role-playing videos posted

---

**1** See, e.g., "Recording Cultural Genocide and Killing Sites in Jewish Cemeteries," accessed January 22, 2022, https://holocaustkillingsites.com/.

on TikTok.[2] Among the most frequently encountered examples are digitized recollections of survivors and other eyewitnesses of the Holocaust. The range of users include high school students, genealogists, descendants of survivors, performing and visual artists, and museum visitors. These resources are also used by an array of scholars in the humanities and social sciences. In addition to those working in Jewish Studies are scholars of genocide, trauma, and memory practices, among other topics.[3]

Given this extensive and varied engagement, examining the nature of these resources' mediation offers insight into the impact of the digital turn in scholarship and contemporary culture generally. Though digitized memories of the Holocaust include many written texts, this essay focuses on audio and visual examples. These resources pose particular challenges with regard to their preservation, cataloging, and accessing, even as they offer distinctive opportunities for engaging with Holocaust remembrance.

Most of this material was not "born digital" but was converted from various analog mediums to a digital format. Therefore, digitizing these resources constitutes a remediation, which shapes how they are accessed through different means of distribution – whether compact disks, downloadable files, postings on websites or social media, or online streaming video. And this distribution, in turn, shapes how these resources are studied. The following discussion considers "born analog" examples first and then turns to more recent materials that were created in digital mediums.

Digitization has made some of the earliest remembrances of Holocaust survivors newly available, sometimes after decades of absence. Among these are the pioneering audio interviews conducted by American psychologist David Boder in 1946. To document survivors' wartime experiences "in their own voice," Boder traveled to Europe to conduct over 100 interviews with survivors in nine different languages, using a magnetic wire recorder (then state-of-the-art equipment).[4] In 1949, Boder published a book on these recordings, *I Did Not Interview the Dead*, and continued to work on them, but his interviews fell into obscurity

---

**2** On video games, see, e.g., Claudio Fogu, "Digitalizing Historical Consciousness," *History and Theory* 48, no. 2 (2009): 103–21. On TikTok, see, e.g., Nicole Froio, "We Asked TikTokers Why They're Pretending to Be Holocaust Victims," *Wired*, August 21, 2021, accessed August 24, 2020, https://www.wired.co.uk/article/tiktok-holocaust-pov.

**3** For a bibliography of this scholarship, see Victoria Grace Walden, "Reading about Digital Holocaust Memory," accessed January 22, 2021, https://digitalholocaustmemory.wordpress.com/2020/07/10/reading-about-digital-holocaust-memory/.

**4** See Alan Rosen, *The Wonder of Their Voices: The 1946 Holocaust Interviews of David Boder* (New York: Oxford University Press, 2010).

following Boder's death in 1961.[5] Then, due to growing interest in Holocaust survivor memories in recent decades, the interviews were digitized and posted on a website by the Illinois Institute of Technology, Boder's institutional home and now the keeper of some of his original recordings.[6] The website sorts the interviews by various criteria, including the language and location of the interview and the gender, religion, and wartime experiences of the respective interviewee. In addition, the site facilitates the use of these recordings, which are not of the clearest sound quality, with transcriptions, translations, and annotations, plus background information on Boder and his project. Thus, the website appears to fulfill Boder's original aspirations to make these materials available for further study, something he could not realize in his lifetime. Yet the recordings' digitization transforms how they are encountered, and the website's additional content reflects the distinct interests of a later generation of scholars.

At the same time that Boder undertook these interviews, folklorists started collecting various materials from Holocaust survivors. These early efforts include Yiddish folksongs documented by Ruth Rubin from informants in Canada and the United States, beginning in 1946, and by Ben Stonehill, who made over 1,000 wire recordings of songs performed by Holocaust survivors in New York in 1948.[7] Some songs pertain specifically to the Holocaust, such as those composed and sung in wartime ghettos. But these collections documented the wider repertoires of Jewish folksingers, as part of a larger range of efforts to preserve something of the cultures decimated and dispersed by the genocide. Both Rubin's and Stonehill's collections are now online. Their websites catalog the songs according to a thematic taxonomy and provide translations of lyrics and information on the singers and sources when known. As with Boder's interviews, these online resources are remediations, shaped by the input of scholars and technicians, to present resources to a new generation of users, for whom the material has new value.

Efforts to document prewar Jewish culture include several projects focused on Yiddish language and culture, some of which have been digitized. The earliest of these is the *Language and Culture Atlas of Ashkenazic Jewry*, initiated by linguist Uriel Weinreich at Columbia University in the 1950s. Audiotaped interviews with over 600 native speakers of Yiddish from across northern Europe have subsequently been digitized and posted online, as have scans of this project's extensive

---

**5** David P. Boder, *I Did Not Interview the Dead* (Urbana: University of Illinois Press, 1949).
**6** "Voices of the Holocaust," accessed January 22, 2022, http://voices.iit.edu/.
**7** "The Ruth Rubin Legacy Archive of Yiddish Folksongs," accessed December 24, 2018, https:// exhibitions.yivo.org/exhibits/show/ruth-rubin-sound-archive/home; "The Stonehill Jewish Song Collection," accessed June 3, 2018, http://www.ctmd.org/stonehill.htm.

paper archive.[8] More recent projects documenting native Yiddish speakers born in prewar Europe include the Archives of Historical and Ethnographic Yiddish Memories, known by the acronym AHEYM, which means "homeward" in Yiddish. This project was launched in 2002 by linguist Dov-Ber Kerler and historian Jeffrey Veidlinger at Indiana University. AHEYM includes video interviews with over 300 individuals identified as "Eastern Europe's last native speakers of Yiddish," all born before World War II. The project's interview topics are wide-ranging, including "linguistic ... data, oral histories of Jewish life in Eastern Europe, Holocaust testimonials, musical performances ..., folk narratives, ... reflections on contemporary Jewish life in the region, and guided tours of [local] sites of Jewish memory."[9] Weinreich's *Atlas*, which made early use of computers to correlate data, focused on gathering information for research on the historical foundation of modern Yiddish dialects. AHEYM, by contrast, is more wide-ranging in content, while it centers on the recollections of Holocaust survivors who returned to their prewar homes in provincial East European towns as embodiments of a vestigial "shtetl" culture.

This impulse to document what is often called the "lost world" of East European Jewry as an act of Holocaust remembrance includes collecting prewar photographs. Though made before the war, these images acquire new significance as records of lives destroyed by genocide. Some photography collections that have been digitized and placed online focus on a particular locale – for example, Mike Marvin's assembly of photographs from Szczuczyn, Poland, where his grandfather, Zalman Kaplan, operated the town's only photography studio from the 1890s though the 1930s.[10] Other collections are more wide-ranging, having been gathered and digitized by research organizations. Among these is "People of a Thousand Towns," an indexed catalog of the extensive holdings of photographs of Jewish life in prewar Eastern Europe presented online by the YIVO Institute for Jewish Research. This project began in 1981, when the institute placed 15,000 digitized images on a videodisc – then a brand-new technology – linked to a computer catalog, which, at the time, could only be used on site at YIVO's New

---

**8** "Language and Culture Atlas of Ashkenazic Jewry," accessed June 3, 2018, http://library.columbia.edu/locations/global/jewishstudies/lcaaj.html. For audio recordings from the Atlas, see also "EYDES: Evidence of Yiddish Documented in European Studies," accessed June 3, 2018, http://www.eydes.de/.

**9** "AHEYM: The Archives of Historical and Ethnographic Yiddish Memories," accessed June 3, 2018, http://www.iub.edu/~aheym/.

**10** "The Zalman Kaplan Collection of Pre-war Photos of Szczyczyn," accessed January 22, 2021, http://www.szczuczyn.com/kaplan.htm. See also Louis D. Levine, ed., *Lives Remembered: A Shtetl through a Photographer's Eye* (New York: Museum of Jewish Heritage, 2002).

York headquarters.[11] In 2002 the photographs and catalog were placed online and made publicly accessible.[12] This move is typical of the evolving of digitized materials; as technology advances, the possibilities for accessing resources expands.

An essential component of digitizing these photographs for researchers to access is the creation of databases of the images' metadata. This information, extrinsic to the photographs themselves, provides the point of entry to their use, shaping how they can be searched. Databases of photographs typically include generic descriptions of their content, as well as specific information, when known: their locations and dates, the names of their subjects and photographers. Less often recorded, but no less valuable, is information about photographs' provenance, given their sometimes charged trajectory during and after the war. Including this information distinguishes the collecting project known as "*I ciągle widzę ich twarze* / And I Still See Their Faces." It was initiated in the 1990s by Fundacja Shalom, an organization dedicated to promoting Jewish culture in Poland. The project gathered photographs of Jews in the possession of Polish citizens, yielding hundreds of images. Many came from people who did not know the identity of these photographs' subjects – only that they were Jews. Therefore, "And I Still See Their Faces" derives much of its value not from the information that the images provide about Polish Jewish life, but from the disrupted history of the physical photograph. Sometimes this history is evinced by the photograph's damaged condition, thereby materializing the haunting power of the loss of people and their culture.[13]

The photographs and recollections that Fundacja Shalom gathered for "And I Still See Their Faces" have been presented in a book, exhibition, and online.[14] More recently, the project expanded and transformed its digital presence on a website called Żydzi Polscy (Polish Jews). In addition to providing materials the Foundation had collected, the website invites visitors to post their own memories of Jewish relatives as well as letters, diaries, or photographs. In this way, the Foundation asserts, the website – and the story of Polish Jewry – will not "have a last page." Żydzi Polscy explains that it "is open for everyone, history lovers and

---

**11** Jeffrey Shandler, "'People of a Thousand Towns': YIVO's Videodisc Project," *Jewish Folklore and Ethnology Review* 10, no. 1 (1988): 18.

**12** "People of a Thousand Towns," YIVO Institute, accessed January 22, 2021, http://yivo1000towns.cjh.org/search_tips.asp.

**13** *And I Still See Their Faces: Images of Polish Jews / I ciągle widzę ich twarze: Fotografia Żydów polskich* (Warsaw: Fundacja Shalom, 1996).

**14** "And I Still See Their Faces," Fundacja Shalom, accessed January 22, 2021, http://shalom.org.pl/en/i-ciagle-widze-ich-twarze-2/.

witnesses, scientists and research centers, who would like to join the ... effort to recreate the world of Polish Jews." This "continuously updated material," renders Żydzi Polscy "a living panorama of the Jewish community in Poland."[15] This internet platform, interactive and renewable, fosters a virtual community of people interested in the topic of Polish Jewry and models cultural liveliness for the present activity of memory work, as this undertaking strives to animate a "lost world."

Institutional collections of Holocaust survivors' memories house tens of thousands of their personal histories, many now available in digital form. These efforts began in the war's immediate aftermath and continue to the present, undertaken in a variety of mediums. For example, Yad Vashem started archiving survivors' written accounts in the late 1940s, shortly after the State of Israel's national center of Holocaust commemoration and documentation had been established. Yad Vashem began recording survivors on audiotape in 1954 and videotape in 1989.[16] Overall, dozens of projects to make recordings of Holocaust survivors have been undertaken in multiple countries in the Americas and Europe, as well as in Israel and Australia. The United States Holocaust Memorial Museum currently houses over 70,000 audio and video interviews with Holocaust survivors and eyewitnesses, whether online or in its Washington DC headquarters. These recordings have been gathered from almost 100 different sources, ranging from large-scale collecting projects to local community efforts and individual undertakings.[17]

Digitization not only facilitates accessing these interviews; it also enables their remediation in new formats. For example, the University of Southern California Shoah Foundation recorded over 51,000 video interviews, mostly in the mid- to late 1990s, for its Visual History Archive (hereafter, VHA).[18] The largest such collection by far, its holdings have been remediated in a variety of formats. Especially significant is an online platform the foundation launched in 2009 called IWitness, which provides a "guided exploration" of over 1,500 videos in the VHA, primarily for use in American secondary schools (most interviews are in English). IWitness explains that it combines learning "first hand from survivors

---

**15** "Żydzi Polscy," accessed January 22, 2021, http://www.zydzipolscy.pl/.

**16** See "About the Yad Vashem Archives," accessed November 20, 2012, www1.yadvashem.org/yv/en/about/archive/about_archive_whats_in_ archive.asp; and Yad Vashem Collection of Testimonies, "Listing of the Record Groups in the Yad Vashem Archives," 12, accessed November 20, 2012, www1.yadvashem.org/yv/en/about/archive/pdf/list_of_record_groups.pdf.

**17** United States Holocaust Memorial Museum, Collections search: audio visual oral histories, accessed January 21, 2021, https://collections.ushmm.org/search/?f%5Bf_audiovisual%5D%5B%5D=testimony&q=oral+histories.

**18** "USC Shoah Foundation, Visual History Archive," accessed January 21, 2021, https://vhaonline.usc.edu/login.

and witnesses of the Holocaust" with "participatory" use of the videos through a "built-in online video editor," which enables students to "build [their] own video projects." Like Żydzi Polscy, IWitness uses the internet to enable interactive engagement with digitized resources – not by contributing to its holdings but by inviting their sampling. The Shoah Foundation explains that IWitness offers students the opportunity to learn about the impact of the Holocaust on individual lives while also learning "important digital media skills, including ... ethical remixing, that will prepare [students to become] digital citizen[s] in the 21st century."[19] Through this platform, the Foundation promotes its videos as moral touchstones, even as it invites students to sample the collection to create new cinematic narratives. IWitness characterizes the exercise as an opportunity for ethical instruction in its own right, extending the Foundation's conviction that viewing the Archive's interviews has a morally galvanizing power.

One of the first major video archives of Holocaust memories established in the United States recently initiated a different kind of remediation of its digitized holdings. Begun in the late 1970s, the Fortunoff Video Archive for Holocaust Testimonies recorded over 4,000 videos, now housed at Yale University.[20] Starting in 2019, the Archive has produced a series of podcasts titled "Those Who Were There: Voices from the Holocaust." The series presents excerpts from audio tracks of video interviews, supplemented with narration and musical scoring. Similar to IWitness, this remediation addresses a moral imperative. The podcasts' website explains that "it is our duty to listen and share, so that the horrific events of the Holocaust do not fade from memory."[21] Also like IWitness, "Those Who Were There" appears to have embraced newer media in order to engage younger audiences. As one reviewer of the project observes, it "applies a twenty-first century social media strategy to reach out to a generation ... weaned on Instagram and Snapchat."[22] The pedagogical aims of "Those Who Were There" are evident in supporting materials on its website: interview transcripts, vintage photographs and documents, additional readings and historical background on the Holocaust.

By dint of the interviews' digital format, their remediation in these projects entails both reducing and supplementing the original recordings: in IWitness,

---

**19** "About/IWitness," accessed January 3, 2015, http://iwitness.usc.edu/SFI/About.aspx.

**20** "Fortunoff Video Archive for Holocaust Testimonies," accessed January 21, 2021, https://fortunoff.library.yale.edu/.

**21** "Those Who Were There: Voices from the Holocaust," accessed January 21, 2021, https://fortunoff.library.yale.edu/education/podcasts/.

**22** Jonathan C. Friedman, "*Those Who Were There: Voices from the Holocaust*, Yale University Fortunoff Video Archive for Holocaust Testimonies" [review], *The Oral History Review* 47, no. 1 (2020): 112–13.

students select clips from VHA videos; each podcast episode of "Those Who Were There" presents up to 25 minutes of audio culled from longer interviews. At the same time, both projects integrate additional elements: music and narration are added in the podcasts, and the IWitness editing platform enables students to integrate these and other elements into their videos. As these examples demonstrate, the digital remediation of works of Holocaust memory can have epistemological consequences. The original resource, itself a mediation of recollections, can be segmented and incorporated with other material, to create new works of remembrance.

Similarly, the means of accessing digitized works of Holocaust memory through cataloging, transcribing, translating, and tagging can inform how these resources are engaged, as exemplified by the Shoah Foundation's Visual History Archive.[23] Among the many projects undertaken to record interviews with survivors and eyewitnesses of the Holocaust, the VHA stands out on several counts: not only as the largest such collection, but also as the most diverse, with interviews conducted in 56 countries and 32 languages, encompassing a range of Holocaust survivors and eyewitnesses. In recent years the Foundation has expanded the Archive's holdings to include interviews with Holocaust survivors collected by other organizations as well as films and videos of survivors or eyewitnesses of other atrocities. These range from the Armenian genocide, which began in 1915, to Rohingya survivors of mass violence in Burma in 2017. The VHA is also the most widely accessible major collection of Holocaust interviews, available in a variety of formats, including documentary films, educational DVDs, the aforementioned IWitness platform, and an online archive housing the entire collection, which facilitates searching this extensive body of material with an elaborate indexing system. As a result of its pioneering work on indexing and cataloging videos, as well as creating a digital platform for their viewing online, the Shoah Foundation holds multiple patents on its technological innovations.[24]

The scope of the VHA's index has also evolved as the archive's holdings have expanded – adding, for example, indexing terms specific to the interviews with survivors or eyewitnesses of atrocities other than the Holocaust (e.g., "Armenian culture," "attitudes toward Hutu," "foot binding"). In addition, indexing terms that are relevant to multiple atrocities – such as "political identity," "post-conflict justice," and "trauma-related dreams" – now enable researchers to make comparisons among these accounts of different historical events. The index indicates as

---

**23** See Jeffrey Shandler, *Holocaust Memory in the Digital Age: Survivors' Stories and New Media Practices* (Stanford, CA: Stanford University Press, 2017), ch. 1.

**24** USC Shoah Foundation, *Testimony: The Legacy of* Schindler's List *and the USC Shoah Foundation* (New York: Newmarket Press and HarperCollins, 2014), 331.

well when search terms have been added to the VHA's search mechanism, evincing an evolving understanding of what topics researchers may wish to explore.

The VHA is especially noteworthy because it was created on the cusp of the transition from analog to digital media. Since its inception, the Shoah Foundation has grappled with the possibilities and challenges posed by an expanding array of new technologies. The Foundation recorded its interviews in the mid-1990s on Betacam SP format analog videotapes, then the broadcast industry standard. Concurrently, digital media and the internet were coming into widespread use, fostering new possibilities for creating and engaging with information. Digital and online technologies soon became central to the Archive's agenda, facilitating the preservation, indexing, and accessibility of its collection.

Thus, the VHA straddles both the temporal boundary marked by the impending loss of living witnesses to the Holocaust and the transition from what has been called the "video age" to the "digital age." The Shoah Foundation set out to use state-of-the-art technology to preserve the memories of survivors as their passing approached. Yet its archive evinces how mutable memory practices are – how quickly the mediums employed have changed, plus the fact that newer mediums are less stable than older ones. In this respect, the VHA exemplifies a challenge that Holocaust remembrance has confronted from the start. As Holocaust studies scholar Alan Rosen notes, Holocaust memory projects have always "been bound up ... with technological advance and obsolescence."[25]

Digitizing the Shoah Foundation's videos facilitated their cataloging and, especially, their indexing. Like other projects discussed above, digitization enables searching this large collection of videos by dint of their metadata. Moreover, the VHA's index expands greatly the possibilities for searching within and across interviews, thereby transforming how they can be used. The Shoah Foundation initially decided to index rather than transcribe the videos, as a more expedient undertaking and a more useful aid to researchers.[26] As taping interviews proceeded, the Foundation began developing a matrix of search terms specific to the Archive.

Like the VHA itself, the archive's index is vast, with over 50,000 search terms. They are keyed to particular interview segments during which interviewees discuss these terms. In addition to looking for names of particular people, places, events, or institutions, VHA users can search the content of videos through a graduated taxonomy of subject terms. The index facilitates both expanding and constraining searches. Its subject terms are generalized, enabling users to cross-ref-

---

**25** Rosen, *The Wonder of Their Voices*, 20.

**26** See USC Shoah Foundation, *Testimony*, 221.

erence interviews conducted in various locations and languages and sometimes across different genocides. Each search term can also be narrowed by the criteria of interview language, gender of interviewee, and "experience group" (such as Roma survivors or liberators). Searching coincident terms (for instance, finding interviews that discuss both "early personal aspirations" and "school antisemitism") can further winnow interview selections.[27] Therefore, exploring this extensive body of material relies to a great deal on the archive's choice of index terms and how individual interviews are tagged with these terms.

Though the Shoah Foundation originally decided against transcriptions, it has recently begun to add them as a searchable resource that complements the index. The Foundation explains that "the transcripts ... appear on the screen as interviewees are talking so there will not be any loss of nuance of expression or paralinguistic cues."[28] The on-screen juxtaposition of video, indexing terms, transcription, plus mapping of locations mentioned, situates the recordings within a complex of search tools and supporting materials. This presentation also serves to remind viewers that the individual interviews they watch are part of a larger project.

At the same time that it has expanded the scope of the VHA, the Shoah Foundation continues to explore new digital technologies for documenting and presenting Holocaust memories. Starting in 2012, the Foundation collaborated with University of Southern California's Institute for Creative Technologies to develop New Dimensions in Testimony. This project strives to enable people "to interact personally with testimony through holographic display" of a Holocaust survivor. Questions people ask the animate image of the survivor trigger prerecorded responses, by means of a "technology called Natural Language Understanding." New Dimensions in Testimony both extends and alters the Shoah Foundation's mission to document and disseminate survivors' life stories. Interactive holography shifts the mode of engaging survivors' memories from listening to and watching the VHA's "talking-head" video interviews to interacting with three-dimensional, full-body apparitions of survivors.

Interactive mediums are widely hailed for enhancing audience encounters with information and thereby empowering them to further action. According to the Institute for Creative Technologies, this project's goal is to offer people "simulated, educational conversations with survivors though the fourth dimension

---

**27** "USC Shoah Foundation, Visual History Archive: Search," accessed September 17, 2015, http://vha.usc.edu/keywordsearch/keywordSearch.aspx.

**28** USC Shoah Foundation, "Partnership with ProQuest to Increase Access to Visual History Archive," accessed March 9, 2016, https://sfi.usc.edu/news/2016/03/11051-partnership-proquest-increase-access-visual-history-archive.

of time. Years from now," the institute writes, "long after the last survivor has passed on, the New Dimensions in Testimony project can provide a path to enable young people to listen to a survivor and ask their own questions directly." At the same time that its creators champion the project's use of "the latest technologies," they also assert that it "advances the age-old tradition of passing down lessons through oral storytelling."[29]

Notwithstanding these claims, there are telling differences between conversing with an actual person and the simulation this project offers. Unlike video interviews of survivors, the technology used for New Dimensions in Testimony requires that the exchange between interviewer and informant be atomized into discrete questions and answers. Thus, Holocaust Studies scholar Rachel Baum argues, an interactive holograph of a survivor is, in effect, "the visual representation of a database" of information.[30] Therefore, the exchange of questions and answers with the holograph is transactional rather than developmental. This alters the nature of the information provided as well as how it is received. Communications scholar Ekaterina Haskins argues that, when using interactive digital media, "the audience no longer acts as a consumer of a linear story [but] takes part in the experience by making choices."[31] Yet the holograph's prerecorded life story in units of information, intended to answer individual questions, precludes the possibility of asking follow-up questions that probe or develop the survivor's narrative, which can be done when interviewing a living informant, as documented in audio and video interviews.

This simulation of an actual conversation is limited both by the technology's occasional failure to offer apposite answers to questions and by the fact that the holograph cannot respond to a question that was not anticipated when recording the survivor's database of responses.[32] Not hearing answers to one's questions is, of course, a limitation that viewers of video interviews with survivors might also experience. However, those recordings are manifestly closed works, documenting an interaction one can audit but not enter. By contrast, the interactive holograph purports to provide the equivalent of a conversation with a living survivor.

---

**29** USC Institute for Creative Technologies, "New Dimensions in Testimony," accessed March 9, 2019, http://ict.usc.edu/prototypes/new-dimensions-in-testimony/.

**30** Rachel N. Baum, "Remediating the Body of the Witness: Holocaust Testimonies in New Media," Paper presented at the International Conference on Bearing Witness More Than Once, Humboldt-Universität, Berlin, March 16, 2016, n.p. Cited with permission.

**31** Ekaterina Haskins, "Between Archive and Participation: Public Memory in a Digital Age," *Rhetoric Society Quarterly* 37, no. 4 (2007): 406.

**32** See, e.g., Anna Russell, "A Holocaust Survivor's Digital Doppelgänger," *The New Yorker*, January 22, 2018, accessed January 20, 2020, https://www.newyorker.com/magazine/2018/01/22/a-holocaust-survivors-digital-doppelganger.

Video interviews provide an unfolding narrative, enabling the opportunity to observe the dynamics of memory. As literature scholar James Young notes, video foregrounds the process of remembrance by recording "both the witnesses as they make their testimony and the understanding and meaning of events generated in the activity of testimony itself." Consequently, observers of these videotapes "witness ... the making of testimony."[33] By contrast, interactive holographs offer multiple potential points of entry to the survivor's life narrative, prompted by the asker's particular initiative. One need not, for example, start at the beginning of the survivor's life or of World War II, as some video interviews do, in keeping with the protocols for their respective projects. This open-ended interaction with the holograph relies on askers' knowledge of what to ask, as they must initiate each exchange. The challenge this interactive format poses to users is not limited to holography. Media scholar Anna Reading observes that visitors using interactive computer displays in the Simon Wiesenthal Center's Multi-media Learning Center in Los Angeles soon lose interest after making random choices of topics to pursue and generally "prefer to make interactive choices based around 'what they already know.'" Interactivity, she notes, "is not the same as agency."[34] In the case of the holographs, the technology's novelty predominates. Engaging with it can prove enticing or daunting, as its presentation of a virtual person may seem either a wondrous or an unsettling phenomenon.

Given this challenge, the Shoah Foundation recently announced a more directly pedagogical approach to engaging with survivor holographs. In an activity called "A Conversation with Pinchas Gutter," students first learn about this survivor's background and practice interviewing skills; then they are invited to ask questions of Gutter's "interactive biography." The pedagogy for this project is equally concerned with the process of interrogation as it is with the information students acquire from engaging with the holograph. As the Foundation explains, "Students will learn the techniques for having a conversation with a survivor and how to construct questions appropriately to elicit personal, historical and universal thematic responses."[35]

In yet another use of new media technology by the Shoah Foundation, Gutter is the subject of "The Last Goodbye," described as an "immersive virtual reality testimony experience" that "represents unprecedented advances in storytelling

**33** James E. Young, *Writing and Rewriting the Holocaust: Narrative and the Consequences of Interpretation* (Bloomington: Indiana University Press, 1988), 159, 171.

**34** Anna Reading, "Digital Interactivity in Public Memory Institutions: The Uses of New Technologies in Holocaust Museums," *Media, Culture & Society* 23 (2003): 79, 73.

**35** USC Shoah Foundation, "A Conversation with Pinchas Gutter," accessed January 22, 2021, https://iwitness.usc.edu/sfi/Activity/Detail.aspx?activityID=5391.

through technology." In this 20-minute presentation, which debuted in 2017, users don virtual reality headsets to simulate accompanying Gutter on his "final return" to the former death camp at Majdanek, where he was imprisoned and his parents and sister were murdered.[36] The technology employed in "The Last Goodbye" provides a distinct mode of accessing survivor storytelling, with its own assets and limitations. Here, connecting with the survivor is not interactive. Unlike "New Dimensions in Technology," "The Last Goodbye" does not proffer a conversation with Gutter. Instead, the project's structure follows his monologue, organized by the sites visited. "The Last Goodbye" resembles survivors recounting their wartime experiences to tour groups visiting death camps. It thereby constitutes an effort to use new technology to simulate another waning practice of face-to-face encounters with survivors. Thus, when "The Last Goodbye" was presented at the Museum of Jewish Heritage in New York, the museum explained that, "as Pinchas recounts his experiences, you walk alongside him – seeing what he sees, hearing what he hears, and learning as he guides you through an account of his own history."[37]

As a consequence of this cascade of new digital technologies, students, teachers, museum goers, and others are offered multiple approaches to engage with the life stories of Holocaust survivors. Each technology facilitates a different kind of encounter, with its own possibilities and limitations. The fact that some survivors, such as Gutter, offer their personal histories in multiple formats provides opportunities to reflect on the impact of each medium and format for recording and sharing a life story on both the survivors and their audiences.

<p style="text-align:center">*</p>

Beyond their role as conveyers of information about the Holocaust and its remembrance, these many resources are noteworthy in their own right as phenomena of digital humanities, especially for scholars who are interested in this evolving interrelation of new technologies with the documentation and study of the past. In addition, all users of digital materials such as the aforementioned examples can benefit greatly from observing that their mediation and remediation are intrinsic to these resources:

First, by examining the protocols for how the materials, whether analog or digital, were collected: what mediums were used and how were the possibilities and the value of these mediums understood. Consider, for example, decisions

---

**36** USC Shoah Foundation, "The Last Goodbye," accessed January 22, 2021, https://sfi.usc.edu/lastgoodbye.

**37** Museum of Jewish Heritage, "The Last Goodbye," accessed March 23, 2019, https://mjhnyc.org/exhibitions/the-last-goodbye/.

to make video recordings of Holocaust survivors as talking heads against a blank background, in their homes, or in environments associated with their wartime experiences; the decision to include photographs, documents, or other personal objects in the interview and how they are incorporated. Such decisions as these shape how the survivors' narratives are mediated, as the narratives themselves are mediations of recalled experience.

Second, by examining how digitized resources are remediated when they are cataloged and indexed, noting how the cataloging or indexing criteria shape the ways resources are located in a collection as well as how these criteria enable or constrain what is findable within individual resources. Though this metadata and the rubrics through which it is searched are extrinsic to collections of photographs and audio or video recordings, this information serves as the point of entry to the resources themselves, defining their value.

Third, by examining the implications of remediation: for example, the transcription or translation of audio recordings, the addition of annotations, images, maps, etc., to provide context for the original resources. Consider what these additions reveal about expectations of who will use these resources and to what ends, as well as what new research possibilities these tools prompt. Moreover, consider how the remediation of so many resources on the internet enables relationships among them, linking collections within one online mega-archive, while also imposing on the delivery of these wide-ranging materials the common medium of a website.

Fourth, by examining the means used to access these resources, including the possible venues for this engagement and how they can shape users' encounters with materials. Consider, for example, the differences among viewing the same digitized photograph or listening to the same song recording at home, in a classroom, in a library, or in a museum.

The decisions of how Holocaust memories are documented and how these documentations are remediated have important implications for their scholarly engagement. For example, communications scholar Amit Pinchevski argues that the turn to videography to document Holocaust survivor narratives, beginning in the 1970s, facilitated the advent of a scholarly "discourse of trauma and testimony." Pinchevski posits that the medium of videotape served as "the technological unconscious," by enabling closer scrutiny of survivors' oral accounts than had previously been possible.[38] In my own work on the VHA, its index and search mechanism enabled me to locate hundreds of instances in which survi-

---

**38** Amit Pinchevski, "The Audio Visual Unconscious: Media and Trauma in the Video Archive for Holocaust Testimonies," *Critical Inquiry* 39, no. 1 (Autumn 2012): 144.

vors discuss the role of Yiddish in their lives, from which I constructed a survey of the dynamics of Yiddish as a presence in Holocaust survivors' lives over the course of the 20th century. The result is a composite account, constructed from segments of different survivors' narratives, none of whom provides such an overview. Rather, survivors discuss Yiddish only on occasion, usually just once, in the course of relating their life histories. In contrast to the narrative constructed from the aggregation of these discussions, the limited, inconstant presence of Yiddish in these narratives is telling in its own right, demonstrating the linguistic disruption in 20th-century European Jewish life that is, in turn, emblematic of great cultural, political, geographic, and demographic upheavals.[39]

The larger context of the digital turn is especially resonant for works of Holocaust remembrance, as this moment coincides with the aging and passing away of survivors. Mounting concerns about how Holocaust education and remembrance can proceed without survivors have prompted turns to these various mediations of their remembrances in order to extend survivors' presence, even striving to endow them with a kind of immortality.[40] Yet the notion that digital media ensure stable preservation of survivors' memories is a false friend. Digital media are distinctly mutable, composed of long strings of binary code that can easily become corrupted. These media are dependent on rapidly changing technologies, whether equipment that quickly becomes obsolete or online platforms that suddenly change or disappear without a trace.

The turn to ever newer technologies in the digital age highlights the speed with which new media become old. Eventually, their technology, sensibility, or aesthetic will appear less than state-of-the-art – just as Boder's interviewing methodology can now sound outdated, or videos made in the last decades of the 20th century are evidently of an earlier era, marked by survivors' clothing, speech, and home furnishings. At the same time, the cascade of innovations in digital media can also present new opportunities for documenting Holocaust memories, and not only to institutions, but also to individuals. Just as the first videotaped life histories of survivors were made by their families almost a half-century ago, similar videos are now recorded on cell phones and posted on social media platforms. With this practice, the memories of older generations are integrated into the posters' ongoing online self-portraiture.

The tension between the desires to preserve memory and the evolving, shifting possibilities of digital media may seem challenging. However, this is emblem-

---

**39** See Shandler, *Holocaust Memory in the Digital Age*, ch. 3.
**40** See Jeffrey Shandler, "The Savior and the Survivor: Virtual Afterlives in New Media Practices," *Jewish Film & New Media: An International Journal* 8, no. 1 (2020): 23–47.

atic of memory generally, which is by its nature subjective, contingent, and relational. Therefore, as we grapple with how to understand and to work with these new kinds of resources, we need to look back in time to consider the dynamics of their development, even as we look forward, as new technologies, research interests, and practices of engagement with digital resources continue to emerge.

The field of Jewish Studies has examined the impact of earlier new mediums and media practices on Jewish life, starting with the advent of writing in relation to oral language in ancient times and extending through print, photography, sound recordings, and broadcasting. Just as reflecting on digital media in Holocaust memory practices redounds to issues of concern regarding these newest technologies in other fields of Jewish Studies, the digital turn in this field resonates with ongoing considerations of how prior engagements with media that were once new have figured in Jewish life and how it has been studied.[41]

# References

"About the Yad Vashem Archives." Accessed November 20, 2012. www1.yadvashem.org/yv/en/about/archive/about_archive_whats_in_ archive.asp.

"About/IWitness." Accessed January 3, 2015. http://iwitness.usc.edu/SFI/About.aspx.

"AHEYM: The Archives of Historical and Ethnographic Yiddish Memories." Accessed June 3, 2018. http://www.iub.edu/~aheym/.

"And I Still See Their Faces." Fundacja Shalom. Accessed January 22, 2021. http://shalom.org.pl/en/i-ciagle-widze-ich-twarze-2/.

*And I Still See Their Faces: Images of Polish Jews / I ciągle widzę ich twarze: Fotografia Żydów polskich*. Warsaw: Fundacja Shalom, 1996.

Baum, Rachel N. "Remediating the Body of the Witness: Holocaust Testimonies in New Media," Paper presented at the International Conference on Bearing Witness More Than Once, Humboldt-Universität, Berlin, March 16, 2016.

Boder, David P. *I Did Not Interview the Dead*. Urbana: University of Illinois Press, 1949.

"EYDES: Evidence of Yiddish Documented in European Studies." Accessed June 3, 2018. http://www.eydes.de/.

Fogu, Claudio. "Digitalizing Historical Consciousness." *History and Theory* 48, no. 2 (2009): 103–21.

"Fortunoff Video Archive for Holocaust Testimonies." Accessed January 21, 2021. https://fortunoff.library.yale.edu/.

Friedman, Jonathan C. "Those Who Were There: Voices from the Holocaust, Yale University Fortunoff Video Archive for Holocaust Testimonies" [review]. *The Oral History Review* 47, no. 1 (2020): 112–13.

---

**41** See Jeffrey Shandler, "What Becomes of Old Media?," *AJS Perspectives*, Spring 2018: 20–21.

Froio, Nicole. "We Asked TikTokers Why They're Pretending to Be Holocaust Victims," *Wired*, August 21, 2021. Accessed August 24, 2020. https://www.wired.co.uk/article/tiktok-holocaust-pov.

Haskins, Ekaterina. "Between Archive and Participation: Public Memory in a Digital Age." *Rhetoric Society Quarterly* 37, no. 4 (2007): 401–22.

"Language and Culture Atlas of Ashkenazic Jewry." Accessed June 3, 2018. http://library.columbia.edu/locations/global/jewishstudies/lcaaj.html.

Levine, Louis D., ed. *Lives Remembered: A Shtetl through a Photographer's Eye*. New York: Museum of Jewish Heritage, 2002.

Museum of Jewish Heritage. "The Last Goodbye." Accessed March 23, 2019. https://mjhnyc.org/exhibitions/the-last-goodbye/.

"People of a Thousand Towns." YIVO Institute. Accessed January 22, 2021. http://yivo1000towns.cjh.org/search_tips.asp.

Pinchevski, Amit. "The Audio Visual Unconscious: Media and Trauma in the Video Archive for Holocaust Testimonies." *Critical Inquiry* 39, no. 1 (Autumn 2012): 142–66.

Reading, Anna. "Digital Interactivity in Public Memory Institutions: The Uses of New Technologies in Holocaust Museums." *Media, Culture & Society* 23 (2003): 67–85.

"Recording Cultural Genocide and Killing Sites in Jewish Cemeteries." Accessed January 22, 2022. https://holocaustkillingsites.com/.

Rosen, Alan. *The Wonder of Their Voices: The 1946 Holocaust Interviews of David Boder*. New York: Oxford University Press, 2010.

Russell, Anna. "A Holocaust Survivor's Digital Doppelgänger." *The New Yorker*, January 22, 2018. Accessed January 20, 2020. https://www.newyorker.com/magazine/2018/01/22/a-holocaust-survivors-digital-doppelganger.

"The Ruth Rubin Legacy Archive of Yiddish Folksongs." Accessed December 24, 2018. https://exhibitions.yivo.org/exhibits/show/ruth-rubin-sound-archive/home.

Shandler, Jeffrey. *Holocaust Memory in the Digital Age: Survivors' Stories and New Media Practices*. Stanford, CA: Stanford University Press, 2017.

Shandler, Jeffrey. "'People of a Thousand Towns': YIVO's Videodisc Project." *Jewish Folklore and Ethnology Review* 10, no. 1 (1988): 18.

Shandler, Jeffrey. "The Savior and the Survivor: Virtual Afterlives in New Media Practices." *Jewish Film & New Media: An International Journal* 8, no. 1 (2020): 23–47.

Shandler, Jeffrey. "What Becomes of Old Media?" *AJS Perspectives*, Spring 2018: 20–21.

"The Stonehill Jewish Song Collection." Accessed June 3, 2018. http://www.ctmd.org/stonehill.htm.

"Those Who Were There: Voices from the Holocaust." Accessed January 21, 2021. https://fortunoff.library.yale.edu/education/podcasts/.

United States Holocaust Memorial Museum. Collections search: audio visual oral histories. Accessed January 21, 2021. https://collections.ushmm.org/search/?f%5Bf_audiovisual%5D%5B%5D=testimony&q=oral+histories.

USC Institute for Creative Technologies. "New Dimensions in Testimony." Accessed March 9, 2019. http://ict.usc.edu/prototypes/new-dimensions-in-testimony/.

USC Shoah Foundation. "A Conversation with Pinchas Gutter." Accessed January 22, 2021. https://iwitness.usc.edu/sfi/Activity/Detail.aspx?activityID=5391.

USC Shoah Foundation. "The Last Goodbye." Accessed January 22, 2021. https://sfi.usc.edu/lastgoodbye.

USC Shoah Foundation. "Partnership with ProQuest to Increase Access to Visual History Archive." Accessed March 9, 2016. https://sfi.usc.edu/news/2016/03/11051-partnership-proquest-increase-access-visual-history-archive.

USC Shoah Foundation. *Testimony: The Legacy of* Schindler's List *and the USC Shoah Foundation*. New York: Newmarket Press and HarperCollins, 2014.

"USC Shoah Foundation, Visual History Archive." Accessed January 21, 2021. https://vhaonline.usc.edu/login.

"USC Shoah Foundation, Visual History Archive: Search." Accessed September 17, 2015. http://vha.usc.edu/keywordsearch/keywordSearch.aspx.

"Voices of the Holocaust." Accessed January 22, 2022. http://voices.iit.edu/.

Walden, Victoria Grace. "Reading about Digital Holocaust Memory." Accessed January 22, 2021. https://digitalholocaustmemory.wordpress.com/2020/07/10/reading-about-digital-holocaust-memory/.

Yad Vashem Collection of Testimonies. "Listing of the Record Groups in the Yad Vashem Archives." Accessed November 20, 2012. www1.yadvashem.org/yv/en/about/archive/pdf/list_of_record_groups.pdf.

Young, James E. *Writing and Rewriting the Holocaust: Narrative and the Consequences of Interpretation*. Bloomington: Indiana University Press, 1988.

"The Zalman Kaplan Collection of Pre-war Photos of Szczyczyn." Accessed January 22, 2021. http://www.szczuczyn.com/kaplan.htm.

"Żydzi Polscy." Accessed January 22, 2021. http://www.zydzipolscy.pl/.

Inna Kizhner, Melissa Terras, Julia Afanasieva,
Diana Pusenkova, Maria Sherer, and Daniil Skorinkin

# The Culture of the Very Rich and Very Poor: Do Digital Museum Collections Tell us Anything about Jewish Culture?

**Abstract:** Digital approaches to Jewish Studies allow collecting data from multiple sources. This enlarges the picture, makes our understanding of historical experiences more complete, and triggers further research questions. This chapter compares samples of artworks from digital online collections of the Metropolitan Museum of Art in New York and the Russian State Catalogue of Museum Collections in the Russian Federation. We show that the difference in their temporal and geographical coverage of Jewish artworks/historical documents often results from tagging, naming and understanding of what is deemed Jewish. The collective decisions made in these datasets, by many individuals, and differing institutions about cataloguing, indexing, and returning searches on Jewish culture in a digital age inform what parts of Jewish culture are accessible. The technological and data-led decisions become a part of multiple layers of decisions that inform how primary sources are formed in a digital age.

**Keywords:** online museum collections, cultural heritage, indexing, cataloguing, bias, Metropolitan Museum of Art, Russian State Catalogue of Museum Collections

# 1 Introduction

Data-driven research in Jewish Studies has been used to analyze historical Hebrew newspapers,[1] manuscripts,[2] how Jewish content appears in cultural heritage aggre-

---

**1** O. Soffer et al., "Computational Analysis of Historical Hebrew Newspapers: Proof of Concept," *Zutot – Perspectives on Jewish Culture* 17 (2020): 97–110.
**2** M. Zhitomirsky-Geffet and Gila Prebor, "SageBook: Toward a Cross-Generational Social Network for the Jewish Sages' Prosopography," *Digital Scholarship in the Humanities* 34, no. 3 (September 2019): 676–95; G. Prebor, Maayan Zhitomirsky-Geffet, and Yitzchak Miller, "A New Analytic Framework for Prediction of Migration Patterns and Locations of Historical Manuscripts Based on Their Script Types," *Digital Scholarship in the Humanities* 35, no. 2 (June 2020): 441–58.

gators such as Europeana,[3] and to evidence the politics of digitization as revealed through textual collections.[4] However, analysis has not yet focused on how Jewish culture is represented within datasets of museum objects. The aim of this chapter is to present the results of an exploratory study to establish a methodology that lets us understand Jewish culture (or other minority cultures) as represented within digital collections in different parts of the world. The comparison of search results for the Metropolitan Museum of Art in the United States and those for the State Catalogue of Museum Collections of the Russian Federation will allow us to understand the effect that a variety of collecting and cataloging approaches may have on research at the crossroads of Digital Humanities and Jewish Studies. This will also help us to examine how the status of a minority population within different societies has affected the institutional and collections' response to their culture. In doing so, we can see a diversity of cataloging approaches for different cultural environments and the effects of data-driven analysis when studying minority cultures.

Bias in cultural heritage data and representations of human knowledge in mass-digitized collections have been widely covered in recent literature.[5] Such work also relates to epistemic complications of data collecting, data cleaning, and model training,[6] and the difficulties imposed by "the infrastructures of knowledge-making."[7] Apart from selection and exclusion bias,[8] these "infrastruc-

---

**3** Dov Winer, "Judaica Europeana: An Infrastructure for Aggregating Jewish Content," *Judaica Librarianship* 18 (2014): 88–115.

**4** G. Zaagsma, "Digital History and the Politics of Digitization," Paper presented at *Digital Archive and Canon Workshop*, March 10, 2021, accessed May 11, 2021, https://www.digitales-archiv-und-kanon.de/contributions/Zaagsma_en.pdf.

**5** See, for example, B.H. Daru et al., "Widespread Sampling Biases in Herbaria Revealed from Large-Scale Digitization," *New Phytologist* 217 (2018): 939–55; N.B. Thylstrup, *The Politics of Mass Digitization* (Boston: MIT Press, 2019); K. Bode, "Why You Can't Model Away Bias," *Modern Language Quarterly* 81, no. 1 (2020): 95–124; A. Liu, "Toward a Diversity Stack: Digital Humanities and Diversity as Technical Problem," *PMLA/Publications of the Modern Language Association of America* 135, no. 1 (2020): 130–51; S. Bagga and A. Piper, "Measuring the Effect of Bias in Training Data for Literary Classification," *Proceedings of LaTeCH-CLfL 2020*, Barcelona, Spain, December 12, 2020, 74–84

**6** A. Bechmann and G.C. Bowker, "Unsupervised by Any Other Name: Hidden Layers of Knowledge Production in Artificial Intelligence on Social Media," *Big Data & Society* 6, no. 1 (2019): 1–11.

**7** B. Mak, "Archaeology of a Digitization," *Journal of the Association for Information Science and Technology* 65, no. 8 (2014): 1515–26, 1519, cited in Bode, "Why You Can't Model Away Bias," 2.

**8** Maria Economou, "Heritage in the Digital Age," in *A Companion to Heritage Studies*, ed. William Logan, Máiréad Nic Craith, and Ullrich Kockel (Chichester, UK: Wiley, 2016), 215–28; Bode, "Why You Can't Model Away Bias"; T. Hauswedell et al., "Of Global Reach Yet of Situated Contexts: An Examination of the Implicit and Explicit Selection Criteria that Shape Digital Archives

tures of knowledge-making" include misrepresentations caused by metadata conventions,[9] cataloging approaches,[10] classification principles,[11] and linguistic issues.[12] How things are named, classified and cataloged, and what contexts are selected to denote the meaning of artworks and documents strongly influences whether users are able to find minority cultures in the wealth of digitized cultural heritage we have. Perspectives and surrounding contexts may vary in different cultural environments as they tend to reflect the values of catalogers during a historical period. Only by understanding how cataloging principles and linguistic approaches influence the results of data retrieval, can we develop a methodology for studying minority cultures in contexts that combine big data approaches and epistemic dependence on the historical perceptions of data. We need to make explicit the principles that govern how information infrastructures shape our perceptions of Jewish culture and other cultures. We are always at risk of losing what we do not count and losing what we do not help users to see.[13] On the other hand, whatever classification instruments and cataloging principles we employ, some cultural objects tend to be always left behind. This happens because multiple "sieves" or instruments have different users in mind.[14] This means that due to their coding cultures and functionality they tend to exclude some objects, and therefore potential users, along the way. These lost objects and associated cultural contexts may be important regarding minority cultures where the risk of losing contexts is especially high.[15]

---

of Historical Newspapers," *Archival Science* 20 (2020): 139–65; Zaagsma, "Digital History and the Politics of Digitization."

**9** M.V. Fernandez, "The Coloniality of Metadata: A Critical Data Analysis of the Archive of Early American Images at the John Carter Brown Library," PhD thesis, University of Texas, 2018; Zaagsma, "Digital History and the Politics of Digitization."

**10** N.J. Bingham and H. Byrne. "Archival Strategies for Contemporary Collecting in a World of Big Data: Challenges and Opportunities with Curating the UK Web Archive," *Big Data & Society*, January 2021.

**11** G.C. Bowker and S.L. Star, *Sorting Things Out: Classification and Its Consequences* (Cambridge, MA: MIT Press, 1999); K. Cotter et al., "'Reach the Right People': The Politics of 'Interests' in Facebook's Classification System for Ad Targeting," *Big Data & Society*, January 2021.

**12** J. Aguilera, "Another Word for 'Illegal Alien' at the Library of Congress: Contentious," *The New York Times*, July 22, 2016, accessed April 25, 2021, https://www.nytimes.com/2016/07/23/us/another-word-for-illegal-alien-at-the-library-of-congress-contentious.html.

**13** G.C. Bowker, "Biodiversity Datadiversity," *Social Studies of Science* 30, no. 5 (2000): 643–83.

**14** J. Likhter, personal communication, May 7, 2021.

**15** J. Likhter, personal communication, May 7, 2021.

## 2 Methods

Our sources of data for analysis were the entirety of the online digital collections of the Metropolitan Museum of Art[16] in New York and the State Catalogue of the Museum Collections in the Russian Federation.[17] Our choice of online digital collections from two different countries and different contexts in cultural heritage was determined by our aim to see the difference (if any) in search results for "jew," "jewish," and related search terms in different settings. We sought to know whether and how the representation of objects reflected political, social, and epistemic attitudes in these different parts of the world, and between a leading Western institution, and a national catalog, to amplify difference in approach to test if this methodology would yield results. There is no national catalog in the USA against which to compare the Russian catalog, so the Metropolitan Museum of Art was chosen given it "collects, studies, conserves, and presents significant works of art across all times and cultures in order to connect people to creativity, knowledge, and ideas."[18] Both collections detail a significant number of objects published online, are accessed by large numbers of users, and allow similar access to metadata, which facilitated comparison. The online digital collection of the Metropolitan Museum of Art included over 400,000 objects in 2017[19] and it attracted 8 million visitors in 2018.[20] The Russian State Catalogue includes images for over 24 million museum objects at the time of writing, which makes it one of the largest national aggregators of cultural heritage across the world. It was built for inventory purposes, representing a third of Russian museum collections covering almost all state museums across the country. This makes it a representative dataset to study Jewish culture in the Russian environment. The comparison of results for the Metropolitan Museum of Art and those for the Russian State Catalogue will allow us to examine how the status of a minority population within different societies and contexts has affected the institutional and collections' response to culture in societies with different political and collections traditions. We did not compare the results for the Russian State Catalogue with those from

---

**16** https://www.metmuseum.org/, accessed April 25, 2021.

**17** https://goskatalog.ru/portal/, accessed April 25, 2021.

**18** Metropolitan Museum of Art, "About the Met," 2021, accessed June 25, 2021, https://www.metmuseum.org/about-the-met.

**19** T. Navarette and E. Villaespesa, "Digital Heritage Consumption: The Case of the Metropolitan Museum of Art," *magazén* 1, no. 2 (December 2020).

**20** Metropolitan Museum of Art, "Met Museum Sets New Attendance Record with More Than 7.35 Million Visitors," 2018, accessed May 12, 2021, https://www.metmuseum.org/press/news/2018/met-museum-sets- new-attendance-record.

a large international aggregator of cultural heritage images, such as Google Arts and Culture or Europeana Collections. This is because we wanted to focus on the epistemic path dependence, and the political and social attitudes that national and institutional digital infrastructures maintain in producing knowledge. The content of aggregators can be studied at a later date so that we see a combined approach to representations of Jewish culture or other minority cultures.

We used the collections' search engines for museum objects matching "jew," "jewish," "hebrew," and "yiddish" search terms. Russian search terms were "еврей" as a lemma because the Catalogue's search results can return results for parts of words and this is important for a language with numerous word forms, such as Russian. Other search terms were "иврит" for "hebrew," "идиш" for "Yiddish," and "древнееврей (ский)" for "Old Jewish," which is an equivalent to "hebrew" for religious, biblical, or literary contexts. We did not use the lemma "иудей" or "Judaic" as opposed to "Christian," although this search term returned over 500 results for the Russian State Catalogue and two results for the Metropolitan Museum of Art's online digital collection. Neither did we use results for "Judaism," although this search term returned 200 results for the Metropolitan Museum of Art and 24 results for the Russian State Catalogue.

As of October 2020, our search yielded approximately 900 results for the Metropolitan Museum of Art's digital collection and about 6,300 results for the Russian State Catalogue (see Tables 1 and 2 for a detailed breakdown of results). To get the results that indeed influence the perception of Jewish culture, we used only the "artworks with images" option on the websites. We used a sample (10%) of the results from the Russian State Catalogue which we made more representative by conducting a random search in the 12 categories that were offered by the Catalogue's interface, such as "paintings," "sculpture," "graphics," and "rare books." After filtering out duplicates and irrelevant results, we were left with 407 records for the Metropolitan Museum of Art, while the sample from the Russian State Catalogue yielded 404 records after excluding duplicates. We tabulated the metadata and data from textual descriptions for the results of the search. The fields included the name and culture of artists, types of objects, dates or time periods, and geographical descriptions. We also included data on whether the artwork was or had been a part of an exhibition in the Jewish Museum in New York or any other results for an exhibition that contextualized search results. This could only be done for the Metropolitan Museum of Art as the Russian State Catalogue does not record this information. We preferred to include the geography that was not directly related to the place where an artwork was produced because it enabled us to contextualize the Jewish culture showing the breadth and depth of linkages. The textual data for the Metropolitan Museum of Art were a combination of metadata, extensive historical contexts revealed in textual descrip-

tions, and exhibition histories that placed an artwork in a variety of cultures (Figure 1). The information for the Russian State Catalogue was obtained from the titles (which were sometimes quite extensive), the date, and the place where an artwork was produced (Figure 2). In this way, the interface of the two digital collections structured, directed, and limited our results (see Section 3 for analysis). Biographical details for the artists whose artworks were returned following our search in the Metropolitan Museum of Art online digital collection were obtained from Wikipedia. The maps showing the distribution of search results and their historical contexts across space were produced using Adobe Illustrator. We obtained contemporary maps through Yandex Maps service.[21] Maps of historical places, such as Mesopotamia, were more difficult to produce due to the uncertainty about their borders. Uncertain borders, derived from popular sources, such as Wikipedia, are shown as blurred lines to reflect this.

**Table 1:** The number of records and metadata returned by the search engine of the Metropolitan Museum of Art's online digital collection when using "jewish," "jew," "hebrew," and "yiddish" as search terms in autumn 2020.

| Search terms | Number of records with metadata and images returned by the search engine |
|---|---|
| "jewish" | 496 |
| "jew" | 210 |
| "hebrew" | 194 |
| "yiddish" | 4 |

**Table 2:** The number of records and metadata returned by the search engine of the Russian State Catalogue's online digital collection when using "jewish," "jew," "hebrew," "yiddish," and "old jewish" as search terms in autumn 2020.

| Search terms | Number of records with metadata and images returned by the search engine |
|---|---|
| "еврей" as a part of the word for "jew" and "jewish" | 4,687 |
| "идиш" for "yiddish" | 1,510 |
| "иврит" for "hebrew" | 96 |
| "древнееврей(ский)" for "old jewish" | 33 |

---

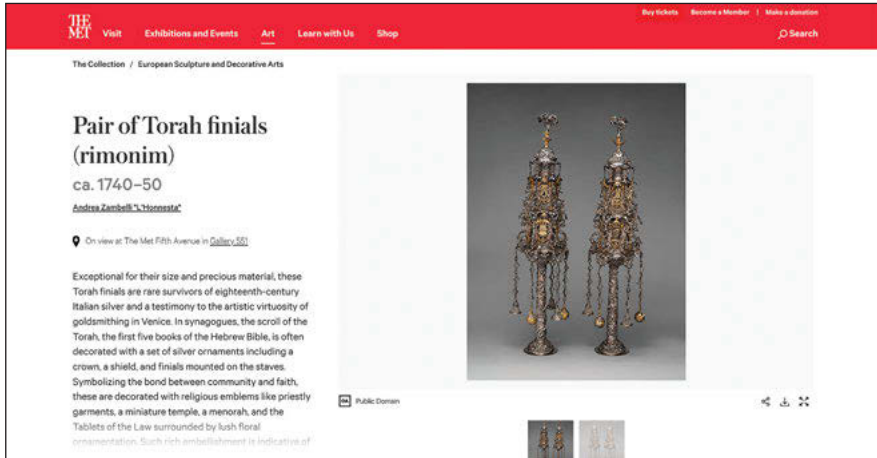**21** Accessed April 25, 2021, https://yandex.ru/maps/?ll=92.852572%2C33.461352&z=2.

**Figure 1:** An example of a record returned by the search engine of the Metropolitan Museum of Art's online digital collection when using "hebrew" as a search term. Images, metadata, and a textual description showing a pair of Torah finials. The Metropolitan Museum of Art, New York. Walter and Leonore Annenberg Acquisitions Endowment Fund, 2016.



**Figure 2:** An example of a record returned by the search engine of the Russian State Catalogue's online digital collection when using "old jewish" as a search term. An image with metadata showing Wilhelm Schickard's book with rules for Old Jewish language. Vologda State Museum of History, Architecture and Art.

# 3 Results

## 3.1 Geographical Distribution of Artworks

Figure 3 shows the geographical distribution of artworks in our sample for the search conducted on the Metropolitan Museum of Art website. We can see a wide geographical coverage. The artworks related to the USA dominate in the sample but we can also see artworks related to Europe, Russia, Egypt, Israel, Northern Africa, and Mesopotamia. Figure 4 shows the results for the geographical distribution in our sample for the search conducted on the website of the Russian State Catalogue. The artworks in the sample are almost exclusively produced in, published in, or related to places within Russia or countries that used to be a part of the Russian Empire or the Soviet Union. This representation is unrelated or weakly related to either historical or contemporary cultures but rather shows Jews as a separate ethnicity in Russian society. Conversely, in the representation of the Metropolitan Museum of Art, Jewish culture is embedded in its relations with neighboring countries and cultures. What is more important, the Metropolitan Museum of Art's online digital collection represents Jewish culture as a part of the Western canon[22] through its relation to biblical geography and the nations that interacted and influenced Jewish culture from the foundations of their society. This evidence supports a wide historical span and is augmented with the evidence on 20th century artists closer to the newer end of the canon (Figure 5). The results for the distribution of the Metropolitan Museum of Art artworks differ dramatically from the results of the Russian State Catalogue. This happens because contexts for "Jewish" and "Hebrew" differ in the two online digital collections and the words that are used as search terms, as a result, often mean different things. This is connected with different histories of Jews, different cultural environments, different attitudes to reconstructing histories, different cataloging approaches and different meanings of the search terms in the two countries. To a large extent, the difference is determined by what is deemed Jewish by catalogers and institutions in the two countries. It is also explained by the structure of the interfaces of the two digital collections and by how (whether) the artworks are contextualized through exhibitions or titles (see Section 2).

---

**22** N. Frye, *The Great Code: The Bible and Literature* (San Diego: Harcourt Brace Jovanovich, 1982); H. Bloom, *The Western Canon* (New York: Harcourt Brace, 1994).
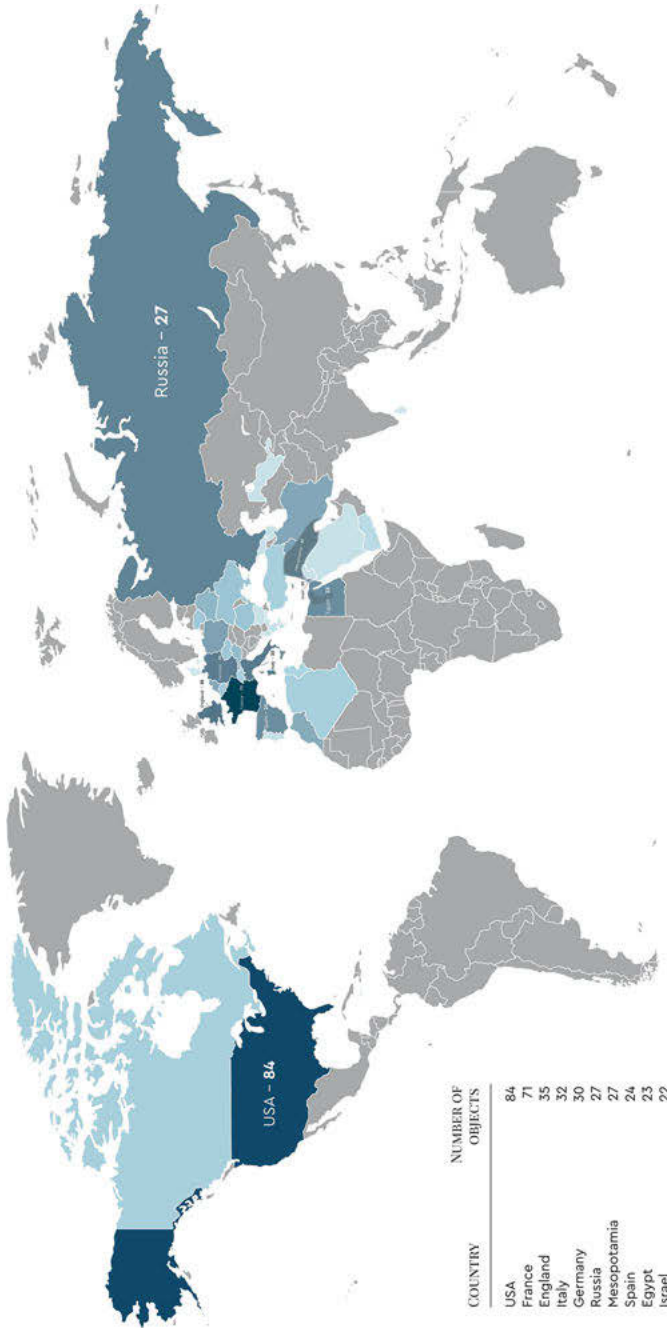
**Figure 3:** Geographical distribution of the records related to Jewish heritage in the Metropolitan Museum of Art, New York. Map: Julia Afanasieva, Inna Kizhner, Melissa Terras, Diana Pusenkova, Maria Sherer, and Daniil Skorinkin.
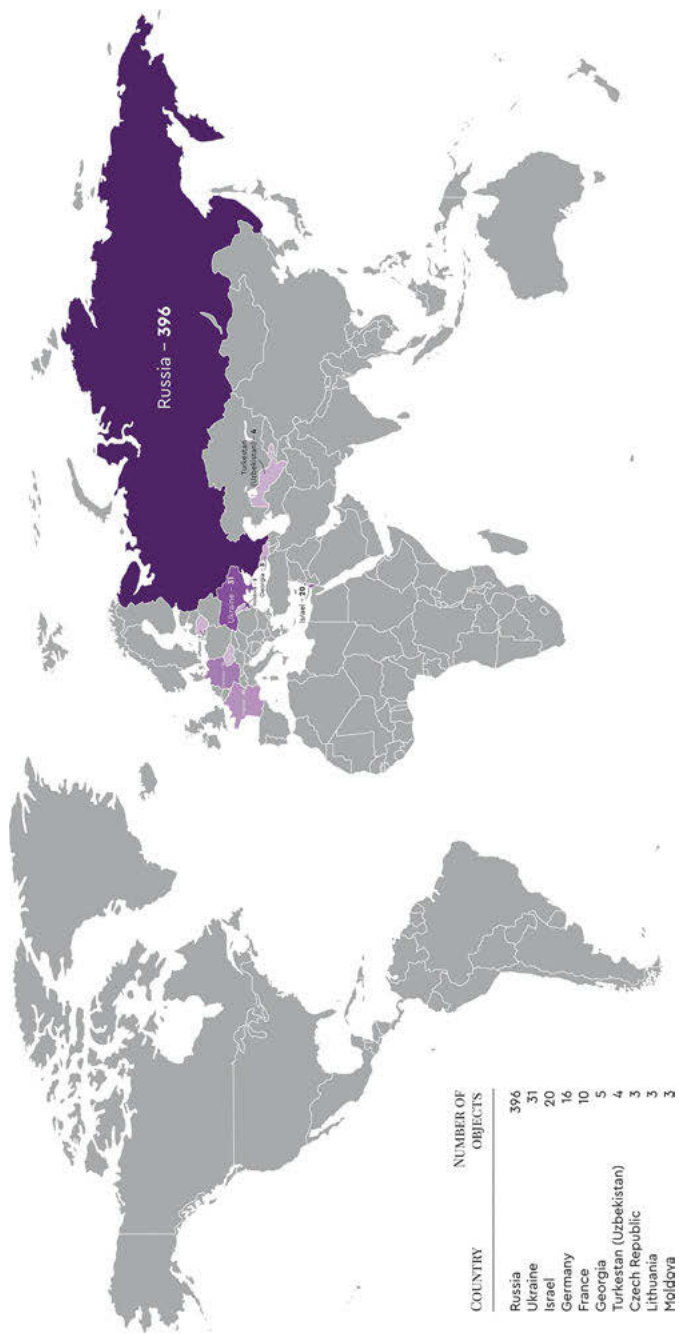
| COUNTRY | NUMBER OF OBJECTS |
|---|---|
| USA | 84 |
| France | 71 |
| England | 35 |
| Italy | 32 |
| Germany | 30 |
| Russia | 27 |
| Mesopotamia | 27 |
| Spain | 24 |
| Egypt | 23 |
| Israel | 22 |

| COUNTRY | NUMBER OF OBJECTS |
|---|---|
| Russia | 396 |
| Ukraine | 31 |
| Israel | 20 |
| Germany | 16 |
| France | 10 |
| Georgia | 5 |
| Turkestan (Uzbekistan) | 4 |
| Czech Republic | 3 |
| Lithuania | 3 |
| Moldova | 3 |

**Figure 4:** Geographical distribution of the records related to Jewish heritage in the State Catalogue of Russian Museum Collections. Map: Julia Afanasieva, Inna Kizhner, Melissa Terras, Diana Pusenkova, Maria Sherer, and Daniil Skorinkin.

## 3.2 Distribution of Artworks Across Time

Figure 5 shows the difference in the distribution of artworks across time for the two samples. It demonstrates a pronounced peak in early 20th-century artworks from the Metropolitan Museum of Art, and an even distribution going back to the fourth millennium BC. Combined with a significant number of avant-garde Jewish artists in this dataset (see Section 3.3), the results show a span across time and space that ranges from the "exemplary ancient" to "model moderns."[23] This happens because the Metropolitan Museum of Art has a metadata field that mentions all the places where artworks were exhibited. If such a place has "jewish" in its title, such as the Jewish Museum in New York, the artwork will be returned as a search result by the Metropolitan Museum of Art search engine. As a result, artworks are linked to the exhibitions at the Jewish Museum in New York in the period between 1956 and 2018,[24] and the Biblical Archaeological Exhibition at the University of Wisconsin-Madison Department of Hebrew and Semitic Studies (1975).[25] The links to the "Colmar Treasure" exhibition at the Metropolitan Museum of Art in 2019[26] show the possessions of a Jewish community in medieval Germany. The results linked to such exhibitions relate to 157 search results or about 40% of the dataset. Tagging systems and classification approaches mean that artworks from Assyrian, Babylonian, Sumerian, medieval, and contemporary cultures are included in the artworks returned from our queries and are therefore present in the dataset and analysis. Conversely, the Russian State Catalogue does not have this functionality and, consequently, the temporal and geographical spans of the sample are much narrower due to a lack of information and context. Figure 5 shows that artworks related to Jewish identity in the Russian State Catalogue date back to the 17th century and have little relation to biblical contexts or to the part of the European cultural canon that is associated with the history of Mesopotamia and ancient civilizations. The trough in the temporal distribution in the 1940s–1960s in the artworks from the Russian State Catalogue, which is most

---

**23** Yuri Slezkine, *The Jewish Century* (Princeton, NJ: Princeton University Press, 2002).

**24** See, for example, Jewish Museum, New York, "Russian Jewish Artists in a Century of Change, 1890–1990," September 21, 1995–January 28, 1996.

**25** University of Wisconsin-Madison Department of Hebrew and Semitic Studies, "The Book and the Spade" (Biblical Archaeological Exhibition), April 13–May 4, 1975.

**26** Metropolitan Museum of Art, "The Colmar Treasure: A Medieval Jewish Legacy," July 22, 2019–January 12, 2020, accessed April 30, 2021, https://www.metmuseum.org/exhibitions/listings/2019/colmar-treasure-medieval-jewish-legacy.
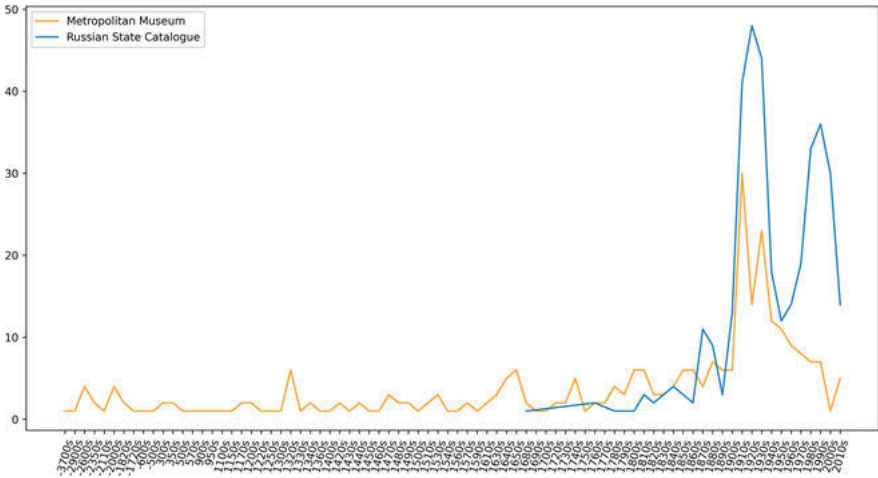
**Figure 5:** Temporal distribution of collections related to Jewish Studies in the Metropolitan Museum of Art and the Russian State Catalogue. Graph: Daniil Skorinkin, Inna Kizhner, Melissa Terras, Julia Afanasieva, Diana Pusenkova, and Maria Sherer.

pronounced directly after World War II, is explained by the famous anti-Jewish campaigns in the Soviet Union at the end of the 1940s.[27]

## 3.3 Distribution of Artworks by Topics and Types of Art

As shown in Figure 6, Jewish identity in the sample of search results returned by the Metropolitan Museum of Art's search engine is linked to contemporary artists, biblical contexts, and Mesopotamia (ancient civilizations). The perception of "Jewishness" is broadened by the inclusion of relations between Jews and Muslims or Sephardic Jews. On the other hand, Figure 6 shows that Jewish identity in the sample of search results returned by the Russian State Catalogue relates to high Yiddish culture and is linked to Jewish (Yiddish) theater, literature, and music. This is because the Russian State Catalogue returns numerous results for the Jewish State Theatre (over 30% of the search results for "Jewish" as a search term – over 1,500 records – where plays were performed in Yiddish). Other results from the sample are linked to books translated from (into) Yiddish and musical scores where texts were written in Yiddish.

---

**27** See, for example, B. Pinkus, *The Jews of the Soviet Union: The History of a National Minority* (Cambridge: Cambridge University Press, 1988).

Figure 7 shows the distribution of records by type of art for the most frequent types in the Metropolitan Museum of Art and Russian State Catalogue. We can see the prevalence of books for the Russian State Catalogue and the dominance of paintings for the Metropolitan Museum of Art. This cannot be explained by the fact that the Russian museums and, consequently, the Russian State Catalogue collect books rather than paintings. Indeed, paintings exceed books in the collection of the Metropolitan Museum of Art but there are twice as many paintings as books in its online digital collection.[28] However, the dominance of paintings over books from this online digital collection is much more pronounced in the artworks related to Jewish identity in our sample. The share of books related to Jewish artworks in the sample of search results from the Russian State Catalogue is much greater than the share of all books in the Russian State Catalogue.[29] This happens because texts accompanying images in the State Catalogue are descriptive and rarely contextualize artworks, while books come with extensive titles, quite often with the note "translated from Jewish (Yiddish)." It means that biblical paintings come without an explanation of who is depicted and what biblical story inspired the artwork. In addition, in the Russian State Catalogue textual descriptions that are not a part of the title are not included as a field that is used for retrieval. It means that even if Judith, the Jewish heroine, is presented as such in the textual description, the image and metadata will not be returned as search results for the terms "Jewish" or "Hebrew." Conversely, the Metropolitan Museum of Art's search engine retrieves artworks if the search term is mentioned in any of the metadata fields displayed to the user. This explains why "paintings" is the most frequent category for the topical distribution of search results for the Metropolitan Museum of Art. A third of the paintings from the Metropolitan Museum of Art in our sample involve biblical and Christian themes and contexts. Two-thirds of the 93 paintings in the sample are 20th-century artworks, and half of them are by American Jewish artists, such as Max Weber (1881–1961). A quarter of them are by French Jewish artists, such as Chaim Soutine (1893–1943) or Mark Chagall (1887–1985). On the other hand, famous contemporary artists in the Russian State Catalogue are rarely retrieved as Jewish. While the Metropolitan Museum of Art's search engine retrieves 11 of Chagall's paintings tagged as Jewish in data annotations (19% of 59 Chagall works in the online digital collection), the Russian State Catalogue's search engine produces four works tagged as Jewish or 2% of 185 records related to Chagall in this online digital collection). In the search results for the Metropolitan Museum of Art,

---

**28** Metropolitan Museum of Art, "Explore the Collection," 2021, accessed June 25, 2021, https://www.metmuseum.org/art/collection.

**29** Russian State Catalogue (The State Catalogue of Museum Collections of the Russian Federation), 2021, accessed June 25, 2021, https://goskatalog.ru/portal/#/.

a significant proportion of contemporary paintings (42 records out of 57 artworks, or about 70% of the sample) was produced by artists who were born in the Russian Empire, such as Mark Chagall or Max Weber. This group also includes artists whose parents emigrated from the Russian Empire at the end of the 19th and the beginning of the 20th century. The textual descriptions for these records could have mentioned Yiddish, since these painters were from families where the Yiddish language and culture were a part of their life and a significant source of influence.[30] However, users can rarely retrieve the word Yiddish in the curatorial texts and data annotations of the Metropolitan Museum of Art's online digital collection. The "Yiddish" search term returns four artworks or about 1% of all search results for the sample of search results for the Metropolitan Museum of Art. This means that until recently Yiddish meant "wide association with marginality, mutability, or obsolescence,"[31] and the museum community preferred to focus on other representations of Jewish culture. Conversely, searching for "Yiddish" in the Russian State Catalogue produces 1,510 results or 24% of results for the three search terms. However, nearly all the results (95% or 1,414 objects out of 1,510 results for this search term) are retrieved from the Regional Museum of the Jewish Autonomy in the Far East, reducing Yiddish culture to the idea of Jewish territorialism.[32]

## 3.4 Linguistic Differences Related to Search Results

One of the results reflecting the difference in linguistic approaches is the use of "Old Jewish" (древнееврейский) to write about "Hebrew" and the use of "Jewish" (еврейский) to write about "Yiddish" in the 20th century in the results from the Russian State Catalogue. This observation agrees with literature showing that "Jewish" was used in Russian/Soviet publication standards to denote "Yiddish"[33] and with literature on the eradication of Hebrew publishing in the 1920s in

---

**30** See, for example, B. Harshav and M. Chagall, *Marc Chagall and His Times: A Documentary Narrative* (Palo Alto, CA: Stanford University Press, 2004); J. Shandler, *Yiddish: Biography of a Language* (Oxford: Oxford University Press, 2020).

**31** J. Shandler, *Adventures in Yiddishland: Postvernacular Language and Culture* (Berkeley: University of California Press, 2006).

**32** Ellen Eisenberg, *Jewish Agricultural Colonies in New Jersey, 1882–1920* (Syracuse, NY: Syracuse University Press, 1995).

**33** L. Kogan and S. Loesov, "Old Jewish Language," in *World Languages: Semitic Languages. The Akkadian Language. Northwest Semitic Languages*, ed. A. Belova, L. Kogan, S. Loesov, and O. Romanova. Russian Academy of Sciences. Institute of Linguistics (Moscow: Academia, 2009), 296–375 [Russian].
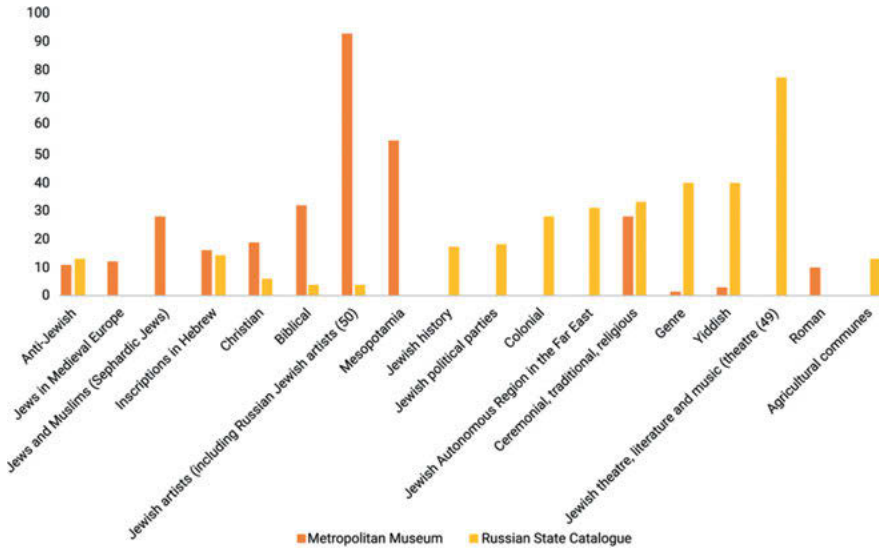
**Figure 6:** Distribution of records by topics for the most frequent types in the Metropolitan Museum of Art and the Russian State Catalogue. Graph: Julia Afanasieva, Inna Kizhner, Melissa Terras, Diana Pusenkova, Maria Sherer, and Daniil Skorinkin.
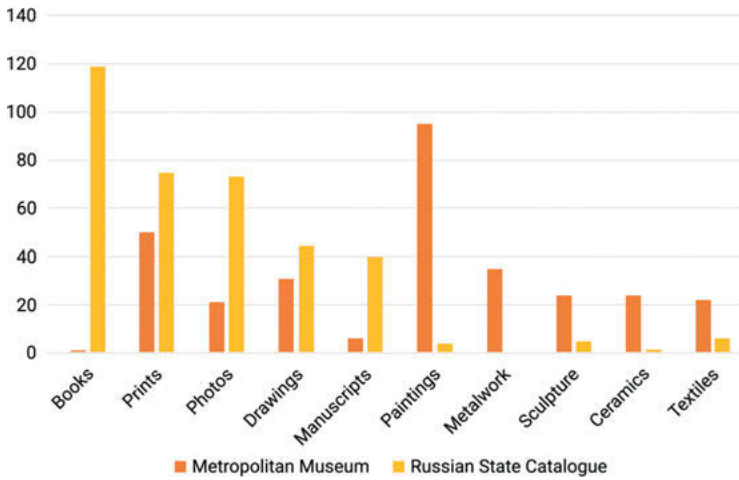


**Figure 7:** Distribution of records by type of art for the most frequent types in the Metropolitan Museum of Art and the Russian State Catalogue. Graph: Julia Afanasieva, Inna Kizhner, Melissa Terras, Diana Pusenkova, Maria Sherer, and Daniil Skorinkin.

Russia.[34] The latest item described with the expression "translated into Old Jewish" in the Russian State Catalogue is a book of Chekhov's stories published in 1918. The author of the earliest book in our dataset that uses "Jewish," meaning "Yiddish," includes a caveat, "written in Jewish-German dialect." The book was published in 1913. This means that the search for "Jewish" for the results dating back to the 20th century from the Russian State Catalogue will almost always return results that mean "Yiddish," and they do not include biblical contexts, religious contexts, and the contexts of ancient civilizations. Such confusion of words and terminology related to the circumstances and environments of Jewish culture definitely influenced what was deemed Jewish, and what aspects of Jewish life were emphasized.

Different contexts in which the results of the search are displayed imply that they reflect different attitudes and, probably, different understandings of the Jewish identity in the two digital collections. The contexts in the Russian State Catalogue are very much about the Yiddish speaking population of the Russian Empire and the Soviet Union, using "the language of the Jewish street,"[35] while the contexts of the Metropolitan Museum of Art are very much about Jewish identity as a part of the European cultural canon and post-colonial perception of Jews and Muslims or Jews in Northern Africa, such as in the exhibition "The Cairo Geniza: Jews & Muslims in the Mediterranean World 800–1500" about Jews in Egypt.[36] However, sometimes the results related to this part of Jewish identity are mixed with the representation of colonialism and orientalism.[37] In this sense, the results from the Metropolitan Museum of Art are in line with the results from the Russian State Catalogue displaying colonial photography from Georgia, Dagestan, and Uzbekistan at the turn of the 20th century.

Our results show different ways of representing Jewish identity in metadata and collection management systems. In this way, Russian cataloging approaches or cataloging voices show differences in cultural or social attitudes that date back to the very first years after the Russian Revolution. They are very much about the formal policy and representation of Jewishness through the Yiddish language and literature. In the case of the Metropolitan Museum of Art, Jewish identity is given a broader context that encompasses biblical culture, ancient civilizations, medieval European culture, Northern Africa, and an age of modernity showing contemporary Jewish art in Europe and the United States (mostly in New York).

---

**34** A. Blyum, "Hebrew Publications and the Soviet Censor in the 1920s," *East European Jewish Affairs* 23, no. 1 (1993): 91–99.

**35** Blyum, "Hebrew Publications and the Soviet Censor in the 1920s," 98.

**36** Jewish Museum, New York. "The Cairo Geniza: Jews and Muslims in the Mediterranean World 800–1500." January 1, 1997–October 12, 1997.

**37** Edward W. Said, *Orientalism* (New York: Random House, 1978).

# 4 Conclusion

This chapter showed that comparing metadata and catalog searches in this manner provides a methodology that lets us understand the representation of Jewish culture (or other minority cultures) by digital collections in different parts of the world, and in different contexts. Our results demonstrate that digital cultural heritage searches are complicated by a variety of cataloging voices and classification principles. The evidence that supports the argument of this chapter demonstrates how the layers of debates on "what is Jewish?" intertwine through collecting, tagging, linking, and publishing artworks online in very different contexts. Generalizations that can follow such analysis at scale vary widely, depending on political, cultural, and social perspectives. These perspectives are reflected in tagging principles, word usage, and classification approaches. What constitutes datasets under analysis, what texts, links and associating concepts are involved in producing contexts, determines our conclusions. Generalizations that emerge as a result produce contrasting, absolutely different representations of objects under study as shown in this chapter. One cannot but agree with Andrew Piper's argument about our "failure to generalize *well*" and about the crisis of reproducibility in the humanities based on "case-driven research."[38] Our chapter argues, however, that even analysis at scale does not contribute to better generalizations. To a certain extent, this happens because of a variety of cataloging conventions. Of course, the difference in results between the two chosen sources can also be assigned to different collections practices, since this exploratory research is comparing the curated collection of a large institution (the Metropolitan Museum of Art) with the portal to the collections of an entire nation (the State Catalogue of Museum Collections of the Russian Federation). Further work is needed on collections practices and processes, and how these approaches may have affected individual or national collections at different points of time, and we stress that the choice of which collections to compare is a crucial part of this design methodology.

The existence of multiple views and multiple approaches to museum annotation is not in itself a problem if our aim is to show that the truth is constituted by numerous contexts and does not depend on singular political and historical circumstances of building a collection. What is important, however, is making these circumstances and their influence explicit. In doing so we can demonstrate

---

**38** A. Piper, *Can We Be Wrong? The Problem of Textual Evidence in a Time of Data* (Cambridge: Cambridge University Press, 2020), 4–5.

how data models at the foundation of information systems depend on "ethical and political values, modulated by local administrative procedures."[39]

One of the ways of avoiding the complications of subjective interpretations and consequent difference in data models is reducing classifications to a list of categories from pre-developed ontologies as it is done in sciences. However, literature on biological databases in recent years has been devoted to the importance of contexts, the relevance of time and space, the historiography (provenance) of data diversity,[40] and semantic relations of synonyms and overlapping taxonomic classes.[41] What holds for biological data relates to social and cultural data where the contexts of data production are highly relevant. The complexity of metadata standardization in this case is augmented by the necessity to have standards reflecting a variety of types of data content. While Iconclass,[42] for example, gives a wide range of classes related to Jewish culture, they are mostly about religion, rituals, and religious institutions.[43] Participation in politics, Jewish welfare, the histories of Jewish territorialism, lay Yiddish and Hebrew culture, literature, and theater are difficult to catalog using the Iconclass perspective. In addition, the contexts of data provenance (how, when, and where data were produced) influence cataloging approaches and, consequently, affect the compatibility of data produced under different conditions. The layers of intertwined perspectives demonstrated in this chapter show how much work is to be done to achieve the compatibility of digital infrastructures for minority cultures.

What we have here is not only data produced in different spatial and temporal contexts. These are also data produced in cultural environments that were very different from (and sometimes hostile to) those of a minority culture. Such dispersal of attitudes and perspectives does not make easier the work of standardizing the approaches to cataloging Jewish culture. Future work may concentrate on how to approach the "overly optimistic" task of standardizing data structures and formats[44] for minority cultures. A possible problem here is how to select features to build classes if the size of a sample is too small or if a sample is not

---

**39** Bowker and Star, *Sorting Things Out*, 321.

**40** See, for example, Bowker, "Biodiversity Datadiversity"; L.M. Schriml et al., "COVID-19 Pandemic Reveals the Peril of Ignoring Metadata Standards," *Scientific Data* 7 (2020): article 188.

**41** Beckett W. Sterner, Nico M. Franz, and J. Witteveen, "Coordinating Dissent as an Alternative to Consensus Classification: Insights from Systematics for Bio-ontologies," *History and Philosophy of the Life Sciences* 42, no. 1 (2020): article 8.

**42** Iconclass, 2009, accessed April 25, 2021, http://www.iconclass.nl/home.

**43** Accessed April 25, 2021, http://www.iconclass.org/rkd/1/?q=jewish&q_s=1; accessed April 25, 2021, http://www.iconclass.org/rkd/12A/.

**44** Bowker, "Biodiversity Datadiversity," 661n4, cited in Claire Waterton, "From Field to Fantasy: Classifying Nature, Constructing Europe," *Social Studies of Science* 32 (2002): 177–204

representative enough.[45] Possible directions for complicated tasks of building classifications for the analysis of minority cultures is an exciting research area that can provide foundations for finding, sustaining, and disseminating cultures in today's world of statistical dominance and historical generalizations that privilege empires rather than local data.[46] This, however, does not solve the problem of the authoritative voices behind these classifications and the issues of applying standard classifications by curators educated within a certain school of thought or research community. Principles that will guide data models will rely on ethical norms, educational institutions, academic research, reading lists, and users to whom cataloging voices will be addressed. These agents will determine the data models and the accessibility of cultures we will see in the near future.

What is clear from this research is that the technological and data-led decisions made about cataloging, indexing, and returning searches on Jewish culture in a digital age inform what parts of Jewish culture are *accessible* in a digital age. The choices made in these platforms and data structures form online Jewish historical culture, as reflected in, supported by, and delivered from galleries, libraries, archives, and museums. The collective decisions made in these datasets, by many individuals, and differing institutions, often over long timeframes, will affect how users can find and navigate Jewish culture for years to come. We therefore suggest that the data structures which underpin content management systems are worthy of future study, comparison, and critique, when understanding how minority cultures are delivered to online users of cultural heritage.

---

**45** L.I. Kuncheva et al., "Feature Selection from High-Dimensional Data with Very Low Sample Size: A Cautionary Tale," August 28, 2020, accessed May 12, 2021, https://arxiv.org/pdf/2008.12025.pdf.
**46** G.C. Bowker, Foreword, in *All Data Are Local: Thinking Critically in a Data-Driven Society*, ed. Yanni Alexander Loukissas (Cambridge, MA: MIT Press, 2019); Yanni Alexander Loukissas, *All Data Are Local: Thinking Critically in a Data-Driven Society* (Cambridge, MA: MIT Press, 2019).

# References

Aguilera, J. "Another Word for 'Illegal Alien' at the Library of Congress: Contentious." *The New York Times*, July 22, 2016. Accessed April 25, 2021. https://www.nytimes.com/2016/07/23/us/another-word-for-illegal-alien-at-the-library-of-congress-contentious.html.

Bagga, S. and A. Piper. "Measuring the Effect of Bias in Training Data for Literary Classification." *Proceedings of LaTeCH-CLfL 2020*, 74–84. Barcelona, Spain, December 12, 2020.

Bechmann, A. and G.C. Bowker. "Unsupervised by Any Other Name: Hidden Layers of Knowledge Production in Artificial Intelligence on Social Media." *Big Data & Society* 6, no. 1 (2019): 1–11.

Bingham, N.J. and H. Byrne. "Archival Strategies for Contemporary Collecting in a World of Big Data: Challenges and Opportunities with Curating the UK Web Archive." *Big Data & Society*, January 2021.

Bloom, H. *The Western Canon*. New York: Harcourt Brace, 1994.

Blyum, A. "Hebrew Publications and the Soviet Censor in the 1920s." *East European Jewish Affairs* 23, no. 1 (1993): 91–99.

Bode, K. "Why You Can't Model Away Bias." *Modern Language Quarterly* 81, no. 1 (2020): 95–124.

Bowker, G.C. "Biodiversity Datadiversity." *Social Studies of Science* 30, no. 5 (2000): 643–83.

Bowker, G.C. Foreword. In *All Data Are Local: Thinking Critically in a Data-Driven Society*, edited by Yanni Alexander Loukissas. Cambridge, MA: MIT Press, 2019.

Bowker, G.C. and S.L. Star. *Sorting Things Out: Classification and Its Consequences*. Cambridge, MA: MIT Press, 1999.

Cotter, K., M. Medeiros, C. Pak, and K. Thorson. "'Reach the Right People': The Politics of 'Interests' in Facebook's Classification System for Ad Targeting." *Big Data & Society*, January 2021.

Daru, B.H., D.S. Park, R.B. Primack, C.G. Willis, D.S. Barrington, T.J.S. Whitfeld, T.G. Seidler, et al. "Widespread Sampling Biases in Herbaria Revealed from Large-Scale Digitization." *New Phytologist* 217 (2018): 939–55.

Economou, Maria. "Heritage in the Digital Age." In *A Companion to Heritage Studies*, edited by William Logan, Máiréad Nic Craith, and Ullrich Kockel, 215–28. Chichester, UK: Wiley, 2016.

Eisenberg, Ellen. *Jewish Agricultural Colonies in New Jersey, 1882–1920*. Syracuse, NY: Syracuse University Press, 1995.

Fernandez, M.V. "The Coloniality of Metadata: A Critical Data Analysis of the Archive of Early American Images at the John Carter Brown Library." PhD thesis, University of Texas, 2018.

Frye, N. *The Great Code: The Bible and Literature*. San Diego: Harcourt Brace Jovanovich, 1982.

Harshav, B. and M. Chagall. *Marc Chagall and His Times: A Documentary Narrative*. Palo Alto, CA: Stanford University Press, 2004.

Hauswedell, T., J. Nyhan, M. Beals, M. Terras, and E. Bell. "Of Global Reach Yet of Situated Contexts: An Examination of the Implicit and Explicit Selection Criteria that Shape Digital Archives of Historical Newspapers." *Archival Science* 20 (2020): 139–65.

Iconclass. 2009. Accessed April 25, 2021. http://www.iconclass.nl/home.

Jewish Museum, New York. "The Cairo Geniza: Jews and Muslims in the Mediterranean World 800–1500." January 1, 1997–October 12, 1997.

Jewish Museum, New York. "Russian Jewish Artists in a Century of Change, 1890–1990." September 21, 1995–January 28, 1996.

Kogan, L. and S. Loesov. "Old Jewish Language." In *World Languages: Semitic Languages. The Akkadian Language. Northwest Semitic Languages*, edited by A. Belova, L. Kogan, S. Loesov, and O. Romanova, 296–375. Russian Academy of Sciences. Institute of Linguistics. Moscow: Academia, 2009. [Russian]

Kuncheva, L.I., C.E. Matthews, A. Arnaiz-Gonzalez, and J.J. Rodrıguez. "Feature Selection from High-Dimensional Data with Very Low Sample Size: A Cautionary Tale." August 28, 2020. Accessed May 12, 2021. https://arxiv.org/pdf/2008.12025.pdf.

Likhter, J. Personal communication, May 7, 2021.

Liu, A. "Toward a Diversity Stack: Digital Humanities and Diversity as Technical Problem." *PMLA/Publications of the Modern Language Association of America* 135, no. 1 (2020): 130–51.

Loukissas, Yanni Alexander. *All Data Are Local: Thinking Critically in a Data-Driven Society*. Cambridge, MA: MIT Press, 2019.

Mak, B. "Archaeology of a Digitization." *Journal of the Association for Information Science and Technology* 65, no. 8 (2014): 1515–26.

Metropolitan Museum of Art. "About the Met." 2021. Accessed June 25, 2021. https://www.metmuseum.org/about-the-met.

Metropolitan Museum of Art. "The Colmar Treasure: A Medieval Jewish Legacy". July 22, 2019 – January 12, 2020. Accessed April 30, 2021. https://www.metmuseum.org/exhibitions/listings/2019/colmar-treasure-medieval-jewish-legacy.

Metropolitan Museum of Art. "Explore the Collection." 2021. Accessed June 25, 2021. https://www.metmuseum.org/art/collection.

Metropolitan Museum of Art. "Met Museum Sets New Attendance Record with More Than 7.35 Million Visitors." 2018. Accessed May 12, 2021. https://www.metmuseum.org/press/news/2018/met-museum-sets- new-attendance-record.

Navarette, T. and E. Villaespesa. "Digital Heritage Consumption: The Case of the Metropolitan Museum of Art." *magazén* 1, no. 2 (December 2020).

Pinkus, B. The Jews of the Soviet Union. The History of a National Minority. Cambridge: Cambridge University Press, 1988.

Piper, A. *Can We Be Wrong? The Problem of Textual Evidence in a Time of Data*. Cambridge: Cambridge University Press, 2020.

Prebor, G., Maayan Zhitomirsky-Geffet, and Yitzchak Miller. "A New Analytic Framework for Prediction of Migration Patterns and Locations of Historical Manuscripts Based on Their Script Types." *Digital Scholarship in the Humanities* 35, no. 2 (June 2020): 441–58.

Russian State Catalogue (The State Catalogue of Museum Collections of the Russian Federation). 2021. Accessed June 25, 2021. https://goskatalog.ru/portal/#/.

Said, Edward W. *Orientalism*. New York: Random House, 1978.

Schriml, L.M., M. Chuvochina, N. Davies, et al. "COVID-19 Pandemic Reveals the Peril of Ignoring Metadata Standards." *Scientific Data* 7 (2020): article 188.

Shandler, J. *Adventures in Yiddishland: Postvernacular Language and Culture*. Berkeley: University of California Press, 2006.

Shandler, J. *Yiddish: Biography of a Language*. Oxford: Oxford University Press, 2020.

Slezkine, Yuri. *The Jewish Century*. Princeton, NJ: Princeton University Press, 2002.

Soffer, O., Z. Segal, N. Greidinger, S. Rusinek, and V. Silber-Varod. "Computational Analysis of Historical Hebrew Newspapers: Proof of Concept." *Zutot – Perspectives on Jewish Culture* 17 (2020): 97–110.

Sterner, Beckett W., Nico M. Franz, and J. Witteveen. "Coordinating Dissent as an Alternative to Consensus Classification: Insights from Systematics for Bio-ontologies." *History and Philosophy of the Life Sciences* 42, no. 1 (2020): article 8.

Thylstrup, N.B. *The Politics of Mass Digitization*. Boston: MIT Press, 2019.

University of Wisconsin-Madison Department of Hebrew and Semitic Studies. "The Book and the Spade" (Biblical Archaeological Exhibition). April 13–May 4, 1975.

Waterton, Claire. "From Field to Fantasy: Classifying Nature, Constructing Europe." *Social Studies of Science* 32 (2002): 177–204.

Winer, Dov. "Judaica Europeana: An Infrastructure for Aggregating Jewish Content." *Judaica Librarianship* 18 (2014): 88–115.

Zaagsma, G. "Digital History and the Politics of Digitization." Paper presented at *Digital Archive and Canon Workshop*, March 10, 2021. Accessed May 11, 2021. https://www.digitales-archiv-und-kanon.de/contributions/Zaagsma_en.pdf.

Zhitomirsky-Geffet, M. and Gila Prebor. "SageBook: Toward a Cross-Generational Social Network for the Jewish Sages' Prosopography." *Digital Scholarship in the Humanities* 34, no. 3 (September 2019): 676–95.

Jakub Mlynář, Jiří Kocián, and Karin Hofmeisterová

# How "Tools" Produce "Data": Searching in a Large Digital Corpus of Audiovisual Holocaust Testimonies

**Abstract:** The field of Jewish Studies is facing many new challenges as a result of ongoing digitization. This chapter focuses on digital oral histories of the Holocaust. Following the digital revolution in oral history, many institutions now provide access to multiple collections at once. One of the new challenges is thus related to the simultaneous availability of several archives, as well as various search engines which apply different methods to browse their content. The aim of this chapter is to identify and describe participants' practices for working with a large corpus of audiovisual Holocaust testimonies, especially in terms of locating relevant results within the collection by using three different search systems. We have conducted an empirical study in an experimental setting designed to emulate work with various search engines. Three pairs of novice users solved ten tasks over video-conferencing software, utilizing three different search "tools" (USC Shoah Foundation's Visual History Archive, Amalach, and Pixla). Our main findings consist of formulating a fundamental structure and elements of participants' collaborative work, composed of three complementary actions: testing, sharing, and implementing. Furthermore, users obtained the search results by two main approaches: aggregation and query refinement. Interestingly, they did not upgrade the searching skills progressively, but rather used the current "best knowledge" for all the tasks and search engines at once. The participants' emergent competence was continuously developed on the basis of collaborative work with the search engines and the results obtained so far through their work on the previous tasks.

**Keywords:** Holocaust testimonies, database searching, digital ethnography, oral history, social interaction, video analysis

# 1 Introduction

Research in the field of Jewish Studies is facing a number of new challenges as a result of ongoing digitization.[1] In this chapter, we focus on the specific domain of digital oral histories of the Holocaust.[2] Following the digital revolution in oral history research,[3] many institutions now provide access to several divergent collections at once. One of the new challenges is thus related to the simultaneous availability of multiple archives, as well as various search engines which apply different methods to browse their content. In this context, our chapter aims at providing a methodological and epistemological reflection of the common approach to the qualitative research praxis. This approach consists of using search tools to obtain data that respond to predefined research questions. Nonetheless, in our chapter, we aim to explore how "tools" create "data," how these two notions intertwine in the practical organization of "search" in large digital corpora of audiovisual materials, and how these issues might project onto research design and the formulation of research questions.

Since the onset of the digitization wave at the turn of the millennium, which engulfed archival sources of various kinds, creators of digital collection systems and their respective user interfaces were posing questions on how the digital turn is reflected in the interaction between users and sources. Search for answers is mostly conducted in the methodological and conceptual domain of user studies, a subfield of human-computer interaction research, which takes into account aspects and variables highly relevant for our research as well, such as the diversity of users, their level of expertise, search tools at hand, terminology representing data and tools, and many more.[4] Audio and video recordings of interaction sequences have been a fundamental method for obtaining relevant data for user

---

**2** See, e.g., Cord Pagenstecher, "Testimonies in Digital Environments: Comparing and (De-)Contextualising Interviews with Holocaust Survivor Anita Lasker-Wallfisch," *Oral History* 46, no. 2 (Autumn 2018): 109–18, accessed February 10, 2021, http://www.jstor.org/stable/44993579; Victoria Grace Walden, "What Is 'Virtual Holocaust Memory'?," *Memory Studies*, November 2019, doi:10.1177/1750698019888712.
**3** Alistair Thomson, "Four Paradigm Transformations in Oral History," *The Oral History Review* 34, no. 1 (2007): 49–70, accessed February 10, 2021, http://www.jstor.org/stable/4495417.
**4** Wendy M. Duff, "User Studies in Archives," in *User Studies for Digital Library Development*, ed. Pierluigi Feliciati, Andy O'Dwyer, and Milena Dobreva (London: Facet Publishing, 2012), 199–207.

analysis and in this sense, our chapter can be considered closely related to this field of study as it is informed by the same type of source material.[5] Nevertheless, unlike the more typical approach of user studies, our chapter neither puts prevalent emphasis on the technological dimension of this issue, nor does the opposite – observe the achieved results through an epistemological prism of research in history, oral history, or social sciences. It rather positions itself at an intersection of both domains and seeks to contribute by approaching the topics in question primarily as situated social practices. In the analysis, we reach our research aims by simultaneously confronting novice user pairs with a set of tasks archetypical for the field of digital oral history, and draw on their collaborative work with multiple tools while focusing on the interactional process of reaching the solutions.[6]

This text draws largely from our experience gained as the staff of the Malach Centre for Visual History (CVHM) at the Charles University in Prague.[7] Over the last decade, CVHM has been providing access for students, researchers, and the general public to several established collections of oral history interviews. Since 2009, CVHM has been an access point to the University of Southern California Shoah Foundation's *Visual History Archive* (VHA), which is an ever-growing collection of interviews with witnesses and survivors of genocides, especially the Holocaust. At the present moment, the VHA contains almost 56,000 audiovisual recordings of oral history interviews in more than 40 languages. Since 2018, the *Fortunoff Video Archive for Holocaust Testimonies* of the Yale University Library with more than 4,400 audiovisual recordings of oral history interviews is also available at CVHM. In addition, users in CVHM can work with smaller collections lacking an integrated user interface such as the *Refugee Voices* archive (150 English interviews), and a small portion of interviews from the Melbourne Holocaust Museum, formerly known as the Jewish Holocaust Center in Melbourne (15 interviews with people of Czechoslovak origin). One of our tasks as employees of the CVHM is therefore to

---

**5** For instance Joyce C. Chapman, "Observing Users: An Empirical Analysis of User Interaction with Online Finding Aids," *Journal of Archival Organization* 8, no. 1 (2010): 4–30, accessed February 10, 2021, doi:10.1080/15332748.2010.484361.

**6** Numerous studies present a structurally similar design, but their main interest relies in large part on the evaluation of finding correct solutions. Instead, the process of the solution method negotiation itself is utterly central to us. For instance: Sadegh Kharazmi, Sarvnaz Karimi, Falk Scholer, and Adam Clark, "A Study of Querying Behaviour of Expert and Non-expert Users of Biomedical Search Systems," in *Proceedings of the 2014 Australasian Document Computing Symposium* (ADCS '14), Association for Computing Machinery, New York, NY, USA, 10–17, doi:10.1145/2682862.2682871.

**7** See Jakub Mlynář, "Malach Center for Visual History," in *Sborník Semináře o digitálních zdrojích a službách ve společenských a humanitních vědách (WDH 2015)*, ed. Jaroslava Hlaváčová (Prague: Charles University, 2015), 83–89; Jiří Kocián, Jakub Mlynář, and Petra Hoffmannová, eds., *Malach Center for Visual History on Its 10th Anniversary* (Prague: Matfyzpress, 2020).

assist and advise researchers in their pursuit of audiovisual materials relevant to their interests.

In this chapter, we examine some characteristic problems that emerge during work with large digital archives by focusing on the example of the Czech-language subsection of the VHA. We first provide a background and rationale for our efforts (in Section 2), problematizing the common-sense link between the "tools" and the "data." We then move in Section 3 to the description of an experiment which was conducted to make visible some of the users' intrinsic practices in working with the database systems available at CVHM. Analysis of the video recorded experimental sessions yielded several main findings, which we present in Section 4. In conclusion, we discuss the findings in a broader context and address the question of how they relate to the use of digital oral history resources such as the VHA in Holocaust research.

## 2 Background and Rationale

Rather frequently, current research praxis in digital environments is conceived in terms of using "tools" upon "data." For example, researchers use search systems ("tools") that allow them to identify relevant units in a corpus of materials ("data"). In the case of the materials available at CVHM, incoming researchers as "our users" ultimately expect to watch interviews (or segments of interviews) that are related to "their research topics." In this sense, for searching within the contents of the VHA, researchers can use several search systems ("tools"). (1) *The integral VHA search systems*: People Search (approx. 1 million personal names), Index Search (around 67,000 hierarchically ordered keywords), Biographical Search (date of birth, place of birth, experience, etc.), Places Search (utilizing indexing terms with Google Maps), and Quick Search (combining all of the above). (2) *Amalach search*: a phonetic fulltext search engine created at the University of West Bohemia (Pilsen, Czechia).[8] Amalach has been available at CVHM in beta-testing since 2012, with many new versions introduced since then, which also incorporate comments and suggestions from the CVHM visitors and staff. (3) *Pixla search*: A phonetic fulltext search similar to Amalach, but voice-controlled, developed also at the University of West Bohemia.[9] Pixla has

---

**8** Jan Švec et al., "On the Use of Grapheme Models for Searching in Large Spoken Archives," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB*, 2018, 6259–63, doi:10.1109/ICASSP.2018.8461774.

**9** Adam Chýlek, Luboš Šmídl, and Jan Švec, "Question-Answering Dialog System for Large Audiovisual Archives," in *Text, Speech, and Dialogue: TSD 2019*, ed. Kamil Ekštein, 385–97 (Cham:

been available at CVHM since May 2020 for user testing, which was, however, heavily hindered by the COVID-19 pandemic.

Despite the rich variety and admirable effectiveness of these search systems, the awareness of their mutual differences in generating the sets of possibly relevant interviews is crucial for a successful implementation in research. A reasonable common-sense presupposition of a researcher-user would be that all these research tools allow users to search within "the same data" – in our case, the complete corpus of 558 interviews in the Czech language. However, we argue that in practice, the "tools" effectively produce the "data." As Ørmen puts it, the "search results are made in the act of searching."[10] Not only that each search system requires very different search mindsets at the input, but the plurality of three fundamentally diverse search tools renders it nearly impossible to arrive at "the same results" by using either of them. Using a "tool" therefore requires a fine-tuned way of formulating the search query, and the results provided by the search engine are ontologically framed by the boundaries of this formulation. Ultimately, the users' knowledge of the search engines and resulting "datasets," gained hermeneutically through numerous iterations of processing their research requests, also projects onto the way in which they pose their research questions and assess the feasibility of related research designs. At the center of this chapter, therefore, we put the users' situated practices rather than the technically intended features of the "tools," following the apt advice given by Egon Bittner to social scientists already in 1965: "It seems reasonable that if one were to investigate the meaning and typical use of some tool, one would not want to be confined to what the toolmaker has in mind."[11]

## 3 Experiment

To illustrate our point and inspect our assumptions, we focused on three types of research topics at three different levels of concreteness. Accordingly, we designed a typology of research questions characteristic to the domains of Jewish and Holo-

Springer, 2019); Adam Chýlek, Luboš Šmídl, and Jan Švec, "Multimodal Dialog with the MALACH Audiovisual Archive," in *Proceedings from Interspeech 2019*, 3663–64, accessed February 10, 2021, doi:10.21437/Interspeech.2019.

**10** Jacob Ørmen, "Googling the News: Opportunities and Challenges in Studying News Events through Google Search," *Digital Journalism* 4, no. 1 (2016): 107–24, accessed February 10, 2021, doi:10.1080/21670811.2015.1093272.

**11** Egon Bittner, "The Concept of Organization," *Social Research* 32, no. 3 (Autumn 1965): 249, accessed February 11, 2021, https://www.jstor.org/stable/40969788.

caust studies based on which we formulated a set of ten specific tasks that served as a guiding framework for the observed experimental interaction (see Table 1). The tasks were provided to the experiment participants in Czech language as an online form which also included blank fields to fill in the results of their work (names of narrators).

**Table 1:** Search tasks overview.

| | | |
|---|---|---|
| Geographical terms | Reaching from supra-localized reference points such as buildings, geographically located institutions or street names to macro-level concepts | Q1: Find narrators who mention Pankrac prison |
| | | Q2: Find narrators who mention Vinohradská street |
| | | Q3: Find narrators born in the territory of interwar Czechoslovakia |
| | | Q4: Find narrators born in Carpathian Ruthenia between 1919 and 1939 |
| Paralinguistic phenomena | Including functional components of oral and visual history (nonverbal cues, visual demonstrations, emotions, sounds) | Q1: Find narrators who interact with their relatives (during the interview) |
| | | Q2: Find narrators who show their tattoos (during the interview) |
| | | Q3: Find narrators who show military decorations (during the interview) |
| | | Q4: Find narrators who show Jewish religious objects (during the interview) |
| Abstract concepts | Having an implicit or explicit verbal representation, such as identity (a typical relevant topic in human and social sciences) | Q1: Find narrators who mention transformation of their religious identity |
| | | Q2: Find narrators who mention the loss of their identity |

In order to emulate and uncover the fundamental user practices in solving qualitatively different types of research questions with various search engines, we have conducted three experimental sessions with six novice users (university students). Reflecting on the intersection of disciplines in Digital Humanities, we selected participants of the experiment based on their educational background. Accordingly, three of them were from the IT sphere while the others were from humanities and social sciences. They were working in pairs through the video-conferencing platform ZOOM. They had approximately 40 minutes to explore the "tools," deal with the experimental tasks, and complete the form with "data"

considered relevant. For analytical purposes, we treated each pair as a collective actor operative as an element in the correlative research triad: user – tool – search query.

# 4 Findings

The recordings of the video-mediated interactions were analyzed through the perspective of qualitative sociological analysis and multimodal interaction analysis. A methodological note is in place here. We are aware of the fact that the experiment setting and available corpus of recordings could be possibly seen as insufficient from a cognitivist point of view, because we purposely did not obtain access to the participants' individual work with their computers (via screen-capture apps, eye-tracking, etc.[12]). However, our analytical approach is grounded in a praxeological point of view and in naturalistic video-based studies of human sociality,[13] exemplified in earlier studies of video-mediated interaction.[14] We aim to describe the participants' own practices through which they methodically and obviously achieve the completion of their tasks. Thus, we record and analyze those aspects of the video-mediated interaction that are observably consequential for the participants in their collaborative work as a pair. In short, as the participants manage to do their assignments without the need of accessing each other's private on-screen conduct or locating the precise position of their interlocutor's on-screen gaze, we should also be able to do without it in our analyses. Everything that the participants themselves need is already there. The subject matter of our research is the *witnessable* social order,[15] and we take into account what the members of the pair themselves observably orient to.

---

**12** Cf. Robert J. Moore, "A Name Is Worth a Thousand Pictures: Referential Practice in Human Interactions with Internet Search Engines," in *Mobile Speech and Advanced Natural Language Solutions*, ed. Amy Neustein and Judith A. Markowitz (New York: Springer, 2013), 259–86; Robert J. Moore and Elizabeth F. Churchill, "Computer Interaction Analysis: Toward an Empirical Approach to Understanding User Practice and Eye Gaze in GUI-based Interaction," *Computer Supported Cooperative Work* 20, no. 497 (2011): 497–528, accessed February 10, 2021, doi:10.1007/s10606-011-9142-2.
**13** See, e.g., Charles Goodwin, *Co-Operative Action* (Cambridge and New York: Cambridge University Press, 2018).
**14** A review is provided by Jakub Mlynář, Esther González-Martínez, and Denis Lalanne, "Situated Organization of Video-Mediated Interaction: A Review of Ethnomethodological and Conversation Analytic Studies," *Interacting with Computers* 30, no. 2 (2018): 73–84, accessed February 10, 2021, doi:10.1093/iwc/iwx019.
**15** Harold Garfinkel, *Ethnomethodology's Program: Working Out Durkheim's Aphorism* (Lanham, MD: Rowman & Littlefield, 2002); Eric Livingston, "Context and Details in Studies of the Wit-

Through our research, we have identified three basic sequential practices of collaborative work in the experimental setting (*testing*, *sharing*, *implementing*). Furthermore, the participants seem to employ either *aggregation* or *refinement* as two general strategies for obtaining relevant results. Amongst most participants, we observed a tendency towards establishing universal solutions utilizable for a larger number of tasks. "Tools" seem to effectively produce "data" through the practice of *querying*, which consists of breaking down the task (question at hand) into searchable units: either through "keywording" (transforming the question into possible keywords in metadata) or through "discoursing" (transforming the question into possible phrases in the speech) – the former dominating.

We should state from the outset that we are conscious of the limits of our study, which is primarily intended as explorational. The experimental setting is a very specific (and indeed unusual) situation. Some of the practices described below could therefore be a residue of the experimental design. For instance, if participants would not have access to all ten questions for the whole duration of their work, their methods of task solution could develop in a quite different way. This is a conjecture that can only be evaluated by conducting further empirical studies where these specific conditions are modified. With this in mind, more research is needed to confirm and elaborate our findings. In the future, we plan to conduct follow-up experiments, this time also providing instructions to the participants and observing any possible changes in their search practices. Nevertheless, we believe that the results presented below have merit and can serve as a useful point of departure for further work. Concurrently, we hope to inspire other researchers to conduct similar studies in both experimental and – perhaps more importantly – naturalistic everyday settings.

In the following subsections, we present the main observations and findings from our analysis of the recordings of the experimental sessions. We describe and conceptualize the participants' practical approach to searching and solving the task(s) by employing the available search systems. First, we focus on the practice of testing, sharing, and implementing in the course of searching. Then we move to aggregation and refinement as two typical general approaches to search querying. Third, we describe how participants transform the experimental search tasks into searchables by "keywording" and "discoursing."

nessable Social Order: Puzzles, Maps, Checkers, and Geometry," *Journal of Pragmatics* 40, no. 5 (2008): 840–62, accessed February 10, 2021, doi:10.1016/j.pragma.2007.09.009.

## 4.1 Testing, Sharing, Implementing

As a basic structure within the sequential development of the participants' collaborative work, we have identified the triad of *testing*, *sharing*, and *implementing*. Testing consisted of experimenting with the search systems and trying them out, typically in a solitary manner. Because of the predominantly individualistic nature of testing, the next phase of sharing has to do with intersubjective orientation, and learning from each other about the knowledge that emerged from the separate testing. The next synthetical step is implementing, which has to do with a specific use of the search systems for different experimental tasks. Abstraction and generalization of findings, in terms of practical procedures for working with the search systems and their relation to the displayed set of results, happened quite often within this step as well.

Our initial understanding of testing, sharing, and implementing was to conceive of them as three subsequent steps or phases in the temporal structure of collaborative work with the search engines. Progressively, through more refined analysis, we have arrived at a dynamic understanding of these concepts as labels for mutually interdependent work practices which are recurrently combined throughout the session. We believe that the second conception is more useful and closer to the reality of the users' actual work with the digital archives under scrutiny. However, the labels "testing," "sharing," and "implementing" remain approximate glosses, which only serve to underscore certain aspects of the participants' work and provide a general framework for it. In praxis, they consist of various verbal and nonverbal practices, including not only talk-in-interaction[16] and embodied action[17] but also the observable work with the software interfaces (such as demonstrations on a shared screen). In the following two subsections, we will describe and illustrate some of the more nuanced practices as components of the setting-specific actions of testing, sharing, and implementing.

## 4.2 Aggregation and Refinement

This subsection describes and illustrates the two main approaches to task solution, identified in our analysis of the video recorded experiments. The first method used by the participants is *aggregation*. This approach seems to aim at generat-

---

**16** Emanuel A. Schegloff, *Sequence Organization in Interaction: A Primer in Conversation Analysis* (Cambridge: Cambridge University Press, 2007).

**17** Christian Meyer, Jürgen Streeck, and J. Scott Jordan, eds., *Intercorporeality: Emerging Socialities in Interaction* (Oxford: Oxford University Press, 2017).

ing "any" results, even with a low level of relevance – i.e., typically yielding high numbers of interviews. Obtained results are then manually sorted and some of them selected as "proper" results which are reasonably answering the question at hand. As an example, consider Extract 1 below, in which Participant 1 (P1) and Participant 2 (P2) work together on the question "Find narrators born in the territory of interwar Czechoslovakia," already more than 30 minutes into the session. P2 has been sharing her screen throughout the whole session. They decided to use the VHA system upon which P1 commented that "there you can search for the years of birth." According to P1's suggestion, P2 types "1918–1939" in the Quick Search field and then they use a suggested search query. The excerpt begins when they have just clicked "Search" and are waiting for the results.[18]

*Extract 1: First group / 34:05–34:40*

(1)  ((results appear, see Figure 1))
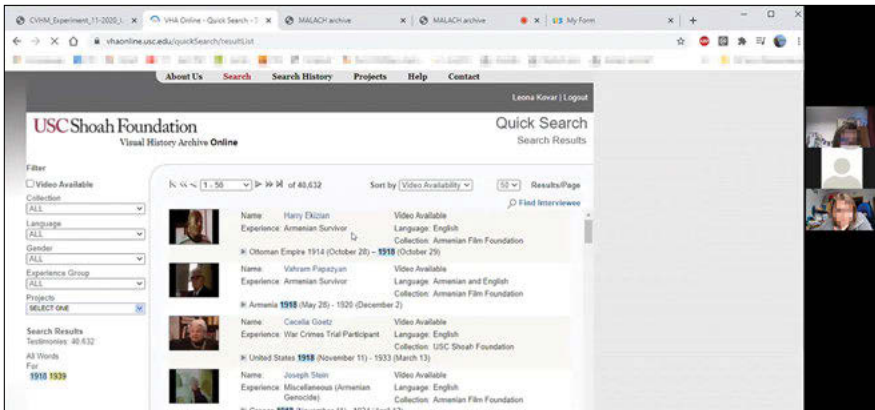


**Figure 1:** Shared screen after the search results for "1918–1939" appeared. Webpage: http://vhaonline.usc.edu, date: 27 November 2020.

---

**18** Screenshots displayed below as Figures 1–5 and 7 are not illustrative, but come from the obtained video recordings and constitute our research data as documents of social interaction. In our case, the meeting in the video conferencing platform, including the use of shared screen, is the work environment used by the participants in our study. The figures show the on-screen appearance of the particular moment of the interaction. This accounts for the slightly impaired resolution of the webpages, as they have been shared in real time during the video call. The participants' faces and names have been anonymized.

(2)        ((1 second pause))
(3)  **P2:  Well …**
(4)  **P1:  Try to use the 'collection,' what is in that 'collection'?**
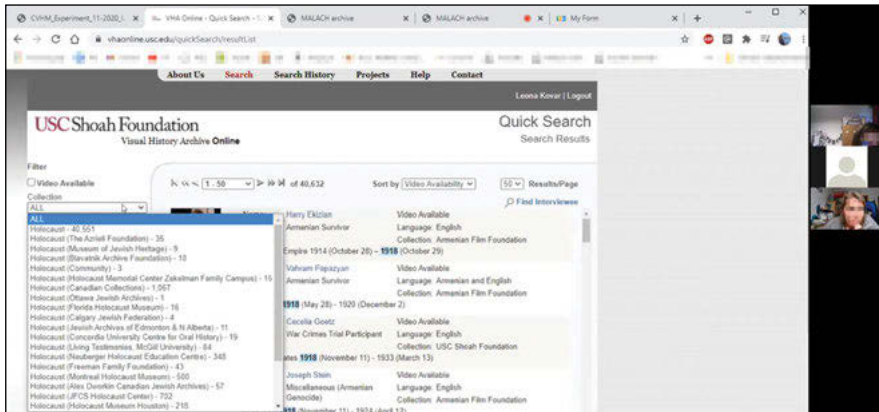(5)  **P2:** ((clicks on the *Collection* menu, see Figure 2))



**Figure 2:** Shared screen after P2 clicked on the Collection menu. Webpage: http://vhaonline. usc.edu, date: 27 November 2020.

               **Well … Probably not. Probably no.**
(6)  **P1:  Uuuuhhhh …** ((silently reading aloud)) **Museum of Jewish …**
(7)  **P2:** ((closes *Collection* menu, clicks on *Language* menu))
(8)  **P1:  Language … Yeah set it to Czech language just so we see.**
(9)  **P2:** ((clicks on *Czech*, results are loading))
(10)        ((0.8 second pause))
(11) **P1:  Aah, I hope that this could filter those …** ((looks at her second screen))
(12)        ((results appear on P2's shared screen, see Figure 3))
(13) **P2:  Hm …** ((scrolls down))
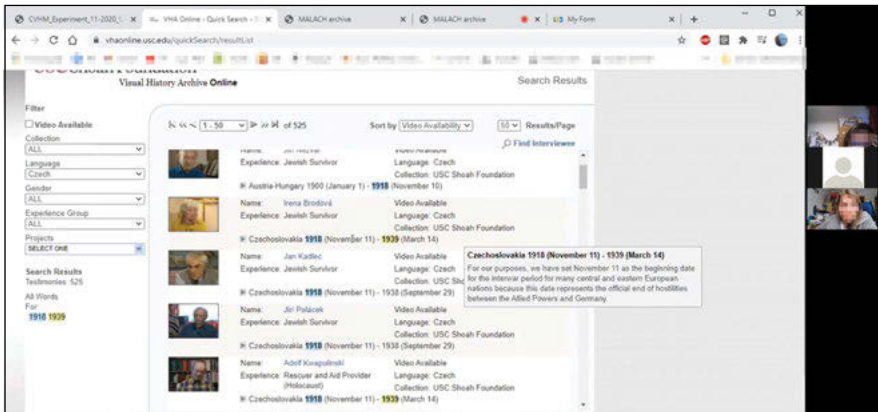
**Figure 3:** Shared screen after the search results filtered by language appeared. Webpage: http://vhaonline.usc.edu, date: 27 November 2020.

(14) **P1:** ((looks back on the shared screen)) **Yes. Great.**

(15) **P2: Yes?**

(16) **P1: And that's … OK. So.** ((starts writing a name into the form)) **Irena Brodová …**

(17) **P2: Next one Jan Kadlec …**

(18) **P1:** ((types on her other computer)) **Mhmm.**

As we can see, although P2 shares her screen, she is depending – at least in this sequence – mostly on guidance and advice from P1 (lines 4 and 8). Furthermore, the crucial moment when the results on screen are determined to be proper answers which can be considered solutions to their task, is also decided by P1 (line 14). The procedure started with typing a range of years ("1918–1939"), which generated a set of results displayed in Figure 1. These results are not treated as adequate (lines 3 and 4), as they visibly are not related to Czechoslovakia. Thus the next step consists of finding a way to "filter out" the interviews related to Czechoslovakia, which is done first by an attempt to use the "Collections" filter (see Figure 2) and later by setting the language of interview to Czech (lines 8–11). The results that appear thereafter (see Figure 3) are treated as satisfactory by P1, who produces a "jubilatory 'yes'"[19] and a positive assessment of the results ("great") in line 14. Then she moves to writing down the displayed results in the online form which is open on her second computer screen. She reads the first

---

**19** Philippe Sormani, "The Jubilatory YES! On the Instant Appraisal of an Experimental Finding," *Ethnographic Studies* 12 (2011): 59–77, accessed February 10, 2021, doi:10.5449/idslu-001104716.

name aloud for herself; but perhaps not just for herself, as P2 picks up this practice and reads out the following name in the list (line 17), providing P1 with a next writable,[20] which is confirmed and typed by P1 in line 18. After another 30 seconds (not included in the transcript) and three more names written in the form they decide that "perhaps this is enough."

Along with aggregation, participants have also used query refinement as the second approach to obtaining search results. Its aim is to make the search as specific as possible, in order to receive a low number of very relevant interviews, which could be directly copied to the form as "proper" results answering the question. An example is provided in Extract 2, which shows Participant 3 (P3) and Participant 4 (P4) working on the task "Find narrators who mention transformation of their religious identity." P3, who is sharing her screen, had used Pixla to search for a *textual* query "religious identity" (although Pixla is intended as a *voice* interface). As the excerpt begins, they are discussing the two results obtained (see Figure 4).

*Extract 2: Second group / 20:00–20:40*
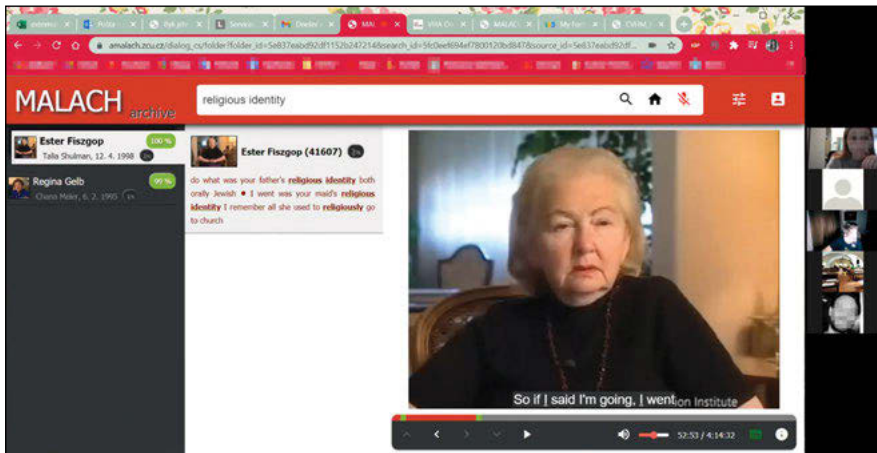


**Figure 4:** Shared screen with the search results of a textual query "religious identity" in Pixla. Webpage: http://amalach.zcu.cz, date: 27 November 2020.

---

**20** Lorenza Mondada, "Going to Write: Embodied Trajectories of Writing of Collective Proposals in Grassroots Democracy Meetings," *Language and Dialogue* 6, no. 1 (2016): 140–78, accessed February 10, 2021, doi:10.1075/ld.6.1.05mon.

(1)  **P4: Just I think that in this task it's something else well, ehm, that ...** ((P3 changes the tab to the online form; P4 reads out part of the task)) **transformation of *their* religious identity, so I think those that changed their religion, or ... their beliefs.**

(2)  **P3: Yeah?**

(3)  **P4: So I'd rather formulate the query like uhh, 'change religion'** ((said in English)). **Like change of religion rather than religious identity.**

(4)  **P3: I have put it in this formulation into another one but ... There it has ... It didn't find much** ((types 'change' in the search field)) **but it found 'change politics'** ((said in English))**, so maybe this ...**

(5)      ((4 second pause, P3 types 'identity' in the search field))

(6)  **P3: Yes.** ((submits query, results start loading))

(7)  **P3: Into this ...** ((switches tab to VHA, see Figure 5))
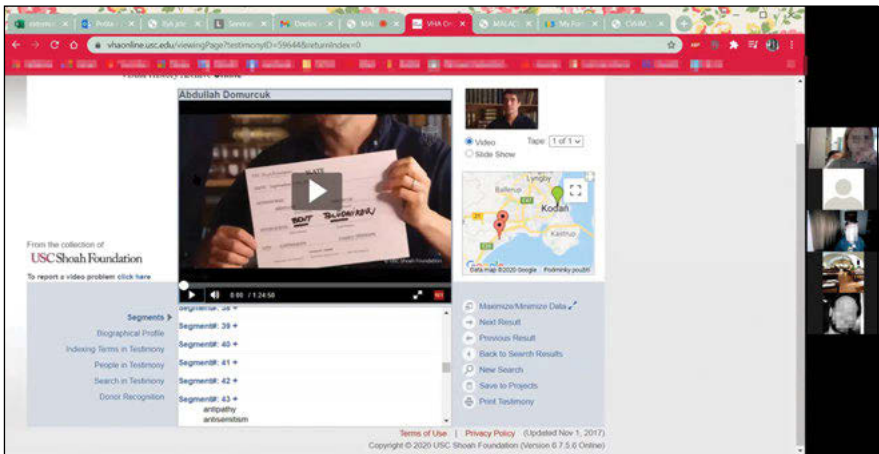


**Figure 5:** Shared screen after P3 switches from Pixla to VHA Webpage: http://vhaonline.usc.edu, date: 27 November 2020.

(8)  **P3: I have put it into this one.**

(9)  **P4: Mhm.**

In line 1, P4 mentions the task formulation, and after he produces several hesitation markers, P3 switches the tab in her browser to the form with the tasks. Now P4 uses the on-screen text as a resource and reads aloud the second part of the task. Thereafter he suggests that they should reformulate the search query in a way which would be in a better accord with the task: "change religion" (line 3). In line 4, P3 responds that she already tried that earlier in the VHA system, but

she agrees that they can also try here in Pixla, and in lines 5 and 6 she types and submits the query, providing online commentary on her ongoing activity (lines 6 and 7). Next, while they are waiting for the results to appear (lines 7 to 9), she switches the tab again, this time to the VHA interface (see Figure 5), where she then proceeds to show P4 the results of her previous attempt (not displayed in the transcript). Contrary to Excerpt 1, P3 and P4 are reformulating and specifying the search query rather than handsorting the relevant results, possibly just because the number of obtained results is too low and not responding to the task (see Figure 4). Note that in lines 3 and 4, both participants resort to code-switching – i.e., alternation between two languages in the course of a single interactional sequence.[21] Here, rather than switching between two languages for the sake of mutual understanding or expressing oneself, the language choice is a "significant aspect of talk organization"[22] in a different sense: English is used because it appears to be taken as the language of the search system. The participants in this strip of interaction seem to operate with the assumption that the search query must be written in English. Therefore, they formulate the query in their talk precisely as it should be typed in – i.e., they use words from the English language.[23]

## 4.3 Querying: Keywording and Discoursing

After describing the findings on the interactional practices of collaborative work which emerged in the observed experimental settings, in this section, we will focus in some detail on the relation of the work praxis to the software search systems. Responding to our central question posed in the title of the chapter, we proceed to the argument that "tools" produce "data" through the process of querying (see Figure 6). This consists of breaking down the question into searchable units by way of two practices: "keywording" and "discoursing." They were not

---

**21** See, e.g., Monica Heller, ed., *Codeswitching: Anthropological and Sociolinguistic Perspectives* (Berlin: Mouton de Gruyter, 1988).

**22** Joseph Gafaranga, "Language Choice as a Significant Aspect of Talk Organization: The Orderliness of Language Alternation," *Text* 19, no. 2 (1999): 201–225, accessed February 11, 2021, doi:10.1515/text.1.1999.19.2.201.

**23** It can be noted that although the participants seem to employ this assumption in their work, it is not quite correct, because the search system could also process queries in Czech, and in fact it would be the right approach as their overall task was to find interviews in Czech language. However, our chapter did not set out as an evaluative undertaking, and we aim at describing and explicating the participants' action rather than assessing it.

equally present as practical methods in the recordings of the three user pairs, but participants appeared to orient to them and verbalize aspects of them. We will describe and illustrate these practices in the following paragraphs.
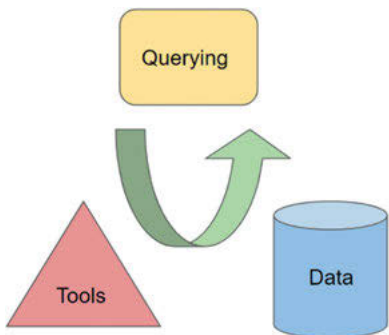


**Figure 6:** Querying as the interplay of "tools" and "data." Graph by the authors.

The practice of "keywording" – turning the task/question into possible keywords in archival metadata – is illustrated by Extract 3. We encounter the same pair as in Extract 1, but this time they are at the very beginning of the session, discussing how to organize their collaborative work. Before the excerpt begins, P1 agrees that she will be sharing her screen and P2 informs that she has the online form open on her second screen.

*Extract 3: First group / 4:07–4:30*

(1) **P1:** ((switches tabs in her browser)) **Hm hm hmm. Yep.**
(2) **P2: I would start with some ... Some task which looks like, maybe the last one ...**
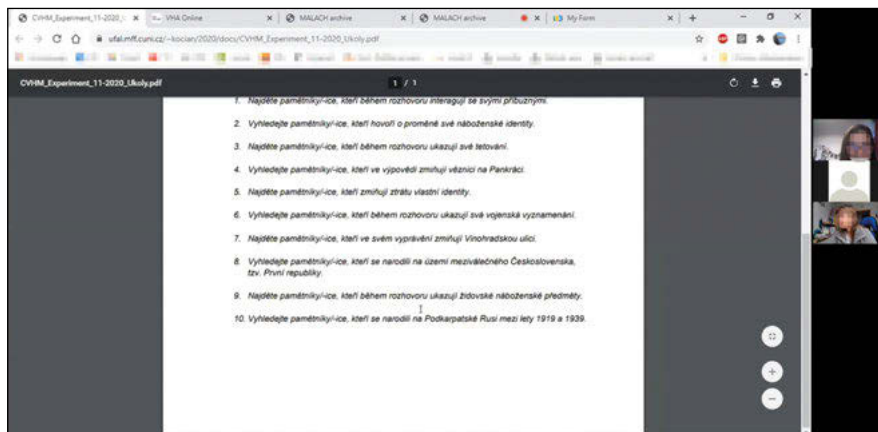(3) **P1:** ((switches to VHA login screen, then to her mailbox, and to the list of the tasks, see Figure 7))

**Figure 7:** Shared screen with the list of the tasks sent in PDF to the participants for their collaborative work.

(4)   **P2:**   **That looks like there are enough keywords which could be used for the search.**

(5)   **P1:**   **Yeah.** ((0.5 second pause)) **Alright so ... Should we try it?**

(6)   **P2:**   **There you search like ...** ((P2 looks at her second screen; P1 points with the cursor to 'Podkarpatské Rusi' [Carpethian Ruthenia] in the text)) **Well, ehm, 'Podkarpatská Rus' and birth ...**

(7)   **P1:**   ((switches to tab with VHA))

While P1 prepares her screen and produces a conventionalized melodic triad of fillers or placeholders[24] which indicates waiting (line 1), P2 turns to her second screen where the online form with the list of tasks is displayed. In line 2, she suggests that they should start with a specific type of question – one that "looks like" something – but doesn't finish the phrase and rather changes to a more concrete designation ("the last one"). P2 seems to take this utterance as an instruction to look at the last question in the list, as she switches tabs to the PDF with the tasks (which the participants received by e-mail just before the experiment; see Figure 7), locating the last question: "Find narrators born in Carpathian Ruthenia between 1919 and 1939." Meanwhile, having already specified the exemplar instance of a more general question type, P1 repeats "that looks like" in line 4 and then makes explicitly relevant the use of keywords as a search method. P1 aligns in

---

**24** Nino Amiridze, Boyd Davis, and Margaret Maclagan, eds., *Fillers, Pauses and Placeholders* (Amsterdam and Philadelphia: John Benjamins, 2010).

line 5 ("Yeah"), holding her cursor under the word "Podkarpatská" ("Carpathian") on her shared screen, and then suggests that they could open one of the search systems to "try it." P2 does not align but continues her explanation, describing very precisely how the process of "keywording" actually works: the question is transformed into two searchable items, "Carpathian Ruthenia" ("Podkarpatská Rus" in Czech) and "birth" ("narození" in Czech). Note that none of these terms are present as such in the question: P2 lemmatizes "in Carpathian Ruthenia" ("na Podkarpatské Rusi") and instead of using "born in" she suggests the noun "birth."

The alternative practice of "discoursing" – turning the task/question into possible phrases in the archived speech – is illustrated by Extract 4. Twenty minutes into the session, the pair consisting of Participant 5 (P5) and Participant 6 (P6) works on the question "Find narrators who show their tattoos (during the interview)." This pair was the only one that decided not to share one of the participants' screens for the whole duration of the experimental session (although they did share it occasionally in an *ad hoc* manner). Just before the excerpt starts, P5 summarizes that they have done three questions out of ten, being around 15 minutes into the job.

*Extract 4: Third group / 19:52–20:58*

(1) **P5: Well, now I think we need to find out how to search for those that are about showing something. Because that's, apart from the Carpathian Ruthenia, those are all the remaining questions.** ((Laughs))

(2) **P6:** ((Laughs)) **So according to what is going on in the video?**

(3) **P5: (What is) going on in the interview, here, right** ((starts reading from the list of tasks)) **– interact with their relatives, show their tattoo, show their military decorations, show Jewish religious objects.**

(4) ((1.2 second pause))

(5) **P5: And then** ((continues reading aloud)) **mention the loss of their identity, mention transformation of their religious identity, and then there is the Carpathian Ruthenia.**

(6) **P6: Mhm. So there must be ... For this there must be some special tool. Right, probably?**

(7) **P5: We must somehow find out how to use it.**
((2 second pause, two hearable clicks from P6))

(8) **P5: So I will try ... I'll try that Amalach, and I'll try for instance tattoo, just like ... Like 'look at my tattoo'?**

(9) **P6:** ((Laughs)) **'Look at my tattoo!' Heh heh heh. 'Watch this!' Heh heh heh heh.**

(10) **P5:** ((Laughs))

In line 1, P5 states that the remaining questions are questions of a certain type – they have to do with "showing something." P6 joins her in laughter and asks a follow-up question about the nature of this question type, which P5 confirms and specifies by reading out aloud all relevant parts of the questions that they still need to do (in lines 3 and 5). Note how precisely her categories are overlapping our own typology displayed in Table 1. First, in line 3, she lists the questions that include a visible feature of the narrator or their environment. After a pause she lists a next category of questions, which have to do with identity, and then she mentions the question about Carpathian Ruthenia, which does not fit either of those categories, but they have decided earlier (after spending some time with attempting a search) that they will put this question aside for later. P6 responds by producing first an acknowledgment token[25] and then a formulation of the gist[26] of P5's previous turns (line 6), which specifies that there must be a "special tool" for these categories of questions. P5, perhaps in a corrective manner, responds that they must "find out how to use it" – i.e., they already have the tool but they need to acquire the competence to use it efficiently. After 2 seconds of silence she suggests that she will use Amalach and type in an imagined speech phrase which could possibly accompany a video-recorded scene of someone showing a tattoo (line 8). Although her suggestion is then treated (first by P6 and then also by P5 herself) as laughable, even somewhat ironicized by P6 ("Watch this!" in line 9),[27] and we are indeed not sure whether she has typed the phrase into the search field (her screen is not shared and she doesn't account for it), this sequence clearly shows participants' orientation to the practice of "discoursing" as a form of querying. The fact that it is treated as laughable might indicate that it is more unusual in comparison to the practice of "keywording," which is utilized much more routinely (see Extract 3).

It seems that in the VHA, perhaps given the abstraction required to produce some results via search query, search results obtained by way of "keywording" are treated as less "certain" and require further checking. On the other hand, Amalach (and Pixla), in its "concreteness," provides results through "discoursing" practices with higher certainty which can also be checked more easily (and

---

**25** Gail Jefferson, "Notes on a Systematic Deployment of the Acknowledgement Tokens 'Yeah' and 'Mm hm'," *Papers in Linguistics* 17 (1984): 197–206, accessed February 12, 2021, doi:10.1080/08351818409389201.

**26** John Heritage and D. Rod Watson, "Formulations as Conversational Objects," in *Everyday Language: Studies in Ethnomethodology*, ed. George Psathas (New York and London: Irvington, 1979), 123–62.

**27** Cf. Elizabeth Holt, "On the Nature of 'Laughables': Laughter as a Response to Overdone Figurative Phrases," *Pragmatics* 21, no. 3 (2011): 393–410.

are less opaque). Such predilection for keywording or discoursing contingent upon particular search systems seems true, however, only to a certain extent. In the analyzed interactions, the participants often used the same querying practices independently from the type of search engine. For example, querying by keywording in VHA generated not only a set of resulting interviews, but also offered new keywords that came up as results of the first keyword search. These newly "discovered" keywords were then typed into other search systems such as Amalach – i.e., used as resources in the "discoursing" practice.

# 5  Concluding Remarks

The aim of this chapter was to identify and describe participants' practices for working with a large corpus of audiovisual Holocaust testimonies, especially in terms of locating relevant results within the collection by using three different search systems. We have started from the assertion that various search devices and systems, rather than working as non-problematical "tools," dynamically produce and practically construe what can be conceived as "data." Instead of formulating further insights on a theoretical or conceptual basis, we have tackled our subject matter through a low-scale empirical study. It consisted of an experimental setting where three pairs of novice users solved ten tasks over video-conferencing software, utilizing three different search "tools" (VHA, Amalach, and Pixla). The experiment was designed to emulate work with various search engines, such as that of the researchers working in the Malach Center for Visual History.

Our main findings presented in this chapter consist of formulating a fundamental structure and elements of participants' collaborative work, which appears to be composed of three complementary actions: *testing*, *sharing*, and *implementing*. The tasks were solved, and relevant results obtained, by two main approaches: *aggregation* and *query refinement*. Each singular search act is taken as an instance or an example of some – thus far unknown (to the users) – general features of the systems, which are to be discovered and identified. The systems are then discussed regarding their utility for solving questions of "a certain kind" (cf. Table 1). In order to conduct searching, the participants seek to transform the textual task into a working query that returns a set of relevant results. They do this mostly by *keywording* (turning the question/task into a set of keywords), and much less often by *discoursing* (turning the question/task into a possible expression in natural language). This might be partially caused by the force of habit as searching keywords is the prevalent practice when using database systems in

contrast to direct interaction with such a system in natural language, though the Pixla system is specifically designed for this purpose. However, it became apparent that participants were able to differentiate the effectiveness of both practices in different environments. The practice of "keywording" seems to stimulate abstraction and is used more often in regard to the VHA search interface, while the practice of "discoursing" stimulates concretization and is used more often in regard to Amalach and Pixla systems. All the concepts as presented in this chapter were drawn inductively and illustrated by examples from the empirical materials, aiming to capture the dynamic process of development and utilization of progressively improving shared knowledge of the workings of the search systems.

Nevertheless, we have realized that the participants surprisingly did not upgrade their searching skills progressively step by step. Their emergent knowledge of the search systems rather seemed to be implemented hermeneutically, again and again, for the whole set of experimental tasks, leading to an improvement of the whole list of answers – each time a new attribute of the search systems had been discovered. Most of the time, they used the current "best knowledge" for all the tasks and all search engines at once. The systems and the tasks were taken for all practical purposes not as separate entities, but as parts of a whole. Also, the "best knowledge" did not alter depending on the qualitative differences between the search questions. Accordingly, the "best knowledge" has universalizing tendencies: the participants aim to establish practices which are utilizable for solving multiple tasks. Still, such knowledge is not a static entity, but it is continuously improved on the basis of participants' collaborative work with the search engines and the results obtained so far, in and through their work on the previous tasks.

One of the important findings of our study points to what we call the "googling paradigm," indicating thus the user's orientation to the search process as not requiring a knowledge of the search tool's inner workings.[28] The practices of breaking down questions into searchable queries (keywords or discursive units) establish the horizon of the materials to be searched. Furthermore, it also seems to structure the participants' practical engagement with the user interfaces – e.g., in the case of VHA, during the time dedicated to the experiment, our participants (as untrained novice users) very rarely moved beyond the simple "Quick Search," which is visually highlighted and designed in a way that resembles a Google search field. In 2010, Lee et al. called Google "one of the most influential

---

**28** Hillis, Petit, and Jarrett find the naturalization of "knowledge of" search without "knowledge about" how it actually works a direct consequence of the Google "magic box." Ken Hillis, Michael Petit, and Kylie Jarrett, *Google and the Culture of Search* (New York: Routledge, 2013), 14–15.

symbols of the new Internet paradigm" since the turn of the century.[29] Another ten years later, we are witnessing a googling paradigm – to which new members of society are ordinarily introduced[30] – as it operates in a broad cultural environment beyond simple web searches and structures the way we routinely approach the very procedures of locating relevant information.

Though the last mentioned observations reflect the recent digitalization of society in general, one must consider the specificities of the Holocaust research domain. Our study is inherently set in this field as well, already by default owing to the nature of the sources we use. It is necessary to bear in mind the singularity of the Holocaust as the ultimate cultural trauma, constant memento, and a negative point of reference for the contemporary "Western" value system. In this respect, among the primary imperatives (both research and ethical) are adequate source representation and interpretation, which is perhaps even more crucial when working with survivor testimonies. As we argued, search tools have a direct impact on displayed data, while the process of searching can lead to de-contextualization and re-contextualization of the original archival recordings. Not only can the user overlook the broader context of the found "segment" within an individual's entire personal life story, but imperfectly formulated search queries may also cause omission of some important aspects of the historical reality. Although any user interface generates a certain learning curve, we observed a considerable lack of adjustment in participants' actions that would respond to the particularities of the systems in use. This seemed to have possibly dissuaded them from discovering more about the tools as well as the resulting sources. We believe this is a critical goal for the future development of technologies for accessing Holocaust related sources. Whichever the "tools" will be, the users should be clearly made aware of how they arrived at their "data," what those results represent, and how they thus ultimately affect their research.

## Abbreviations

CVHM – Malach Center for Visual History at the Charles University
PDF – Portable Document Format
VHA – Visual History Archive of the University of Southern California Shoah Foundation

---

**29** Sang Hoon Lee et al., "Googling Social Interactions: Web Search Engine Based Social Network Construction," *PLoS ONE* 5, no. 7 (2010): e11233, accessed February 10, 2021, doi:10.1371/journal.pone.0011233.

**30** Sandra Houen et al., "Web Searching as a Context to Build on Young Children's Displayed Knowledge," in *Children's Knowledge-in-Interaction: Studies in Conversation Analysis*, ed. Amanda Bateman and Amelia Church (Cham: Springer, 2017), 57–72.

# References

Amiridze, Nino, Boyd Davis, and Margaret Maclagan, eds. *Fillers, Pauses and Placeholders*. Amsterdam and Philadelphia: John Benjamins, 2010.

Bittner, Egon. "The Concept of Organization." *Social Research* 32, no. 3 (Autumn 1965): 239–55. Accessed February 11, 2021. https://www.jstor.org/stable/40969788.

Chapman, Joyce C. "Observing Users: An Empirical Analysis of User Interaction with Online Finding Aids." *Journal of Archival Organization* 8, no. 1 (2010): 4–30. Accessed February 10, 2021. doi:10.1080/15332748.2010.484361.

Chýlek, Adam, Luboš Šmídl, and Jan Švec. "Multimodal Dialog with the MALACH Audiovisual Archive." In *Proceedings from Interspeech 2019*, 3663–64. Accessed February 10, 2021. doi:10.21437/Interspeech.2019.

Chýlek, Adam, Luboš Šmídl, and Jan Švec. "Question-Answering Dialog System for Large Audiovisual Archives." In *Text, Speech, and Dialogue: TSD 2019*, edited by Kamil Ekštein, 385–97. Cham: Springer, 2019.

Duff, Wendy M. "User Studies in Archives." In *User Studies for Digital Library Development*, edited by Pierluigi Feliciati, Andy O'Dwyer, and Milena Dobreva, 199–207. London: Facet Publishing, 2012.

Gafaranga, Joseph. "Language Choice as a Significant Aspect of Talk Organization: The Orderliness of Language Alternation." *Text* 19, no. 2 (1999): 201–25. Accessed February 11, 2021. doi:10.1515/text.1.1999.19.2.201.

Garfinkel, Harold. *Ethnomethodology's Program: Working Out Durkheim's Aphorism*. Lanham, MD: Rowman & Littlefield, 2002.

Goodwin, Charles. *Co-Operative Action*. Cambridge and New York: Cambridge University Press, 2018.

Heller, Monica, ed. *Codeswitching: Anthropological and Sociolinguistic Perspectives*. Berlin: Mouton de Gruyter, 1988.

Heritage, John and D. Rod Watson. "Formulations as Conversational Objects." In *Everyday Language: Studies in Ethnomethodology*, edited by George Psathas, 123–62. New York and London: Irvington, 1979.

Hillis, Ken, Michael Petit, and Kylie Jarrett. *Google and the Culture of Search*. New York: Routledge, 2013.

Holt, Elizabeth. "On the Nature of 'Laughables': Laughter as a Response to Overdone Figurative Phrases." *Pragmatics* 21, no. 3 (2011): 393–410.

Houen, Sandra, Susan Danby, Ann Farrell, and Karen Thorpe. "Web Searching as a Context to Build on Young Children's Displayed Knowledge." In *Children's Knowledge-in-Interaction: Studies in Conversation Analysis*, edited by Amanda Bateman and Amelia Church, 57–72. Cham: Springer, 2017.

Jefferson, Gail. "Notes on a Systematic Deployment of the Acknowledgement Tokens 'Yeah' and 'Mm hm'." *Papers in Linguistics* 17 (1984): 197–206. Accessed February 12, 2021. doi:10.1080/08351818409389201.

Kharazmi, Sadegh, Sarvnaz Karimi, Falk Scholer, and Adam Clark. "A Study of Querying Behaviour of Expert and Non-expert Users of Biomedical Search Systems." In *Proceedings of the 2014 Australasian Document Computing Symposium* (ADCS '14). Association for Computing Machinery, New York, NY, USA, 10–17. doi:10.1145/2682862.2682871.

Kocián, Jiří, Jakub Mlynář, and Petra Hoffmannová, eds. *Malach Center for Visual History on Its 10th Anniversary*. Prague: Matfyzpress, 2020.

Lee, Sang Hoon, Pan-Jun Kim, Yong-Yeol Ahn, and Hawoong Jeong. "Googling Social Interactions: Web Search Engine Based Social Network Construction." *PLoS ONE* 5, no. 7 (2010): e11233. Accessed February 10, 2021. doi:10.1371/journal.pone.0011233.

Livingston, Eric. "Context and Details in Studies of the Witnessable Social Order: Puzzles, Maps, Checkers, and Geometry." *Journal of Pragmatics* 40, no. 5 (2008): 840–62. Accessed February 10, 2021. doi:10.1016/j.pragma.2007.09.009.

Meyer, Christian, Jürgen Streeck, and J. Scott Jordan, eds. *Intercorporeality: Emerging Socialities in Interaction*. Oxford: Oxford University Press, 2017.

Mlynář, Jakub. "Malach Center for Visual History." In *Sborník Semináře o digitálních zdrojích a službách ve společenských a humanitních vědách (WDH 2015)*, edited by Jaroslava Hlaváčová, 83–89. Prague: Charles University, 2015.

Mlynář, Jakub, Esther González-Martínez, and Denis Lalanne. "Situated Organization of Video-Mediated Interaction: A Review of Ethnomethodological and Conversation Analytic Studies." *Interacting with Computers* 30, no. 2 (2018): 73–84. Accessed February 10, 2021. doi:10.1093/iwc/iwx019.

Mondada, Lorenza. "Going to Write: Embodied Trajectories of Writing of Collective Proposals in Grassroots Democracy Meetings." *Language and Dialogue* 6, no. 1 (2016): 140–78. Accessed February 10, 2021. doi:10.1075/ld.6.1.05mon.

Moore, Robert J. "A Name Is Worth a Thousand Pictures: Referential Practice in Human Interactions with Internet Search Engines." In *Mobile Speech and Advanced Natural Language Solutions*, edited by Amy Neustein and Judith A. Markowitz, 259–86. New York: Springer, 2013.

Moore, Robert J. and Elizabeth F. Churchill. "Computer Interaction Analysis: Toward an Empirical Approach to Understanding User Practice and Eye Gaze in GUI-based Interaction." *Computer Supported Cooperative Work* 20, no. 497 (2011): 497–528. Accessed February 10, 2021. doi:10.1007/s10606-011-9142-2.

Ørmen, Jacob. "Googling the News: Opportunities and Challenges in Studying News Events through Google Search." *Digital Journalism* 4, no. 1 (2016): 107–24. Accessed February 10, 2021. doi:10.1080/21670811.2015.1093272.

Pagenstecher, Cord. "Testimonies in Digital Environments: Comparing and (De-)Contextualising Interviews with Holocaust Survivor Anita Lasker-Wallfisch." *Oral History* 46, no. 2 (Autumn 2018): 109–18. Accessed February 10, 2021. http://www.jstor.org/stable/44993579.

Schegloff, Emanuel A. *Sequence Organization in Interaction: A Primer in Conversation Analysis*. Cambridge: Cambridge University Press, 2007.

Sormani, Philippe. "The Jubilatory YES! On the Instant Appraisal of an Experimental Finding." *Ethnographic Studies* 12 (2011): 59–77. Accessed February 10, 2021. doi:10.5449/idslu-001104716.

Švec, Jan, Josef V. Psutka, Jan Trmal, Luboš Šmídl, Pavel Ircing, and Jan Sedmidubsky. "On the Use of Grapheme Models for Searching in Large Spoken Archives." In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB*, 2018, 6259–63. doi:10.1109/ICASSP.2018.8461774.

Thomson, Alistair. "Four Paradigm Transformations in Oral History." *The Oral History Review* 34, no. 1 (2007): 49–70. Accessed February 10, 2021. http://www.jstor.org/stable/4495417.

Walden, Victoria Grace. "What Is 'Virtual Holocaust Memory'?" *Memory Studies*, November 2019. doi:10.1177/1750698019888712.

Anna Bonazzi

# N-gram-based Content Indexing: Semiautomated Analysis of Holocaust Testimonies

**Abstract:** Holocaust interview collections are key resources to preserve and share the memories of genocide survivors. As the size of testimonial collections makes manual analysis unwieldy, Digital Humanities (DH) methods can help develop a comprehensive view of the collections' contents and narratives. Developing nuanced and comprehensive methods to index and summarize the content of these interviews is central to ensuring survivor testimonies are accessible and searchable even in large-scale archives. This chapter develops a content indexing system that is not limited to the identification of keywords to summarize content, but rather includes longer phrases and emotional expressions. In particular, the chapter presents a semiautomated DH approach based on N-grams, i.e. frequent sentence chunks of 'N' words in a row, tested on David Boder's corpus of Holocaust testimonies. The results show that this approach allows us to identify speech patterns and narrative structures that go unseen in traditional keyword-based indexing of the survivors' testimonies, including structural, non-verbal categories like uncertainty, reticence, emotional insistence on time references. This method also helps shift our perspective from the narrative we expect survivors to use to the actual narrative they use, which may go unnoticed if we are not actively searching for it or if we only focus on major keywords.

**Keywords:** Holocaust, testimony, memory, Digital Humanities, interview archives, N-grams, content indexing, narrative structures

## 1 Digital Archives of Holocaust Testimonies

Digital archives of Holocaust and genocide testimonies are invaluable resources to preserve the histories and memories of survivors through large collections of interview recordings and transcripts. As the number of living Holocaust witnesses decreases and the scale of digital interview collections increases, these archives raise a pressing question of access to memory: what can we learn from these testimonies? What are we looking for when we consult them? The answers we give to these questions determine the technical means we develop to concretely access, read, or listen to survivors' testimonies, especially when the size of interview col-

lections makes manual analysis impossible. This chapter addresses the question of content indexing: in other words, how to annotate and categorize the content of thousands of hours of interviews in order to grasp their meaning at a glance or find specific topics across the whole collection. The approach I propose is a frequency-based analysis of interview transcripts based on N-grams aiming to identify recurring themes, expressions, and speech patterns. In particular, this chapter will focus on the interviews from the Boder corpus, a collection of 119 survivor testimonies in about nine languages recorded by Russian-born American psychologist David Boder in Displaced Persons camps in France, Switzerland, Italy, and Germany in 1946, just a year after the end of World War II. Boder interviewed both Jewish and non-Jewish witnesses, survivors, and displaced persons, collecting the earliest known archive of oral Holocaust history.[1]

To contextualize my work,[2] let us take a look at the way content indexing happens in interview archives such as the USC Shoah Foundation[3] archive and Boder's collection.[4] These archives typically develop a keyword-based indexing process: they assemble a controlled vocabulary of relevant keywords and manually annotate interviews in regular units (for example, minute by minute, like the Shoah Foundation, or page by page, like the Boder collection). This indexing approach is useful to quickly summarize certain aspects of interview contents, taking advantage of the regular and scalable nature of controlled vocabularies: it gives researchers a map to navigate the collection, find important factual references, and track certain topics across multiple interviews. The type of access granted by keywords highlights the role of testimonies as repositories of evidence, such as facts, names, and crucial historical references.

However, like all manual approaches, a keyword-based indexing system inevitably has a few blind spots. The main limitation of keywords is that they are predetermined: this approach tends to decide which concepts are significant and worth recording before hearing what survivors have to say. This decision is often influenced by existing research interests or expectations, for example preserving the memory of trauma or details about concentration camps. The risk of this method

---

**1** Alan Rosen, *The Wonder of Their Voices: The 1946 Holocaust Interviews of David Boder* (Oxford: Oxford University Press, 2010).
**2** My research was realized as part of the Holocaust Studies Digital Humanities Lab at UCLA, a research group including Prof. Todd Presner, Dr. Rachel Deblinger, Dr. David Shepard, Lizhou Fan, Michelle Lee, Wanxin Xie, Omar Hassan, and others. The group works on various DH-based approaches to the analysis of genocide interviews (primarily Holocaust, Rwandan Genocide, and Nanjing Massacre testimonies).
**3** "Visual History Archive," USC Shoah Foundation.
**4** "Voices of the Holocaust," Illinois Institute of Technology.

is that it centers the researcher's expectations more than the survivor's actual testimony. This can lead to a narrow gaze that mostly notices what already belongs to the canonical Holocaust narrative, which is problematic because survivors' testimonies exist before and outside of the Holocaust narrative as we know it today: they contribute to it but are not entirely defined by it.

David Boder's indexing system is a good example of a deductive, conceptually predetermined approach. Boder undertook an extensive interview project with Holocaust survivors in Western Europe as early as 1946, publishing some of the interviews as curated autobiographies.[5] With the help of his wife and graduate students, he experimented with various analytic methods, including an early quasi-computational approach. As a pioneer in quantitative psychology, he was interested in the effects of trauma on catastrophe survivors, so he transcribed, annotated, and analyzed his recordings with the aim to quantify the survivors' response to trauma. One of the four main categories of his indexing system was dedicated to nuanced expressions of trauma.[6] While this indexing system proved helpful for Boder's research goals, it offers a good example of a content classification approach that tends to find only what it set out to look for.

Another crucial limitation of keywords is that this method tends to focus on what survivors *say*, mostly in terms of nouns and verbs, and does not easily allow us to understand *how* they say a certain thing, or what they choose *not to say*. In addition to keywords, there are other ways we may define what "content" is in a survivor's testimony. We may want to look at interviews not only as permanent repositories of historical evidence, but also as personal stories that center the survivor's state of mind and help us understand how mainstream genocide narratives developed and what other narratives we might be overlooking. Pace, language choices, narrative structures, insistence, uncertainty, silence, all play a role in the expression of a unique and historically situated memory. Text is far from the only aspect of an interview that we may want to consider. For example, in the case of an audio archive like Boder's, much can be understood by paying attention to the speed of the dialogs, specific moments of silence, annotations of emotion, and the change in pitch in the voice of the interviewees.[7] Boder himself tried to preserve some of this paratextual and non-verbal information in his manual processing of the interviews. While his focus was primarily on the spoken

**5** David P. Boder, *Topical Autobiographies of Displaced People Recorded Verbatim in Displaced Persons Camps with a Psychological and Anthropological Analysis* (Chicago: D.P. Boder, 1950–1957).
**6** Alan Rosen, *The Wonder of Their Voices*.
**7** See for example Todd Presner's analysis of the audio track of some USC Shoah Foundation interviews in Todd Presner, "From Wire Recorder to Database." Meyerhoff Lecture, United States Holocaust Memorial Museum, October 18, 2018, accessed June 30, 2021, https://youtu.be/2aNufsZFfkM.

word, he took care to annotate details like long pauses, laughter, crying, and other visible expressions of emotion.[8] The voice and body languages of survivors, together with the outbursts of emotion that punctuate their testimonies, have the potential to enrich or even completely alter our understanding of interview transcripts. Ideally, a comprehensive indexing system of audio or audiovisual sources could offer its users not only references to spoken content, but also to emotion and body language. In an attempt to reimagine content indexing, this study only focuses on spoken word transcripts. However, by exploring additional options of what an index is supposed to record, it points to the presence of multiple layers of expression and meaning in recorded interviews.

## 2 N-grams

In order to expand the descriptive ability of indexing to include more aspects of a testimony, I experiment with inductive content annotation based on N-grams, meaning short segments of *n* words or characters (in this case, words). N-grams are a classic tool of computational linguistics and have long been used for text analysis and content mining in large data collections.[9] They are particularly useful for text categorization and the identification of topic similarity, including in multilingual text collections, as they are language-independent.[10]

---

**8** The graph in the following link provides an idea of the kind of emotions Boder transcribed – and raises questions concerning our assumptions of the kind of emotions a survivor is "supposed" to display: accessed June 30, 2021, https://drive.google.com/file/d/1y68uqxwIhf3ZeaG_7Aq1Inr_j0-nIju7/view?usp=sharing (UCLA Holocaust Studies DH Lab Work, 2019).

**9** See, e.g., Anne Burdick et al., *Digital_Humanities* (Cambridge, MA: MIT Press, 2012); Susan Schreibman, Ray Siemens, and John Unsworth, *A Companion to Digital Humanities* (New York: John Wiley & Sons, 2008).

**10** See for example Riyad Al-shalabi and Rasha Obeidat, "Improving KNN Arabic Text Classification with N-Grams Based Document Indexing," in *Proceedings of the 6th International Conference on Informatics and Systems INFOS2008* (Cairo, 2008), 108–12; Marc Damashek, "Gauging Similarity with N-Grams: Language-Independent Categorization of Text," *Science* 267, no. 5199 (1995): 843–48; Armand Joulin et al., "Bag of Tricks for Efficient Text Classification," *arXiv:1607.01759 [cs]*, August 9, 2016, accessed June 30, 2021, http://arxiv.org/abs/1607.01759; Artur Šilić et al., "N-Grams and Morphological Normalization in Text Classification: A Comparison on a Croatian-English Parallel Corpus," in *Progress in Artificial Intelligence. EPIA 2007*, ed. J. Neves, M.F. Santos, and J.M. Machado, 671–82. Lecture Notes in Computer Science, vol. 4874 (Berlin: Springer, 2007), accessed June 30, 2021, https://doi.org/10.1007/978-3-540-77002-2_56; Zhihua Wei et al., "N-Grams Based Feature Selection and Text Representation for Chinese Text Classification," *International Journal of Computational Intelligence Systems* 2, no. 4 (December 2009): 365–74, accessed June 30, 2021, https://doi.org/10.1080/18756891.2009.9727668.

In short, I split interview transcripts into snippets of two to six words, compute their frequency, and annotate information about the words and parts of speech each N-gram contains. Then I cluster them in a semiautomated way and manually analyze them to see which themes or recurring structures emerge from the text. This method essentially implies reading through text in a non-linear way: looking at groups of text units longer than words but shorter than paragraphs or pages helps us identify expression patterns and insistence on certain phrases. Most importantly, allowing an unbiased process[11] to decide for us which text features seem significant can help us notice categories we may not have been looking for at the outset, either because they have to do with structure more than with content, or because the content they point to may be unexpected.

This approach is situated somewhere between close reading and unsupervised topic modeling:[12] it resembles the former in its attention to sentence structures and language choices, and it resembles the latter in its ability to quickly identify the main themes of a text. As a semiautomated method, the N-gram approach presents the strengths and weaknesses of both automated categorization and manual analysis: while the identification of significant categories follows more traceable and objective criteria than a manual approach, it still requires subjective judgment in the selection and naming of those categories. Moreover, this method relies heavily on frequency, which can be a drawback: in a text corpus, not all that is frequent is relevant and not all that is relevant is frequent. Still, my analysis suggests that sometimes elements of a testimony which are both frequent and relevant don't easily stand out in a manual or keyword-based analysis, while they do through an N-gram-based approach.

The Boder corpus, given its small size as well as its idiosyncratic and multilingual character, provides a useful case study to test out various indexing methods which can then be scaled up to larger collections.[13] In what follows, I describe two uses of N-gram-based indexing: one for a corpus-wide analysis and one for the analysis of a single interview, to demonstrate the value of the proposed semiautomated DH approach.

---

**11** Unbiased is meant here in the computational sense (in an analogous machine intelligence sense, unsupervised), not the human one.
**12** See for example Hamed Jelodar et al., "Latent Dirichlet Allocation (LDA) and Topic Modeling: Models, Applications, a Survey," *Multimedia Tools and Applications* 78, no. 11 (2019): 15169–211.
**13** Topic modeling is not very helpful on a relatively small and multilingual corpus like this one, but other computational methods proved effective, such as semantic triplets (under development by Lizhou Fan from the Holocaust Studies DH Lab at UCLA) and N-grams.

# 3 Corpus-wide Analysis

Using N-gram-based indexing on an entire interview corpus can precede or complement close reading of specific samples, as it offers an insight into the hundreds of interviews we may not get to examine manually. Identifying the main structural and thematic patterns of a collection is also useful to enable a comparison between different testimony collections, analyze the evolution of testimony as a genre, and understand how a certain Holocaust narrative was established.

To begin with, each digitized interview transcript was transposed to XML format and annotated with Boder's original information about the interview's date, location, and language, and the survivor's name, age, and religion. Each line of the interview was then automatically annotated with data about the speaker, language,[14] and speed of the dialog (as Boder's interviews are highly multilingual, it was important to label the languages used in every line).[15] The text was annotated with Treetagger,[16] a multilingual Part-Of-Speech (POS) tool that identifies grammatical parts of speech such as nouns, verbs, adverbs, etc. In order to filter out semantically weak N-grams (such as "*and then we*"), expressions that did not contain at least a verb, noun, or adjective were excluded from further analysis. POS annotation relies on context and frequency of a word, so it works better on longer texts and does not yield sensible results with short text snippets like an N-gram. For this reason, the text was first annotated as a whole and then segmented into N-grams, with each word carrying its POS tag (e.g., "*meine*-adjective *Mutter*-noun"). The text was segmented into N-grams of two to six words: for example, a sentence like "*the Jewish Council got a call from the police*"[17] would produce 3-grams "*the Jewish Council*," "*Jewish Council got*," "*Council got a*," "*got a call*," 4-grams "*the Jewish Council got*," "*Jewish Council got a*," and so on. These N-grams were marked with their interview and line ID number (so that they can be traced), then they were sorted by frequency to identify the most common ones: while a segment like "*call from the*" may appear once or twice in the whole text and not carry much meaning, N-grams like "*the Jewish Council*" or "*from the police*" may be more frequent and tell us more about the testimony.

---

**14** Two Python packages were used: *TextBlob* for lines below 11 characters (high accuracy but daily usage limitations) and *langdetect* for all other lines (high accuracy but only on longer strings; no usage limitations).

**15** This was also useful to analyze the use of linguistic code switching by survivors.

**16** Helmut Schmid, "Probabilistic Part-of-Speech Tagging Using Decision Trees," in *Proceedings of International Conference on New Methods in Language Processing* (Manchester, 1994).

**17** Boder corpus, interview with Abraham Kimmelmann, August 27, 1946 (transcript n. 56, line n. 77, translated).

N-grams with a minimum frequency of five were manually examined and categorized. To begin with, only N-grams in the largest corpus languages (German, French, English) and Yiddish were considered, while N-grams in other languages (Russian, Polish, Spanish) were only included if they had a high frequency and seemed related to existing categories. N-grams were labeled for their content (for example "camp" or "family") or for their structure (for example "narration" or "uncertainty") across different languages. To do so, N-grams from all languages were manually labeled in English to group together related expressions in different languages.[18] For example, expressions like "*zwölf 12 Uhr*," "*in der Nacht*," "*o'clock in the*," "*at that time*," "*de ce moment*" were all labeled as "time," while expressions like "*nisht gevust vos*," "*I don't know*," "*ich weiß nicht*" were labeled as "don't know / don't remember / don't want [to tell]." I followed a conservative labeling process, only attributing a label to reasonably unambiguous expressions and marking N-grams as "undefined" when in doubt about their meaning, for example in cases like "*no such thing as*" or "*auf der Strasse*," in order to avoid overinterpreting sentence snippets that were inevitably out of context and complicated by their multilingual nature.

This labeling process is in no way conclusive: it is necessarily subjective and prone to errors, just like keyword-based indexing, and it benefits from recursive corrections and feedback, for example in categories like "narration" or "opinion" that seemed marked and expressive enough to be named but could not be considered objective or clearly defined. In addition, an accurate analysis of frequent N-grams requires a few more processing steps (described in the next section) not yet included at this stage. Still, even an approximate analysis of the most frequent N-gram categories is productive to map out the main areas mentioned by interviewees: *camp / suffering* (678), *don't know / don't remember / don't want* (402), *year* (390), *time* (348), *movement* (224), *Jews* (141), *Palestine* (99), *coercion* (96), *work* (83), *Germans* (51), *war* (51), *street* (44), *organizations* (41), *deportation* (38), *future / plans / family abroad* (36), *date* (32), *SS* (23), *liberation* (22), *ghetto* (20), *Russians* (19), *hunger* (12), *Americans* (10), *opinion* (10), *resistance* (10), *hiding* (10).

Of these, perhaps the most surprising is the second largest category: a high number of recurring phrases across the four languages considered here are variations of *I don't know*, *I don't remember*, *I don't want to* [most frequently: *go / say / tell*]. These sentences point to a category of uncertainty that is not conclusively explained by simple question-answer exchanges where the interviewer

---

**18** In cases where I didn't have a direct knowledge of the language (Polish and Russian N-grams), particularly frequent expressions were translated with Google Translate and only labeled if their meaning seemed unambiguous and clearly related to an existing category, for example the many instances of Russian N-gram "я не знаю," "I don't know."

asks about a detail that the interviewee doesn't know. They also appear in the middle of the interviewees' narration, they interrupt the flow of a sentence, they give us a look into the interviewees' perspective as they explain that something was happening without their knowledge, or that they didn't know yet something was about to happen, or that they simply don't want to talk about something. Boder's interviews were conducted only about a year after the end of the war and in displaced persons camps, which means that most interviewees were not in a safe or stable environment and lived in a situation of uncertainty concerning their future, their livelihood, the whereabouts of their family. Many do not seem to have developed a cohesive narrative of their experience yet, and most of them had not yet had a chance to talk to people unfamiliar with what happened (as Boder was). It also seems likely that at this early stage, interviewees saw their experiences as individual pieces of a puzzle. While they have a sense of the scale of the atrocities that happened around them, there are still many elements of the larger historical picture they hadn't heard of yet, and many Holocaust terms we would all become familiar with in later decades are not part of the public discourse at the time of their interviews. This category of uncertainty calls our attention to the unscripted and complex nature of the memories that are being narrated and reminds us that uncertainty is only one of the many emotions survivors display in their interviews. While the N-gram-based indexing system is not able to fully reference the emotions and body language of interviewees, it does point to their presence, and may encourage the researcher to take a closer look at specific interviews or corpus-wide interview passages.

Another N-gram category that stands out in the Boder corpus concerns time expressions (as opposed to chronological dates). While keyword-based indexing often classifies all references to time under the same label, the N-gram-based approach shows that survivors seem to be doing two very different types of memory operations when they refer to time. Sometimes they give specific dates and factual historical information either in response to Boder's questions or to describe their experience for external listeners. They talk about dates or years like "on the 28th of May" or "on the 30th of June 1941" or "in Summer 1943." At other times, though, their time references are highly subjective: they include vague durations, impressions of time passing, narrative constructs like "very late," "every night," "for a long time," or "it was always like that." The space these expressions take up in the survivor's story is often disproportionate to the historical time they reference. At times, an overabundance of both subjective and objective time references signals episodes etched into the survivor's memory which arguably deserve to be marked as significant during the indexing process because that is the way the survivor talks about them. For example, 29-year-old Ephraim Gutman uses six different time expressions to tell Boder about one of the massacres he witnessed: "They

killed them on Wednesday in the night and on Thursday in the night, it was the first day of the month of Tammuz."[19] In this case, the N-gram-based approach highlights the need for a nuanced model of representation and interpretation of time to address what Johanna Drucker has described as "the complexities of lived, reported, constituted, and produced temporality in human documents."[20] More in general, episodes like these offer us a glimpse into what survivors really wanted to convey and show that indexing their words for their mere semantic content ("they're talking about time") might obscure something important about their testimony. For reference, let us take a look at the indexing systems developed by Boder and by the Shoah Foundation. Based on an analysis of three of Boder's typewritten and annotated interview transcripts,[21] Boder indexed his interviews with at least 1,023 keywords, each of which was used on average nine times. His indexing was sparse and relatively inconsistent, and many of the words he annotated referred to people or location names. With his indexing system, expressions of subjective time, emotional intensity, and uncertainty are not visible, while concepts that were central to his research interests are much more evident (for example "TAT test," a psychological test he developed, or "vermin" and "dead-handlers," concepts he may have tracked because of his interest in trauma).

As the purpose and scale of a project change, so does the format of indexing. For example, the USC Shoah Foundation Archive, which started a large-scale testimony videorecording project in 1994, adopted a two-tiered system with more specific terms and broader parent terms. This allows for the systematic annotation of numerous interviews with a wide range of speakers, languages, contents, and time of recording. If we look at Boder's corpus through the lens of the Shoah Foundation's indexing system,[22] what we see is yet another picture. Figure 1 compares the categories we may use to describe Boder's collection with Shoah Foundation keywords and N-gram-based categories.

---

**19** "Dort haben sie ermordet Mittwoch in der Nacht und Donnerstag Nacht, es war der Erste des Monats Tammuz." Boder corpus, interview with Ephraim Gutman, September 12, 1946 (transcript n. 37, line n. 15, translated).

**20** See Drucker's discussion on the need for overlapping and non-homogeneous models of time mapping in data-based humanities work in: Johanna Drucker, *Visualization and Interpretation: Humanistic Approaches to Display* (Cambridge, MA: MIT Press, 2020), 114.

**21** The digitization and categorization of typewritten keywords from Boder's notes was carried out by Lizhou Fan (UCLA Holocaust Studies DH Lab work, 2019).

**22** To compare N-gram-based keywords with Boder's, I use a version of Boder's keywords grouped under the USC Shoah Foundation's controlled vocabulary of parent terms (fewer in number and broader in meaning than Boder's heterogeneous and sporadic keywords). Lizhou Fan and Todd Presner matched Shoah Foundation parent keywords to Boder's terms with the aim to make the two indexing systems comparable (UCLA Holocaust Studies DH Lab work, 2019).
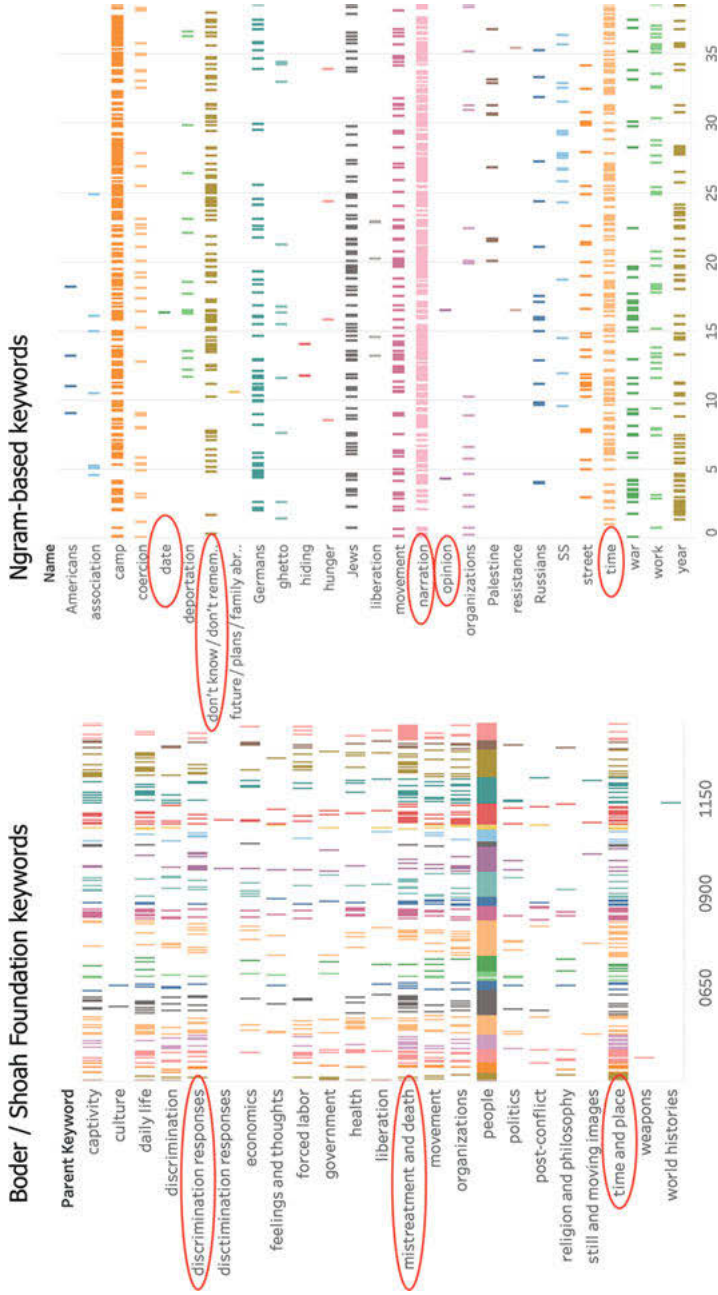
**Figure 1:** Shoah Foundation and N-gram keywords applied to the Boder corpus.

*Source:* Full interactive diagrams on Tableau Public (Lizhou Fan). Boder Keywords: accessed June 30, 2021, https://web.archive.org/web/20210630012904/https://public.tableau.com/app/profile/lizhou/viz/Boder_page_index_dispersion/Pagedispersionofparentkeywords bytestimonies; Ngram Keywords: accessed June 30, 2021, https://web.archive.org/web/20210630013711/https://public.tableau.com/app/profile/lizhou/viz/Ngram-basedPseudoIndexofBoderTestimonies/Ngram-basedPseudoIndexofBoderTestimonies.

Each indexing system highlights different aspects of the interview collection. While the keyword-based one is helpful to track major thematic categories like "discrimination responses" or "mistreatment and death," the N-gram-based one offers an insight into more structural categories of the survivors' testimony, such as their uncertainty, reticence, point of view, and emphasis on specific episodes.

# 4  Interview Analysis

Among all his interviews, Boder dedicated particular attention to his encounter with Abraham Kimmelmann, an 18-year-old survivor he spoke to in 1946. His interview is exceptional on all accounts: it is personal, philosophical at times, elaborate, and detailed. While most interviews range between 20 minutes and an hour, Kimmelmann spoke to Boder for almost 4 hours (and Boder planned to interview him again, although he never managed to). In his testimony, Kimmelmann makes several personal observations of moral and philosophical character about the events he experienced, which is unique given Boder's habit of insisting his interlocutors only focus on precise facts and events they witnessed personally. This is also one of the few interviews that Boder processed for publication in his *Topical Autobiographies*[23] and annotated with his indexing system. The unusual features of this interview, together with the amount of work Boder dedicated to it, make it a useful example to compare the keyword-based indexing system with the N-gram-based one.

For the analysis of Kimmelmann's interview, the N-gram system was improved with some consideration for the exact frequency of each N-gram and the similarity between comparable expressions. Only Kimmelmann's lines were considered, while Boder's questions and comments were left aside. After filtering out semantically poor segments mostly consisting of conjunctions, pronouns, and auxiliary verbs, two drawbacks of raw N-grams needed to be addressed for an accurate frequency count: frequency underestimation in some cases and overestimation in others. On the one hand, the presence of slightly different versions of the same N-gram (such as "*ein jüdischer Miliz*" vs. "*die jüdische Miliz*") leads to the risk that a really frequent expression might get lost among many variants that are not counted together just because they are not identical word by word. On the other hand, the overlap of N-grams of different sizes obtained from the same sentence distorts the final count and makes rare expressions appear more frequent than they really are. For instance, we might be led to believe that the town of Bensburg

---

**23** Boder, *Topical Autobiographies*.

is of incredible importance in Kimmelmann's testimony because it appears in 212 N-grams, while in fact he only mentions the town 25 times. Most N-gram instances of the word are snippets of the same sentence: for example, for the sentence "*Essen war billiger als in Bensburg*" ("food was cheaper than in Bensburg"),[24] we may obtain N-grams "*war billiger als in Bensburg,*" "*billiger als in Bensburg,*" "*als in Bensburg,*" etc.

To improve the reliability of the frequency count, similar N-grams (such as "*jüdische Gemeinde*" and "*bei der jüdischen Kultusgemeinde,*" both designating Jewish religious communities) were grouped together using a similarity algorithm.[25] Any two N-grams were considered passably similar if they fulfilled these conditions: their similarity score was ≥ 0.75; they appeared in different lines of the interview; or if they appeared in the same line, they were not subsets of each other (as "*und meine Mutter*" compared to "*ich und meine Mutter*"). The resulting list was manually filtered to exclude expressions that were clearly unrelated despite matching these criteria on a formal level, like "*meine Mutter*" and "*ein Meter.*"

The remaining N-grams were manually analyzed and grouped to identify frequent units that potentially indicate the main themes or expressions of the interview. A comparison between N-gram-based results and Boder's own annotations of this interview clearly shows the benefits of adding an alternative content indexing system (Figures 2 and 3).

Boder annotated each page of this interview with nouns he found particularly significant, like objects, people, concepts abstracted from Kimmelmann's narration. A look at the terms he recorded (grouped under the Shoah Foundation project's parent keywords) reveals his focus on factual details, like personal and geographical names, and on traumatic categories of mistreatment or life under military occupation. This choice is sensible when one thinks of the aims Boder was trying to pursue: a quantification of survivor trauma and a historical report of the events. However, these terms do not fully capture what the survivor, Kimmelmann, is talking about. A look at the results of N-gram-based indexing done on the same interview shows that several significant aspects of the interview escape a predetermined keyword-based approach.

N-gram-based keywords return a completely different image of what the interview is "about." First, we see a strong narrative component to Kimmelmann's testimony. Unlike other survivors, he is not just answering questions, but build-

---

**24** Boder corpus, interview with Abraham Kimmelmann, August 27, 1946 (transcript n. 56, line n. 277).
**25** *SequenceMatcher* algorithm of Python's *difflib* package. A 0.75 threshold was set after a few tries and may be improved.
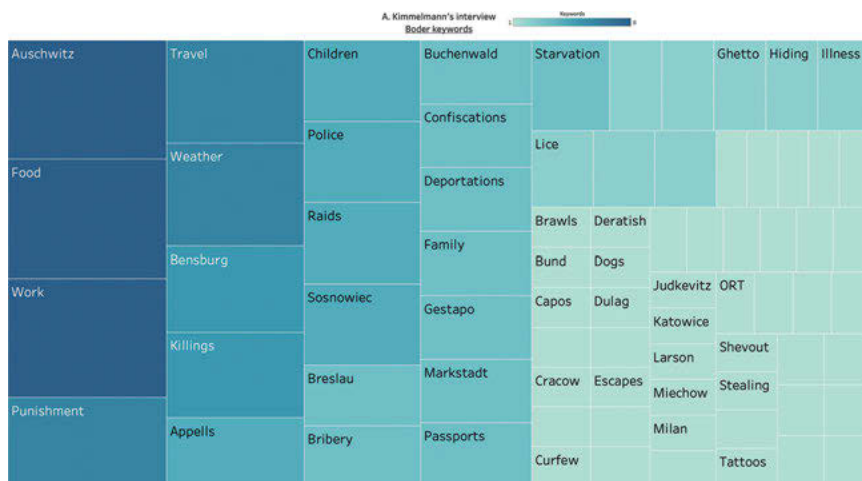
**Figure 2:** Boder's indexing of Kimmelmann's interview.
*Source*: Full interactive diagram on Tableau Public (Anna Bonazzi). Boder's indexing of Kimmelmann's interview: accessed June 30, 2021, /web/20210630014531/https://public.tableau.com/app/profile/anna5558/viz/KimmelmannkeywordsBodervsN-grams/Dashboard1.
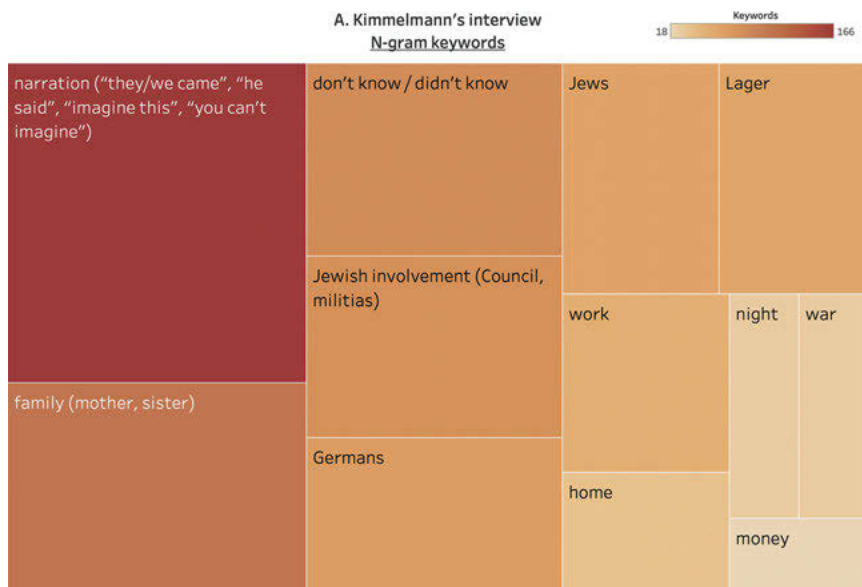


**Figure 3:** N-gram-based indexing of Kimmelmann's interview.
*Source*: Full interactive diagram on Tableau Public (Anna Bonazzi). N-gram-based indexing of Kimmelmann's interview: accessed June 30, 2021, /web/20210630014531/https://public.tableau.com/app/profile/anna5558/viz/KimmelmannkeywordsBodervsN-grams/Dashboard1.

ing a long narration that draws the listener into his experience (as we can easily confirm with a close reading of the interview). He often invites Boder to see things from his perspective, repeating expressions such as "imagine this," "picture this," "you can't imagine how ...." While defining this category does require subjective judgment, N-grams help us notice Kimmelmann's recurring turns of phrase.

Another interesting category of this visualization is Kimmelmann's interest in the agency and moral involvement of other Jews in the events he experienced. He talks about Jewish militias, he talks about the ambiguous role of the Jewish Council, and he is really concerned with the morally gray area that many Jews found themselves in when they had to decide what they were willing to do to save their lives, or what they felt enabled to do under the circumstances. Boder does not index this category (he indexes separate references to people and organizations but does not view this as a topic on its own). While a topic like this might go unseen in manual indexing given its borderline taboo character, the frequency-based approach makes it easier to notice it, name it, and index it (close reading confirms this is in fact a key part of this testimony).

A third category that stands out in this interview is the "I don't know / I didn't know" cluster. The strong presence of this category even in an interview that is mostly self-regulated and contains comparatively few direct questions by the interviewer confirms the narrative role of uncertainty and reticence in early testimonies, as already observed in the corpus-wide analysis from the previous section. These three examples represent central aspects of Kimmelmann's testimony, but they don't easily stand out in a manual keyword-based analysis like that of Boder.

# 5  Conclusion

This chapter discussed an alternative method to index the content of large interview collections based on N-grams (as opposed to manually assigned keywords), particularly in the case of genocide survivors' testimonies. This method helps us expand our understanding of what "content" is and what parts of a testimony should be annotated in addition to verbal, mostly noun-based concepts. This chapter showed that N-gram-based indexing helps identify categories like narration, uncertainty, reticence, emphasis, emotional insistence on time references, and topics we may not be predisposed to manually index, like morally gray areas. This method strongly emphasizes frequency as a marker of the importance of an expression, an approach that is not free from drawbacks: it is important to remember that frequency can be misleading, and that the frequency of an expres-

sion is not the only way to determine what is important in an interview, especially one that deals with difficult or even traumatic personal memories. Yet, with its focus on recurring sentence units, N-gram-based indexing helps us center the survivors' words and expression over our expectation of a certain narrative and expand indexing to include structural, non-verbal aspects of their testimony.

This method touches upon questions at the heart of the digital turn for Jewish Studies, and the humanities more broadly: how can we use digital research methods in an ethical way when we approach the study of Holocaust interviews, and more generally the testimonies of genocide survivors? What pitfalls should we avoid to ensure our computational approaches do not reduce survivors and their life stories to points or numbers in a spreadsheet? As this chapter has tried to show, a thoughtful use of computational research methods in fact has the potential to humanize its objects of study and bring them closer to the spectator in ways that are sometimes not accessible to manual analysis in the case of large archives. By combining qualitative analysis with computational text analysis, N-gram-based indexing offers the spectator a richer and more nuanced access to survivor's memories, and by highlighting the relevance of interview features like emotion and dialogical interaction, it points to multiple potential directions for computational research on testimonies also beyond Holocaust and genocide studies.

# References

## Archives

"Visual History Archive." USC Shoah Foundation. Updated November 1, 2017. Accessed February 15, 2021. https://vhaonline.usc.edu/login?returnUrl=/Search.

"Voices of the Holocaust." Illinois Institute of Technology. Accessed February 15, 2021. https://iit.aviaryplatform.com/collections/231.

## Sources

Al-shalabi, Riyad and Rasha Obeidat. "Improving KNN Arabic Text Classification with N-Grams Based Document Indexing." In *Proceedings of the 6th International Conference on Informatics and Systems INFOS2008*, 108–12. Cairo, 2008.

Boder, David Pablo. *Topical Autobiographies of Displaced People Recorded Verbatim in Displaced Persons Camps with a Psychological and Anthropological Analysis*. Chicago: D.P. Boder, 1950–1957.

Burdick, Anne, Johanna Drucker, Peter Lunenfeld, Todd Presner, and Jeffrey Schnapp. *Digital_ Humanities*. Cambridge, MA: MIT Press, 2012.

Damashek, Marc. "Gauging Similarity with N-Grams: Language-Independent Categorization of Text." *Science* 267, no. 5199 (1995): 843–48.

Drucker, Johanna. *Visualization and Interpretation: Humanistic Approaches to Display*. Cambridge, MA: MIT Press, 2020.

Jelodar, Hamed, Yongli Wang, Chi Yuan, Xia Feng, Xiahui Jiang, Yanchao Li, and Liang Zhao. "Latent Dirichlet Allocation (LDA) and Topic Modeling: Models, Applications, a Survey." *Multimedia Tools and Applications* 78, no. 11 (2019): 15169–211.

Joulin, Armand, Edouard Grave, Piotr Bojanowski, and Tomas Mikolov. "Bag of Tricks for Efficient Text Classification." *ArXiv:1607.01759 [Cs]*, August 9, 2016. Accessed June 30, 2021. http://arxiv.org/abs/1607.01759.

Presner, Todd. "From Wire Recorder to Database." Meyerhoff Lecture, United States Holocaust Memorial Museum, October 18, 2018. Accessed June 30, 2021. https://youtu.be/2aNufsZFfkM.

Rosen, Alan. *The Wonder of Their Voices: The 1946 Holocaust Interviews of David Boder*. Oxford: Oxford University Press, 2010.

Schmid, Helmut. "Probabilistic Part-of-Speech Tagging Using Decision Trees." In *Proceedings of International Conference on New Methods in Language Processing*. Manchester, 1994.

Schreibman, Susan, Ray Siemens, and John Unsworth. *A Companion to Digital Humanities*. New York: John Wiley & Sons, 2008.

Šilić, Artur, Jean-Hugues Chauchat, Bojana Dalbelo Bašić, and Annie Morin. "N-Grams and Morphological Normalization in Text Classification: A Comparison on a Croatian-English Parallel Corpus." In *Progress in Artificial Intelligence. EPIA 2007*, edited by J. Neves, M.F. Santos, and J.M. Machado, 671–82. Lecture Notes in Computer Science, vol. 4874. Berlin: Springer, 2007. Accessed June 30, 2021. https://doi.org/10.1007/978-3-540-77002-2_56.

Wei, Zhihua, Duoqian Miao, Jean-Hugues Chauchat, Rui Zhao, and Wen Li. "N-Grams Based Feature Selection and Text Representation for Chinese Text Classification." *International Journal of Computational Intelligence Systems* 2, no. 4 (December 2009): 365–74. Accessed June 30, 2021. https://doi.org/10.1080/18756891.2009.9727668.

## Spatiality

# Sinja Clavadetscher, Stefanie Mahrer, and Stefanie Salvisberg
# Mapping Forced Academic Migration

**Abstract:** In spring and early summer 1933, after the "Law for the Restoration of the Professional Civil Service" was passed on April 7, some 3.000 academics, the largest part of them Jewish, lost their positions in German universities and other research institutions. Despite the strict immigration laws, which were a result of World War I, over 50% of the dismissed academics (the numbers vary greatly among the different studies) left Germany and tried to continue or rebuild their academic careers in exile. The largest part migrated to the UK and the US, but also France, Sweden, Turkey, the British mandate of Palestine as well as Switzerland were target countries.

We are digitally mapping the geographical movement of academics triggered by mass expulsions from Germany, Austria, and other countries occupied by the Third Reich and shedding light on inner-institutional changes in respect to both academic staff and the academic standing of research institutions. In our contribution we shed light on the case of the mathematicians at Göttingen University in the early years of the "Third Reich" and the impact of National Socialist university politics on the field of mathematics. These are initial results of our larger study which help to identify patterns and ruptures of forced academic migration.

**Keywords:** forced academic migration, migration studies, exile, national socialism, World War II, university history

## 1 Introduction

The German philosopher Theodor W. Adorno (1903–1969), who, as a "non-Aryan", was forced into exile by the National Socialists, wrote in his book *Minima Moralia*, "Every intellectual in emigration is, without exception, mutilated, and does well to acknowledge it to himself, if he wishes to avoid being cruelly apprised of it behind the tightly-closed doors of his self-esteem."[1] He was speaking of his own

---

[1] Theodor W. Adorno, *Minima Moralia: Reflections from Damaged Life* (London: Verso, 2005), 33.

---

biographical disruptions, the expropriation of language,[2] and political isolation and career breaks. But this experience was not his alone. Over 3,000 German academics, most of them Jewish either by self-identification or the racial laws of Nazi Germany, were dismissed from their positions, based on paragraph 3 of the Law for the Restoration of the Professional Civil Service. After April 7, 1933, Jews, other "non-Aryans", and political opponents (under paragraph 4) could no longer serve as teachers, professors, and judges or in other governmental positions. Around 50% of dismissed scientists emigrated and tried to continue their careers abroad. This was likely the largest migration of academics in human history. The emigrees faced, as we have seen with Adorno, alienation, racism, impediments, and often, psychological difficulties in exile, in both their private and their professional lives. However, precisely because emigration emerges from a crisis, it can bring about innovations and result in scientific and social achievements which, according to the historian Dan Diner, "would have hardly been expected under conditions of steadiness and continuity of life plans."[3]

The biographical perspective of the impact of this forced migration during the 1930s and 1940s on individual careers is one of three main objectives of our research project titled "Science Transnational. Switzerland and the Academic Forced Migrants from 1933 to 1950."[4] We are equally interested in the way the transnational academic network and the academic landscape as such changed due to the expulsion of academics. Therefore, we are digitally mapping the geographical movement of academics triggered by mass expulsions from Germany, Austria, and other countries occupied by the Third Reich and shedding light on inner-institutional changes in respect to both academic staff and the academic standing of research institutions.[5] Researchers agree that the mass expulsion of academics, some of them Nobel laureates, was an "intellectual decapitation of Germany"[6]

---

**2** On this topic, also see: Stephan Braese, "Deutsche Sprache, jüdisches Exil – Optionen von 'Identität' nach 1933," in *Exilerfahrung und Konstruktionen von Identität 1933 bis 1945*, ed. Hans O. Horch, Hanni Mittelmann, and Karin Neuburger (Berlin: De Gruyter, 2013), 7–16.

**3** Dan Diner, "Einleitung," in *Tel Aviver Jahrbuch für deutsche Geschichte* 27 (1998), 3 (Translated from German by StM).

**4** This project was funded by a five-year PRIMA-grant (No. 179819) from the Swiss National Science Foundation.

**5** While we have included screenshots of the data-visualization throughout this article, on the project website, we have prepared corresponding scenarios for more explorative and dynamic data analyses, and the website data is updated daily.

**6** Helge Pross, "Die geistige Enthauptung Deutschlands. Verluste durch Emigration," in *Nationalsozialismus und die deutsche Universität. Universitätstage 1966* (Berlin: De Gruyter, 1966), 143.

and a "dismantling of German science,"[7] particularly the modern disciplines and sub-disciplines, such as economics, political science, sociology, atomic physics, and biochemistry. In these areas, the dismissal rate was up to 50% since the new research areas were represented to a higher degree by younger and often Jewish scholars.[8] The universities "surrendered without a fight"[9] to the drastic curtailment of their autonomy,[10] and there was no public opposition to the forced dismissals from either colleagues or other universities. On the contrary, in April 1933, the boards of directors of German universities and, a little later, the universities expressed their approval of Adolf Hitler and the National Socialist state,[11] and at some universities, around 25% of all university teachers had joined the party by the summer of 1933.[12] At that time, among the party members, there were relatively few tenured professors, while the number of assistants, private lecturers, and associate professors was comparatively high. There were two reasons for this: first, the National Socialist university policy clearly represented the interests of a non-established younger generation against the tenured professors (*Lehrstühle*), and second, the mass dismissals improved the previously miserable career opportunities of young academics.[13]

While German research institutions suffered a (self-inflicted) loss of researchers, knowledge, and academic reputation, other countries had the rare opportunity to gain some of the best scientists worldwide. Our large-scale data collection will allow for a broad transnational evaluation of different states and universities with a special focus on the position of Swiss universities in the changing academic landscape. There has been a longstanding connection between German and Swiss universities with a strong intellectual as well as personal exchange. This, as well as a shared and a similar structural design of the university system, made Switzerland a likely emigration destination for many expelled scholars. Thus far, however, neither scholars of forced academic migration nor scholars of Swiss history have paid attention to emigree scholars in Switzerland or the dealing of Swiss universities with them. In this article as well as in our larger

---

**7** Karl Dietrich Bracher, *Die deutsche Diktatur: Entstehung, Struktur, Folgen des Nationalsozialismus*, (Frankfurt am Main: Verlag Ullstein, 1980), 294.

**8** Claus-Dieter Krohn et al., eds., *Handbuch der deutschsprachigen Emigration 1933–1945* (Darmstadt: WBG, 2008), 681–82.

**9** Michael Grüttner, "Die deutschen Universitäten unter dem Hakenkreuz," in *Zwischen Autonomie und Anpassung: Universitäten in den Diktaturen des 20. Jahrhunderts*, ed. John Conelly (Paderborn: Schöningh, 2003), 67.

**10** Grüttner, "Die deutschen Universitäten unter dem Hakenkreuz," 74.

**11** Grüttner, "Die deutschen Universitäten unter dem Hakenkreuz," 74–76.

**12** Grüttner, "Die deutschen Universitäten unter dem Hakenkreuz," 73.

**13** Grüttner, "Die deutschen Universitäten unter dem Hakenkreuz," 76.

project, we will be putting Swiss universities and exiled scientists in Switzerland on the map of forced academic migration.

We are working with nodegoat,[14] a web-based research environment for the humanities, that allows for relational modes of data analysis with spatial and chronological forms of contextualization. Our research question-driven data model enables us to store all known personal and professional data of the dismissed scholars, including not only their career steps at academic institutions across the globe but also their publications, presentations, and research project collaborations, which allows us to illustrate transnational academic migration and shifts in global academic networks, shifting scientific centers, and the transfer of knowledge. This article contains links to interactive data visualizations; we recommend reading in conjunction with the online visualizations.[15]

In this article we will shed light on the case of the mathematicians at Göttingen University in the early years of the "Third Reich" and the impact of National Socialist university politics on the field of mathematics. These are initial results of our larger study which help to identify patterns and ruptures of forced academic migration. At a later stage, it will be part of a comparative study on forced academic migration in which we will be able to make comparisons between different academic fields, answer questions regarding the extent to which age and seniority had an influence on successfully continuing a career in exile, as well as shed light on the situation of female scientists after their expulsion from their positions.

In the first section, we will digitally trace in some detail the direct results of the declaration of the Law for the Restoration of the Professional Civil Service on the mathematicians at Göttingen University and focus on the changing academic reputation of the Göttingen Institute for Mathematics.[16] In the second section, focusing on Switzerland and based on two examples, we explore the interdepend-

---

**14** "nodegoat," lab 1100, accessed February 9, 2021, https://nodegoat.net.

**15** Accessed January 11, 2022, https://forced-academic-migration.net/datapublications/datapublications.p/293.m/tag/Mathematics.

**16** The history of mathematics in Germany and in Göttingen has been studied extensively. See also: David E. Rowe, "Jewish Mathematics at Göttingen in the Era of Felix Klein," *Isis* 77 (1986): 427–49; Arnold Dresden, "The Migration of Mathematicians," *The American Mathematical Monthly* 49, no. 7 (1942): 415–29, accessed February 9, 2021, doi:10.2307/2303266; Colin R. Fletcher, "Refugee Mathematicians: A German Crisis and a British Response, 1933–1936," *Historia Mathematica* 13 (1986): 13–17; Louise Grinstein and Paul J. Campbell, eds., *Women of Mathematics: A Biographical Sourcebook* (New York: Greenwood Press, 1987); Max Pinl and Lux Furtmüller, "Mathematicians under Hitler," *The Leo Baeck Institute Yearbook* 18, no. 1 (1973): 129–82; David E. Rowe, "Klein, Hilbert and the Göttingen Mathematical Tradition," *Osiris* 2nd ser., 5, no. 1 (1989): 186–213.

encies of Swiss universities and/or professors working at Swiss universities within the transnational network of science and show to what extent our digital approach also stimulates questions about Swiss appointment policy(s). Swiss universities had a long history of appointing German scientists as a means to strengthen their academic reputation and therefore the Swiss and German academia were closely interlinked. In the course of World War I however, Swiss immigration policies became more restrictive. The field of mathematics serves as an example to study the changing interlinking of Swiss universities in the transnational network of academia before and during the National Socialist period, this also helps us to understand, to what extent Swiss universities were able and willing to hire exiled mathematicians.

## 2 A Period of Transition? The Göttingen Institute for Mathematics in the 1930s

The Institute for Mathematics at the Georg-August-Universität Göttingen has a long history of excellence, reaching back to the 18th century. Along with Berlin, Göttingen was one of Germany's, and one could argue the world's, main centers for mathematical research. Despite prejudice against Jews before Hitler's rise to power, the University of Göttingen hired several Jewish mathematicians, some of whom were full professors. In April 1933, when Law for the Restoration of the Professional Civil Service was passed, their positions were in jeopardy. So, in 1934, a mere year after the mass-dismissal, David Hilbert (1862–1943), the former head of the Göttingen Institute and the "grand old man of German mathematics" was asked by Bernhard Rust (1883–1945), the German minister of education, how mathematics at Göttingen was, now that it was free from Jewish influence. Hilbert allegedly replied: "There is no mathematics in Göttingen anymore."[17]

In this first section, we explore the question of whether and how this much-quoted statement by Hilbert can be verified based on our relational database, which not only includes the personal data of academics and their career steps linked to institutions but also information on the reasons for their changing positions. The question of the Göttingen Institute's global standing after 1933 thematically overarches this analysis.

---

**17** Constance Reid, *Hilbert: With an appreciation of Hilbert's Mathematical Work by Hermann Weyl*, 4th ed. (New York: Springer, 1970), 205.
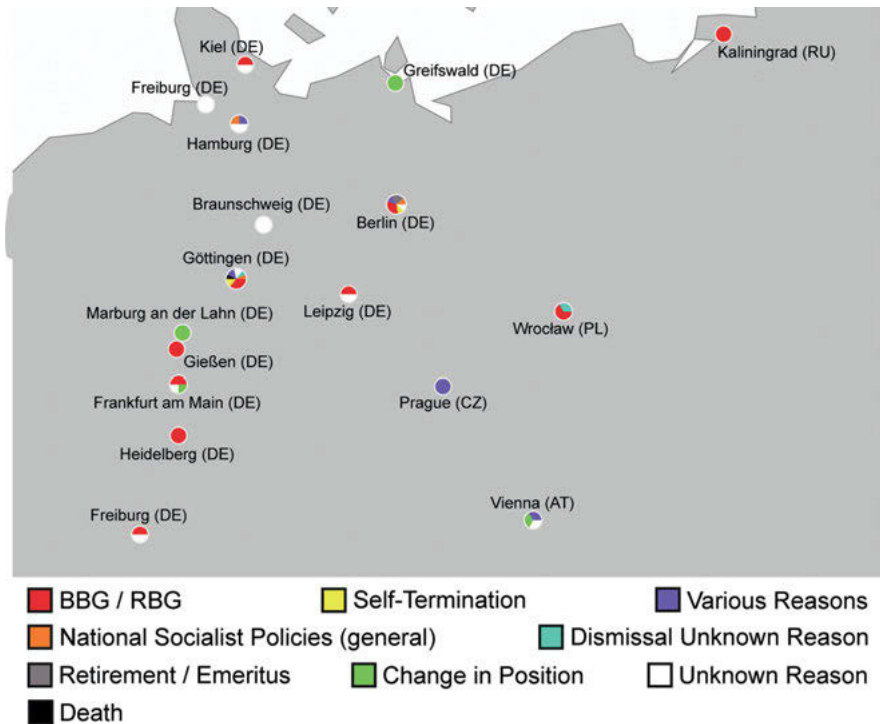
**Figure 1:** Reasons for termination at German universities, 1933–1939.

Figure 1 shows the various reasons for changes in professional positions at German universities between 1933 and the invasion of Poland in 1939.[18] Using pie charts in these figures allows an immediate assessment of not only the different reasons for quitting a position during this period but also of the differences between universities. For increased clarity, the various reasons for change have been grouped thematically as colored pie slices.

---

**18** Based on our classifications in nodegoat, the reasons for termination were grouped thematically. This is a brief explanation of the Figure 1 and 2 legends. BBG/RBG: Dismissals on the basis of the Law on the Restoration of Professional Civil Service of April 7, 1933 and the Reich Citizenship Law of September 15, 1935. National Socialist Policies (General): Cases attributable to National Socialist policies in general. Self-Termination: All persons who resigned from their positions. Dismissal Unknown Reason: Reason for dismissal could not yet be determined. Retirement/Emeritus: retired or emeritus. Change in Position: This category includes terminations due to a change of academic position; for example, if one had completed a qualification (PhD, habilitation, etc.). Various Reasons: Individual cases that do not belong to any of the above-mentioned categories (health reasons, outbreak of war, etc.). Reason Unknown: Cause of termination is unknown.

Particularly noticeable are the white areas, which indicate cases where the causes are still unknown, and the red slices, which stand for dismissals due to the Law for the Restoration of the Professional Civil Service (BBG) in 1933 and the Reich Citizenship Law (RBG) enacted in 1935. To increase the readability and enable us to classify and compare Göttingen to other German universities, the following map (Figure 2)[19] only focuses on reasons related to the National Socialists' seizure of power.[20] From the simplified visualization of the data in Figure 2, we can conclude that whereas a considerable number of universities, such as Frankfurt am Main or Heidelberg, lost their mathematicians due to the new legislation (i.e., the BBG and the RBG), in Göttingen the reasons were more varied.[21]

From Figure 2, we can indeed confirm the hypothesis that many mathematicians left Göttingen due to National Socialist policies in general, but not all of them left with manifest reference to the BBG or RBG.

Given its reputation, we have to assume that the departures of most of the faculty members and senior researchers at the Göttingen Institute of Mathematics had an impact on the academic standing of mathematics at the University of Göttingen. One way to trace this potential shift in the significance of the Göttingen Institute is to display the researchers' career steps before and after 1933. For this purpose, the career paths of the mathematicians were traced on a world map and presented in chronological sequence (Figures 3–5). The colored dots indicate career positions, meaning places of graduation or scientific appointments. Their size corresponds to the frequency of career positions at this location – the larger the dot, the more mathematicians' career steps are located at this institution. The lines connect the temporal and geographic dimensions, creating an interactive

---

**19** Interactive version of this figure is the same as Figure 1. By clicking on the different "reasons" in the legend it is possible to show or hide them: https://forced-academic-migration.net/datapublications/datapublications.p/293.m/17/mathematicians-reason-for-termination.

**20** Lemercier and Zalc emphasize the importance of simplifying visualizations as much as possible to facilitate their readability, which is true for the maps in this article. Claire Lemercier and Claire Zalc, *Quantitative Methods in the Humanities: An Introduction* (Charlottesville: University of Virginia Press, 2019), 127–28. However, simplifications also carry the risk of misinterpretation or overestimation of the factors presented. It is therefore extremely important to draw the reader's attention to the way in which the presentation has been simplified and what it does or does not show. For example, Figure 2 focuses only on certain reasons, and it should not be forgotten which other reasons for dismissal also occurred but are not essential for the current research interest.

**21** State of work in progress as of January 2021: the focus lies on Göttingen mathematicians. The online version of the visualization is continuously evolving.
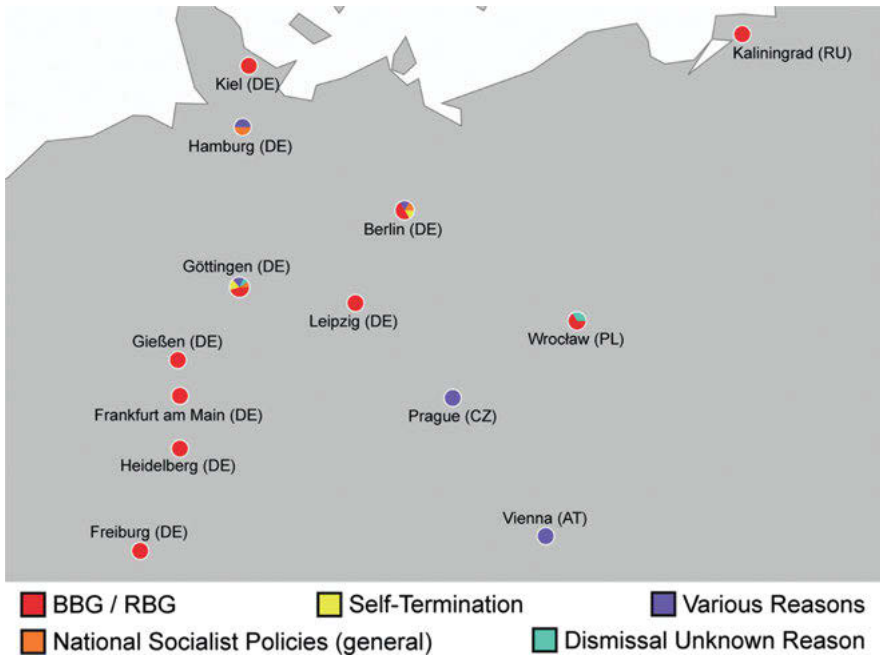
**Figure 2:** Selected reasons for termination at German universities, 1933–1939.

map of mathematicians' movements. In the online version of the map the movements are changing over time and the directions are visible too.[22]

Until the end of 1932 (Figure 3), many mathematicians' careers passed through Göttingen, represented on the one hand by the relatively large dot at the University of Göttingen and on the other hand by the many lines leading to and from Göttingen. In 1933, however, the picture changed abruptly (see Figure 4). Although Göttingen is still mapped, its position becomes marginal. The lines now tend to move away from Göttingen toward Switzerland, the United Kingdom, and, especially, the USA. It is worth noting that after 1933, many mathematicians, as shown in Figure 4, immigrated directly to the USA. Figure 5 confirms this trend also for the years 1935–1945, albeit with slight geographical shifts. Rather than leading from Germany directly to other states, first and foremost to the USA, the lines also connect dots in Switzerland and Great Britain with the USA.

---

**22** The interactive versions of Figures 3–5, accessed January 11, 2022: https://forced-academic-migration.net/datapublications/datapublications.p/293.m/16/geographic-movement-of-gttingen-mathematics.
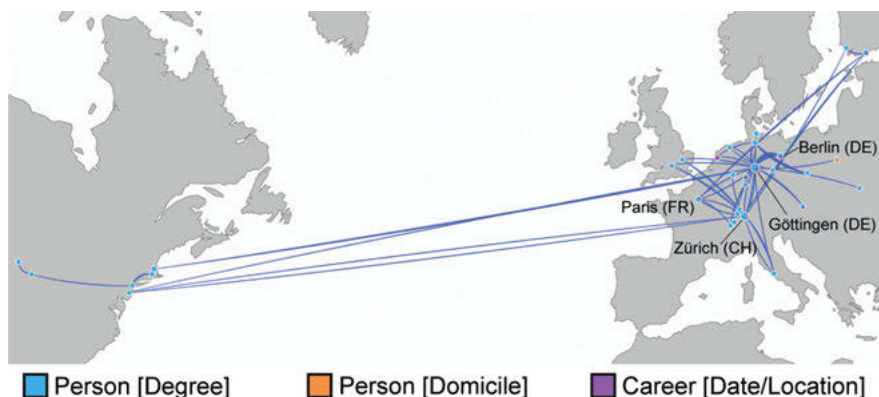
**Figure 3:** Movements of Göttingen mathematicians, 1925–1932.
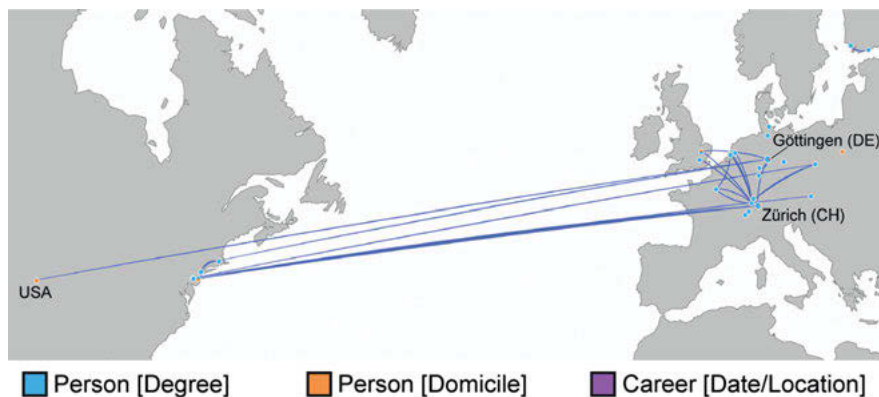


**Figure 4:** Movements of Göttingen mathematicians, 1933–1934.

Thus, Figure 4 suggests that the mathematicians' career paths often did not end at their first station of exile. Often, the academics had to move several times before settling down definitively, which follows from the shown movements in Figures 3–5. The questions of how many mathematicians managed to proceed directly to the USA and why this was the case remain to be answered in future research. However, connections between Göttingen and research institutions in the USA before 1933, as shown in Figure 3, can provide first insights: personal networks played a significant role in the process of (forced) academic
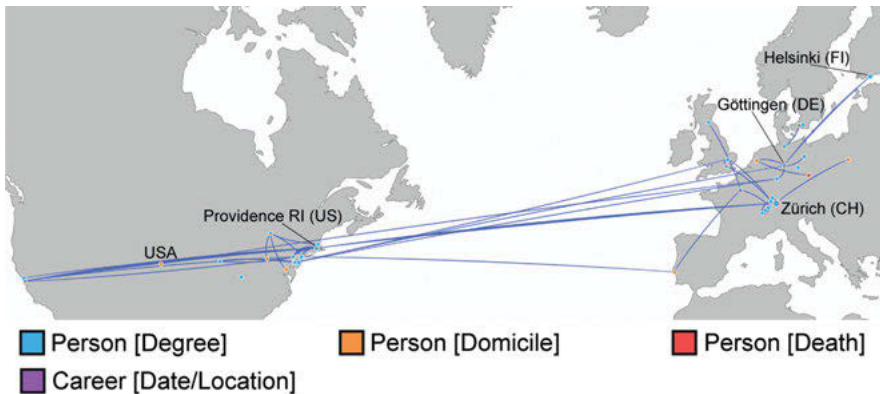
**Figure 5:** Movements of Göttingen mathematicians, 1935–1945.

migration.[23] The visualization emphasizes that as well as research institutions in the USA, universities in Switzerland seemed to be, at least temporarily, a center of attraction for mathematicians. This raises questions about the determining pull factors for the mathematicians' immigration decisions, such as immigration policies, aid organizations, personal and professional networks, and university policies. The relevance of these factors, particularly that of university policies in the emigration process of dismissed scientists, in this case, mathematicians, will be discussed in the following section based on the example of Switzerland.

# 3 Transnational Linkages and Local Idiosyncrasies: Mathematics Professors at Swiss Universities

In geographical, linguistic, and cultural terms, Switzerland was an obvious and important emigration destination for German-speaking academics. After April 7, 1933, Swiss university members received numerous personal visits and letters from German scholars (and from Swiss professors employed in Germany), who had been dismissed for racial and/or political reasons, hoping to take up teaching and research activities in Switzerland. How Swiss universities dealt

---

**23** The contact existing between Göttingen and the USA, visualized by gray lines between the University of Göttingen and various places in the USA, originates from Roland Richardson, who was a postdoctoral student with David Hilbert in Göttingen (nodegoat database).

with these forced academic migrants is the subject of this section, again focusing on the field of mathematics. In the following, we will focus on two examples: first, the interconnections of Swiss universities and professors working at Swiss universities within the transnational network of academia and, second, the development of Swiss appointment policy(s) in regard to exiled scholars. As in the first section, our analysis is built upon a nodegoat database, in this case with information about academics in Switzerland.

Figure 6 is a geographical visualization of the career paths of all mathematicians who held a full professorship at a Swiss university between 1900 and 1950.[24] Consequently, this map visualizes the (inter)national mobility of this group of mathematicians in the first half of the 20th century. The green dots symbolize the university cities, and the larger the dot, the more career path moves mathematicians made through the university. The sequences of the moves are indicated by the blue arrows.[25]

Due to the focus on Switzerland, the Swiss university cities are recognizable as the largest dots, particularly Zurich, since it is home not only to a university but also to the Swiss Federal Institute of Technology (ETH). This visualization illustrates the close ties between German and Swiss universities over the entire period. Particularly



□ University Cities          ■ Career Paths of Mathematicians

**Figure 6:** Career paths of full professors in mathematics at Swiss universities, 1900–1950.

---

**24** All academic positions (and, if known, the place of study) of professors at the universities of Basel, Bern, Fribourg, Geneva, Lausanne, Neuchâtel, Zurich, and the Swiss Federal Institute of Technology in Zurich are included; however, the data are not yet complete.

**25** Link to interactive visualization: Mathematics (full professors) at Swiss Universities: https://forced-academic-migration.net/datapublications/datapublications.p/293.m/18/mathematics-full-professors-at-swiss-unviersitites.

striking is the exchange with the important Göttingen Institute of Mathematics. A large number of mathematicians who were full professors at a Swiss university had completed part of their studies in Göttingen or held a faculty position there. Among them was the German mathematician Hermann Weyl (1885–1955). From 1913, he was a full professor at the ETH. He then returned to Göttingen in 1930 upon receiving a call from his former alma mater. Due to the increasing threat of National Socialism, Hermann Weyl, his wife Helene Weyl-Joseph (1893–1948), a Jewish writer and translator, and their children migrated to the USA. Through mediation, by Albert Einstein among others, he managed to obtain a position at Princeton University.[26]

The visualization also makes it apparent that there was not only movement to the USA, but also back to Europe. One reason for this is temporary research stays, such as the one of George Pólya (1887–1985), a mathematician whose parents converted from Judaism to Catholicism before his birth. A full professor of mathematics at the ETH, he went to Princeton University as a Rockefeller Fellow in 1933 and then returned to his position in Switzerland the same year. Like other scholars working in Switzerland, George Pólya used the connections established during his time abroad to finally emigrate to the USA in 1940 in response to the political situation in Europe and the widespread fear amongst people of Jewish descent in Switzerland that Nazi Germany might invade the country. There, he taught and researched at Stanford University until 1953.[27]

In addition to studying these transnational networks by means of geographical visualizations, developments in the appointment policy(s) of Swiss universities can also be made accessible through chronological visualizations, as Figure 7 illustrates. It depicts the total number of full professors in mathematics in Switzerland (vertical blue lines) and their distribution among various universities (horizontal colored lines/dots) between 1914 and 1950. Personnel changes in the mathematical institutes – appointments as well as departures – can be quickly

---

**26** Hermann Weyl studied mathematics in Munich and Göttingen. There he also received his doctorate and habilitation. From 1913 to 1930 he was a full professor at the ETH, and from 1930 to 1933 at the University of Göttingen. Preempting his dismissal, Weyl submitted his application for discharge in October 1933. He then worked at the Institute for Advanced Study in Princeton until 1951. Reinhard Siegmund-Schultze, *Mathematicians Fleeing from Nazi Germany: Individual Fates and Global Impact* (Princeton, NJ: Princeton University Press, 2009), 56.
**27** George Pólya was born in Budapest, Hungary, in 1887. He studied and received his doctorate in Budapest. From 1914, Pólya was a private lecturer at the ETH in Zurich and a full professor from 1928. He then emigrated to the USA, where he first worked at Brown University and then Stanford University, including previous research stays in Vienna, Paris, Cambridge, and Princeton. David Gugerli, Patrick Kupper, and Daniel Speich, *Die Zukunftsmaschine: Konjunkturen der ETH Zürich 1855–2005* (Zurich: Chronos-Verlag, 2005), 242; Gerald L. Alexanderson, *The Random Walks of George Pólya* (Washington, DC: Mathematical Association of America, 2000).

grasped through the visual representation. After a steady increase in the number of full professors in mathematics until 1933, years of stagnation and decline followed. From 1944 onwards, a renewed increase can be observed: vacant positions were filled and additional positions were created. This breakdown by university also allows the disciplinary focus of the universities to be filtered and compared. In addition, it is possible to make a subdivision according to gender to show the changing positions of women within the professoriate.[28]
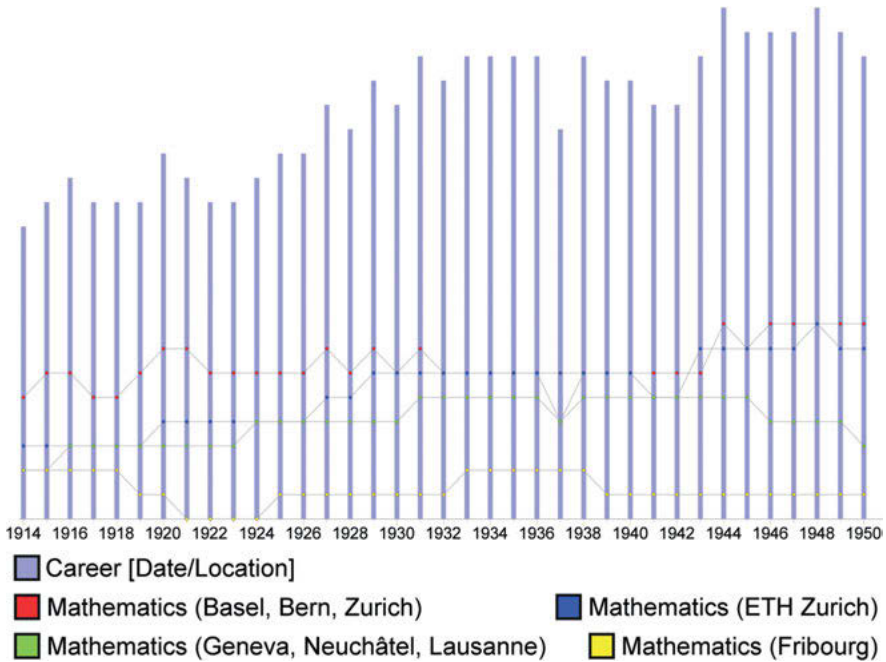


**Figure 7:** Total number of full professors in mathematics and their distribution amongst Swiss universities, 1914–1950.

Regarding the question of what appointment policy(s) the Swiss universities pursued from 1933 to 1945 and what positions were taken toward academic

---

**28** Sophie Piccard became the first female full professor in mathematics at a Swiss university in 1943. Christine Riedtmann, "Wege von Frauen: Mathematikerinnen in der Schweiz," in *Math. ch/100: Schweizerische Mathematische Gesellschaft 1910–2010*, ed. Bruno Colbois, Christine Riedtmann and Viktor Schroeder (Zurich: European Mathematical Society Publishing House, 2010), 403–21.

migrants, the visualization in Figure 7[29] indicates that there was only one appoint-ment of a full professor in the field of mathematics in 1933, that of the Austrian mathematician named Anton Huber (1897–1975), who was not an academic migrant. Huber had been an associate professor at the University of Freiburg in Switzerland since 1928. In 1935, he joined the *Nationalsozialistische Deutsche Arbeiterpartei (NSDAP)*-Switzerland,[30] and in 1938, after the *Anschluss* of Austria, he accepted a position at the University of Vienna.[31] Huber's appointment as a full professor was the last of a foreign mathematician until after the end of the war, which indicates that none of the mathematicians dismissed in Germany since 1933 could continue their scientific activities in Switzerland – at least not in the position of full professor. Support was provided, for example, in the form of temporary teaching assignments. Such was given to the Jewish-Swiss mathe-matician Paul Bernays by the ETH, upon being dismissed from his position at the University of Göttingen in 1933. Despite interventions by Weyl on Bernays behalf, the ETH did not promote Bernays until after the end of the war. Bernays received tenure in 1945 when he was appointed to associate professor. His academic posi-tions, however, did not correspond to his scientific importance.[32]

How these examples can be located within the appointment policy(s) of Swiss universities or to what extent they indicate a discipline- or university-specific appointment practice can only be answered once all the necessary data have been recorded. What is already apparent, however, is that the use of digital methods not only generates new research questions but also promotes stimulating and con-structive exchanges between different research projects.

---

[29] To obtain a clearer visualization, the universities were divided into groups: 1. German-speak-ing universities (Basel, Bern, and Zurich); 2. French-speaking universities (Geneva, Lausanne, and Neuchâtel); 3. the only Catholic and German–French speaking university, Fribourg; and 4. the Federal Institute of Technology in Zurich. The University of St. Gallen, which did not re-ceive the right to award doctorates until 1939, was not (yet) involved in these investigations. Link to interactive visualization: Distribution of mathematics (full professors) at Swiss univer-sities: https://forced-academic-migration.net/datapublications/datapublications.p/293.m/20/distribution-of-mathematics-op-at-swiss-universities.

[30] In 1932, the Swiss National Group of the NSDAP was formed and subordinated into the NSDAP Foreign Department in Germany. Its purpose was to integrate Germans living abroad into the National Socialist system, but it also served intelligence purposes. The NSDAP was only banned in Switzerland shortly before the end of the war. Catherine Arber, "Frontismus und Nationalso-zialismus in der Stadt Bern: Viel Lärm, aber wenig Erfolg," *Berner Zeitschrift für Geschichte und Heimatkunde* 65, no. 1 (2003): 7–8.

[31] Roman Pfefferle, *Glimpflich entnazifiziert: Die Professorenschaft der Universität Wien von 1944 in den Nachkriegsjahren* (Göttingen: V&R Unipress, 2014), 291.

[32] Future research will show if and how academic migrants gained access to other academic positions. Gugerli, Kupper, and Speich, *Die Zukunftsmaschine*, 240f.

# 4 Conclusion

Data visualizations, as shown in this article, are invaluable in answering "traditional" research questions about the consequences of Nazi politics for German academia with respect to emigration, routes of migration, changing reputations of research institutions, networks, and transfer of knowledge; they bring together all the different aspects of academic forced migration into a larger picture, as the example of mathematicians has shown. The forced dismissal of the scientists had tremendous impact on German institutions, as the Göttingen Institute of Mathematics lost its academic standing in the world, on individual academic biographies, and on institutions outside Germany. Personal and institutionalized networks play a key role in the process of academic migration as well as in exile. Yet, data visualization also enables us to detect unexpected events and patterns and thus generate new questions in the field of *Exilforschung*, such as differences between academic disciplines, the extent of multiple migrations, or the relevance of personal vis-à-vis institutional networks.

Moreover, digital tools such as nodegoat allow us to study the macro-level of academic forced migration as well as its micro-historical aspects, as our data model allows us to zoom in on the career paths and personal networks of individual scholars. We have employed a research-question driven digital approach, which exemplifies Miriam Rürup's plea, in the concluding roundtable of the January 2021 #DHJewish conference, to bring an "analogue perspective" to the digital. However, there are limitations to the digital approach. If we want to address more biographical issues, such as the experiences of individual exiled scholars or personal contingency management strategies, as raised by Adorno, we need to also integrate more traditional historical methods and the kind of qualitative approach that a human close reading of primary sources offers.

Migration is a core element of Jewish existence, from ancient history to modern days, and is thus a central research interest of academic Jewish Studies. Migration is being studied in the context of religious, literary, and cultural studies because the individual and collective experiences of migration of the Jewish people have found expression in religious and profane texts, in art, and in memoirs. Historical research into the topic covers a broad field of questions including political, social, and cultural changes as well as questions about gender and age, to mention just a few.

We understand our project on forced academic migration to be part of the field of *Exilforschung*. *Exilforschung* primarily deals with single and collective biographies of German-speaking émigrés, with their personal and professional experience in exile as well as their contribution to the cultural, industrial, and academic sphere. A tremendous amount of research has been conducted in this

field thus far and has helped us to understand the complexities and ambiguities of exilic existence. Our data-driven projects add to the field by making interconnected comparisons possible. So far, the history of exiled scientists is divided in studies on aid organizations or selected disciplines, there is research on questions like gender, publications on universities, academic associations, and there are books dealing with the consequences the expulsion had on German academia. Through a digital approach it is now possible to integrate these aspects in a more comprehensive picture, and thereby greatly enhance our understanding of the phenomenon of *Exil*.

# References

Adorno, Theodor W. *Minima Moralia: Reflections from Damaged Life*. London: Verso, 2005.

Alexanderson, Gerald L. *The Random Walks of George Pólya*. Washington, DC: Mathematical Association of America, 2000.

Arber, Catherine. "Frontismus und Nationalsozialismus in der Stadt Bern: Viel Lärm, aber wenig Erfolg." *Berner Zeitschrift für Geschichte und Heimatkunde* 65, no. 1 (2003): 3–62.

Bracher, Karl Dietrich. *Die deutsche Diktatur: Entstehung, Struktur, Folgen des Nationalsozialismus*. Frankfurt am Main: Verlag Ullstein, 1980.

Braese, Stephan. "Deutsche Sprache, jüdisches Exil – Optionen von 'Identität' nach 1933." In *Exilerfahrung und Konstruktionen von Identität 1933 bis 1945*, edited by Hans O. Horch, Hanni Mittelmann, and Karin Neuburger, 7–16. Berlin: De Gruyter, 2013.

Diner, Dan. "Einleitung." In *Tel Aviver Jahrbuch für deutsche Geschichte: Historische Migrationsforschung* 27, 1–4. Gerlingen, 1998.

Dresden, Arnold. "The Migration of Mathematicians." *The American Mathematical Monthly* 49, no. 7 (1942): 415–29. Accessed February 9, 2021, doi:10.2307/2303266.

Fletcher, Colin R. "Refugee Mathematicians: A German Crisis and a British Response, 1933–1936." *Historia Mathematica* 13 (1986): 13–17.

Grinstein, Louise and Paul J. Campbell, eds. *Women of Mathematics: A Biographical Sourcebook*. New York: Greenwood Press, 1987.

Grüttner, Michael. "Die deutschen Universitäten unter dem Hakenkreuz." In *Zwischen Autonomie und Anpassung: Universitäten in den Diktaturen des 20. Jahrhunderts*, edited by John Conelly, 67–100. Paderborn: Schöningh, 2003.

Gugerli, David, David Kupper, and Daniel Speich. *Die Zukunftsmaschine: Konjunkturen der ETH Zürich 1855–2005*. Zurich: Chronos-Verlag, 2005.

Krohn, Claus-Dieter, Gerhard Paul, Lutz Winckler, and Patrik von Zur Mühlen, eds. *Handbuch der deutschsprachigen Emigration 1933–1945*. Darmstadt: WBG, 2008.

Lemercier, Claire and Claire Zalc. *Quantitative Methods in the Humanities: An Introduction*. Charlottesville: University of Virginia Press, 2019.

Pfefferle, Roman. *Glimpflich entnazifiziert: Die Professorenschaft der Universität Wien von 1944 in den Nachkriegsjahren*. Göttingen: V&R Unipress, 2014.

Pinl, Max and Lux Furtmüller. "Mathematicians under Hitler." *The Leo Baeck Institute Yearbook* 18, no. 1 (1973): 129–82.

Pross, Helge. "Die geistige Enthauptung Deutschlands. Verluste durch Emigration." In *Nationalsozialismus und die deutsche Universität. Universitätstage 1966*, 143–55. Berlin: De Gruyter, 1966.

Reid, Constance. *Hilbert: With an Appreciation of Hilbert's Mathematical Work by Hermann Weyl*. 4th ed. New York: Springer, 1970.

Riedtmann, Christine. "Wege von Frauen: Mathematikerinnen in der Schweiz." In *Math.ch/100: Schweizerische Mathematische Gesellschaft 1910–2010*, edited by Bruno Colbois, Christine Riedtmann, and Viktor Schroeder, 403–21. Zurich: European Mathematical Society Publishing House, 2010.

Rowe, David E. "Jewish Mathematics at Göttingen in the Era of Felix Klein." *Isis* 77 (1986): 427–49.

Rowe, David E. "Klein, Hilbert and the Göttingen Mathematical Tradition." *Osiris* 2nd ser., 5, no. 1 (1989): 186–213.

Siegmund-Schultze, Reinhard. *Mathematicians Fleeing from Nazi Germany: Individual Fates and Global Impact*. Princeton, NJ: Princeton University Press, 2009.

Maja Hultman

# The GIS prism: Beyond the Myth of Stockholm's *Ostjuden*

**Abstract:** In this chapter, I argue that the GIS approach holds the potential to challenge historiographical master narratives in Jewish urban history. Using Stockholm's modern Jewish population as a case study, I propose that the digital, quantitative studies associated with GIS can be used as analytical prisms through which to explore qualitative sources. In the case of Stockholm's Jewry, this methodology allows for a re-examination of spatially inscribed tropes, particularly the so-called *Ostjude*.

I begin the article by describing the largely unchallenged historiographical idea that Stockholm's Jewish pre-1939 population was divided into two groups: the integrated, Reform, and northern-residing Jews, and the Eastern European, poor, orthodox, and southern-residing Jews – the *Ostjuden*. Introducing the analytical possibilities and methodological challenges of the GIS approach, I thereafter use ArcGIS to digitally map Jewish economic engagement with Stockholm's urban topography in relation to members of two synagogues, one Reform and one orthodox. The results show that the two religious groups utilized a unified geographical integration and created communal connections across religious barriers.

With this new framework in mind, I lastly turn to a newspaper article, written by a reformed Jew in 1905, that describes a shabbat service in the orthodox synagogue. Textual analysis reveals the author's construction of the spatially inscribed stereotypes previously mentioned, in particular the ostracized trope of the *Ostjude*, and their loose ties to the Jewish community's social reality. Thus, this chapter shows that the GIS approach is vital for understanding the Swedish Jewish community's creation of tropes to sustain inner-communal hierarchies.

**Keywords:** digital, settlement patterns, inner-communal relations, urban topography, Swedish Jews

---

# 1 Introduction

Exiting a staircase and standing on the threshold into a room on the right-hand side, a reporter from the Swedish daily newspaper *Daily News* encounters the following scene in the shul of Adass Jisroel on March 5, 1905:

> They read the prayer with fierce waddles and bow with their backs towards the congregation. They turn around, and now I see these two shapeless figures with head and a bigger part of the body covered by the white-blue striped shawl, and I feel closer to the Orient when they slowly, murmuring, waddle to and fro. There is something mystic, something ancient about it all, which makes me feel, despite the fact that I should find it all ridiculous, partly moved by these ceremonies, which the faithful have kept for centuries despite oppression, persecution, ridicule and contempt, ceremonies that might have been performed by my own forefathers some millennia ago in Solomon's own temple. But as I get back to the vestibule again, I am yet again in the twentieth century. The electric tram turns down the street, and unabashedly, I light my cigarette, even though it is Sabbath.[1]

The writer, using the signature "H V-n," describes a shabbat service at the shul, located in a rented room in a former Pietist girls' orphanage on the slum and industrial island of Södermalm in southern Stockholm. In the article, Hugo Vallentin – the Jewish editor behind the signature[2] – depicts members of Adass Jisroel as fundamentally different from "acclimatized" Jews, who supposedly attend the "grand" and "sophisticated" services in the purpose-built Great Synagogue in the northern, central parts of the Swedish capital. Indeed, according to Vallentin, members of Adass Jisroel and the Great Synagogue are socially and culturally dichotomized, divided by ethnic background, religious practices, and economic status. He inscribes these differences into the city's topography, locating integrated, Reform, and modern Jews in the northern district, and Eastern European, poor, and orthodox Jews in the southern district. On the shabbat morning of March 5, 1905, as Adass Jisroel's members with "dark eyes" and "long beards" pray, Hugo Vallentin, Swedish-born and among the Great Synagogue's third richest members,[3] exits the building, lights a cigarette, and hops onto a tram to get back to his home in northern Stockholm,[4] trusting that readers will under-

---

**1** H V-n, "De rättrogna," *Dagens Nyheter*, March 5, 1905, 3, at Swedish Royal Library (henceforward referred to as SRL). All subsequent quotes from "H V-n" originate from this article.
**2** Register of pseudonyms and signatures, SRL.
**3** Hugo Vallentin was born in 1860 in Gothenburg. In 1909, he paid 120 Swedish *kronor* for his membership to the Mosaic Congregation in Stockholm; see: Register of taxpayers for 1910, SE/RA/730128/01/A1a79, Swedish State Archive (henceforward referred to as SSA).
**4** According to the register of Stockholm's taxation records, available at Stockholm's City Archive (henceforward referred to as SCA), Hugo Vallentin lived on Tegnérgatan 37 in 1905 and 1909.

stand that he is breaking a shabbat rule, and, therefore, is not a member of the Jewish "colony" in southern Stockholm.

Is Vallentin's depiction of a dichotomized, spatially inscribed Jewish community in Stockholm in the beginning of the 20th century correct? The narrative has not lost its potency in the last hundred years and has been reproduced and cemented in Swedish Jewish historiography and historical memory. Although some researchers have noted that Stockholm's small Jewish population of some 3,000 people never established an Eastern European, Jewish urban district comparable to, for example, Scheunenviertel, the East End, Marais, and Leopoldstadt in Berlin, London, Paris, and Vienna respectively, the supposed division between integrated, Reform, northern-residing Jews and Eastern European, poor, orthodox, southern-residing Jews – Stockholm's version of the *Ostjude* – persists. In this chapter, I show that geographical analysis through a GIS (Geographic Information Systems) approach challenges the established historiographical narrative and accentuates a "more chaotic, more contingent, more fluid, more uncertain, more ambiguous, more immediate – in other words, more fully human"[5] experience of Jewish life in Stockholm. Facilitating an analysis based on quantitative sources, GIS stresses the pivotal role of Stockholm's topography, uniquely defined by islands, islets, peninsulas, inlets, bays, streams, and straits, in shaping Jewish negotiations on social integration, religious practices, and internal relations. As I will argue, a GIS approach is highly useful to debunk spatially inscribed, historiographical master narratives and highlight the complexity of the Jewish, urban experience.

In order to show the promise of GIS I firstly introduce the historiographical narrative of Stockholm's dichotomized Jewish population pre-World War II. The synagogues mentioned in Vallentin's article, Adass Jisroel and the Great Synagogue, receive particular emphasis, since their divergent geographical positions and religious affiliations have previously served as a foundational framework for understanding Jewish life in the Swedish capital as divided. The potential of the GIS approach to challenge historiographical discourses, despite current discussions on its limitations, is secondly highlighted, and I argue that GIS functions as a prism through which qualitative sources, such as Vallentin's article, should be evaluated and examined. I also introduce some of the multiple methodological challenges associated with the digital software tool ArcGIS, and how I have

---

**5** David J. Bodenhamer, "Chasing Bakhtin's Ghost: From Historical GIS to Deep Mapping," in *The Routledge Companion to Spatial History*, ed. Ian Gregory, Don DeBats, and Don Lafreniere (London: Routledge, 2018), 539.

approached them in my own study. Having thus explained the historiographical setting and the digital methodology, I revisit the Jewish community in Stockholm and compare the economic status of members of Adass Jisroel and the communal organ the Mosaic Congregation, which mainly supported the Reform community belonging to the Great Synagogue. As I will show, the GIS approach efficiently determines their economic and geographical similarities, raising the question whether they should indeed be regarded as two groups. With this new framework in mind, I lastly return to Vallentin's article and his, and the Swedish Jewish historiography's, narrative of a dichotomized Jewish community, and argue that Stockholm's *Ostjude* was an ostracized trope created as a result of the community's internal hierarchy, shaped by the local, urban fabric.

## 2  Stockholm's Jewry: A Narrative of Dichotomy

Echoing editor Hugo Vallentin's narrative, scholars in Swedish Jewish history have ever since the publication of *Judarnas historia i Sverige* (The History of Jews in Sweden) by Jewish historian Hugo Valentin (not to be mistaken for the above-mentioned Hugo Vallentin) in 1924, crudely divided Jewish migration groups moving to Sweden from the 18th century until the Holocaust into two groups. The first Jew to be allowed to practice Judaism settled in Stockholm in 1775, and relatives, business colleagues, and friends subsequently moved from mainly Mecklenburg to Sweden. Danish Jews simultaneously extended their businesses into Sweden. This first migration group from central and northern Europe is known for aiding the development of Sweden's commercial consumption through sugar and textile productions, the foundation of banks and modern industries, and involvement in the construction of railroads and extraction of ore.[6] Later generations became influential in cultural spheres as well, being artists, writers and intellectuals, financial donors to the capital's Concert Hall, constructers of Stockholm's modern architecture, publishers of the national literary canon, and owners of department stores and famous restaurants.[7] During the decades before the emancipation in

---

**6** See, for example: Fredric Bedoire, *Ett judiskt Europa: Kring uppkomsten av en modern arkitektur, 1830–1930* (Stockholm: Carlsson, 1998), 18–32; Anna Brismark and Pia Lundqvist, "Sidensjalar och socker: Judiska näringsidkares betydelse för konsumtionsrevolutionen i Sverige," in *Från sidensjalar till flyktingmottagning: Judarna i Sverige – en minoritets historia*, ed. Lars M. Andersson, and Carl Henrik Carlsson (Uppsala: Historiska institutionen, 2013), 17–47; Hugo Valentin, *Judarna i Sverige* (Stockholm: Bonniers, 1964), 85–95.

**7** See, for example: Mia Kuritzén Löwengart, *En samhällelig angelägenhet: Framväxten av en symfoniorkester och ett konserthus i Stockholm, cirka 1890–1926* (Uppsala: Acta Universitatis

1870, many Jews in Stockholm moved from the medieval district on the island of Stadsholmen, today called Old Town, into modern developments on the northern mainland.

The purpose-built Great Synagogue was built in the vicinity of this newly established commercial and cultural city center. Constructed during the 1860s, the new synagogue received an architectural design linked to the emerging Reform Judaism, aimed at modernizing Jewish religion in Europe. In contrast to its central location in the previous synagogue in Old Town, the bimah was instead placed at the eastern end of the ship in the new synagogue, close to the Aron Hakodesh, facing an organ and space for a choir at the other end. Its inauguration ceremony on September 16, 1870 included sermons and hymns in Swedish, the latter accompanied by the organ. Throughout the 19th and 20th centuries, the rabbi's attire was furthermore inspired by Christian priests, "confirmation" for girls was added, and many members discarded the practices of kashrut and shabbat. The Great Synagogue became a space for modernized religious practices, inspired by the Protestant environment in Sweden.

The second group of Jewish migrants included some 3,000 to 4,000 Jews migrating from mainly Grodno, Kovno, Suwalki, Vilna, and Vitebsk to Sweden between the 1860s and 1917.[8] With the arrival of Eastern European Jews, Stockholm's Jewish population grew from 900 people in 1870, to 1,250 Jews in 1890, and 2,600 Jews in 1910.[9] Although many worked as peddlers,[10] Eastern European Jews were not a static group of poor Jews. Rita Bredefeldt shows that they used both industrial and trade sectors to advance economically and integrate socially.[11] Individuals from Eastern Europe were also successful in commerce, and some were famous artists,

---

Upsaliensis, 2017); Jacqueline Stare, *Porträtt: Speglingar av svensk judisk kultur* (Stockholm: Fallmarks, 1993), 26–52.

**8** These areas had not experienced widespread pogroms but rather crop failures, famine, and military conscriptions. In 1890, about a fifth of those who had arrived in the last 30 years had used Sweden as a transit nation, continuing their travels towards America. The migration to Sweden largely stopped in 1917 with the introduction of passports. See: Carl Henrik Carlsson, "Immigrants or Transmigrants? Eastern European Jews in Sweden, 1860–1914," in *Points of Passage: Jewish Migrants from Eastern Europe in Scandinavia, Germany, and Britain, 1880–1914*, ed. Tobias Brinkmann (New York: Berghahn, 2013), 55–56; Carl Henrik Carlsson, *Medborgarskap och diskriminering: Östjudar och andra invandrare i Sverige, 1860–1920* (Uppsala: Acta Universitatis Upsaliensis, 2004), 31.

**9** Ingvar Svanberg and Mattias Tydén, *Tusen år av invandring* (Stockholm: Dialogos, 1992), 237.

**10** For a description of life as a Jewish peddler in Sweden, see: Jacqueline Stare, ed., *Judiska gårdfarihandlare i Sverige* (Stockholm: Judiska Museet, 1996).

**11** Rita Bredefeldt, *Judiskt liv i Stockholm och Norden: Ekonomi, identitet och assimilering, 1850–1930* (Stockholm: Stockholmia, 2008), 59.

such as painter Isaac Grünewald. Reconstructing an ordinary day on Södermalm at the beginning of the 20th century, Mats Franzén notes that although Eastern European Jews and Italians were both considered as strangers in the streetscape, the former integrated occupationally and culturally faster.[12] Putting together these historical fragments, it is clear that Eastern European Jews arriving in Stockholm were generally determined to integrate into the capital's society.

In response to the construction of the Reform-aligned Great Synagogue, Adass Jisroel was created in 1870, and as mentioned above, it was set up in a Pietist orphanage on the island of Södermalm, south of Old Town. Hugo Vallentin describes in his article that the bimah was placed at the center of the room rented on the first floor, and a women's balcony was constructed. Correspondence from Adass Jisroel's leaders throughout the first decades of the 20th century describes the place as a "shul," "synagogue," "minyan," and "chevra" respectively, and its religious orientation is ambiguous, ranging from "traditional" and "orthodox" to "conservative." Dedicated male members, however, participated in morning and evening prayers, families kept kosher, but also had to navigate the rules of shabbat. Some, for example, used the tram to get from their home to shabbat services in the synagogue,[13] although the payment of the tickets was a break of shabbat rules. Despite the unclear religious stance of Adass Jisroel, it was noticeably more traditional than the Great Synagogue.

Research has only scratched the surface of these two groups' relationship beyond Vallentin's and Valentin's suggested dichotomy. Carl Henrik Carlsson's study on Eastern European Jewish acquisition of Swedish citizenship at the turn of the 20th century shows that the first group, as leaders of the Mosaic Congregation, more often than not wrote favorable character descriptions to Stockholm's Police to assist the process.[14] Studies on Jewish philanthropy in Stockholm similarly show that wealthier Jews aided some of the poorer families arriving from Eastern Europe. They built cheap apartments, established a youth leisure center, provided clothes, shoes and food for children, and formed a summer youth camp on an island in Stockholm's archipelago.[15] Anna Besserman argues that "the only

---

12 Mats Franzén, *Den folkliga staden: Söderkvarter i Stockholm mellan krigen* (Lund: Arkiv förlag, 1992), 144–54.

13 Interview with Henry Blideman (October 21, 2014).

14 Carlsson, *Medborgarskap och diskriminering*.

15 See, for example: Anna Besserman, "'…Eftersom nu en gång en nådig försyn täckts hosta dem upp på Sveriges gästvänliga stränder': Mosaiska Församlingen i Stockholm inför den östjudiska invandringen till staden, 1860–1914," *Scandinavian Jewish Studies* 5 (1984): 32; Svante Hansson, *Flykt och överlevnad: Flyktingverksamhet i Mosaiska församlingen i Stockholm, 1933–1950* (Stockholm: Hillelförlaget, 2004), 62.

existent contact space was between the needy and the philanthropists."[16] Other scholars and public educators have adopted and circulated this dichotomized image between established Jews and Eastern European Jews, inscribing supposed religious, ethnic, and economic differences into the urban space. Some echo Vallentin and argue, without much research evidence, that there existed a territorial division between the groups, mainly due to Eastern European Jews' "particular Jewish life and institutions, which, together with their foreignness and their relatively poorer economic circumstances, contrasted with the more established Jews of the cities."[17] The discourse was reconfirmed in scholarly work as late as 2004.[18] In late 2021, the City Museum in Stockholm offered guided tours on "Northern Jews and Southern Jews," describing the Jewish community as "divided into two groups into the late 1930s" in their marketing.[19]

On the other hand, Carl Henrik Carlsson notes that "the real correlation 'Western Jew'/Reformed Jew respectively 'Eastern Jew'/orthodox Jew is not necessary as strong as it seems or is usually emphasized."[20] In an article on Jewish migrants making Sweden their new home as a consequence of World War I, he calls orthodox yet economically affluent Jacob Ettlinger, chairman of Adass Jisroel from the end of the 1910s, a "typically untypical Jew" and an example of how the dichotomized identity "templates are not always true."[21] Suggesting that there were more "typically untypical" Jews, Carlsson was the first to question the historical memory of Stockholm's Jews.

The internal structure of the Mosaic Congregation also connected the majority of Jews living in Stockholm, whether rich or poor, reform or orthodox. The Mosaic Congregation was the only official Jewish institution and up until 1910 legally responsible for providing information for the taxation of Stockholm's Jews to the city's municipality. Its membership was, therefore, the only way for a Jewish individual to join a community with societal status in the non-Jewish sphere. The membership itself could, however, be too expensive for less wealthy Jews, as it was calculated in proportion to one's income, and difficult for migrants

---

**16** Besserman, "Eftersom nu en gång," 29.

**17** Joseph Zitomersky, "The Jewish Population in Sweden, 1780–1980: An Ethno-demographic Study," in *Judiskt liv i Norden*, ed. Gunnar Broberg, Harald Runblom, and Mattias Tydén (Uppsala: Acta Universitatis Upsaliensis, 1988), 114.

**18** Hansson, *Flykt och överlevnad*, 44.

**19** "Norrjudar och söderjudar," *Stadsmuseets kalendarium*, accessed September 13, 2021, https://stadsmuseet.stockholm.se/kalendarium/2021/11/16/norrjudar-och-soderjudar-sodermalm/.

**20** Carlsson, *Medborgarskap och diskriminering*, 34.

**21** Carl Henrik Carlsson, "Judiska invandrare i Sverige under första världskriget. Fyra fallstudier," in *Första världskriget i svenska arkiv: Årsbok för Riksarkivet och Landsarkiven 2014*, ed. Carl Henrik Carlsson (Stockholm: TMG Tabergs, 2014), 168–70.

to obtain, since Swedish citizenship was a prerequisite for membership from 1882 on. Because the Swedish government discriminated against Eastern European Jewish naturalization processes, Swedish citizenships were difficult for migrants to achieve.[22] Consequently, the presence and influence of poorer, migrant Jews in the Mosaic Congregation were inhibited.

My previous study on the construction of the Great Synagogue in the 1860s relatedly portrays the power relation between the two groups, and how they negotiated the future sacred space. As leaders of the Mosaic Congregation, established Jews did not invite Eastern European Jews into the planning process, and even ignored and belittled their wishes on the inclusion of a mikveh in the new synagogue.[23] The negotiation, although unbalanced, shows that the two groups had more complex roles to play than those of philanthropists and the needy, with Eastern European Jews actively striving to influence how the community's shared spaces would be shaped. As I will show, the Swedish *Ostjude* presented in Hugo Vallentin's article from 1905 – the poor, orthodox, Eastern European Jew living on Södermalm – which has been reproduced and cemented by both Swedish Jewish historiography and Stockholm's heritage industry, does not align with the Jewish community's actual topographical engagement with Stockholm. Instead, I argue that the use of a GIS approach as framework for understanding the Jewish community's relationship to urban topography firstly, helps us to locate spatial stereotypes, and secondly, in combination with qualitative sources, suggests the importance of said stereotypes for internal structures and communal relations.

# 3 GIS: Analytical Possibilities and Methodological Challenges

Although GIS is fundamental for developing my argument, critics have pointed out its limitations in exploring human complexity. GIS tools involve the capture, plotting, and analysis of geographical data through collected, quantifiable information at a specific time and place. They have, therefore, been accused of producing "static residential spaces,"[24] trapping mobile environments and actors within the boundaries of a chosen date and location of mapping, yielding precise,

---

**22** Carlsson, *Medborgarskap och diskriminering*.

**23** Maja Hultman, "The Construction of the Great Synagogue in Stockholm: A Space for Jewish and Swedish-Christian Dialogues," *Arts: Synagogue Art and Architecture* 9 (2020): 33.

**24** Don Lafreniere and Jason Gilliland, "'All the World's a Stage': A GIS Framework for Recreating Personal Time-Space from Qualitative and Quantitative Sources," *Transactions in GIS* 19 (2015): 226.

organized snapshots, void of lived experience or agency.[25] Encountering these methodological problems, researchers in digital and spatial humanities have in recent years further fine-tuned the use of GIS to support digital analyses of human ambiguity. Some add physical attributes, relational information, and narrative sources to the various software programs to analyze social networks in the streetscape over time. Others use PGIS (Participatory Geographic Information Systems) or deep mapping to involve the public in studying temporal multilayers of meanings attached to a small geographical area.[26] I, however, dispute a quick dismissal of GIS and argue that by using it as a prism for examining qualitative sources, human complexity can indeed be captured and explored.

Within Jewish Studies, the GIS approach has proved vital in contradicting accepted historical narratives. The assumed predominance of urban Jewries in the Byzantine Empire has, for example, been disproved with a GIS-supported analysis.[27] Mapping population patterns of Jewish communities in pre-World War II Poland, Malgorzata Hanzl finds individual motives for migration and settlement.[28] Other studies have used GIS to explore the shape of Jewish/non-Jewish relations and inner-communal relations in urban settings.[29] Although providing a static picture of human life, GIS can clearly be used to debunk historical myths and deliver a geographical grid to be filled with human experiences, encounters, practices, and meanings. By placing a GIS-informed, quantitative analysis at the forefront of this study, the stereotypical trope of the Swedish *Ostjuden*, found in qualitative sources, is revealed as a constructed, reproduced, and cemented image, a product of social webs and internal hierarchies. The use of GIS, in tandem

---

**25** Trevor M. Harris, "From PGIS to Participatory Deep Mapping and Spatial Storytelling: An Evolving Trajectory in Community Knowledge Representation in GIS," *The Cartographic Journal* 53 (2016): 319–21.

**26** Harris, "From PGIS to Participatory Deep Mapping and Spatial Storytelling."

**27** Gethin Rees, Nicholas de Lange, and Alexander Panayotov, "Mapping the Jewish Communities of the Byzantine Empire Using GIS," in *Migration and Migrant Identities in the Near East from Antiquity to the Middle Ages*, ed. Justin Yoo, Andrea Serbini, and Caroline Barron (New York: Routledge, 2019), 116.

**28** Malgorzata Hanzl, "Jewish Communities in Pre-war Central Poland as an Example of a Self-Organising Society," in *Computational Science and Its Applicants – ICCSSA 2017: 17th International Conference, Trieste, Italy, July 3–6, 2017, Proceedings, Part 3*, ed. Osvaldo Gervasi et al. (Cham: Springer, 2017), 231.

**29** See, for example: Mary Anne Poutanen and Jason Gilliland, "Mapping Work in Early Twentieth-Century Montreal: A Rabbi, a Neighbourhood, and a Community," *Urban History Review* 45 (2017): 7–24; Máté Rigó, "Ordinary Women and Men: Superintendents and Jews in the Budapest Yellow-Star Houses in 1944–1945," *Urban History* 40 (2013): 71–91.

with qualitative sources, sheds light on the ambiguity and complexity of human experiences.

In this chapter, I will use GIS to map Jewish residences in Stockholm in 1909 and 1935. The geographical distribution of Jewish homes is analyzed in relation to economic and religious aspects. The specific geographical locations of Adass Jisroel and the Great Synagogue are shown to be part of a complicated and unbalanced web of communal relations, and as such, they stand as an example of the religious, ethnic, and economic complexity that existed within Stockholm's Jewish community before World War II. As the economic links and spatial resemblances between the Mosaic Congregation and Adass Jisroel become clear, I will argue that their members were more similar than previously believed.

Using a GIS approach is not without practical and methodological challenges, however. Choosing to work with the software program ArcGIS, the georeferencing of historical maps, placing historical images on top of today's geographical grid, was a particularly time-consuming task, and the final result received a rather high RMS (Root Mean Square) error number, which potentially offsets the exact location of Jewish residences.[30] Furthermore, the application of ED (Enumeration District) boundaries, which allow the map to be divided into urban districts, and thereby for subsequent comparisons between different districts, proved too time-consuming for this project.

A second challenge was the collection of quantitative data from a variety of sources, located in different archives. Data was collected from the Mosaic Congregation's membership list in 1939, records of "alien faith believers" in local church parishes, and local city taxation records from 1935.[31] Together, they provide information on residential address, name, gender, marital status, age, occupation, birthplace, and, in case of a membership in the Mosaic Congregation, internal taxation. To make sure that I included only people who defined themselves as Jewish, I firstly collected names from the Mosaic Congregation's membership list and those who were not members of the Mosaic Congregation but defined themselves as "Mosaic" in public records related to local church parishes. Although I came across other Jewish-sounding names in the taxation list, I did not include them, since I had no proof of their Jewish identification. Different name spellings in handwritten sources made the manual cross-reference between the list of Jews,

---

**30** The historical map that was used is: H. Hellberg, and A.E. Påhlman, "1934 års karta över Stockholm med omgivningar (1917–1934)," SE/SSA/Stockholmskartor, SCA. The RMS error number is 29,1225.

**31** Mosaic church books at SSA; Christian parish books of Adolf Fredrik, Engelbrekt, Gustav Vasa, Jakob, Johannes, Katarina, Matteus, Oscar, Sofia, and S:t Göran at SCA; Stockholm's taxation records at SCA.

gathered from the membership list and church parishes, and their addresses from the city taxation records difficult, time-consuming, and likely incomplete. Still, I was able to localize 49.8% (2,488 adults) of Stockholm's approximately 5,000 Jews – adults and children – in 1935,[32] which gives me a representative sample of the whole Jewish population.

It was not easy to locate which of the 2,488 Jewish individuals that were affiliated to Adass Jisroel. Going through the private archive related to Adass Jisroel's chairman Jacob Ettlinger, I came across an undated document that listed members of the shul.[33] It includes two lists, one typewritten and one handwritten, possibly suggesting that the individuals named had variously strong or limited affiliations. From its position in the archival folder, as well as some cross-referenced addresses in city taxation records, it is likely that the lists were compiled in the 1940s. Using this document to deduce Adass Jisroel's members, I found the addresses of those who lived in Stockholm and were adults in 1935. I am thereby able to compare their settlement pattern in relation to that of the Mosaic Congregation's members.

The last methodological hurdle was related to the lack of material. Whether it is due to Stockholm's historically non-unitary approach towards the administration of non-Protestant inhabitants in the beginning of the 20th century, or the loss of archival material during a later stage, records of "alien faith believers" do not exist for all local parishes. Some parish books were, furthermore, unavailable for research in 2018, when I collected data for the project, due to Swedish law on public confidentiality.[34] This resulted in missing data from five of Stockholm's 15 geographically defined parishes in 1935.[35] Many Jewish individuals were, however, found multiple times in the ten available records, meaning that non-members of the Mosaic Congregation frequently moved across parish borders. Tests on the robustness of digital, quantitative network analysis using pools of data with missing components also show that results are not as affected by missing data as previously presumed.[36] Although voices from Jews in poorer

---

**32** Clemens Maier-Wolthausen gives Stockholm a Jewish population of 7,000 individuals in 1939 in his book *Zuflucht im Norden: Die Schwedischen Juden und die Flüchtlinge, 1933–1941* (Göttingen: Wallstein Verlag, 2018), 300; and Pontus Rudberg states that 3,063 Jewish refugees had arrived in Sweden in 1939 in *The Swedish Jews and the Holocaust* (London: Routledge, 2017), 111. The Jewish population can therefore be deduced to some 5,000 Jews in 1935.

**33** List of possible members of Adass Jisroel, SE/RA/720483/5/1, SSA.

**34** The Swedish law 2009:400 states that archival documents must be older than 70 years to be published online.

**35** The unavailable parishes were: Högalid, Maria, Hedvig Eleonora, Storkyrkan, and Kungsholmen.

**36** Yann C. Ryan and Sebastian E. Ahnert, "The Measure of the Archive: The Robustness of Network Analysis in Early Modern Correspondence," *Journal of Cultural Analytics* 7 (2021): 57–88.

circumstances and/or lacking a Swedish citizenship are missing, the missing data potentially has only a modest effect on the analytical results.

These methodological challenges notwithstanding, I use ArcGIS to compare and contrast the geographical distribution of members of both the Mosaic Congregation and Adass Jisroel, as well as determine their connections. The results from the analysis of quantitative, geographical data reveal insights that overturn scholars' previous understanding of Jewish traditional life in Stockholm. This geographical analysis of a Jewish population is, to my knowledge, the first within Jewish Studies that endeavors to use GIS to explore religious minorities and internal hierarchies in urban environments. Using class as the analytical entry point into the Jewish population's relationship with Stockholm's urban fabric and the Jewish community's religious spaces,[37] the following section will showcase the potency of GIS to question historical myths, reveal communal webs, and define internal hierarchies.

# 4 Revisiting Stockholm's Jewry: Connections Across the City

In the beginning of the 20th century, the city of Stockholm was composed of 14 islands in Sweden's third largest lake. Naturally, the inlets, bays, and channels divided urban districts from each other, and although bridges were continuously built, the topographical divisions exacerbated and cemented socio-economic differences. The islands of Södermalm in the south and Kungsholmen to the west became increasingly associated with industry and slum. In contrast, modern Haussmann-inspired urban developments changed the north-eastern parts of the city. The topography of Stockholm accentuated socio-economic division and was used by Hugo Vallentin, among others, to describe the *Ostjuden* who attended Adass Jisroel, located on Södermalm, as poor. But who were Adass Jisroel's members? Did they differ that much from Hugo Vallentin and the other members of the Mosaic Congregation? Were they mainly poor and Eastern European? And did they cluster on Södermalm?

---

**37** Discussing Max Weber's distinction between class and status, Till van Rahden uses the concept of class to research income and actual economic possibilities and limitations among Breslau's Jewish community. See: Till van Rahden, *Jews and Other Germans: Civil Society, Religious Diversity, and Urban Politics in Breslau, 1860–1925* (Madison: University of Wisconsin Press, 2008), 23.

The first analysis (Figure 1) shows the residential distribution of the Mosaic Congregation's taxed members in 1935 in relation to their economic capabilities. The Mosaic Congregation's membership fee was, as previously mentioned, proportionally determined by each member's income, and an analysis of the fee can therefore reveal the economic dimension of urban relationships. As can be observed in the map, residences belonging to members paying 0–100 *kronor*, 68% of the total population, were located across the whole city. Clearly, apartments for the working-classes existed in all districts, making no urban space impenetrable for the lower classes. In this sense, economic standing did not automatically lead to topographical segregation. On the other hand, the Mosaic Congregation's wealthiest members did not set up homes in any urban district. The residences were located mainly in the modern, north-eastern area of Stockholm, but also along waterfront promenades connected to the main water bodies. Some also lived on Kungsholmen and Södermalm, in areas with views of the capital's natural features and urban skyline. Consequently, if members of the Mosaic Congregation could afford it, they would establish homes in urban districts that allowed for an immersion into the modern or uniquely natural environment of Stockholm.
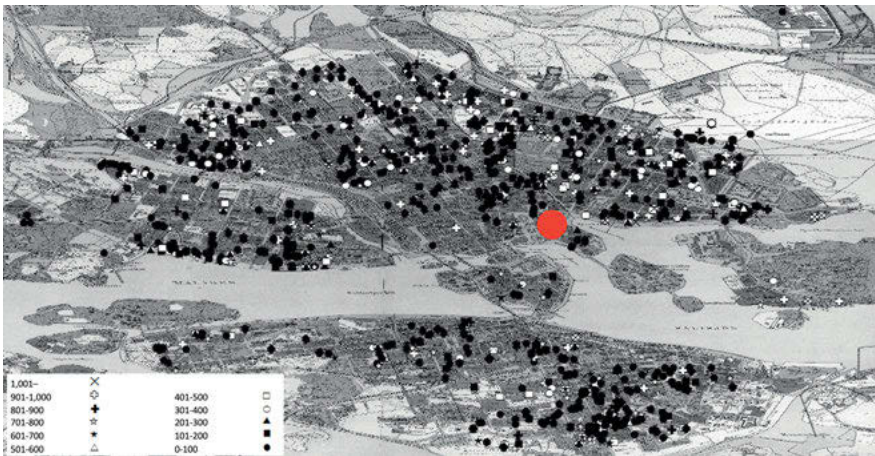


**Figure 1:** The residential distribution of the Mosaic Congregation's taxed members (in Swedish kronor) in 1935. The red circle represents the location of the Great Synagogue.
*Source*: H. Hellberg and A.E. Påhlman, 1934, City Archive, Stockholm.

A similar pattern is found among Adass Jisroel's members in 1935 (Figure 2). Curiously, members of Adass Jisroel were oftentimes also members of the Mosaic Congregation. As mentioned earlier, although the Mosaic Congregation was mainly run by descendants of the first migration group at the beginning of the 20th century, and was decidedly linked to Reform Judaism, its membership was the only way for an individual to join a Jewish community with legal, societal status. The social status of the membership seemingly encouraged traditional practitioners to join, connecting people to two religious places. Therefore, the links between the Mosaic Congregation and traditional practitioners were many. Adass Jisroel's wealthiest members also lived in the modern, north-eastern area of Stockholm. While members untaxed by the Mosaic Congregation – meaning that they were either too poor to afford to pay the membership fee, were not yet Swedish citizens, or did not want to be affiliated to either the Reform orientation or the leadership – indeed clustered on the southern parts of Södermalm, the rest of the members were scattered all over Stockholm. It would have taken between 30 and 45 minutes for some to reach the shul for morning and evening prayers. Indeed, the residential distribution of traditional practitioners essentially mirrors the pattern of the larger community, showing that neither class nor the location of religious institutions were large influences on the membership of the shul. Instead, both Reform and traditional Jews adopted the same relationship to the city and settled in more modern and developed areas if economic means allowed.
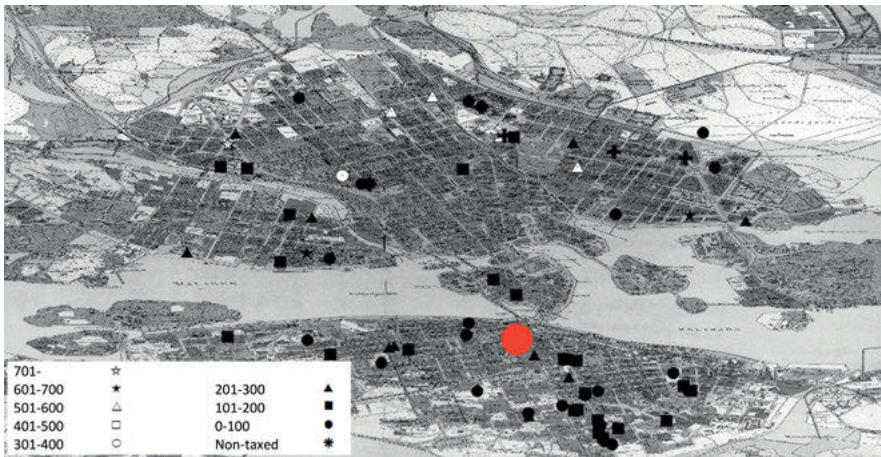


**Figure 2:** The residential distribution of Adass Jisroel's members, taxed or non-taxed (in Swedish kronor) by the Mosaic Congregation, in 1935. The red circle represents the location of Adass Jisroel.
*Source*: H. Hellberg and A.E. Påhlman, 1934, City Archive, Stockholm.

Class clearly shaped Jewish life in Stockholm. While not limiting Jews from settling wherever they wanted, it was simultaneously used to gain access to modern and status-infused areas in the city. The ArcGIS analysis shows that the geographical distribution of Jewish residences was not influenced by religious orientation. Many attending Adass Jisroel in the 1930s were not only living off Södermalm, but also had comparably comfortable economic means. In other words, Adass Jisroel's members were not only poor Jews living on the southern island. Was this also the case when Hugo Vallentin visited Adass Jisroel in 1905? No membership list for Adass Jisroel exists from the time, but taking Hugo Vallentin's, and Swedish historians', argument that the shul's members had no links to the Mosaic Congregation, I have also analyzed the residential pattern of Jews not taxed by, and therefore without membership in, the Mosaic Congregation in 1909. Again, an equal residential distribution across the whole city emerges (Figure 3). Whether the untaxed people were poor, did not hold a Swedish citizenship, or did not want to be a part of a community synonymous with Reform Judaism, they were living in diverse urban districts. In contrast to, for example, Viennese Jews,[38] Jews in Stockholm did not settle or resettle in the vicinity of other Jewish individuals with a similar economic status or religious affiliation. Instead, when setting up
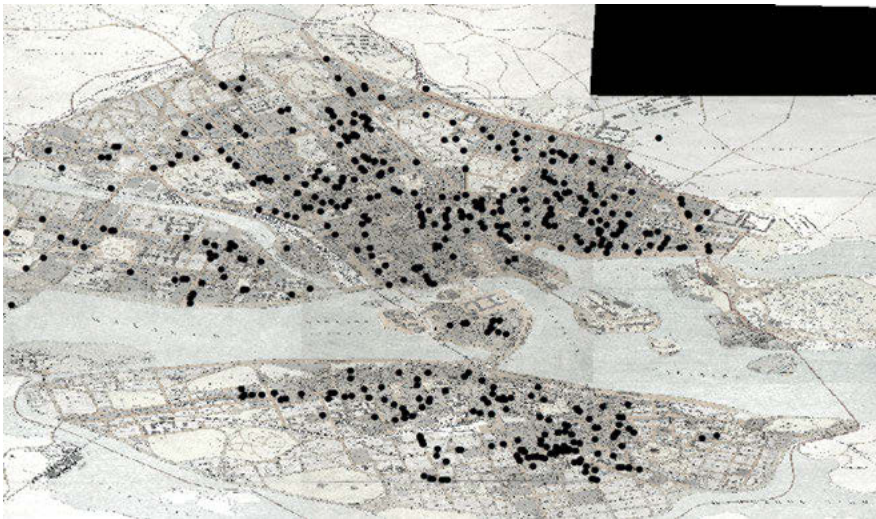


**Figure 3:** The residential distribution of Jews not taxed by the Mosaic Congregation in 1909.
*Source*: Alfred Bentzer, 1909, City Archive, Stockholm.

---

**38** Marsha L. Rozenblit, *The Jews of Vienna, 1867–1914: Assimilation and Identity* (Albany: State University of New York Press, 1984), 71–98.

their private homes in Stockholm, Jews aligned to class divisions topographically inscribed into the urban landscape.

Accordingly, there is a stark difference between the image of the *Ostjude* delivered by Hugo Vallentin, and the cemented historiographical narrative of a dichotomized population, and the result of the ArcGIS mappings conducted in this research. While previous assumptions have crudely inscribed two divergent stereotypes into Stockholm's topography, the digital analysis shows that there were connections and similarities between traditional Jews and the Reformist Mosaic Congregation. Not only had members of both synagogues a similar distribution of economic means, but many individuals were also members of both communities. Although the synagogues were positioned in two different urban districts, their members were both socially interlinked and formed similar relationships to the urban landscape. Despite understanding and practicing Judaism differently, the members of Adass Jisroel and the Mosaic Congregation thus had more in common than has previously been communicated through the image of the *Ostjude* and the idea of a dichotomized Jewish population.

With this spatial framework in mind – a unified geographical integration and communal connections across religious barriers – Hugo Vallentin's exotified description of Adass Jisroel's members gains new meaning. In his article, Vallentin argues that the general, traditional practitioner exclaims "He is leading us towards baptism!" when the Mosaic Congregation's rabbi makes "trivial simplifications and swedicizes the [religious] ritual." Many members of Adass Jisroel did, however, belong to both religious communities. Similarly, Vallentin describes that "many of its members are retailers, with or without a shop. There is also one or two artisans, and their numbers are probably growing due to the commendable efforts of the upper class to lead the younger generation onto new paths." Although his and earlier scholarly studies' emphasis on the traditional Jew as a Jew in need of philanthropic aid, sources show that Adass Jisroel had members of similar economic caliber as the Mosaic Congregation. Lastly, Vallentin notes that most members of Adass Jisroel lived on Södermalm, while my spatial analysis has determined the obvious similarities of unclustered residential patterns between the Mosaic Congregation and Adass Jisroel.

Vallentin was, clearly, not relaying the social reality of Jewish life in Stockholm, but rather a constructed, imaginary image. As portrayed by Steven Aschheim, an antipathy flourished among integrated German Jews in the 19th and 20th centuries towards the traditional, and often also Eastern European, Jew.[39] The Swedish

---

**39** Steve E. Aschheim, *Brothers and Strangers: The East European Jew in German and German Jewish Consciousness, 1800–1923* (Madison: University of Wisconsin Press, 1982).

*Ostjude* was similarly a trope created by established Jews to belittle poorer Jewish immigrants and their cultural practices. Although members of Adass Jisroel were as economically and spatially integrated as members of the Mosaic Congregation, they were associated with the industrial and slum topography of Södermalm. This created a trope that, in tandem with philanthropy and the exclusion from membership in the Mosaic Congregation, ostracized Eastern European Jews, downplayed the importance of traditional rituals, and belittled the agency of poorer Jews. Using a GIS approach to explore the topographical context of Vallentin's article, an underlying internal hierarchy, determined by the social and religious prejudice of the Mosaic Congregation's bourgeois members, is exposed. The quantitative, digital analysis reveals the connections between Stockholm's Jews across geographical, economic, and religious borders. When exploring qualitative sources relaying personal experiences of Jewish space through this framework, the contemporary condescension and ignorance of this Jewish social fabric instead ignites questions of how and why stereotypical tropes emerged, and how they influenced people's relations with the Jewish community and the larger city. In other words, the GIS approach exposes the complex and ambiguous experience of being Jewish in Stockholm at the beginning of the 20th century, and its prism can guide and direct future scholars' explorations of Jewish urban life.

# 5 Conclusion

Starting with Vallentin's prejudiced depiction of a shabbat service in Adass Jisroel in 1905, this chapter has, with the help of ArcGIS, excavated the topographical reality of the shul's members in the beginning of the 20th century. I argue that today's historiographical narrative of divergent *Ostjuden* and established Jews has little bearing in historical results of social connections and unified economic behaviors, and is a persisting, monolithic image that does not reflect Adass Jisroel's community at the time. Using a historically informed GIS approach allows us to go beyond the imaginary *Ostjude*, a product of class-based power structures inscribed in Stockholm's topography. Through quantitative and data-driven, digital methods, we can explore the complexity and ambiguity that defined Jewish life in Stockholm at the beginning of the 20th century.

Neatly situated in the recent trend of scholarly work on Jewish local, urban communities,[40] this research's emphasis on quantitative data and use of digital

---

**40** To mention only a few examples from the pool of excellent research: Barbara E. Mann, *A Place in History: Modernism, Tel Aviv, and the Creation of Jewish Urban Space* (Stanford, CA: Stan-

methods is relatively rare.[41] While most urban historians in this field base their research on various and multivocal qualitative material, I would argue that topographical patterns derived from quantitative sources provide vital frameworks for contextualizing and understanding such research. My joint use of quantitative data and digital methodology in performing spatial analysis based on archival primary sources can hopefully encourage other researchers of Jewish, urban communities to dive deeper into the Jewish relationship with urban settings across the globe, despite the computational literacy and time-consuming data collection needed. What other – or similar – stereotypical tropes can we uncover, and what patterns of spatial integration do we find across modern Europe? Is there perhaps a Jewish border-crossing, transnational practice of urban integration and inner-communal relations that awaits future exploration with GIS?

Digital humanities indeed hold further possibilities for in-depth explorations of Jewish inner-communal relations through the topographical lens. While the GIS approach reveals internal hierarchies and the use of imagined tropes within Stockholm's Jewish community, deep mapping, as applied by scholars in spatial humanities,[42] can help to further explore disparate Jewish spatial experiences of Stockholm, as well as other cities. Developed to excavate temporal multilayers of meanings attached to a small geographical area through multifarious methodologies and sources, deep mapping is a digital and methodological entry point into "spatially framed identities."[43] It emphasizes the use of non-traditional sources, such as folklore, memories, and art, to reach beyond hierarchies of knowledge. It is a methodology that puts multivocality at the forefront, and promotes a "nuanced, non-reductionist"[44] view of the world, in order to "amplify the voices of marginalized stakeholders, both socially and ecologically."[45]

---

ford University Press, 2006); Natan M. Meir, *Kiev, Jewish Metropolis: A History, 1859–1914* (Bloomington: Indiana University Press, 2010); Devin E. Naar, *Jewish Salonica: Between the Ottoman Empire and Modern Greece* (Stanford, CA: Stanford University Press, 2016).

**41** Van Rahden produces excellent research from his quantitative data, and he also attempted to produce a digital, geographical analysis, but lost his work. See: van Rahden, *Jews and Other Germans*.

**42** William Least Heat-Moon, *PrairyErth (a Deep Map)* (Boston: Mariner Books, 1999); Les Roberts, "Deep Mapping and Spatial Anthology," *Arts* 5, no. 1 (2016): 1–8.

**43** David J. Bodenhamer, John Corrigan, and Trevor M. Harris, "Introduction," in *Deep Maps and Spatial Narratives*, ed. David J. Bodenhamer, John Corrigan, and Trevor M. Harris (Indianapolis: Indiana University Press, 2015), 3.

**44** Harris, "From PGIS to Participatory Deep Mapping and Spatial Storytelling," 320.

**45** Selina Springett, "Going Deeper or Flatter: Connecting Deep Mapping, Flat Ontologies and the Democratizing of Knowledge," *Humanities* 4 (2015): 624.

Since deep mapping not only includes but centers digital research on minorities and individuals positioned in societal peripheries, it would be specifically useful for uncovering and analyzing the different meanings attached to, for example, Adass Jisroel and the Great Synagogue. I have collected a variety of qualitative material, such as newspaper articles, architectural designs, photographs, poems, and private letters, that reveal different uses, emotions, and imaginations of Jewish spaces in Stockholm. The use of a deep mapping approach could enable the visualization of multiple, disparate spatial stories on top of the quantitative, data-driven analysis of the GIS approach and help to find new connections or disconnections between them. Furthermore, the online publications of such mappings, for instance with a tool like ArcGIS, enables us to invite readers (or users) to join in the academic endeavor, and thereby, facilitate open-ended, public-related research.[46] Deep mapping can thus not only help to further destabilize the current master narrative of a dichotomized Jewish population in Stockholm, as in the specific example of this article, but more generally highlight the complexity of Jewish urban life and the experiences of previously marginalized voices.

# References

## Primary Sources

Conducted interviews
    Henry Blideman. Stockholm, October 21, 2014.
Stockholm's City Archive (*Stockholms stadsarkiv*)
    Christian Parish Books (SE/SSA/0001, 0009, 6001, 6003, 6009, 6021, 6024, 6025, 6030)
    Local City Taxation Records from 1909 and 1935
    Maps
Swedish Royal Library (*Kungliga biblioteket*)
    Newspaper *Dagens Nyheter*
    Register of pseudonyms and signatures
Swedish State Archive (*Riksarkivet*)
    Archive of Jewish Community in Stockholm (SE/RA/730128)
    Personal archive of Jacob Ettlinger (SE/RA/720483)

---

**46** As is one of the goals of deep mapping, see: Bodenhamer, Corrigan, and Harris, "Introduction," 3.

## Websites

Stadsmuseets kalendarium. "Norrjudar och söderjudar." Accessed September 13, 2021.
    https://stadsmuseet.stockholm.se/kalendarium/2021/11/16/norrjudar-och-soderjudar-
    sodermalm/.


## Secondary Sources

Aschheim, Steve E. *Brothers and Strangers: The East European Jew in German and German
    Jewish Consciousness, 1800–1923*. Madison: University of Wisconsin Press, 1982.
Bedoire, Fredric. *Ett judiskt Europa: Kring uppkomsten av en modern arkitektur, 1830–1930*.
    Stockholm: Carlsson, 1998.
Besserman, Anna. "'…Eftersom nu en gång en nådig försyn täckts hosta dem upp på Sveriges
    gästvänliga stränder': Mosaiska Församlingen i Stockholm inför den östjudiska
    invandringen till staden, 1860–1914." *Scandinavian Jewish Studies* 5 (1984): 13–38.
Bodenhamer, David J. "Chasing Bakhtin's Ghost: From Historical GIS to Deep Mapping."
    In *The Routledge Companion to Spatial History*, edited by Ian Gregory, Don DeBats, and
    Don Lafreniere, 530–43. London: Routledge, 2018.
Bodenhamer, David J., John Corrigan, and Trevor M. Harris. "Introduction." In *Deep Maps and
    Spatial Narratives*, edited by David J. Bodenhamer, John Corrigan, and Trevor M. Harris,
    1–5. Indianapolis: Indiana University Press, 2015.
Bredefeldt, Rita. *Judiskt liv i Stockholm och Norden: Ekonomi, identitet och assimilering,
    1850–1930*. Stockholm: Stockholmia, 2008.
Brismark, Anna and Pia Lundqvist. "Sidensjalar och socker: Judiska näringsidkares betydelse
    för konsumtionsrevolutionen i Sverige." In *Från sidensjalar till flyktingmottagning:
    Judarna i Sverige – en minoritets historia*, edited by Lars M Andersson and Carl Henrik
    Carlsson, 17–47. Uppsala: Historiska institutionen, 2013.
Carlsson, Carl Henrik. "Immigrants or Transmigrants? Eastern European Jews in Sweden,
    1860–1914." In *Points of Passage: Jewish Migrants from Eastern Europe in Scandinavia,
    Germany, and Britain, 1880–1914*, edited by Tobias Brinkmann, 47–62. New York:
    Berghahn, 2013.
Carlsson, Carl Henrik. "Judiska invandrare i Sverige under första världskriget. Fyra fallstudier."
    In *Första världskriget i svenska arkiv: Årsbok för Riksarkivet och Landsarkiven 2014*, edited
    by Carl Henrik Carlsson, 152–77. Stockholm: TMG Tabergs, 2014.
Carlsson, Carl Henrik. *Medborgarskap och diskriminering: Östjudar och andra invandrare i
    Sverige, 1860–1920*. Uppsala: Acta Universitatis Upsaliensis, 2004.
Franzén, Mats. *Den folkliga staden: Söderkvarter i Stockholm mellan krigen*. Lund: Arkiv förlag,
    1992.
Hansson, Svante. *Flykt och överlevnad: Flyktingverksamhet i Mosaiska församlingen i
    Stockholm, 1933–1950*. Stockholm: Hillelförlaget, 2004.
Hanzl, Malgorzata. "Jewish Communities in Pre-war Central Poland as an Example of a
    Self-Organising Society." In *Computational Science and Its Applicants – ICCSSA 2017:
    17th International Conference, Trieste, Italy, July 3–6, 2017, Proceedings, Part 3*, edited by
    Osvaldo Gervasi, Beniamino Murgante, Sanjay Misra, Giuseppe Burroso, Carmelo M. Torre,

Ana Maria A. C. Rocha, David Taniar, Bernady O. Apduhan, Elena Stankova, and Alfredo Cuzzocrea, 224–38. Cham: Springer, 2017.

Harris, Trevor M. "From PGIS to Participatory Deep Mapping and Spatial Storytelling: An Evolving Trajectory in Community Knowledge Representation in GIS." *The Cartographic Journal* 53 (2016): 318–25.

Hultman, Maja. "The Construction of the Great Synagogue in Stockholm: A Space for Jewish and Swedish-Christian Dialogues." *Arts: Synagogue Art and Architecture* 9 (2020): 22–42.

Kuritzén Löwengart, Mia. *En samhällelig angelägenhet: Framväxten av en symfoniorkester och ett konserthus i Stockholm, cirka 1890–1926*. Uppsala: Acta Universitatis Upsaliensis, 2017.

Lafreniere, Don and Jason Gilliland. "'All the World's a Stage': A GIS Framework for Recreating Personal Time-Space from Qualitative and Quantitative Sources." *Transactions in GIS* 19 (2015): 225–46.

Least Heat-Moon, William. *PrairyErth (a Deep Map)*. Boston: Mariner Books, 1999.

Maier-Wolthausen, Clemens. *Zuflucht im Norden: Die Schwedischen Juden und die Flüchtlinge, 1933–1941*. Göttingen: Wallstein Verlag, 2018.

Mann, Barbara E. *A Place in History: Modernism, Tel Aviv, and the Creation of Jewish Urban Space*. Stanford, CA: Stanford University Press, 2006.

Meir, Natan M. *Kiev, Jewish Metropolis: A History, 1859–1914*. Bloomington: Indiana University Press, 2010.

Naar, Devin E. *Jewish Salonica: Between the Ottoman Empire and Modern Greece*. Stanford, CA: Stanford University Press, 2016.

Poutanen, Mary Anne and Jason Gilliland. "Mapping Work in Early Twentieth-Century Montreal: A Rabbi, a Neighbourhood, and a Community." *Urban History Review* 45 (2017): 7–24.

Rahden, Till van. *Jews and Other Germans: Civil Society, Religious Diversity, and Urban Politics in Breslau, 1860–1925*. Madison: University of Wisconsin Press, 2008.

Rees, Gethin, Nicholas de Lange, and Alexander Panayotov. "Mapping the Jewish Communities of the Byzantine Empire Using GIS." In *Migration and Migrant Identities in the Near East from Antiquity to the Middle Ages*, edited by Justin Yoo, Andrea Serbini, and Caroline Barron, 100–118. New York: Routledge, 2019.

Rigó, Máté. "Ordinary Women and Men: Superintendents and Jews in the Budapest Yellow-Star Houses in 1944–1945." *Urban History* 40 (2013): 71–91.

Roberts, Les. "Deep Mapping and Spatial Anthology." *Arts* 5, no. 1 (2016): 1–8.

Rozenblit, Marsha L. *The Jews of Vienna, 1867–1914: Assimilation and Identity*. Albany: State University of New York Press, 1984.

Rudberg, Pontus. *The Swedish Jews and the Holocaust*. London: Routledge, 2017.

Ryan, Yann C. and Sebastian E. Ahnert. "The Measure of the Archive: The Robustness of Network Analysis in Early Modern Correspondence." *Journal of Cultural Analytics* 7 (2021): 57–88.

Springett, Selina. "Going Deeper or Flatter: Connecting Deep Mapping, Flat Ontologies and the Democratizing of Knowledge." *Humanities* 4 (2015): 623–36.

Stare, Jacqueline, ed. *Judiska gårdfarihandlare i Sverige*. Stockholm: Judiska Museet, 1996.

Stare, Jacqueline. *Porträtt: Speglingar av svensk judisk kultur*. Stockholm: Fallmarks, 1993.

Svanberg, Ingvar and Mattias Tydén. *Tusen år av invandring*. Stockholm: Dialogos, 1992.

Valentin, Hugo. *Judarna i Sverige*. Stockholm: Bonniers, 1964.

Zitomersky, Joseph. "The Jewish Population in Sweden, 1780–1980: An Ethno-demographic Study." In *Judiskt liv i Norden*, edited by Gunnar Broberg, Harald Runblom, and Mattias Tydén, 99–125. Uppsala: Acta Universitatis Upsaliensis, 1988.

Piergabriele Mancuso

# Archival Research, Virtual Reality, and 3D Modeling: Toward a Comprehensive Reconstruction of the Ghetto of Florence

**Abstract:** The Ghetto of Florence was established by Cosimo I de' Medici, Tuscany first Grand Duke, in 1570, a few decades after those of Venice (1516) and Rome (1555). Belonging to the Medici, the Florentine ghetto was under the direct control of the Medici administrators, whose main task was not only to supervise it physically-architecturally but also to ensure its financial and economic feasibility. From a documentary-archival standpoint, this led to the production of an extremely rich set of papers, hundreds of volumes now stored at the National Archive in Florence, consisting of accounting books and financial documents, contracts, maps, and cadastral registers, chronologically stretching from the late 16th century to the late 19th century when the ghetto – like many Jewish ghettos in post-Unitarian Italy – was demolished. Studying and analyzing together all the data that these documents provide (especially maps and textual descriptions), it is possible to reconstruct the ghetto exactly as it was, as well as it changed during the centuries.

**Keywords:** early modern Jewish Italian history, Florentine jewry, ghetto of Florence, De' Medici family, virtual reconstruction

## 1 Introduction and a Preamble

Upon completion of a comprehensive survey at the National Archive in Florence of all cadastral and accounting papers produce by the Medici bureaucracy between the mid-16th and mid-18th centuries and concerning the real estate properties of the Medici family, the Eugene Grant Project Jewish History Program (hereafter EGJHP) at the Medici Archive Project discovered more than one hundred bulky volumes concerning the Ghetto of Florence, probably the richest, most detailed, and chronologically most extended documentation about an Italian ghetto. Consisting of various and heterogeneous materials, as we will see in more detail herein – including financial-accounting papers, blueprints, maps and topographies, and extremely detailed textual descriptions of virtually every single segment of the Florentine ghetto – and shedding potential light on virtually every aspect of Jewish people in the ghetto, in 2018 the EGJHP launched the Ghetto Mapping Project (GMP), an ambitious research project aiming to reconstruct the architectural features, economic-financial trends,

and demographic trends of the Ghetto of Florence. Due to the heterogeneous character of these materials, whose study requires exegetical expertise in different fields, and especially given the uninterrupted flows of additional materials found in other archival sites (expanding the project chronology so as to reach the late 19th century, when the Ghetto of Florence was demolished), the GMP remains a work in progress. The aim of this chapter is to offer a succinct but hopefully comprehensive historical outline of the Ghetto of Florence, with a focus on its establishment in the late 16th century, to describe the archival materials located up until now, ultimately defining the state of the art of our research while also outlining very briefly future research perspectives.[1]

## 2  Florence 1570: Genesis and Development of an Early Modern Ghetto

Established by Cosimo I de' Medici in 1570, the Ghetto of Florence is one of the oldest Italian European ghettos, the third after that of Venice in 1516 and that of Rome in 1555. Located in the very heart of the city center, at an almost equal very short distance from Santa Maria del Fiore and Brunelleschi's Dome and Palazzo Vecchio, the political barycenter of Medici Florence, the ghetto faced Piazza del Mercato Vecchio (Old Market Square), one of the most important popular trading spots of the old city, yet certainly not a marginal portion of the urban fabric. The making of the Ghetto of Florence did not require any major transformation. Consisting mostly of a group of adjacent old buildings whose foundations dated back to the Middle Ages, the ghetto's only really relevant intervention was to turn the two main entrances – one facing the Old Market, the other one leading to Via dei Succhiellinai (via Roma), another major artery of the city center – into two gates that, according to the ghettoization chart of 1570, should have been closed at night and supervised by Christian guards.

---

**1** The provisional results of the ongoing research were presented in P. Mancuso, "The Politics of Segregation in Grand Ducal Florence: The Ghetto of Florence, from Material Fall to Virtual Rebirth – a Short Presentation," in, *Judisches Kuturerbe (re-)praesentieren – Judisches Kulturerbe*, ed. Katrina Kessler et al., vol. 2, 45–66 (Braunschweig: Netwerk judisches Kultuerbe c/o Beit Tfila – Forschnugsstelle fur judische Architektur, 2019); P. Mancuso and L. Vigotti, "From Centuries-Old Squalor: The Ghetto of Florence, from History to Virtual Life. Introduction to the Ghetto Mapping Project," *Rivista di Letteratura Storiografica Italiana* 1 (2017): 123–34; P. Mancuso and L. Vigotti, "Reconstructing a Lost Space: The Ghetto Mapping Project at the MAP," *Materia Giudaica* 22 (2017): 221–32.

The real problems in the making of the ghetto were of a legal and bureaucratic nature. First, ghettoizing the Jews, which meant forcibly gathering them from all over Tuscany and – in the case of Florence – forcibly relocating families that had already developed strong and productive ties with the surrounding Christian majority, could be perceived as a major political and judicial-juridical *vulnus* and weakening of the socio-political prerogatives also by the Christian subjects. Nothing judicially cogent and serious enough to justify such an action had come from the survey of the Jewish population of Tuscany conducted since 1569 by the Magistrato Supremo headed by Carlo Pitti, and the only solution was to "forge a case," pretending that the vast majority of Jewish money lenders had violated the banking charters to such an extent and so seriously as to force the state itself to enact physical and juridical separation between Jews and Christians.

Second, although exerting absolute (today one would say dictatorial) power, the grand ducal role was still under the scrutiny of the old governing bodies, especially the Senate, as in the case of the 1567 vote to impose badges on Jews. To make the ghetto a private Medici property and enable the Medici family to enjoy the profits that would be earned in the future, especially in terms of fiscal revenues (as envisaged by Pitti in his different anti-Jewish plan, although he had initially pushed for complete expulsion of the Jews), the area where the ghetto would be established first had to be confiscated by the state and on behalf of the general interest and only then transferred to the *Scrittoio delle Regie Possessioni*, the official register of the Medici real estate properties. As clearly explained by Siegmund, such a property switch took the form of a financial "triangulation": the Christian owners would be fully refunded not with real currency, but with the financial interests generated by equivalent sums of money (paid by the state) that the owners were forced to deposit into the Florentine Monte di Pietà, at that time no longer a charitable institution but one of the Medici's financial banking tools. This resulted in a double benefit for the Medici, who "acquired" the ghetto and its future fiscal and rent revenues at basically no cost. In addition, the Medici politically aligned with the Roman Curia that in the same year (1570) had granted Cosimo the grand ducal dignity. The economic-financial liabilities of all this mostly fell on the state's coffers (providing funds to turn the Old Market's set of old houses into a Jewish ghetto), and obviously on the Jews forced to reside inside the ghetto and pay house rents (by law about 30% higher than in the rest of the city).

Initially amounting to no more than 500–600 people, the Jewish population of the Florentine ghetto rapidly grew until, in 1704 under the rule of Cosimo III de' Medici (1642–1723), it required the creation of an additional area (Ghetto Nuovo) on the northern side of the old ghetto.

At the end of the 19th century, between 1884 and 1892 and after the process of political unification of the Italian Peninsula, the entire historical city center,

then including the area of the two ghettos, underwent a general, massive, and very debated general urban plan, pompously called the "Risanamento,"[2] in order to change radically the physiognomy of the old city. Since the early 19th century, Jews had gradually left the ghetto and, by the end of the century, it had become one of the most degraded areas of the city, with the small Jewish houses now being occupied by sub-proletariats and criminals of various types. Demolishing the area meant getting rid of the shameful past and present (see Figure 1).[3]

# 3 Archival and Extra-Archival Sources

As with any other Medici private real estate properties, the ghetto was heavily supervised, meticulously controlled by the police authorities, and under the strict and direct control of the *Scrittoio delle Regie Possessioni*, whose main aim was to ensure the economic profitability of the estate and take careful note of anything occurring within and directly or indirectly affecting the physical-architectural and financial conditions of the Medici private properties. The *Scrittoio* work of ghetto supervision stretched from 1588 to 1808, when the French revolutionary army and temporary government lifted all restrictions on Jews (which were then partially re-implemented with the return of the Lorena dynasty), on the whole amounting to 95 volumes that can be divided into five main categories:
1. "Entrate e uscite del ghetto" (Revenues and Expenditures)
2. "Registri di acconcimi" (Registers of Damages and Reparations)
3. "Registi di creditori e debitori del Ghetto" (Debits and Credits)
4. "Amministrazione del ghetto" (Accounting Ledgers)
5. "Descrizioni del Ghetto and Piante" (Textual Descriptions and Maps)

Although the first four categories deal with different aspects of the same topic – basically the balance between expenditures and revenues – the fifth category holds an independent typological position, consisting mostly of textual descriptions of the entire ghetto, of virtually every single space, from the smallest storage

---

**2** Literally "the Healing." The term was an effective "flag" that the pro-demolition groups (especially the local municipality) used in the ideological contention with the opposite front. For more information on the coining of this term, see Maria Sframeli, *Il centrodi Firenze restituito* (Florence: Alberto Bruschi, 1989), especially 15–26.

**3** A well-documented study on the demolition of the Ghetto of Florence is offered by O. Fantozzi Micali, *La segregazione urbana. Ghetti e quartieri ebraici in Toscana* (Florence: Alinea, 1995), especially 15–86; see also L. Cerasi, "Fiorentinità. Percorsi di un'ideologia identitaria fra Otto e Novecento," *Studi Novecenteschi* 28, no. 62 (2001): 311–43.
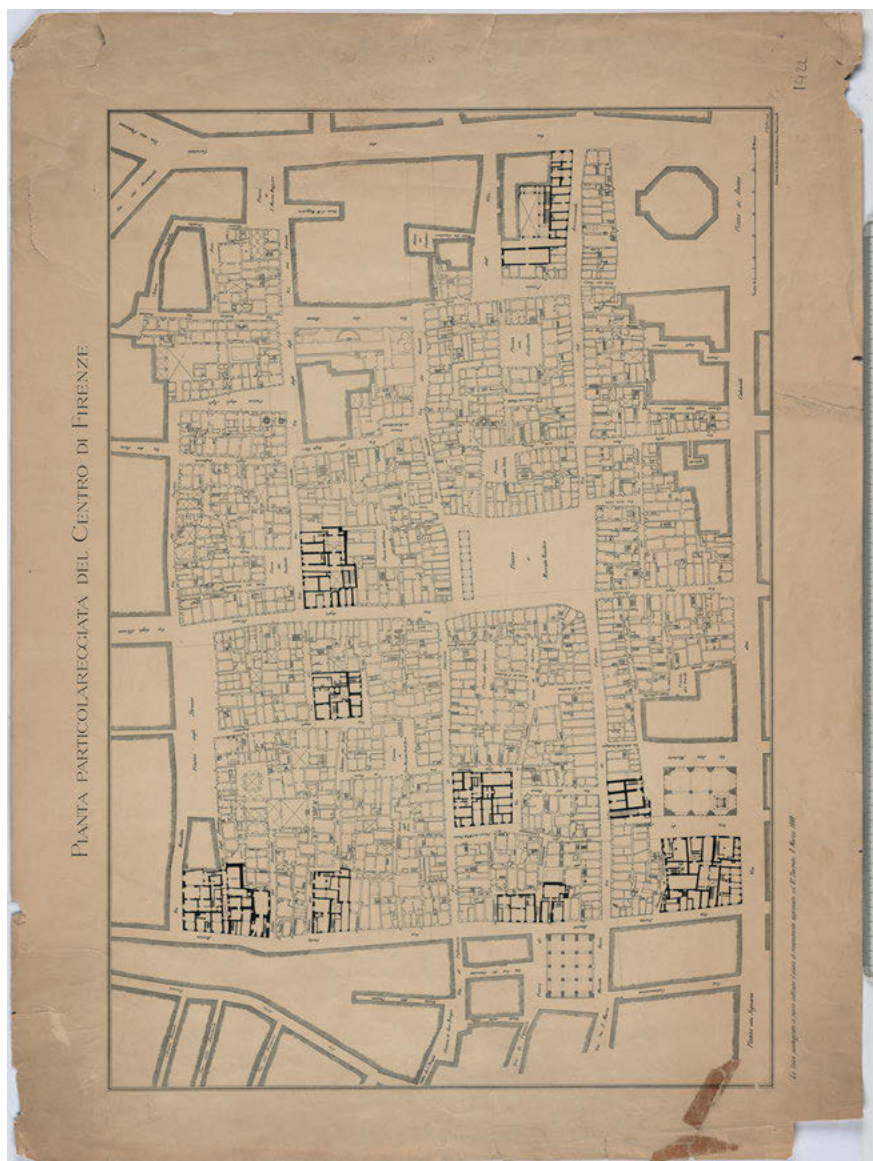
**Figure 1:** Archivio Storico della Città di Firenze, 1422, cass. 49, ins. A. Riordinamento del centro: planimetria generale della zona compresa tra le vie dei Calzaiuoli, dei Cerretani, dei Pescioni e Porta Rossa prima delle demolizioni, con toponomastica, rilievo dei piani terreni e nuovi allineamenti stradali previsti (1888). Map of the center of Florence drawn in 1888: with the exception of only the buildings in black, the entire city center was torn down.

room through the shops and private apartments to the synagogues and other spaces of Jewish sociability. In 1720, a few years after the completion of Ghetto Nuovo, the *Scrittoio* produced one of the most accurate textual descriptions of the ghetto (volume 3559, folder 4) of both Ghetto Vecchio and Ghetto Nuovo, containing more than 200 detailed maps that were subsequently transcribed and technically perfected and gathered in the *Piante dello Scrittoio delle Regie Possessioni* (Maps of the Register of the Royal Possessions), a branch and subcollection of the *Scrittoio delle Regie Possessioni* archival collection. To the best of our knowledge, these are richest, most detailed, and most comprehensive set of iconographic materials on a Jewish ghetto, certainly as far as Italy is concerned (see Figure 2).

The main aim of the textual and iconic materials is to provide the Medici administration with a coherent, consistent, and comprehensive set of information about the architectural features and material conditions of the ghetto spaces, especially after major transformation, from the enlarging of perimetral areas, as in the case of the two main synagogues, through the making of additional floors to host a growing population, to fires, especially the 1670 fire that destroyed part of the Italian synagogues.[4]

Despite the rich amount of information and the preciseness of the technical data that one can draw from these sources, these sources provide us with basically no information about the accessory elements contained in these spaces and nothing that sheds light on everyday life in the ghetto. Since the aim of the *Scrittoio* bureaucrats is to "take a photo," a cold, objective image of the extant situation, distinguishing clearly and in an unequivocal way what belongs to the *Scrittoio* and what is instead owned or what has been done by the tenants, these descriptions deal overabundantly with extremely specific technical features (from the shape and type of wood of the beams, for example, to the type of handles on windows and doors, the materials used for the gutters, and the forms, shapes, and conditions of the nails and other minor supporting elements subject to erosion as well as the type of material used for the internal and external wall plasters, but paradoxically not their color), but not with the height of the spaces, unless something specific and directly dealing with it (for example, the creation of a mezzanine or the demolition of a partition wall) had indeed occurred.

Much of the missing information, especially the height of the floors, can be obtained with a comparative analysis of the historical buildings that survived the *Risanamento* as well as from a rich set of visual materials such as paintings, watercolors, and even photographs produced before the demolition works (see Figure 3), which took place around 1888–1889, when the ghetto, no longer an area

---

4 See Archivio di Stato di Firenze, *Scrittoio delle Regie Possessioni*, vol. 6559.
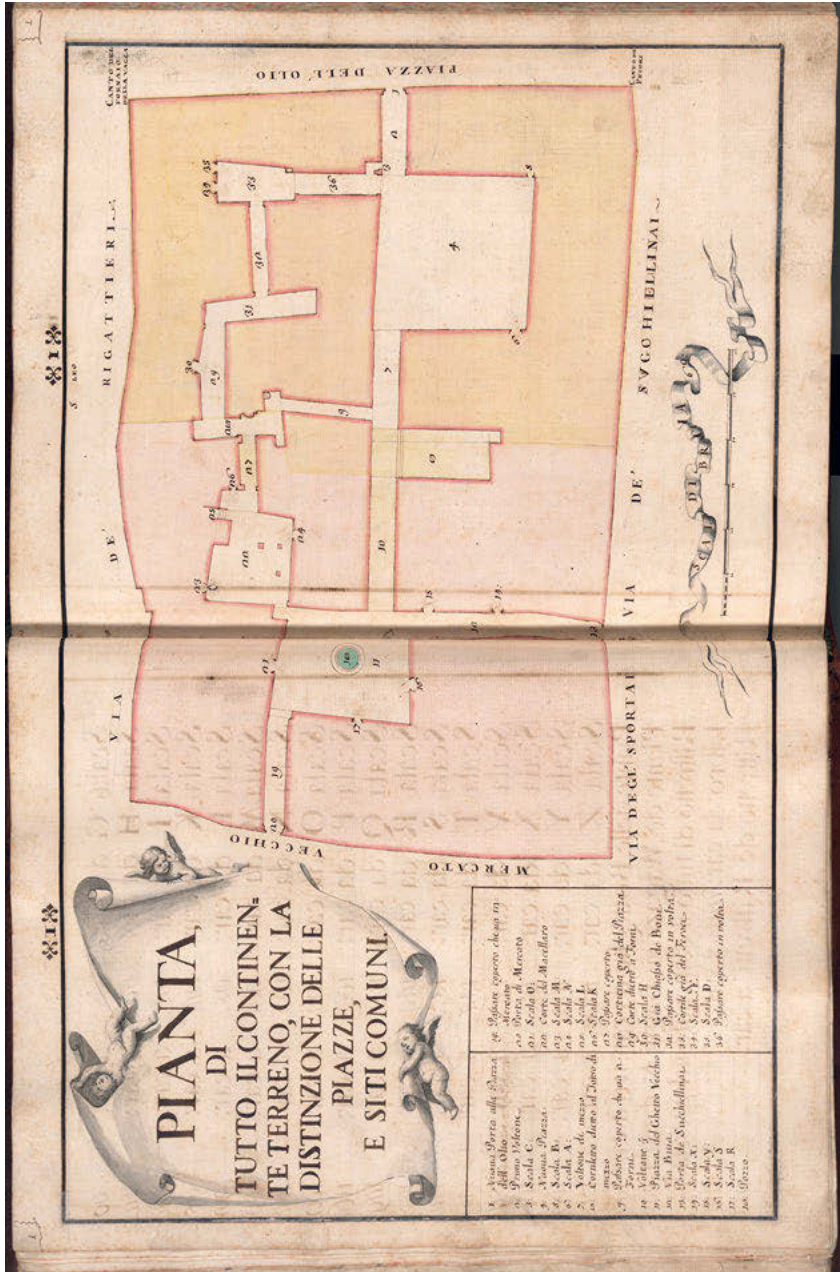
**Figure 2:** Archivio di Stato di Firenze, *Piante dello Scrittoio delle Regie Possessioni*, vol. 26, c. 2. The map (produced in 1721) shows the exact location of the Ghetto Vecchio (Old Ghetto), on the left side in pink color, and the Ghetto Nuovo (New Ghetto), on the right in orange.
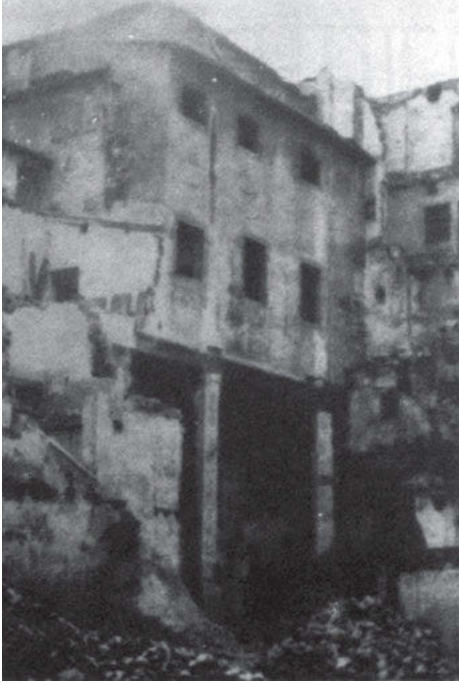
**Figure 3:** Archivio Alinari, Florence. Photo taken during the demolition works of the Old Ghetto. The photo shows the western side of the Italian synagogue, whose additional area was supported by two columns – here clearly visible – set in Corte del Macello (lit. "the Butcher Courtyard").

of forced Jewish residence, had turned into one of the most degraded segments of the city center (from this point of view turning into a "modern ghetto"), inhabited by the local sub-proletariat, an architectural labyrinth, and a tangle of increasingly interconnected spaces (with cellars, hiding places, and underground passages and secret paths) where criminals of all kinds, together with prostitutes and a plethora of socially marginalized individuals, had settled.

Although not dealing directly with the ghetto in architectural-physical terms, a rich set of articles and pamphlets written mostly in support of the demolition of the ghetto, denouncing its filthy standards of life and promoting its physical removal seen as a necessary requirement for the socio-political and even economic rebirth of the old city, provided the reader with valuable information about socio-economic features of the ghetto around the second half of the 19th

century.[5] The goal was eventually achieved, and its ideological tenets are perfectly summarized in the top register of the massive arch that dominates Piazza della Repubblica, the area where the Old Market was located: *L'antico centro della città / da secolare squallore / a vita nuova restituito* (The ancient center of the city, from centuries-old squalor to new life restored).[6]

# 4 Reconstructing a Lost Jewish Space: Sources, Methods, and Works in Progress

The Medici ownership over the ghetto is a unicum, as said, in the history of the Italian Jewry and the real reason for such an extraordinary wealth of documentation. While "enjoying" the status of a Medici private property (from this point of view experiencing equal control as any other component of the patrimony of Florence's dynasty), the ghetto was also a place that contained the Jews, a minority regulated since the early Middle Ages and the making of the Christian canon laws.

Reconstructing a lost Jewish space poses several methodological questions, including why and for which purpose a reconstruction is done. The textual descriptions of the *Scrittoio*, as previously stated, provide us with overabundant pieces of information about elements that today one might find irrelevant, or almost totally insignificant, while simultaneously withholding critical data. Reconstructing the economic and demographic features of the ghetto is a complex but relatively straightforward operation whose complexity comes primarily from the exceptional quantity of data available. While still gathering, deciphering, and processing *Scrittoio* accounting and demographic papers, the GMP decided to move forward in creating a 3D model of the ghetto primarily based on the 1721 set of maps of the *Piante dello Scrittoio delle Regie Possessioni* collection as well as virtual images of reconstruction of the main ghetto synagogues.

---

**5** A stubborn supporter of the demolition of the ghetto and of the *Risanamento* urban plan was the Florentine journalist Giulio Piccini, also known as Jarro, who gathered his numerous articles published in the local newspaper *La Nazione* in *Firenze sotterranea – Appunti – Ricordi – Descrizioni – Bozzetti* (Florence: R. Bemporad & Figlio, n.d.; Florence: Le Lettere, 2017). Worth mentioning is Carolina Invernizio's lengthy novel, *L'orfana del ghetto* (Florence: Adriano Salani Tip. Edit., 1887; Milan: Editrice Lucchi, 1975), a tearful story not without strong anti-Semitic nuances.
**6** For more information about the *Risanamento* and the extent to which this changed the physiognomy of the old city, see Luciano Artusi and Vincenzo Giannetti, *"A vita nuova" – Ricordi e vidende della grande operazione urbanistica che distrusse il centro storico di Firenze* (Florence: Edizioni Lio Terrazzi, 1995).

The first phase was to assemble and collate all single maps into a series of paper sketches (using the 1888 demolition plan maps – see Figure 1 – to define the perimetral borders), each one corresponding to a separate floor (see Figure 4).

These paper sheets were then photographed and elaborated bi-dimensionally using AutoCAD and Adobe Suite. Once a standard height was determined (ca. 2.50 meters, upon comparative analysis of photographic and iconographical materials and a study of surviving historical Florentine buildings), these materials (see Figures 5 and 6) were elaborated using Rhinoceros and Keyshort programs and turned into three-dimensional virtual spaces.

Although extremely detailed, as previously stated, and very carefully drawn, the maps present numerous (obvious and unsurprising) imperfections that appeared only when all the single maps were collated and combined into the different floor charts. Bigger or shorter walls, erroneous correspondences between lower and upper walls, and missing entrances were adjusted and corrected. Once we drew the bi-dimensional floor maps, we noticed that especially on the ground floor, many places around the ghetto perimeter had not been mapped and no information was provided. The *Scrittoio* accounting books labeled this area as *Ghetto Esterno* (External Ghetto): mostly shops on the ground floor but also private apartments on the upper floors that in both ghettos were given to Christian tenants in order to "hide" the Jewish presence, creating a sort of architectural buffer zone all around the areas rented, inhabited, or used by the Jews, which the *Scrittoio* called *Ghetto Interno*, or the Jewish ghetto proper. Being located mostly on the ground floor and being surmounted by spaces of equal size and shape, these missing places could easily be reconstructed and then integrated into the general map. This allowed us to produce a 3D printed model of the ghetto, as it appeared at the time of its maximum expansion in the mid-18th century (Figures 7and 8). A slightly more elaborately rendered image of the model was then inserted into an aerial view of modern Florence, clearly showing how the ghetto had been integrated into the city fabric (Figure 9).

We applied similar criteria for the reconstruction of two specific places in *Ghetto Interno* – namely, the Italian and Spanish synagogues, which were not only the two main places for religious worship and traditional education, but also the main sites of several Jewish confraternities (providing the population with services and forms of social welfare). Due to the demographic growth, the two ghetto synagogues, originally meant to serve a population of about 450–550, underwent several changes over the centuries, mostly enlargements with interventions that had an immediate and significant impact on the outdoor spaces and the general physiognomy of the ghetto. In the case of the Italian synagogue, a wide balcony area was added on the western side and supported by two massive

**Figure 4:** Collation of all the individual maps of the first floor of the ghetto taken from Piante dello Scrittoio delle Regie Possessioni, vol. 26. The missing maps on the left side are the shops and apartments of "Ghetto Esterno" that the Scrittoio authorities had given to Christian tenants to conceal the presence of the Jews from the city center. The Medici Archive Project®.

**Figure 5:** The ground floor of the ghetto with the distinction between Ghetto Interno (in red) and Ghetto Esterno (sky-blue), a strip of shops and apartments stretching from the ground to the upper floors used and inhabited by Christian tenants. These areas were provided with independent external entrances. The Medici Archive Project®.
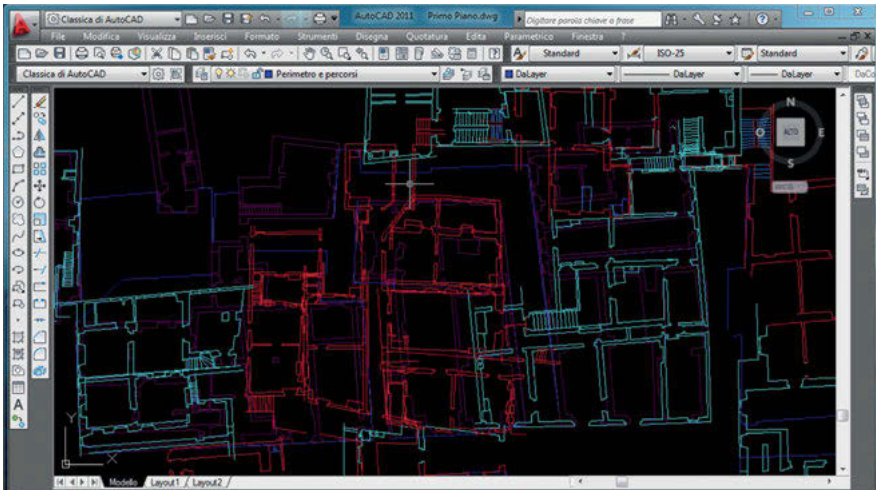


**Figure 6:** Screenshot of AutoCAD image with superimposition of the ground and first floors of the ghetto. The Medici Archive Project©.

**Figure 7:** Work-in-progress of the 3D printing of the Ghetto of Florence.
The Medici Archive Project®.

wooden columns resting in the middle of Corte del Macello (The Butcher Court-yard – see Figure 3).

After careful examination of the maps of the two synagogues produced by the *Scrittoio* officials in 1721, and assuming that the two Florentine synagogues would not divert on a formal level from other synagogues of the same time and of the same rite and that most of the functional elements – e.g., the Bible-prayer reading desk, the position of the benches, the color of the curtains – were in all likelihood the same ones found in other places of Jewish worship belonging to the same cultural and ritual milieus, the data that the *Scrittoio* papers do not provide were obtained from external sources and then implemented into our virtual model.

The black-and-white marble floor, quite commonplace in the central-north Italian synagogues, as well as the color of the curtains, the shape of the benches, and the type of brass chandeliers that the *Scrittoio* officials in no way mentioned in their reports was implemented into the virtual models upon comparative anal-
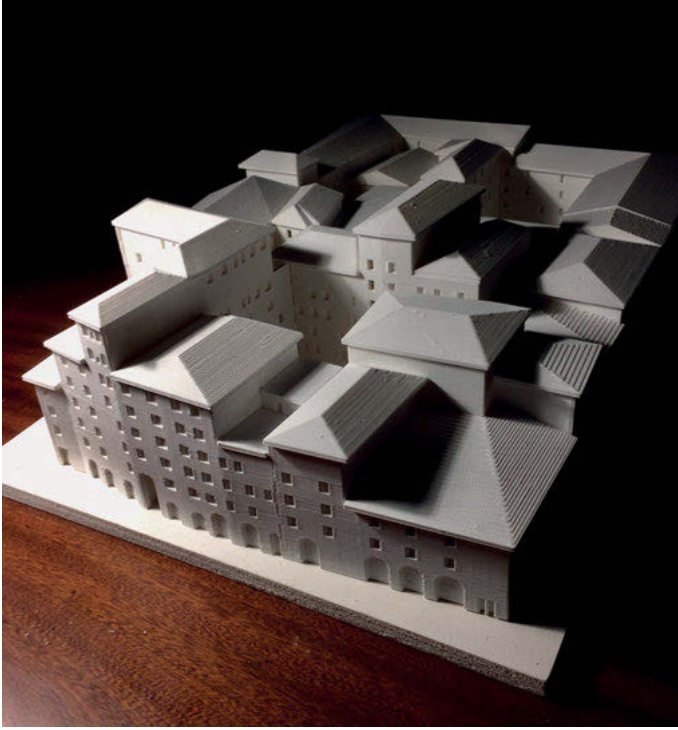
**Figure 8:** The final 3D printed model of the Ghetto of Florence according to maps of 1721. The Medici Archive Project®.

ysis of external sources, mostly the synagogues of Pitigliano (a small hilltop town in the south of Tuscany under Medici rule since the early 17th century), the Levantine, Sephardic, and Italian *scole* of the Venetian Ghetto,[7] as well as an analysis of original pieces of Italian synagogue furniture that had been relocated to various Israeli synagogues soon after World War II.

One of the most significant gaps that we filled thanks to such an analysis are the two tabernacles that the *Scrittoio* documents do not describe at all, only briefly mentioning their position in the synagogue space. As underscored in previous

---

[7] For more information about the Venetian synagogues, see Donatella Calabi, *Venezia e il Ghetto. Cinquecento anni del "Recinto degli ebrei"* (Venice: Marsilio, 2016), also available in an English translation: *Venice, the Jews and Europe – 1516–2016*, ed. D. Calabi (Venice: Marsilio, 2016). About Pitigliano, see Luigi Cerroni, *Breve storia della comunità ebraica di Pitigliano* (with English translation, *A Short History of Jewish Settlement in Pitigliano*) (n.p., n.d.); Roberto G. Salvadori, *La comunità ebraica di Pitigliano: Dal 16mo al 20mo secolo* (Florence: Giuntina, 1991).

**Figure 9:** Aerial view (graphic processing on Google MAP, Imagery © 2022 Maxar Technologies, Map Data, © 2022) of modern Florence with a slightly rendered version of the virtual image of the ghetto, exactly where it was before demolition. The ghetto not only faced the popular Old Market square, but was also very close to the main Christian spiritual sites (Santa Maria del Fiore Basilica and the Baptistry, on the left side) and the political barycenter, Palazzo Vecchio, the Medici's main building, on the right side.

studies and documented by many archival papers, before demolition of the ghetto, the late Renaissance tabernacle of the Levantine/Sephardic synagogue was moved to via delle Oche, where two "oratori" (one for the Levantine/Sephardic, the other for the Italian *minhag*) had been opened soon after the inauguration of the majestic "Tempio Maggiore."[8]

Around the early 1960s, together with many Italian synagogue artifacts, the tabernacle was sent to Israel, becoming part of the synagogue of Keren be-Yavneh, an Orthodox yeshivah. The tabernacle was photographed and a virtual rendition of it inserted into our virtual model. Much more complex was the identification of the tabernacle of the Italian synagogue that, according to the *Scrittoio*, was a lavish and beautifully decorated Baroque structure, done mostly with stucco and presumably also *marmorino* (fake marble done with painted wood) – two materials that cannot be reassembled once removed from their original site. Before

---

**8** Micali, *La segregazione*, 74–75.
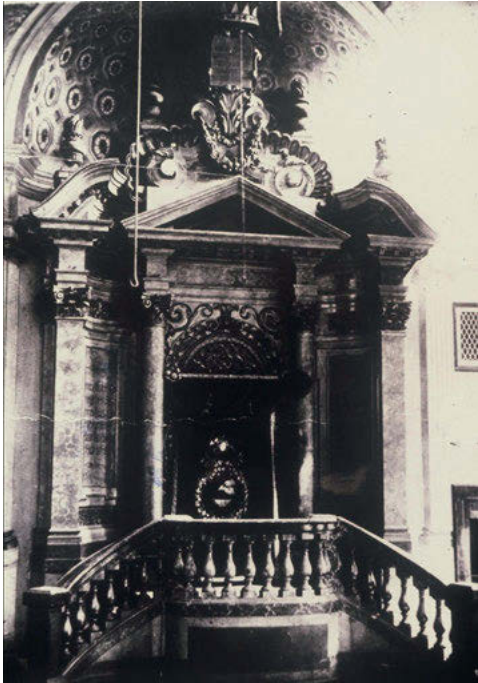
**Figure 10:** Archivio della Comunità Ebraica di Firenze. Photo of the tabernacle of the Italian synagogue shortly before the demolition of the ghetto. This item has not yet been cataloged.

the demolition of the Italian synagogue, the tabernacle was photographed, and copies of the daguerreotype were auctioned for charitable purposes. We found one copy in the Archive of the Jewish Community of Florence where the tabernacle is clearly visible (see Figure 10).[9]

A tentative reproduction of the structure was then inserted into the virtual model. Our main goal for the time being was to determine the tabernacle's size and its position within the ritual space, not its complex aesthetic features, which we hope to reproduce more faithfully and philologically in the near future. An attempt to reconstruct the internal space of the two synagogues was done a couple of years ago, and a new version, implementing further extra-archival data and – most importantly – including a much more faithful reproduction of the tabernacle, will soon be released.

---

**9** Dora Liscia Bemporad, "La scuola italiana e la scuola levantina nel ghetto di Firenze: Prima ricostruzione," *Rivista d'Arte – Studi documentari per la storia delle arti in Toscana* 8, no. 2 (1987): 3–48, especially 19–26. See also S.E. Glicksberg, "Minhage bet ha-knesset shek kehillat kodesh italki be-'ir Firenze" (The Customs of the Italian Jewish Community of Florence), *Alei Sefer: Studies in Bibliography and in the History of the Printed and the Digital Hebrew Book* 24–25 (2015): 253–83 [Hebrew].

**Figure 11:** Virtual reconstruction of the Italian synagogue. The Medici Archive Project®.

# 5 Conclusions

What lies beneath the rise and fall of the Ghetto of Florence is a complex set of historical events. It was the only ghetto in Italy owned by the ruling dynasty. Cosimo I de' Medici made a considerable chunk of the city center his own through a complex political and financial operation, basically depriving the Christian owners of their properties and turning a potentially very expensive political-administrative action in a profitable financial business – at least for his own family. In the mid-18th century, the ghetto, as both a place and a system of social segregation, started showing evident signs of weakness: since the 1680s, more than 100 Jewish families, amounting to approximately 400 people, had moved outside the ghetto, breaking a number of laws, but clearly with the silent assent of local authorities and probably with the sympathetic connivance of several "normal" people perfectly aware that their Jewish neighbors should have been living within the ghetto.

The demolition of the ghetto did not take place with the specific purpose of cleaning up the city and removing the remnants of a shameful past, but as a component of a much wider urban, socio-political, and economic plan to renew the entire city of Florence, a city whose physiognomy was complex, heterogeneous,

**Figure 12:** Virtual reconstruction of the Levantine synagogue. The Medici Archive Project®.

deeply grounded in its medieval and premodern times prior to the *Risanamento*. By the end of the 19th century, the Ghetto of Florence had already long turned into a sub-proletarian, criminal area, a labyrinth of houses and internal passages that time was turning into a possibly more and more intricate social anthill. The *Risanamento* did not change only the external physiognomy of Florence, but also its social fabric, providing many with better housing and improving their overall living standards. The Jews of Florence felt such a hiatus even more dramatically, leaving the ghetto not only materially/physically but also metaphorically, abandoning the small and dark traditional *scole* (synagogues) for a magnificent, glorious, and perhaps bit too big of a "Tempio," a grandiose building clearly inspired by Christian art (the basilica shape is absolutely evident), the only one with a dome able to compete with – but certainly not win against – Brunelleschi's dome. The reconstruction of the ghetto means not only bringing to (virtual) life a place that has been abandoned, but also filling a huge chronological and physical gap, shedding light on what is unknown or wrongly assumed. One of our most recent discoveries, for example, was the map of the lavish apartment of Salomon Levi. Departing from the stereotypical view of the ghetto house as a poor, narrow, and overcrowded space, Levi's apartment appears to be a luxurious space, decorated with paintings and stuccos, a colorful upper-middle-class home and probably the same that Ricciardo Meacci represented in a watercolor

**Figure 13:** Archivio di Stato di Firenze, Piante dello Scrittoio delle Regie Possessioni, vol. 26, c. 29 (1721). The map of the luxurious apartment of Salomon Levi, a prominent textile trader and one of the most influential inhabitants of the Ghetto Nuovo. The house consisted of a central hall (A) and several rooms (B), two bathrooms, and a terrace.
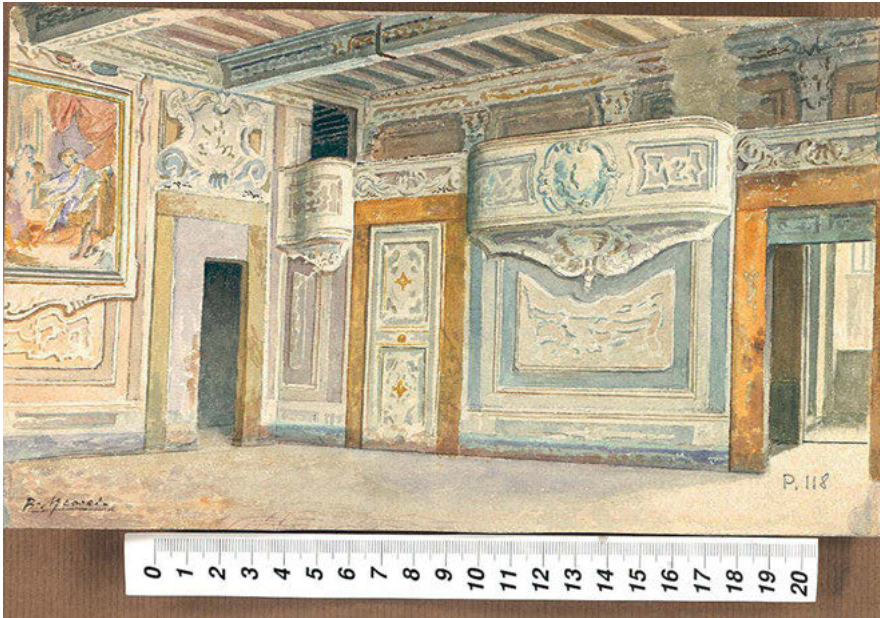
**Figure 14:** Archivio Storico del Comune di Firenze, Ricciardo Meacci, doc. n. 420736. Although the ghetto was commonly described and thought to be a decadent and unhealthy area, some of its apartments were unexpectedly beautiful and decorated with stuccos and paintings. Meacci's watercolor closely reminds us of what the local historian Guido Carocci wrote in his report of the historical city center, *Il Ghetto di Firenze e i suoi ricordi – Illustrazione storica* (Florence: Galletti e Cocci, 1886), 31: "Per rovescio di medaglia, in questo ceppo di case posto fra Via della Nave e l'Arcivescovado erano quartieri eleganti e decorati con lusso non comune. Tuttora si veggono difatti sale adorne di buone pitture e fra le altre è degna di considerazione un'ampia sala da ballo con orchestra e colle pareti adorne di ricche cartelle dove sono dipinti fatti del vecchio testamento" (on the contrary, in this group of houses set in between Via della Nave and the Archbishopric site, there were lavish apartments, decorated with stuccos and paintings of various sizes. Still today one can see rooms with nice paintings, among which worth mentioning is a dancing hall with an orchestra, with adorned walls where scenes of the Bible are represented).

in 1886. Bringing back to life Levi's apartment is one of several potential reconstructing tasks inside the ghetto and one of the many tiny ghetto details that we aim to know more about and understand better in a consistent and trustworthy reconstructed virtual form.

# References

Archivio Alinari, Florence.

Archivio della Comunità Ebraica di Firenze.

Archivio di Stato di Firenze. *Piante dello Scrittoio delle Regie Possessioni*, vol. 26.

Archivio di Stato di Firenze. *Scrittoio delle Regie Possessioni*, vol. 6559.

Archivio Storico del Comune di Firenze, Ricciardo Meacci, doc. n. 420736.

Archivio Storico della Città di Firenze, 1422, cass. 49, ins. A. *Riordinamento del centro: planimetria generale della zona compresa tra le vie dei Calzaiuoli, dei Cerretani, dei Pescioni e Porta Rossa prima delle demolizioni, con toponomastica, rilievo dei piani terreni e nuovi allineamenti stradali previsti* (1888).

Artusi, Luciano and Vincenzo Giannetti. *"A vita nuova" – Ricordi e vidende della grande operazione urbanistica che distrusse il centro storico di Firenze*. Florence: Edizioni Lio Terrazzi, 1995.

Bemporad, Dora Liscia. "La scuola italiana e la scuola levantina nel ghetto di Firenze: Prima ricostruzione." *Rivista d'Arte – Studi documentari per la storia delle arti in Toscana* 8, no. 2 (1987): 3–48.

Calabi, Donatella. *Venezia e il Ghetto. Cinquecento anni del "Recinto degli ebrei."* Venice: Marsilio, 2016.

Calabi, Donatella, ed. *Venice, the Jews and Europe – 1516–2016*, Venice: Marsilio, 2016.

Carocci, Giudo. *Il Ghetto di Firenze e i suoi ricordi – Illustrazione storica*. Florence: Galletti e Cocci, 1886.

Carocci, Guido. *Ricordi del vecchio mercato di Firenze*. Florence: Tip. dell'Arte della stampa, 1887.

Cerasi, Laura. "Fiorentinità. Percorsi di un'ideologia identitaria fra Otto e Novecento." *Studi Novecenteschi* 28, no. 62 (2001): 311–43.

Cerroni, Luigi. *Breve storia della comunità ebraica di Pitigliano*. n.p, n.d.

Fantozzi Micalli, Osanna. *La segregazione urbana. Ghetti e quartieri ebraici in Toscana*. Florence: Alinea, 1995.

Glicksberg, S.E. "Minhage bet ha-knesset shek kehillat kodesh italki be-'ir Firenze" (The Customs of the Italian Jewish Community of Florence). *Alei Sefer: Studies in Bibliography and in the History of the Printed and the Digital Hebrew Book* 24–25 (2015): 253–83.

Invernizio, Carolina. *L'orfana del ghetto*. Florence: Adriano Salani Tip. Edit., 1887; Milan: Editrice Lucchi, 1975.

Mancuso, Piergabriele. "The Politics of Segregation in Grand Ducal Florence: The Ghetto of Florence, from Material Fall to Virtual Rebirth – a Short Presentation." In *Judisches Kuturerbe (re-)praesentieren – Judisches Kulturerbe*, edited by Katrina Kessler, Martha Stellmacher, Alexander von Kienlin, and Ulrich Knufinke, vol. 2, 45–66. Braunschweig: Netwerk judisches Kultuerbe c/o Beit Tfila – Forschnugsstelle fur judische Architektur, 2019.

Mancuso, Piergabriele and Vigotti, Lorenzo. "From Centuries-Old Squalor: The Ghetto of Florence, from History to Virtual Life. Introduction to the Ghetto Mapping Project." *Rivista di Letteratura Storiografica Italiana* 1 (2017): 123–34.

Mancuso, Piergabriele and Vigotti, Lorenzo. "Reconstructing a Lost Space: The Ghetto Mapping Project at the MAP." *Materia Giudaica* 22 (2017): 221–32.

Piccini, Giulio. *Firenze sotterranea – Appunti – Ricordi – Descrizioni – Bozzetti*. Florence: R. Bemporad & Figlio, n.d.; Florence: Le Lettere, 2017.

Salvadori, Roberto G. *La comunità ebraica di Pitigliano: Dal 16mo al 20mo secolo*. Florence: Giuntina, 1991.

Sframeli, Maria. *Il centrodi Firenze restituito*. Florence: Alberto Bruschi, 1989.

Daniel Stein Kokin

# Introducing "Kol ha-Nekudot"/"All the Points"/"Kull al-Nuqaṭ": Interactive, Online Mapping of the Israeli-Palestinian Region (1840–Present)

**Abstract:** This article introduces "Kol ha-Nekudot"/"All the Points"/"Kull al-Nuqaṭ," a digital humanities project and pedagogical resource currently under development. Drawing its name from the classical Zionist tendency to refer to communities as "points on the map," "All the Points" is an interactive series of online maps depicting the Jewish, Palestinian Arab, and other communities that have existed in the Israeli-Palestinian region in any given year from 1840 to the present. "All the Points" provides basic information about each community and, by using colors and shapes, visually distinguishes among the kibbutz, moshav, Arab village, and many other community types and sub-types. Specifically in this piece, I explain various features of the enterprise, e.g. the respective rationales for its time range, geographical purview, and classificatory scheme; the ways in which the user can interact with the maps by adjusting the time bar, selecting the base map, and filtering in and out various settlement types; the manner in which data is obtained; and how various challenges posed by "All the Points" have been addressed. In addition, I discuss the project's specialized, curated maps, focusing in particular on three thereof, namely those depicting the "Wall and Tower" communities of the late 1930s, the modern history of settlement on the Golan Heights beginning in 1878, and settlements of the imagination – i.e., fictional communities that exist only on the page, screen, or stage. Finally, accompanying images provide a visual impression of "All the Points" and its maps.

**Keywords:** Israel, mapping, Palestine, settlement, Zionism

# 1 Getting to the Point(s)

In this article, I introduce and describe "All the Points," an ongoing digital humanities project developing online, interactive maps that explore settlement in the Israeli-Palestinian region from 1840 to the present. Produced with the Geographic Information System mapping software ArcGIS,[1] "All the Points" emerged out of the course "Settlement in Israeli History" that I had the privilege to create and teach as a visiting professor at UCLA. This class considered the history of Zionism and the State of Israel from the perspective of the diverse range of communities founded under their aegis and/or affected by them – i.e., the moshava, the kibbutz, the moshav, Tel Aviv and other cities, the Development Town, the Arab town and village, and – of course – the *hitnaḥaluyot* or "settlements" of current controversy.

The title of this class was deliberately provocative, but for good reason, namely that it is impossible fully to understand the history of Zionism and Israel without considering the approximately one thousand new communities that were founded (and the somewhat more than half as many that were dislodged[2]) in their wake during the past one and a half centuries. Nor would one want to, since these settlements, taken together, constitute a vast and perhaps unprecedented laboratory of social experimentation. No other place in the world, so far as I am aware, features such an astounding array and complexity of ideologically, ethnically, historically, and religiously inflected communities. I wanted my students, most of whom had no prior background on Israeli or Palestinian history, to be able to visualize the rapid changes that have transpired and the diversity in the built environment to which they have given rise. Since most maps are both frozen in a particular historical moment and blind as to community type, I resolved to create an alternative that would address these lacunae, namely by showcasing change over time, and using colors and shapes to distinguish amongst different kinds of communities. Thus was born "All the Points."

In what follows, I shall discuss the basic contours of the project – its title, coverage (both historical and geographical), and component parts – while also addressing the challenges it faces. Alongside the text, select images offer a visual impression of "All the Points," and of the main maps and curated maps that comprise it. See Figure 1 for the project logo.

---

**1** ArcGIS is a product of the Environmental Systems Research Institute (ESRI). Upon consultation with digital humanities experts at UCLA, I determined that the capabilities offered by this software best suited the aims and needs of "All the Points," at least as this initial stage.
**2** This includes communities in traditional Palestine proper as well as on the Golan Heights. See below for further discussion.

Figure 1: "All the Points" Logo, Meira Stein Kokin.

## 2  The Point(s) of "All the Points"

The name of the project was inspired by, and is in dialogue with, the classic Zionist trope of *'od nekudah 'al ha-mapah* ("another point on the map") – i.e., the notion that each new settlement marks progress along the path of forging or strengthening the rebuilt Jewish national home or state.[3] In essence, this rhetoric indicates that the Zionist, and later Israeli, project was about creating, and subsequently completing (to the extent possible), a map. Likely the most famous instance thereof concerns the night of October 5–6, 1946, in the course of which 11 new settlements were founded in the northern Negev. These communities were collectively known as *Aḥad 'asar ha-nekudot* ("The Eleven Points") and played a crucial role in the ultimate incorporation of this region into what became the State of Israel.[4]

---

**3** While I am not aware of any extant work of scholarship that specifically explores the emergence, development, and larger cultural significance of this rhetoric of *points* across the history of Zionism and Israel, its existence and prominence is adequately attested by the title of the following study: Osnat Shiran, *Nekudot 'oz: mediniyut ha-hityashvut be-zikah le-ye'adim politiyim u-vitḥoniyim be-terem medinah uve-reshitah* (Points of Strength: Settlement Policy in Relation to Pre-State and Early-State Political and Security Goals) (Tel Aviv: Department of Defense, 1998).

**4** On the 11 points, see Sarit Okon, ed., *Aḥad 'asar ha-nekudot: sipuro shel mivtsa' no'az* (The Eleven Points: The Story of a Daring Operation) (Alumim: Be'erot ba-Negev, merkaz moreshet, hadrakhah ve-eruaḥ, 2017).

While the celebration of "points on the map" was part and parcel of the Zionist consensus in the early decades of Israeli statehood, by the late 20th century widely contrasting approaches to this discourse had emerged. On the one hand, in the choice of the name *Nekudah.* ("Point.") for the influential monthly magazine that was closely affiliated with, and published during the heyday of, the national-religious settler movement,[5] we discern especially close identification with – and indeed intensification of – the trope, now applied in particular to the territories conquered in the Six-Day War. This is reflected visually in the grammatically superfluous period following the publication's name, lest the reader miss the *point*.

On the other, however, in an important article by the prominent Israeli architect and author Sharon Rotbard, we encounter a starkly contrasting perspective, namely that the tendency to refer to communities as points "hints at the fact that the 'point' on the map was more important than the 'settlement' itself."[6] Rotbard here criticizes the replication of what he calls "the 'settlement point' system … in national master plans throughout Israel's history"[7] and appears to be suggesting of Israeli settlement that – to evoke and subvert a famous English saying – "it's easy to lose the points for their map." Along the same lines, in the 1995 song *Ma'aleh Avak* about a fictional city on Israel's desert periphery, and its lonely and forlorn populace, the famed Israeli rock band Teapacks satirizes the rhetoric and practice of Israeli settlement. The lyrics commence as follows:

> "It doesn't make a good impression," thought the leading figures in the government, "there are empty patches on the map and down there another point is still missing." So the big shots issued an order: "Let's build a city here and we'll also bring some people, who will fill up the new houses with their lives. It's great – lots of points on the map. And, anyway, in the papers they promised a big scoop." Thus the senior ministers commanded in a drowsy voice and ran off to deal with all kinds of emergency situations. A minor deputy came out all the way to bless the new settlement, named Ma'aleh Avak.[8]

Standing in a long line of *reductiones ad absurdum* at once literary and cartographical, this passage can be profitably placed in dialogue with arguably the

---

**5** *Nekudah.* appeared between 1980 and 2010. It was the official publication of the *Amana* settler movement of the national-religious *Gush Emunim* ("Block of the Faithful").

**6** Sharon Rotbard, "Wall and Tower (Homa Umigdal): The Mold of Israeli Architecture," in *A Civilian Occupation: The Politics of Israeli Architecture,* ed. Rafi Segal and Eyal Weizman (London: Verso; Tel Aviv: Babel, 2003), 48.

**7** Rotbard, "Wall and Tower," 49–50.

**8** *Ma'aleh Avak*, translation mine. The name of the song can be rendered as both "Ascent of Dust" and "Gathering Dust," simultaneously reflecting both the lofty ambitions and sordid reality of the many planned communities on the Israeli periphery. For the complete Hebrew text of *Ma'aleh Avak*, see: https://shironet.mako.co.il/artist?type=lyrics&lang=1&prfid=429&wrkid=2205, accessed April 5, 2021.

most famous of these texts, Jorge Luis Borges's 1946 short story "On Exactitude in Science", in which "the Cartographers Guilds struck a Map of the Empire whose size was that of the Empire, and which coincided point for point with it," rendering their map utterly useless.[9] For if *On Exactitude* explores what happens when a map succeeds all too well in imitating the reality it wishes to depict, *Ma'aleh Avak* grapples instead with the deleterious social consequences of the map itself effectively becoming that reality, to the neglect of its constituent communities.

In thus naming my map for its composite points, I attempt to place these settlements themselves front and center. And, in referring to "*All* the Points," I challenge and expand upon the traditional Zionist – and therefore exclusively Jewish – connotations of the *nekudah 'al ha-mapah*, emphasizing that every kind of community counts, and that wherever a new point appears, an old one may have existed before. Indeed, as seen above, the complete title of the project features Hebrew and Arabic alongside English, and my ultimate goal is to make its maps available in all three of these languages.[10]

Unlike many other online mapping initiatives involving this land, "All the Points" is not restricted to any specific geographical portion thereof, nor is it invested in any particular ideological or political plan or perspective. Stepping back from the questions – as important as they are – of who *should* or *should not* live or have lived *where*, it instead asks: who *has* lived *where*, *when*, and in *what kind* of community. My hope is that visitors to the site will be fascinated and engaged by what they find there, and that this will in turn spur further and deeper exploration, learning, and – ultimately – understanding.

What, then, does "All the Points" cover? Its time frame, 1840–present, was deliberately chosen so as to encompass the modern period of intense Jewish settlement, European Christian colonization, as well as evolving Arab presence in as historically sensible and politically neutral a manner as possible. Thus, 1840, the year in which the Ottoman Empire regained political control over Palestine from Muhammad Ali's Egypt, is a date of historical significance, but not one that arouses passion and disagreement today.

---

**9** Jorge Luis Borges, "On Exactitude in Science" (originally: "Del rigor en la ciencia"), trans. Andrew Hurley, accessed April 5, 2021, https://genius.com/Jorge-luis-borges-on-exactitude-in-science-annotated.

**10** The project name in Arabic, "Kull al-Nuqaṭ," is particularly apropos, as the root *n.q.ṭ*, the primary meaning of which is point or dot, can also refer to a location, village, market town, or even military outpost. See Hans Wehr, *A Dictionary of Modern Written Arabic*, ed. J. Milton Cowan. 3rd ed. (Ithaca, NY: Spoken Language Services, 1976), 993. My thanks to Reuven Firestone for "pointing" this out to me.

Geographically speaking, the map documents all modern Jewish settlement in the wider region throughout this period. For example, it shows the five communities that briefly existed in the Hauran region of today's Syrian Arab Republic in the 1890s, as well as the settlements founded during the period of Israeli control over the Sinai Peninsula between 1967 and 1981. By contrast, non-Jewish communities are only covered in the region that remains today the active zone of Israeli control and influence – i.e., internationally recognized Israel, plus the Golan Heights, West Bank, and Gaza Strip. To leave out Jewish attempts to settle the Hauran or Sinai would distort the past on the basis of the present, whereas to attempt to include their non-Jewish communities would discount the marginal role these regions play in the overall story of "All the Points" and pose serious, potentially insurmountable, challenges concerning data collection and accuracy.

# 3 Plotting Points: Classifying Communities

With regard to the classification of settlements, "All the Points" features a two-tiered scheme that enables varying degrees of detail depending upon viewer background and interest, and also allows for potential expansion in the future. Each community *type* is assigned a letter in the project database (spreadsheet) corresponding to a color on the map – e.g., A (green) denotes Arab communities; K (red) is for kibbutzim. In addition, each *sub-type* is indicated by a number, expressed in turn on the map in the form of a shape. Thus, Bedouin Arab communities are indicated by A2 (green square), whereas K7 (red Magen David or Star of David) is for kibbutzim affiliated with the Religious Kibbutz Movement. The basic version of the main map, intended for beginning students or others with little prior knowledge, distinguishes solely among settlement *types*, with all points appearing as small dots, varying only by color. On the advanced or detailed map, by contrast, the points are presented in different shapes and colors, reflecting both *types* and *sub-types*. Table 1 represents a chart of the complete classification system as it exists at present, while Figures 2, 3, and 4 depict the advanced main map in three different years.
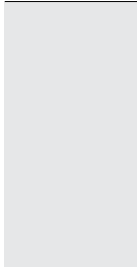
**Table 1:** The "All the Points" Classification System.

| | 1) ARAB COMMUNITIES (GREEN) |
|---|---|
| | A: Arab Community [CIRCLE] |
| | A1: Arab City or Village [CIRCLE] |
| | A2: Bedouin Arab City, Village, or Tribe [SQUARE] |
| | A3: Majority or Entirely Druze Village [STAR] |

**Table 1** (continued)

| | |
|---|---|
| | A4: Unrecognized Arab Community [DIAMOND] |
| | A5: Unrecognized Bedouin Arab Community [TRIANGLE] |
| | A6: Palestinian Refugee Camp [HEXAGON] |
| | A7: Arab, Status to be determined [QUESTION MARK] |
| **2)** | **MAʿABARAH/TRANSIT CAMP (MAROON)** |
| | B: Maʿabarah Transit Camp [CIRCLE] |
| **3)** | **COMMUNITY SETTLEMENTS (PURPLE)** |
| | C: Community Settlement [CIRCLE] |
| **4)** | **EUROPEAN CHRISTIAN COLONIES, E.G. GERMAN, RUSSIAN, AMERICAN, GREEK (PINK)** |
| | E: European Colony [CIRCLE] |
| | E1: German Templer Colony [SQUARE] |
| | E2: Non-Templer German Colony [STAR] |
| **5)** | **KIBBUTZIM (RED)** |
| | K: Kibbutz [CIRCLE] |
| | K1: United Kibbutz Movement [CIRCLE] |
| | K2: United Kibbutz [CIRCLE] |
| | K3: Union of the Collectives and the Kibbutzim [SQUARE] |
| | K4: Kibbutz of the Land [TRIANGLE] |
| | K5: Association of Collectives [SQUARE] |
| | K6: Union of the Kibbutzim [DIAMOND] |
| | K7: Religious Kibbutz Movement [MAGEN DAVID/JEWISH STAR] |
| | K8: Other Religious Kibbutz [STAR] |
| | K9: Kibbutz, Status to be determined [QUESTION MARK] |
| **6)** | **MOSHAVIM (BLACK)** |
| | M: Moshav [CIRCLE] |
| | M1: Moshav ʿOvdim [CIRCLE] |
| | M2: Moshav Shitufi [SQUARE] |
| | M3: Moshav Poʿalim [TRIANGLE] |
| | M4: Moshav, Status to be determined [QUESTION MARK] |
| **7)** | **NAḤAL ARMY/AGRICULTURAL SETTLEMENTS (ORANGE)** |
| | N: Naḥal Settlement [CIRCLE] |

**Table 1** (continued)

| | |
|---|---|
| | **8) MOSHAVA FARMING COLONIES (YELLOW)** |
| | O: Moshava [CIRCLE] |
| | **9) JEWISH AND/OR MIXED CITIES, TOWNS, VILLAGES** |
| | U1: Generic, Secular or Religious Jewish Majority City or Town [BLUE CIRCLE] |
| | U2: Jewish/Arab Mixed City [HALF BLUE/HALF GREEN CIRCLE] |
| | U3: Development Town [BLUE SQUARE] |
| | U4: Ultra-Orthodox/Ḥaredi Jewish Majority City or Town [DARK BLUE CIRCLE] |
| | **10) MISCELLANEOUS COMMUNITIES (GRAY)** |
| | X: Miscellaneous [CIRCLE] |
| | X1: Circassian Village [SQUARE] |
| | X2: Jewish: Miscellaneous or Status to be determined [MAGEN DAVID/STAR OF DAVID] |
| | X3: Youth or Educational Village [TRIANGLE] |
| | X4: Agricultural School or Farm [DIAMOND] |
| | X5: Other, Status to be determined [QUESTION MARK] |

There will be numerous options to customize the main map's appearance. With regard to the base map underlying the points, viewers will be able to choose from among a topographical map; a contemporary map indicating both present borders and armistice lines, as well as highways and streets across the country (and thus the approximate current physical size of communities); historical maps, including maps reflecting the external borders and administrative divisions of the land under Ottoman, British, Israeli, and Palestinian rule; and maps of the various initiatives proposed to resolve the Israeli-Palestinian conflict, including the Peel Commission Plan (1937), UN Partition Plan (1947), Allon Plan (1967), Clinton Parameters (2001), and Trump peace plan (2020), to name but a few. In addition, users of the site will be able to decide whether to view every form of settlement or instead to filter in or out specific types (on the basic map), or types and sub-types (on the advanced map). Concerning the map's timeline, they will choose whether to examine a specific year, set a range of years, or move forwards or backwards in time across the entirety of the covered period, watching as communities emerge, disappear, and change their identities; the speed with which the map moves from year to year is also subject to adjustment. Finally, it will also be possible to zoom in or out of particular regions, and to click on individual points for further information and links to relevant websites.
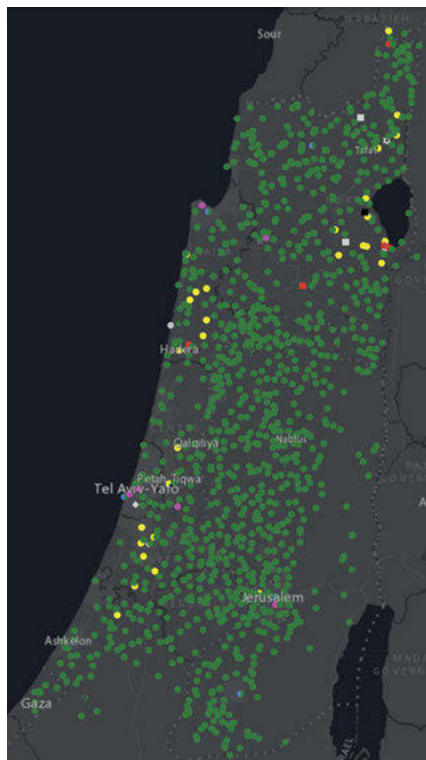
**Figure 2:** Main map, Ottoman Palestine (1917), Solomon Vimal.
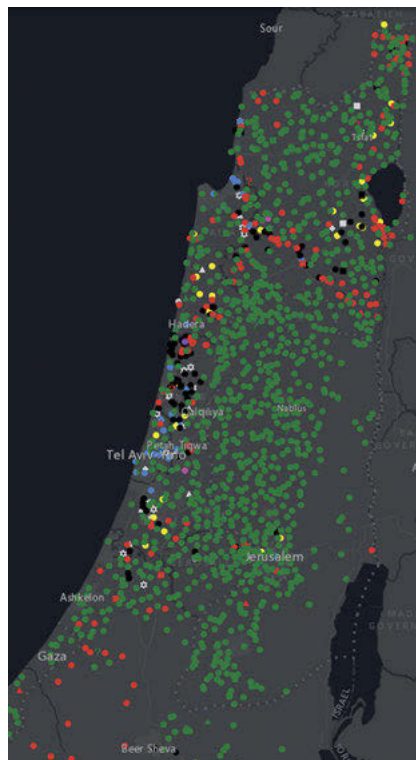


**Figure 3:** Main map, British Mandatory Palestine (1947), Solomon Vimal.

## 4  At What Point a Point?

As indicated above, the *all* in the project title reflects my desire to represent every kind of settlement and to be as comprehensive as possible. But what merits a point on the map? Or, stated otherwise, what counts as a settlement? Our general rule of thumb is to include any group of people beyond an individual family unit that resided together in a specific site for at least the better part of a year and had the sense that they constituted a distinct community. It is of no consequence if the community quickly moved or disbanded, or whether or not it received official recognition or sanction; it existed for a period of time and is therefore worthy of inclusion on the map. Thus, aside from private family farms and nomadic tribes, all other kinds of settlements are ultimately intended to be present, including monasteries, Palestinian refugee camps, the *ma'abarot* transit camps used to
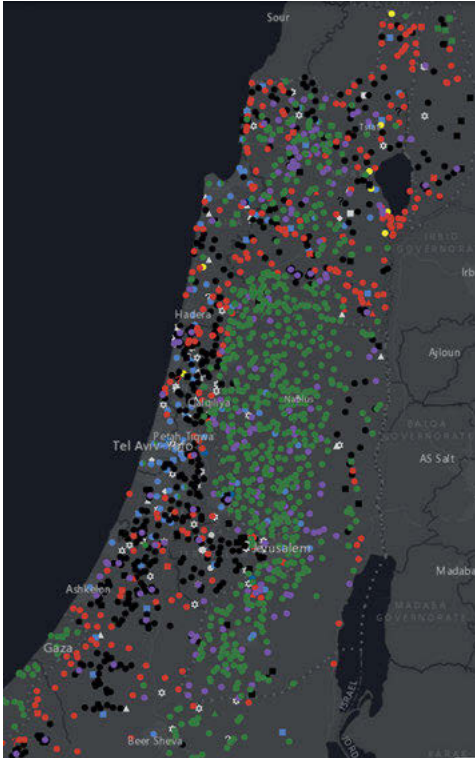
**Figure 4:** Main map, Israel & the Palestinian Territories
(2021), Solomon Vimal.

house immigrants to the new State of Israel in the 1950s, unrecognized Bedouin villages in the Negev, and settler outposts on the West Bank. To be sure, because many of these kinds of communities expose the limits of state oversight and control – and thus also of data collection, it will not always be possible to represent them to our satisfaction. To evoke *Ma'aleh Avak*, there will likely always be points missing from, and incorrectly represented on, the map.

For instance, a *ma'abarah* ("transit camp") may have been *technically* closed in, say, 1958, its erstwhile residents offered permanent quarters elsewhere, but several dozen families may have continued to reside there nonetheless. Similarly, the *official* evacuation of a settler outpost does not exclude some residents remaining or managing to return immediately thereafter. Finally, it may be difficult to determine if a particular area of Bedouin Arab settlement constitutes one or multiple communities. In each case, we will make the best determination possible in light of the available data. Fortunately, an online map is always subject to revision.

With regard to community start dates, we try to get as close as possible to the moment in time at which a site was first actually populated or regarded itself as a distinct community, regardless of when it was officially founded. As for end date, in addition to the problems raised in the previous paragraph, we are often faced with the challenge of appropriately representing the historical trajectory of an individual community's relations with its neighbor(s). For example, if two or more previously independent settlements are combined into a single municipality, should they from that point on be denoted on the map as a single point? Here, too, we believe that case-by-case determination is essential.

Consider, as an example, 'Ein Ganim, the first moshav po'alim or worker's moshav, founded in 1907 and formally annexed by neighboring Petaḥ Tikvah in 1937. At around that time, it ceased to maintain its identity as a distinct community and for this reason does not merit a separate point thereafter. By contrast, the Palestinian Arab village of Silwan was fully annexed by the Jerusalem municipality in 1952, but retains its unique character to this day and thus still warrants a distinct point. Likewise, on the West Bank, the Jewish settlements of Alon and Nofei Prat are officially part of nearby Kfar Adumim, but since these three communities are all geographically distinct, they are each represented by discrete points.

# 5 Procuring Points: Sources and Reliability

The data underlying "All the Points" are stored on the project spreadsheet (a Google "sheet"),[11] from which ArcGIS draws directly. Much of this information was obtained by scraping two large data sets, both of which provided geographical coordinates for the communities they contained: the 1945 *British Survey of Palestine* and the 2017 *List of Localities* produced by the Israel Central Bureau of Statistics (CBS). Drawing upon these two sources brought many advantages. For example, the 1945 survey included the Palestinian villages destroyed between 1947 and 1949 or subsequently, whereas the 2017 list comprised all recognized Israeli communities (Jewish or Arab) extant in that year, including the West Bank settlements.

Nonetheless, the joint use of these two data sets also occasioned a number of challenges, including duplicates and lacunae. Pre-1945 Jewish and Arab communities that still existed in 2017 Israel appear in both sources – often with quite different spellings, rendering the respective entries difficult to identify and combine. And missing are the Jewish communities founded after 1945 that by 2017 had ceased to exist – e.g., the Sinai communities evacuated as a result

---

**11** The spreadsheet exists in multiple copies and is regularly downloaded and backed up.

of the 1978 Camp David Accords between Israel and Egypt, the Gaza and West Bank settlements uprooted in the 2005 Disengagement, as well as other communities that failed, moved, or changed their name or status, etc. Finally, because these sources did not classify settlements in the same manner as "All the Points," further research was required. Nonetheless, scraping these two sources very quickly yielded a rather comprehensive dataset.

Other communities were added or nuanced based on my own research and that of students who have contributed to the project, with *Wikipedia* serving as the primary information source. While by no means the final word, *Wikipedia* has proven to be an excellent starting point: nearly every community present or past has its own entry in English, Hebrew, and often also in Arabic, and the information required for "All the Points" is, frankly, quite minimal, and thus typically provided: name, date of founding (and, if relevant, disappearance and/or destruction), classification, and geographical coordinates. Community and organization websites have also been consulted[12] as has published scholarly research. Viewers can obtain this information and its sources by clicking on individual points, and cases in which additional clarification is required are always noted. Furthermore, as the project is ongoing, the dataset is continually being improved. The goal in this initial stage of the project is, in any case, to produce as extensive and detailed a template map as possible, as a springboard to sufficient funding for a team of researchers. They, in turn, will further nuance, supplement, and – where necessary – correct the information we have thus far obtained. Precisely because of the provisional nature of both the data and the maps, we do not feel comfortable making either publicly available just yet.[13] Parties interested in examining and/or contributing to our data and maps are invited to reach out to us privately.

# 6 Precision Points: Curated Maps

Alongside the basic and detailed primary maps introduced above, "All the Points" involves the production of more specialized, curated map *exhibits*. While also

---

**12** The Palestine Remembered website (https://www.palestineremembered.com) has been a helpful source of data for the depopulated Palestinian villages, while the היאחזויות section of http://www.nahal.co.il/, a website devoted to preserving the memory of fallen Naḥal soldiers and the overall Naḥal heritage, has been used extensively in tracking the dates on which Naḥal outposts were founded and either civilianized, as was typically the case, or removed. An acronym for *No'ar ḥalutzi loḥem* ("Fighting Pioneer Youth"), Naḥal refers to a program that traditionally combined military service with the founding of agricultural settlements, primarily in peripheral areas.
**13** Our goal is formally to launch the open access website of the project in 2023.

designed to be interactive, these maps focus on specific, pre-determined regions or aspects of settlement history that merit more focused examination and/or require more precise chronological scale than is possible with an all-encompassing map. Here is a list of the curated maps currently in development:

1) "'In the Land of Jesus': European Christian Colonies in the Holy Land (1867–1948)"
2) "'Wall and Tower': The Emergence and Spread of an Iconic Settlement Type (1936–1939)"
3) "'From Palestine to Israel': The Disappearance and Appearance of Arab and Jewish Settlements (November 1947–November 1949)"
4) "'Peripheral Points': "From Naḥal Military Outpost to Civilian Settlement (1951–2013)"
5) "'Arab Settlement in the Jewish State': Recognized and Unrecognized Arab Communities Founded in Israel since 1948"
6) "'Home on the Hilltop': The Founding of Galilean 'Mitzpim' and West Bank Settlements and Outposts (1967–2000)"
7) "'Jerusalems': Arab, Jewish, and Other Neighborhoods in the Holy City and Environs"
8) "'Settled, Unsettled, Resettled': The Golan Heights (1878–Present)"
9) "'Points That Never Were': Fictional Communities in Israel/Palestine"

In what follows, I elaborate upon three of these maps: "Wall and Tower," "Settled, Unsettled, Resettled," and "Points that Never Were."

## 6.1 "Wall and Tower"

Alternatively translated as either "Wall and Tower" or "Tower and Stockade," the Hebrew phrase *Ḥoma u-Migdal* refers to the 52 impromptu Jewish settlements that were set up (often overnight) during the Arab Revolt of 1936–1939. Initially, these communities comprised, quite literally, a perimeter wall and lookout tower. As it became increasingly clear in this period that Mandatory Palestine was headed for some form of partition, the founding of these settlements was intended to secure and/or expand the Jewish presence in strategic territory. This particular enterprise occupies a central place in Israeli collective memory, gave birth to the notion of "another point on the map," and has exerted substantial influence on subsequent Israeli architecture and settlement (in terms of both strategy and practice alike), particularly on the post-'67 West Bank.[14] It is thus highly appro-

---

**14** Rotbard, "Wall and Tower," 49–50.

priate to produce a curated map devoted to the "Wall and Tower" communities. This map distinguishes between kibbutzim and moshavim – the two kinds of "Wall and Tower" settlement – and, in light of the brief time frame, showcases the founding of these communities on a month-to-month basis. See Figure 5 for a composite view of the "Wall and Tower" curated map.
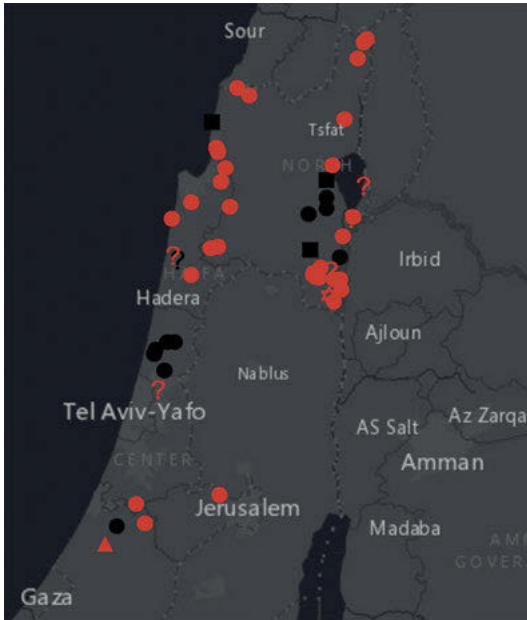


**Figure 5:** Curated Map, "Wall and Tower" (1936–1939), Solomon Vimal.

## 6.2 "Settled, Unsettled, and Resettled"

The purpose of this curated map is to showcase the modern settlement history of the Golan Heights, a history as dramatic as it is poorly known. This map is especially timely, as the Trump Administration's recognition in 2019 of Israeli sovereignty over the majority of the strategic plateau (seized from Syria during the 1967 Six-Day War) subjected this region to an unusual degree of media attention.[15]

---

**15** In February 2021, new U.S. Secretary of State Anthony Blinken partially walked back this decision.

The modern settlement history of the Golan commences only in 1878 with the founding of the Circassian town of Quneitra and other nearby villages.[16] Thereafter, the establishment of new communities continued at a varying, but at times rapid, rate (up to an average of six or eight new settlements per year), such that on the eve of the 1967 war, the Heights were host to 273 communities with a combined population of approximately 150,000. While it is unlikely that we will acquire sufficient data to map the founding of these communities on an annual basis, thanks to the assistance of Israeli Golan Heights scholar Yigal Kipnis, we are already able to produce maps showcasing the communities founded in the periods 1878–1884, 1885–1913, 1914–1945, and 1945–1967, and to denote the predominant ethno-religious group in each settlement, whether Bedouin Arab, Circassian, Turkmenian, Druze, Alawite, or Christian Arab.[17] Two hundred and twenty-three of these communities (all villages, with the exception of Quneitra) were located in the territory conquered by Israel, and all but five of them were abandoned by their inhabitants and ceased to exist.[18] Thereafter, a modest Israeli settlement drive began, resulting in the founding of the 32 Jewish communities extant on the Golan today.[19] In 1981, Israel also formally applied its law to the Golan Heights, effectively annexing the region. Thus, it is anticipated that the individual images of "Settled, Unsettled, and Resettled" will depict the Golan pre-1878, in 1884, 1913, 1945, May 1967, June 1967, 1981, and 2022.

---

**16** The contents of this section draw heavily from Yigal Kipnis, *The Golan Heights: Political History, Settlement and Geography since 1949* (London: Routledge, 2013), 125–48 (Chapter 3, "The Settlement Map of the Syrian Golan").

**17** Bedouin Arabs constituted the majority of the inhabitants, with a significant presence of both Circassians and Druze and, to a lesser extent, of Turkmenians and Christians. See Kipnis, *The Golan Heights*, 136, for a breakdown of the population on ethno-religious lines.

**18** In two cases, a few residents remained behind in other settlements. See Kipnis, *The Golan Heights*, 145.

**19** 33, if one counts *Ramat Trump* ("Trump Heights"), officially founded in honor of the former American president on June 16, 2019. The first of a planned initial 20 families moved to the new community in April 2021. See Guy Varon, "The First Family Has Come to Live in Ramat Trump: 'Strange without Neighbors, but it's an Adventure'" [Hebrew], *N12*, April 7, 2021, accessed November 29, 2021, https://www.mako.co.il/news-israel/2021_q2/Article-2d78a08db4ba871026.htm; "Community Settlement Ramat Trump, Golan" [Hebrew], *Atar ha-bayit: atar ha-megorim ba-kfar* (The Home Site: The Site for Village Residences), undated, accessed November 29, 2021, https://www.homee.co.il/רמת-טראמפ/?s=17932.

## 6.3 "Points That Never Were"

Perhaps the most surprising (and certainly the most creative) aspect of a project committed to depicting "who has lived where, when, and in what kind of community" is "Points That Never Were" ("Nekudot she-lo hayu me-'olam"). The aim of this curated map is to document settlements of the imagination – i.e., communities that exist only on the page, on the screen, or onstage. Probably the best-known example thereof is Beit ha-Tikvah ("House of Hope"), the desert development town mistakenly visited by the Alexandria Ceremonial Police Orchestra in the award-winning musical *The Band's Visit* (2016–) and, previously, in the similarly acclaimed 2007 Israeli film of the same name (*Bikur ha-Tizmoret*, in Hebrew).

Fantasy settlements have, however, long been a staple of artistic engagement with Israel. To offer but a few examples: the renowned Israeli satirist Ephraim Kishon set his 1955 novel '*Ein Kamonim* (*Cumin Spring*), at a fictional moshav bearing the same name deep in the Galilee,[20] while for his 1965 debut novel *Makom aḥer* (*Elsewhere, Perhaps*), Amos Oz did in fact invent an "other place" (the literal translation of the Hebrew title), namely Kibbutz Metzudat Ram ("Lofty Fortress"). Indeed, the novel's very first page offers a rather precise description of its location near the Israel-Syria border north of the Sea of Galilee. Passing to North America, readers of Leon Uris's saccharine *Exodus* or James Michener's sweeping *The Source* may recall that these books go so far as to include maps in which their fictional Israeli communities are fixed in precise locations.[21]

More recently, Assaf Gavron created the outpost Ma'aleh Ḥarmesh Gimel ("Scythe Ascent 3") for *Ha-giv'ah* (*The Hilltop*), his 2013 novel about a West Bank settlement. And even Theodore Herzl, in his 1902 fantasy novel *Altneuland* (*Old-New Land*), dreamt into existence – just outside Nazareth – Neudorf or "New Village," an idealized, cooperative farming community. Finally, it would be remiss not to note the presence on this map of the Ma'aleh Avak mentioned above, here placed way in the south, since – as resounds in the song – *sham le-mata' ḥaserah 'od nekudah* ("down below another point is missing"). See Figure 6 for a composite image of "Points That Never Were."

---

[20] The novel – still set at 'Ein Kamonim – was republished in 1972 under the title *Ha-shu'al be-lul ha-tarnegolim* (*The Fox in the Chicken Coop*). A film bearing the revised title was released in 1977 and a private farm bearing the name "'Ein Kamonim" was founded in 1979–1980 southwest of Safed on Route 85. To be sure, the name is also inspired by the presence nearby of Har Kamon, the highest peak in the Lower Galilee.

[21] Leon Uris, *Exodus* (Garden City, NY: Doubleday, 1958); James A. Michener, *The Source: A Novel* (New York, Fawcett, 1965).

**Figure 6:** Curated Map, "Points That Never Were", Solomon Vimal.

Paradoxically, the sheer number of these fantasy points – there are undoubtedly many more awaiting inclusion on this map – provides perhaps the very best evidence for the importance of thinking about the history of Zionism and Israel from the perspective of the *real* points to which they gave rise. And thus, this most unusual approach to mapping Israel strikes me as a worthy capstone both for the project as a whole and for this exposition thereof.

# 7 A Few Final Points

To be sure, "All the Points" is an ambitious and challenging undertaking, the kind of initiative likely to resist any clear point of completion. Indeed, as of this writing, it is hard to predict what complications lie ahead and to what degree the project will be able to be fully realized as intended. And yet, at the same time, the experience thus far of dynamically mapping Israeli and Palestinian communities raises the possibility of extending this same approach to other regions and periods with contested histories. "All the Points" may thus ultimately come to encompass even more points than initially anticipated.
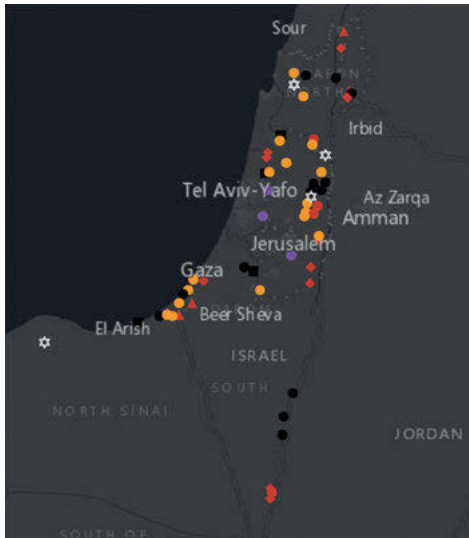
**Figure 7:** Curated map, "Peripheral Points" (here 1951–1977), Solomon Vimal.
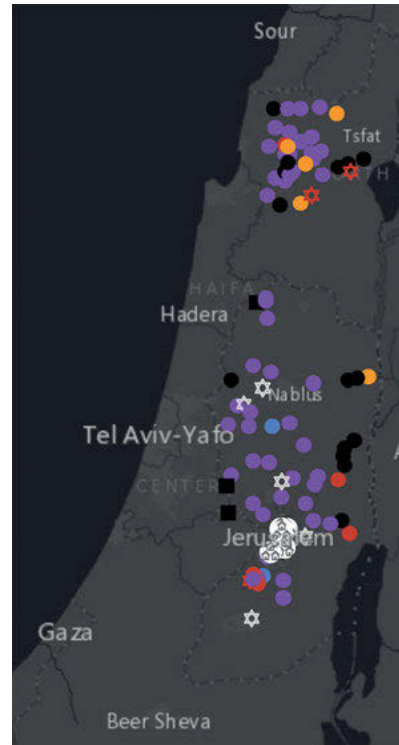
**Figure 8:** Curated map, "Home on the Hilltop" (1981), Solomon Vimal.

Even in its current, pilot state, however, "All the Points" represents a valuable pedagogical tool and points (pun indeed intended!) to new ways of learning and thinking about the modern history of the Israeli-Palestinian region. It is extremely powerful to watch the year-by-year slideshow of the main map: one sees Israel emerge before one's eyes and observes how the predominant settlement types evolve over time; at the same time, the mass disappearance of Palestinian villages in the course of 1948 is also highly visible and makes a vivid impression. In particular, the coding by color and shape gives instant expression to this land's astounding diversity, stimulating the viewer's interest and inviting them to explore further. Many lesser-known features of the region's modern settlement history are likewise rendered visible here, including the surprisingly high number of communities in which both Arab and Jewish populations were present in the 19th and early 20th centuries, and are, once again, in the early 21st; the geographical extent of early Zionist colonization efforts; Arab communities founded in the State of Israel (i.e., post-1948); and the non-Arab, non-Jewish settlements to which this land has also been host.

Perhaps the most significant contribution of "All the Points" is that it invites us to tell a history of Israel not focused on the major centers or iconic locales, but rather from the perspective of the periphery *in*, i.e. from the "out-of-the-way" points that, taken in aggregate, are at least as essential to the story. In this regard, what might at first glance appear to be a limitation of "All the Points," namely that its points reflect neither the number of inhabitants nor the geographical extent of the communities they represent, actually reveals itself, upon closer consideration, to be one of its strengths. Indeed, the fact that each point is the exact same size encourages the viewer to take serious note of settlements that would otherwise easily slip under the radar.

To be sure, I do not exclude the eventual representation of population and area in some manner, perhaps on a decadal, as opposed to annual, basis. In other words, "who has lived where, when, and in what kind of community?" may yet also encompass "how many have lived in a where of what size?" But, for now, the limited focus of "All the Points" redounds nicely to its dual, intertwined purposes, namely (1) to showcase the contribution of each point to the ongoing making and re-making of the Israeli-Palestinian map, and (2) to celebrate each of these points as an experiment in human community worthy of attention in its own right – i.e., to forget, for a moment, that map in favor of each and all of its points.[22]

# References

Borges, Jorge Luis. "On Exactitude in Science". Translated by Andrew Hurley. Accessed April 5, 2021. https://genius.com/Jorge-luis-borges-on-exactitude-in-science-annotated.

"Community Settlement Ramat Trump, Golan" [Hebrew]. *Atar ha-bayit: atar ha-megorim ba-kfar* (The Home Site: The Site for Village Residences). Undated. Accessed November 29, 2021. https://www.homee.co.il/רמת-טראמפ/?s=17932.

Kipnis, Yigal. *The Golan Heights: Political History, Settlement and Geography since 1949*. London: Routledge, 2013.

Michener, James A. *The Source: A Novel*. New York: Fawcett, 1965.

Okon, Sarit, ed. *Aḥad ʿasar ha-nekudot: sipuro shel mivtsaʿ noʿaz* (The Eleven Points: The Story of a Daring Operation). Alumim: Beʾerot ba-Negev, merkaz moreshet, hadrakhah ve-eruaḥ, 2017.

Rotbard, Sharon. "Wall and Tower (Homa Umigdal): The Mold of Israeli Architecture." In *A Civilian Occupation: The Politics of Israeli Architecture*, edited by Rafi Segal and Eyal Weizman, 38–57. London: Verso; Tel Aviv: Babel, 2003.

---

**22** This last point points to various possible *public-facing* applications of "All the Points" – e.g., a project blog or tour, in or on which individual points and their stories could be highlighted.

Shiran, Osnat. *Nekudot ʿoz: mediniyut ha-hityashvut be-zikah le-yeʿadim politiyim u-vitḥoniyim be-terem medinah uve-reshitah* (Points of Strength: Settlement Policy in Relation to Pre-State and Early-State Political and Security Goals]. Tel Aviv: Department of Defense, 1998.

Teapacks. *Maʿaleh Avak*. Accessed April 5, 2021. https://shironet.mako.co.il/artist?type= lyrics&lang=1&prfid=429&wrkid=2205.

Uris, Leon. *Exodus*. Garden City, NY: Doubleday, 1958.

Varon, Guy. "The First Family Has Come to Live in Ramat Trump: 'Strange without Neighbors, but it's an Adventure'" [Hebrew]. *N12*, April 7, 2021. Accessed November 29, 2021. https://www.mako.co.il/news-israel/2021_q2/Article-2d78a08db4ba871026.htm.

Wehr, Hans. *A Dictionary of Modern Written Arabic*, ed. J. Milton Cowan. 3rd ed. Ithaca, NY: Spoken Language Services, 1976.

**Text**

Benjamin Charles Germain Lee

# The Digital Humanities and the Ladino Press: Using Machine Learning to Extract and Analyze Visual Content in Historic Ladino Newspapers

**Abstract:** *La Vara*, *El Tiempo, and La Boz De Oriente* represent three of the major historic Ladino newspapers published across the diasporic Sephardic Jewish world in the twentieth century. While Sephardic Jewish history and culture have received increasing scholarly attention in recent years, the vast corpus of Ladino newspapers largely remains unmined, and the field continues to be marginal from the perspective of Jewish Studies. In this chapter, I apply computational analysis of the visual content to explore the Ladino press at a macroscopic level. Using a machine learning model that I developed for my project, *Newspaper Navigator*, I have constructed a dataset of extracted photographs, illustrations, maps, comics, editorial cartoons, and advertisements from over 15,000 digitized pages of Ladino newspapers. This method represents an emerging approach to digital humanities research with periodicals and presents opportunities to facilitate access and research within Jewish Studies.

With this extracted visual content, it is possible to study the transnational dynamics shaping Sephardic print culture and the broader Sephardic experience at an unprecedented scale. Accordingly, I describe my analysis of this visual content using emerging techniques in order to provide insights related to motifs and temporal trends. I offer this work as a case study in interdisciplinary research in the digital humanities and Jewish Studies. In addition, I offer methodological

reflections related to applying emerging computational techniques to Jewish Studies. I conclude with a reflection on the ethical considerations of applying machine learning techniques to Ladino newspapers and, more generally, to Jewish cultural heritage.

**Keywords:** Newspaper Navigator, machine learning, artificial intelligence, digitized newspapers, Sephardic Studies, Ladino, digital humanities, computing cultural heritage

*La Vara*, *El Tiempo*, and *La Boz De Oriente* represent three of the major historic Ladino newspapers published across the diasporic Sephardic Jewish world in the 20th century – from New York to Constantinople to Istanbul. While Sephardic Jewish history and culture have received increasing scholarly attention in recent years, the vast corpus of Ladino newspapers largely remains unmined, and the field remains marginal from the perspective of Jewish Studies. A new collaboration between the Stroum Center for Jewish Studies' Sephardic Studies Program and the Paul G. Allen School for Computer Science & Engineering at the University of Washington seeks to draw on innovative machine learning techniques to render the visual content of Ladino newspapers more accessible to scholars and students alike and, in so doing, change the trajectory of Sephardic Studies writ large. This chapter reports the findings of this research.

Many Ladino titles have been digitized by the Sephardic Studies Program at the University of Washington. These Ladino newspapers contain not just articles and editorials but also an abundance of rich visual content that sheds light on Sephardic Jewish experiences in modernity. The advertisements appearing within Ladino newspapers have received attention from scholars within Sephardic Studies,[1] and analysis thereof has revealed connections between the American Ashkenazic and Sephardic communities, as well as the ways in which advertisers' attempts to provide remedies speak to "readers' anxiety about the fragility of life under Ottoman rule."[2] Indeed, the visual content within newspapers has proven to be a capacious source for humanists. Within periodicals studies, scholars have utilized the visual content in newspapers to investigate topics as far ranging as the evolution of comedic sensibilities within comic strips to hidden editorial practices

---

**1** Makena Mezistrano, "Why Are These Passover Ads Different from All Other Ads?" Stroum Center for Jewish Studies (2021), accessed May 19, 2021, https://jewishstudies.washington.edu/sephardic-studies/why-are-these-passover-ads-different-from-all-other-ads/.
**2** Sarah Abrevaya Stein, *Making Jews Modern: The Yiddish and Ladino Press in the Russian and Ottoman Empires* (Bloomington: Indiana University Press, 2004), 179.

embedded within newspaper layout.[3] This collective body of work is bolstered by new methodologies being employed within the digital humanities to extract and analyze visual content in historic newspapers.[4]

In this chapter, I scale up this analysis of visual content to explore the Ladino press at a macroscopic level. Using a machine learning model that I developed as part of my project, *Newspaper Navigator*, I have constructed a dataset of extracted photographs, illustrations, maps, comics, editorial cartoons, and advertisements from over 15,000 digitized pages of these Ladino newspapers.[5] This approach of utilizing a machine learning model to extract visual content represents an emerging methodology for digital humanities research with periodicals and presents opportunities to facilitate access and research within Jewish Studies.

With the extracted visual content from the Ladino newspapers, it is possible to study the transnational dynamics shaping Sephardic print culture and the broader Sephardic experience at an unprecedented scale. Accordingly, I describe my results related to analyzing this visual content using emerging visualization techniques in order to provide insights related to recurring motifs and temporal trends. I offer this work as a case study in interdisciplinary research in the digital humanities and Jewish Studies. Throughout the chapter, I offer methodological reflections related to applying emerging computational techniques to Jewish Studies. I conclude the chapter with a reflection on the ethical considerations of applying machine learning and computer vision techniques to these Ladino newspapers and, more generally, to Jewish cultural heritage.

---

**3** Jean Lee Cole, *How the Other Half Laughs: The Comic Sensibility in American Culture, 1895–1920* (Jackson: University Press of Mississippi, 2020); Kevin G. Barnhurst and John Nerone, *The Form of News: A History* (New York: The Guilford Press, 2002).

**4** Andrew Piper, Chad Wellmon, and Mohamed Cheriet, "The Page Image: Towards a Visual History of Digital Documents," *Book History* 23, no. 1 (2020): 365–97, accessed December 10, 2020, https://doi.org/10.1353/bh.2020.0010; Paul Fyfe and Qian Ge, "Image Analytics and the Nineteenth-Century Illustrated Newspaper," *Journal of Cultural Analytics*, October 25, 2018, 11032, accessed December 15, 2020, https://doi.org/10.22148/16.026; Melvin Wevers and Thomas Smits, "The Visual Digital Turn: Using Neural Networks to Study Historical Images," *Digital Scholarship in the Humanities* 35, no. 1 (April 1, 2020): 194–207, accessed December 15, 2020, https://doi.org/10.1093/llc/fqy085.

**5** Benjamin Charles Germain Lee et al., "The Newspaper Navigator Dataset: Extracting Headlines and Visual Content from 16 Million Historic Newspaper Pages in Chronicling America," in *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, CIKM '20 (New York: Association for Computing Machinery, 2020), 3055–62, accessed December 15, 2020, https://doi.org/10.1145/3340531.3412767; Benjamin Charles Germain Lee and Daniel S. Weld, "Newspaper Navigator: Open Faceted Search for 1.5 Million Images," in *Adjunct Publication of the 33rd Annual ACM Symposium on User Interface Software and Technology*, UIST '20 Adjunct (New York: Association for Computing Machinery, 2020), 120–122, accessed December 15, 2020, https://doi.org/10.1145/3379350.3416143.

# 1 The Digital Humanities and Visual Analysis of Newspapers

The visual culture preserved within historic newspapers has proven to be a fruitful and capacious source among scholars across diverse research areas. For example, scholars have studied the embedded editorial cartoons to understand the invocation of historical analogies;[6] comic strips to understand the evolution of humor about ethnicity;[7] illustrations to study the portrayal of identity;[8] maps to assess cartographic practices as well as the spatial thinking abilities of readers;[9] and photographs to study the history of photojournalism.[10] Within Jewish Studies, Sarah Stein's book *Making Jews Modern: The Yiddish and Ladino Press in the Russian and Ottoman Empires* makes a compelling case for the significance of visual culture within Ladino and Yiddish newspapers. Stein's detailed analysis of the advertisements within the Constantinople-based Ladino newspaper *El Tiempo* traces recurring motifs in order to argue how readers sought remedies to the anxieties of modernity under Ottoman rule: advertisements for medicines, insurance, clothes, travel, and lotteries all targeted readers concerned about stability and class.[11] This chapter builds on this already significant body of work in order to consider advertisements and other visual content in the Ladino press at the macroscopic scale.

Scholarship making use of visual culture in historic newspapers has been redoubled by the growing interest in visual analysis within the digital humanities. Though research in the digital humanities has historically centered around text as the primary medium of interest, the field's "visual digital turn" over the

---

**6** Betty H. Winfield and Doyle Yoon, "Historical Images at a Glance: North Korea in American Editorial Cartoons," *Newspaper Research Journal* 23, no. 4 (September 1, 2002): 97–100, accessed June 10, 2021, https://doi.org/10.1177/073953290202300411.

**7** Cole, *How the Other Half Laughs*.

**8** Andrea N. Williams, "Cultivating Black Visuality: The Controversy over Cartoons in the Indianapolis *Freeman*," *American Periodicals* 25, no. 2 (2015): 124–38, accessed June 10, 2021, http://www.jstor.org/stable/24589083.

**9** Pınar Sarın and Necla Uluğtekin, "Analyzing Newspaper Maps for Earthquake News through Cartographic Approach," *ISPRS International Journal of Geo-Information* 8, no. 5 (May 2019): 235, accessed June 10, 2021, https://doi.org/10.3390/ijgi8050235; André Reyes Novaes, *Maps in Newspapers: Approaches of Study and Practices in Portraying War since 19th Century* (Leiden: Brill, 2019), accessed June 10, 2021, https://brill.com/view/title/54806.

**10** Michael Griffin, "The Great War Photographs: Constructing Myths of History and Photojournalism," in *Picturing the Past: Media, History and Photography*, ed. Bonnie Brennan and Hanno Hardt (Urbana: University of Illinois Press, 1999), 122–57.

**11** Stein, *Making Jews Modern*, 153–201.

past decade has begun foregrounding the analysis of visual media, including images and video.[12] This visual digital turn has coincided with methodological advances in machine learning approaches to image analysis due to deep learning. With the democratization of deep learning approaches to image recognition over the past few years via open source libraries and pre-trained models, digital humanities practitioners have begun utilizing these approaches for a wide range of research goals, from enriching the metadata of digitized collections to analyzing sitcoms.[13] As these approaches continue to improve, it is clear that machine learning will occupy an increasingly important role within the digital humanities and the humanities writ large, as well as within the cultural heritage sector, including libraries and archives. Within periodicals studies, researchers have begun utilizing machine learning approaches to study the visual components of newspaper pages, from analyzing visual layouts to classifying and searching the visual content embedded within the pages.[14] Indeed, as a mode of humanistic inquiry, the application of machine learning to the visual analysis of newspapers has much to offer to both Jewish Studies and the digital humanities.

In the case of the Ladino press, the utilization of machine learning methodologies to study the visual content embedded within the newspaper pages is even more urgent due to the extant challenges surrounding the application of optical character recognition (OCR) algorithms to transcribing Ladino texts. Off-the-shelf OCR algorithms have yielded poor performance to date because these algorithms interpret Ladino texts printed in Rashi script as Hebrew; the poor OCR quality in turn restricts the ability of scholars to perform reliable keyword searches or apply digital humanities methodologies for textual analysis. Though new OCR engines

**12** Lev Manovich, "How to Compare One Million Images?," in *Understanding Digital Humanities*, ed. David M. Berry (London: Palgrave Macmillan, 2012), 249–78, accessed December 15, 2020, https://doi.org/10.1057/9780230371934_14; Lev Manovich, "Data Science and Digital Art History," *International Journal for Digital Art History*, no. 1 (June 26, 2015), accessed December 15, 2020, https://doi.org/10.11588/dah.2015.1.21631; Wevers and Smits, "The Visual Digital Turn."
**13** Joshua Gomez et al., "Experimenting with a Machine Generated Annotations Pipeline," *The Code4Lib Journal*, no. 48 (May 11, 2020), accessed December 10, 2020, https://journal.code4lib.org/articles/15209; Matthew Lincoln et al., "CAMPI: Computer-Aided Metadata Generation for Photo Archives Initiative," October 8, 2020, accessed December 15, 2020, https://doi.org/10.1184/R1/12791807.v1; Elizabeth Lorang et al., "Digital Libraries, Intelligent Data Analytics, and Augmented Description: A Demonstration Project," 2020, accessed December 15, 2020, https://labs.loc.gov/static/labs/work/experiments/final-report-revised_june-2020.pdf; Taylor Arnold, Lauren Tilton, and Annie Berke, "Visual Style in Two Network Era Sitcoms," *Journal of Cultural Analytics*, July 19, 2019, 11045, accessed December 15, 2021, https://doi.org/10.22148/16.043.
**14** Piper, Wellmon, and Cheriet, "The Page Image"; Fyfe and Ge, "Image Analytics and the Nineteenth-Century Illustrated Newspaper"; Wevers and Smits, "The Visual Digital Turn"; Lee et al., "The Newspaper Navigator Dataset"; Lee and Weld, "Newspaper Navigator."

are being developed specifically for Ladino texts, a fundamental challenge remains at this time: how do we study the Ladino press at a macroscopic scale beyond close reading?[15] Applying machine learning to study the visual content in these pages affords us a path forward.

The next sections of this chapter concern the methodology employed to extract and analyze this visual content within 15,820 Ladino newspaper pages using machine learning. The chapter then turns to analyzing the extracted visual content.

# 2 Constructing the Dataset of Excavated Visual Content

This chapter explicitly builds on *Newspaper Navigator*, a project that I created in order to develop machine learning methodologies for extracting and analyzing visual content in historic newspapers. Just as the goal of OCR is to unlock the ability to search textual content, the goal of *Newspaper Navigator* is to unlock the ability to search *visual* content embedded within historic newspaper pages. The project utilizes a visual content recognition model to automate the identification and classification of visual content within digitized historic newspaper pages according to seven different categories: photographs, illustrations, maps, comics, editorial cartoons, headlines, and advertisements. The visual content recognition model is an instantiation of a machine learning algorithm known as a neural network, which has achieved state-of-the-art performance across a range of machine learning tasks and domains, from images to text.

The visual content recognition task considered in *Newspaper Navigator* is a form of object detection, a computer vision task formulated with the goal of identifying and localizing objects in images. Accordingly, the visual content recognition model began as an object detection model that had been trained by machine learning practitioners on the oft-used Common Objects in Context dataset to identify the locations of objects such as dogs and bicycles in images.[16] In order to succeed at this more granular task of identifying visual content on digitized newspaper pages, I subsequently trained the model on crowdsourced bounding box annotations of visual content in World War I-era newspaper pages as part of

---

**15** DiJeSt Team, "Our Text Recognition Ground Truth and Model – DiJeSt," accessed March 31, 2021, https://dijest.net/gtmodel/.

**16** Tsung-Yi Lin et al., "Microsoft COCO: Common Objects in Context." *ArXiv:1405.0312 [Cs]*, February 20, 2015, accessed December 10, 2020, http://arxiv.org/abs/1405.0312.

the Beyond Words initiative launched by the Library of Congress.[17] This framework of utilizing a model that had already been trained on one task and training it on a more specialized, domain-specific dataset is known within machine learning as *pre-training* and *finetuning*. An example of predictions made by the fully trained model on a sample newspaper page from *Chronicling America* is shown in Figure 1.

In the construction of the *Newspaper Navigator* dataset, this trained visual content recognition model was utilized to extract visual content from 16.3 million newspaper pages in the *Chronicling America* database. The visual content recognition model, as well as the code and training dataset utilized to train it, is available at the Library of Congress's GitHub repository for the project.[18] The model is also pip-installable via the LayoutParser library.[19] For more information on the visual content recognition model, I refer the reader to the technical paper on the *Newspaper Navigator* dataset, as well as the corresponding data archaeology.[20]

In this section, I describe the process of utilizing this visual content recognition to extract visual content from Ladino newspapers in more detail. The corpus of Ladino newspapers consists of 15,820 Ladino newspaper pages from eight titles published between 1890 and 1948, amounting to 63.3 gigabytes (GB) of image data.[21] Table 1 presents a breakdown of the Ladino corpus included in this analysis according to title and date of publication. Manually assembling a dataset of extracted visual content across this full corpus would require hundreds of human annotation hours. However, the automated *Newspaper Navigator* visual content recognition model can process multiple pages per second on a single graphics processing unit (GPU), making it possible to process the full Ladino corpus in just a few hours.

To begin the processing of these pages, I first moved the high-resolution images of digitized Ladino newspaper pages to a private Amazon AWS S3 bucket, a form of cloud storage that facilitates fast computing against the corpus. I then wrote code to process these pages using the existing *Newspaper Navigator* visual

---

**17** LC Labs, "Beyond Words," accessed March 31, 2021, http://beyondwords.labs.loc.gov/#/.

**18** Benjamin Charles Germain Lee, *LibraryOfCongress/Newspaper-Navigator*, GitHub Repository, Library of Congress, 2020, accessed December 10, 2020, https://github.com/LibraryOfCongress/newspaper-navigator.

**19** Zejiang Shen et al., "LayoutParser: A Unified Toolkit for Deep Learning Based Document Image Analysis," *ArXiv:2103.15348 [Cs]*, March 29, 2021, accessed June 10, 2021, http://arxiv.org/abs/2103.15348.

**20** Benjamin Charles Germain Lee, "Compounded Mediation: A Data Archaeology of the Newspaper Navigator Dataset," Digital Humanities Quarterly 15, no. 4, 2021, accessed January 10, 2022, http://digitalhumanities.org/dhq/vol/15/4/000578/000578.html.

**21** Lee et al., "The Newspaper Navigator Dataset."

**Figure 1:** An example of the visual content recognition model's predictions on a page from *Chronicling America*. Chronicling America, *The Ogden Standard*, April 16, 1918, accessed December 10, 2020, https://chroniclingamerica.loc.gov/lccn/sn85058396/1918-04-16/ed-1/seq-6/. For each bounding box, the predicted category is shown in the top left, along with the machine learning model's confidence score.

content recognition model that had been trained on the Beyond Words annotations. Because machine learning models can be stored as single files known as "weights files," which can be loaded onto a computer and utilized for processing data with just a few lines of code, the majority of this code was devoted to handling the downloading of images from the cloud and processing the images in parallel. To deploy this code, I ran the processing pipeline on a rented Amazon AWS g4dn.12xlarge EC2 instance consisting of 48 CPUs and four NVIDIA T4 GPUs.[22] In total, the pipeline extracted six classes of visual content across the corpus of Ladino newspapers: photographs, illustrations, maps, comics, editorial cartoons, and advertisements.[23] I then saved the resulting extracted images, as well as metadata from the machine learning model, to the AWS S3 bucket, making it straightforward for us to download the full dataset and relevant subsets as necessary. I am currently in the process of investigating options for making this dataset of extracted visual content available to researchers and the public alike.

A breakdown of identified visual content is presented in Table 2. Because the machine learning model returns a confidence score with each predicted bounding box, and because one's choice of threshold cut on confidence score affects one's tradeoff between false positives and false negatives and thus changes the size of the resulting dataset, I include three cuts on confidence score in the table: 90%, 70%, and 50%.

Notably, this visual content recognition model was trained on annotated World War I-era newspaper pages in *Chronicling America*, rather than annotated Ladino pages. Consequently, the resulting dataset contains a nontrivial number of false positives and false negatives, as evidenced by the map class, which largely consists of false positives. It should be noted that the performance of the visual content recognition model is dependent on a confluence of factors, ranging from page layout to time period, typeface, language, and even subtleties of the digitization pipeline, such as the scanner used to image the pages. For an analysis of the effects of time period on the generalization performance of the visual content recognition model, I refer the reader to the *Newspaper Navigator* dataset paper.[24] It is undoubtedly the case that the utilization of a model trained on annotations for *Chronicling America* pages, rather than Ladino pages, impacts the resulting

---

**22** The total computing costs for testing and deploying the pipeline amounted to less than 50 USD.

**23** Because the pages did not have corresponding OCR (for reasons described earlier in this chapter), this pipeline did not attempt to extract any textual captions within the predicted bounding boxes, and predicted headlines were also omitted from the resulting dataset.

**24** Lee et al., "The Newspaper Navigator Dataset."

dataset. This effect can be quantified by evaluating the performance of the visual content recognition model on a hand-labeled test sample of Ladino pages (an evaluation left for future work). However, as evidenced by the analysis of the dataset presented in the next section, it is clear that the resulting dataset is of more than sufficient quality for supporting downstream exploration and research pertaining to questions of humanistic inquiry.

**Table 1:** Ladino newspaper titles with corresponding number of images and digitized pages processed using *Newspaper Navigator*. In the case of *La Vara*, each image contains two newspaper pages. In bold are the statistics for all digitized pages for a given title.

| Newspaper Title | # of Digitized Images | # of Pages |
|---|---|---|
| *El Instruktor revista siyentifika i literaria* | **331** | **331** |
| *El jugeton, Jurnal umoristiko* | **4** | **4** |
| *El Kirbatch Americano* **(1915–1917)** | **206** | **206** |
| *El Luzero Sefaradi* (October 1926) | 28 | 28 |
| *El Luzero Sefaradi* (May 1927) | 28 | 28 |
| *El Luzero Sefaradi* **(Total)** | **56** | **56** |
| *El Progresso / Yosef Daat* | **332** | **332** |
| *El Tiempo* (1890–1891) | 926 | 926 |
| *El Tiempo* (1896–1897) | 1,138 | 1,138 |
| *El Tiempo* (1900) | 584 | 584 |
| *El Tiempo* (1900–1901) | 981 | 981 |
| *El Tiempo* **(Total)** | **3,629** | **3,629** |
| *La Boz de Oriente* **(April 1931–April 1932)** | **860** | **860** |
| *La Vara* (January 9, 1922–June 4, 1923) | 141 | 282 |
| *La Vara* (May 1, 1927–December 17, 1929) | 675 | 1,350 |
| *La Vara* (January 3, 1930–December 20, 1932) | 701 | 1,402 |
| *La Vara* (December 7, 1932–April 25, 1941) | 667 | 1,334 |
| *La Vara* (January 6, 1933–December 27, 1935) | 636 | 1,272 |
| *La Vara* (January 3, 1936–August 26, 1938) | 704 | 1,408 |
| *La Vara* (September 2, 1938–April 25, 1941) | 681 | 1,362 |
| *La Vara* (May 2, 1941– December 29, 1944) | 648 | 1,296 |
| *La Vara* (January 5, 1945–February 13, 1948) | 348 | 696 |
| *La Vara* **(Total)** | **5,201** | **10,402** |
| **Total** | **10,619** | **15,820** |

**Table 2:** A breakdown of extracted visual content in the Ladino newspaper titles processed. Three different cuts on the visual content recognition model's confidence score (90%, 70%, and 50%) are presented to show the effect of the cut choice when favoring false positives or false negatives.

| Visual Content Type | # > 90% | # > 70% | # > 50% |
|---|---|---|---|
| Photographs | 348 | 770 | 1,060 |
| Illustrations | 52 | 378 | 960 |
| Maps | 27 | 182 | 300 |
| Comics | 10 | 39 | 184 |
| Editorial Cartoons | 8 | 31 | 111 |
| Advertisements | 18,381 | 31,523 | 42,505 |
| **Total** | **18,826** | **32,923** | **45,120** |

# 3 Analyzing the Excavated Visual Content

To begin the macroscopic analysis of the excavated visual content, I created cluster-based visualizations of advertisements and photographs, grouped according to their semantic content. In this step, I generated image embeddings for all of the extracted visual content; to accomplish this, I modified the *Newspaper Navigator* pipeline code, available in the Library of Congress GitHub repository for *Newspaper Navigator*, and utilized img2vec, a library for the streamlined generation of image embeddings from image files.[25] The image embeddings utilized in this analysis are lower-dimensional representations of the images extracted from the hidden layers of ResNet-18 and ResNet-50, two neural image classification models.[26] Originally trained on ImageNet, these models can classify images according to their content (e.g., "dog" or "cat").[27] Because these models capture the semantics of images, image embeddings generated by feeding images into these models capture semantic similarity: if the distance between two image embeddings (vectors that are each hundreds or thousands of dimensions in length) is

**25** Lee, GitHub Repository; Safka, Christian, *Christiansafka/Img2vec*, Python, 2021, accessed December 10, 2020, https://github.com/christiansafka/img2vec.

**26** K. He et al., "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, 770–78, accessed December 10, 2020, https://doi.org/10.1109/CVPR.2016.90.

**27** J. Deng et al., "ImageNet: A Large-Scale Hierarchical Image Database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009, 248–55, accessed December 10, 2020, https://doi.org/10.1109/CVPR.2009.5206848.

small, the corresponding images likely have similar semantics. Thus, generating visualizations of photos and advertisements clustered based on the image embeddings can provide an informative summary of the landscape of visual content. To make the high-dimensional clustering visible in two-dimensional visualizations, I have utilized T-SNE, a dimensionality reduction algorithm that preserves close clusters of points, meaning that clustered points in the visualization are also clustered in the high-dimensional embedding space. It should be noted that long distances are not preserved in T-SNE, so the relative positions of clusters should not be taken into consideration.[28]

To start, I generated visualizations of the 348 photographs identified by the *Newspaper Navigator* visual content recognition model with confidence scores greater than 90%. Figure 2a presents the cluster visualization of these 348 photographs. In this visualization, I present a summary view that would have ordinarily required manually inspecting and analyzing 15,820 pages. Examining the visualization, it is immediately apparent that many photographs depict people. Portrait shots, such as the ones clustered together and depicted in Figure 2b, are one of the most common types of image. Other notable clusters include wartime photographs (shown in Figure 2c) and crowds and groups of people (shown in Figure 2d).

As shown in Table 2, the critical mass of identified visual content consists of advertisements. The advertisements embedded within the Ladino press attest to daily life within Sephardic culture from Constantinople to New York. By studying these extracted advertisements at a macroscopic scale, it is possible to augment the extant historiography, including Sarah Stein's detailed analysis of advertisements in the Ladino and Yiddish Press.[29]

Given the number of advertisements identified, I chose to generate cluster visualizations of advertisements for *La Vara* for different temporal regions. Figure 3a shows a cluster visualization of 2,812 advertisements extracted from *La Vara* issues published between January 3, 1936, and August 26, 1938. The visualization reveals numerous distinct clusters of the same advertisements reproduced multiple times throughout multiple issues of *La Vara* within the given temporal range. Many of these clusters contain dozens of the same advertisement, reflecting businesses that chose to advertise consistently within the pages of *La Vara*. Figure 3b shows six of these clusters, along with magnified versions of

**28** Laurens van der Maaten and Geoffrey Hinton, "Visualizing Data Using T-SNE," *Journal of Machine Learning Research* 9, no. 86 (2008): 2579–605. For more information on T-SNE, see: Martin Wattenberg, Fernanda Viégas, and Ian Johnson, "How to Use T-SNE Effectively," *Distill* 1, no. 10 (October 13, 2016): e2, accessed December 10, 2020, https://doi.org/10.23915/distill.00002.
**29** Stein, *Making Jews Modern*, 153–201.

**Figure 2a:** A cluster visualization of the 348 photographs identified by the visual content recognition model with confidence scores greater than 90%. I constructed this visualization using ResNet-50 embeddings and T-SNE for dimensionality reduction.

the reproduced advertisements. Examples of advertisers with reproduced advertisements in the dataset include Brockman Monument Works, Meyer London's Matzos, Standard Truss Co., Golden Wine & Liquor Co., Paradise Interior Decorators, Aristocratic Imported Virgin Olive Oil, Harem Oriental Pastry, Constantinople Oriental Pastry Shop, Macedonia Importing Co., Royal Hall, Mid-Bronx Used Car Exchange, the Luxor Food Market, the Luxor Restaurant, the Sephardic Jewish Center, Inc., Joseph Levy (furniture, radios, and oilcloths), Louis J. Opal (counselor at law), Simon S. Nessim (counselor at law), Dr. J. Feitelson (dental surgeon), Irving Matalon, P. Vladeff, and Madame Gilda Malky. The overwhelming majority of these advertisements are for local New York City businesses and also feature prominent English text, both of which reflect *La Vara*'s role as an American Sephardic press within New York City.

**Figure 2b:** A magnified cluster within Figure 2a consisting of portrait shots of people.

Analysis of the surfaced advertisements that have been reproduced many times over reveals similar advertising patterns to those uncovered within *El Tiempo* by Sarah Stein: a preponderance of advertisements of sartorial nature, as well as for doctors, dentists, medical treatments for ailments, and legal counsel.[30] In the case of Constantinople-based *El Tiempo*, Stein argues that these advertisements speak to readers' anxieties under the precarity of Ottoman rule, including class and economic anxieties and aspirations. The apparent resonances between the advertisements in *El Tiempo* and those in the New York-based *La Vara* suggest an even broader pattern of Sephardic Jewish experiences in response to social and economic uncertainty and change during the late 19th and early 20th centuries, whether in the United States or the Ottoman Empire.

---

**30** Stein, *Making Jews Modern*, 185–87.
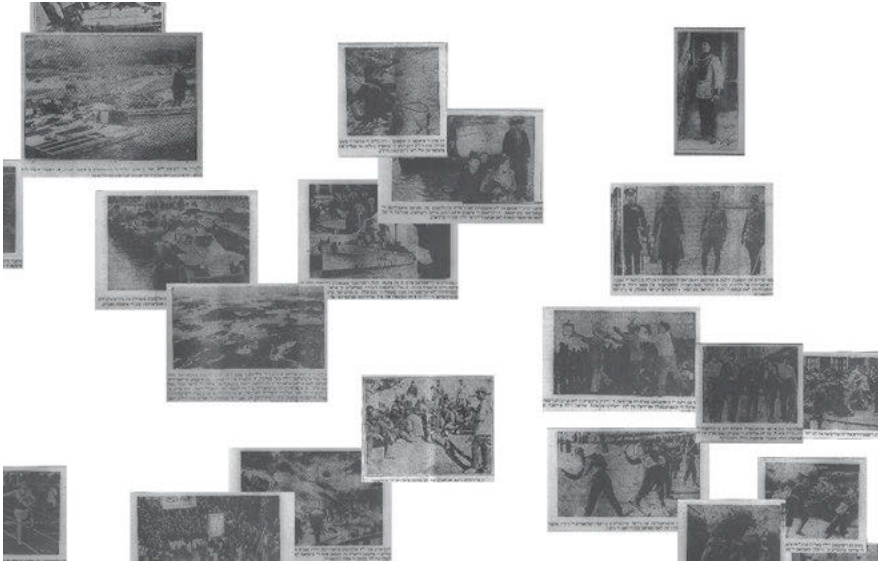
**Figure 2c:** A magnified cluster within Figure 2a consisting of wartime photographs.

Of particular interest are the recovered advertisements for Meyer London's Matzos (entry b in Figure 3b), which appeared concurrently in American Yiddish newspapers. As described by Makena Mezistrano, "Matsa advertisements in the American Yiddish and Ladino presses offer a rare opportunity to place these two communities in dialogue with one another, instead of only positioning them as separate or in bitter conflict – two common assumptions about intra-Jewish relationships in twentieth century New York."[31] Thus, the advertisements uncovered through this cluster-based analysis speak to not only individual Sephardic communities but also relationships between and across communities, as embedded within cultural practices.

However, not all clusters correspond to advertisements. In Figure 3c, magnified clusters of extracted photographs and full newspaper pages from *La Vara* are shown. These clusters are false positives, reflecting the imperfect performance of the visual content recognition model utilized for visual content extraction. These clusters are an important reminder that algorithmic approaches to extracting visual content are inevitably imperfect. However, using clustering and other machine learning techniques, it is possible to remove many of these false positives quickly.

---

**31** Mezistrano, "Why Are These Passover Ads Different from All Other Ads?"

**Figure 2d:** A magnified cluster within Figure 2a consisting of photographs of crowds and groups of people.

# 4 Future Work

Ongoing work consists of continuing to explore the extracted visual content via both macroscopic analysis and close analysis within the page-level context. In terms of macroscopic analysis, I plan to expand the study of the reprinting patterns of advertisements in order to examine the network of advertisers that funded the Ladino press. By building on the provocations offered in this chapter, one can ask questions such as: did advertisers purchase advertising space in different titles? And what does this tell us about the interconnectedness of the Ladino press? Moreover, I plan to expand this analysis to different temporal slices of *La Vara* along with quantitative assessments of different photograph types in order to understand the evolution of the visual content. I will also expand this analysis to include a greater exploration of the other Ladino titles present in the
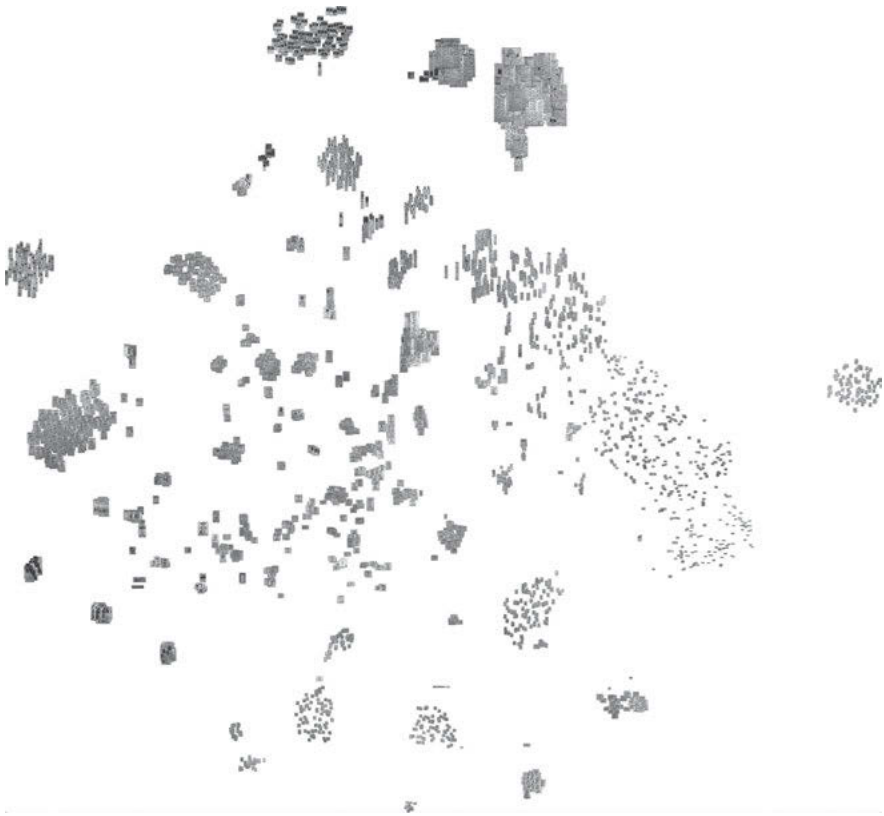
**Figure 3a:** A cluster visualization of 2,812 advertisements identified by the visual content recognition model with confidence scores greater than 90% within issues of *La Vara* published between January 3, 1936 and August 26, 1938. I constructed this visualization using ResNet-50 embeddings and T-SNE for dimensionality reduction.

dataset (as enumerated in Table 1). With this future analysis, I can begin to ask questions surrounding the intended audiences of the advertisements and how they changed over time as well as by title, building on Sarah Stein's analysis of the advertisements in *El Tiempo*.[32]

Because the analysis of the visual content in this chapter has focused on the extracted dataset, future work also entails understanding the visual content by recontextualizing it at the page level within the broader *mise en page*. What types of articles accompany visually similar photographs? What advertisements appear next to one another? What do the captions reveal about the visual content?

---

**32** Stein, *Making Jews Modern*, 153–201.

**Figure 3b:** Six different magnified clusters within Figure 3a showing advertisements reprinted throughout different issues of *La Vara* (top), along with magnified versions of the reproduced advertisements (bottom). The advertisements are for Brockman Monument Works (a), Meyer London's Matzos (b), Standard Truss Co. (c), and Aristocratic Imported Virgin Olive Oil (d, e, f).

From a computational perspective, future work with the dataset of extracted visual content entails evaluating the generalization of the visual content recognition model on the Ladino newspaper pages. This evaluation will require manually annotating enough pages across different titles and temporal slices in order to derive reliable statistics. With this in-depth evaluation across many different
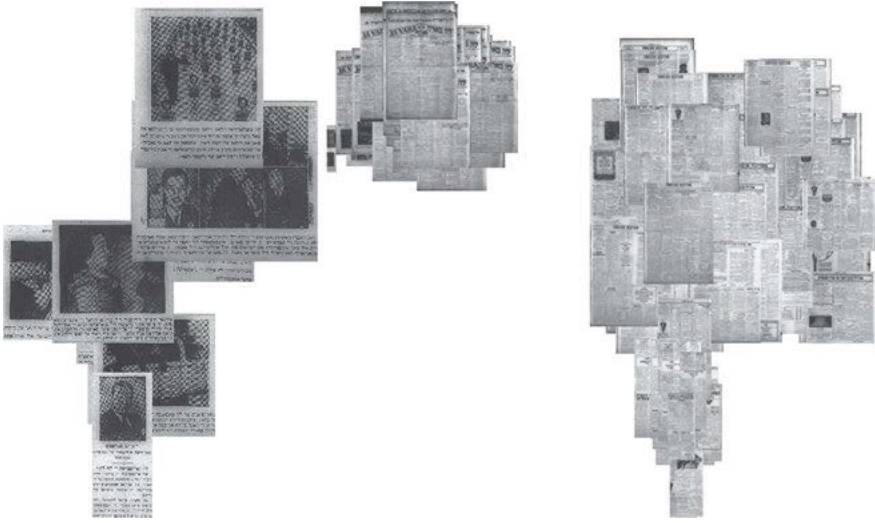
**Figure 3c:** Magnified clusters of photographs and full pages, showing false positives among the identified advertisements from *La Vara*.

newspapers at varying time periods, it is possible to better understand the bias of the visual content recognition model, which will, in turn, inform the results of this macroscopic analysis even further. Lastly, given that so many of the advertisements had captions written in English, future work entails running English OCR engines on the advertisements utilizing the results for textual analysis of the captions.

Other potential work includes cross-matching the visual content in the Ladino press with the extracted visual content from other newspaper corpora, such as the visual content from 16 million pages in *Chronicling America* contained within the *Newspaper Navigator* dataset. Identifying reproduction patterns among the visual content within these different presses could indeed speak to the proximity or marginal position of the Ladino press in relation to broader American newspaper syndicates. Moreover, Makena Mezistrano's discovery of Meyer London's Matzos advertisements in both the American Sephardic press and American Ashkenazic press suggests that this future direction of cross-matching visual content across different presses has the capacity to enrich our understanding of how cultural practices change across different communities.

Lastly, in regard to the dataset of extracted visual content from Ladino titles, I have two primary goals for future work. First, as articulated earlier in this chapter, I plan to make this dataset of extracted visual content publicly available to encourage re-use among scholars and the public. Second, I hope to expand

the Ladino titles that have been processed in order to further excavate the visual content across the Ladino press.

# 5  Ethical Considerations

Given the profound implications of machine learning perpetuating marginalization and erasure through algorithmic bias and other mechanisms, any application of machine learning to cultural heritage collections would be remiss without a discussion of the ethical considerations surrounding doing so. Within the library, archive, and museum ("LAM") community, there has been a growing effort to consider a critical, sociotechnical lens surrounding machine learning, data science, and cultural heritage. This effort has culminated in the development of responsible operations and best practices, as well as surveys of projects in this liminal space.[33] In the case of difficult and understudied histories, extra precaution must be taken, and a rich discourse in the scholarly community has explored the ethics of datafication and the application of machine learning methodologies within this context.[34] In this section, I draw from these emerging bodies of work and explicitly build on the *Newspaper Navigator* data archaeology, which I wrote in order to detail the implications of machine learning for search and discovery from a sociotechnical perspective.[35]

Though often overlooked, the marginal position of Sephardic Studies within Jewish Studies has been amplified by machine learning, having altered the discoverability of the Sephardic historical record via digitization. As detailed earlier in this chapter, off-the-shelf OCR algorithms consistently fail to transcribe Ladino

---

**33** Ryan Cordell, "Machine Learning + Libraries: A Report on the State of the Field," 2020, accessed July 22, 2020, https://labs.loc.gov/static/labs/work/reports/Cordell-LOC-ML-report.pdf; Thomas Padilla, "Responsible Operations: Data Science, Machine Learning, and AI in Libraries," OCLC, August 26, 2020, accessed December 15, 2020, https://www.oclc.org/research/publications/2019/oclcresearch-responsible-operations-data-science-machine-learning-ai.html; Eileen Jakeway et al., "Machine Learning + Libraries Summit Event Summary," 2020, accessed February 12, 2020, https://labs.loc.gov/static/labs/meta/ML-Event-Summary-Final-2020-02-13.pdf.
**34** Presner Todd. "The Ethics of the Algorithm: Close and Distant Listening to the Shoah Foundation Visual History Archive," in *Probing the Ethics of Holocaust Culture*, ed. Fogu Claudio, Wolf Kansteiner, and Todd Presner (Cambridge, MA: Harvard University Press, 2020), 175–202; Benjamin Charles Germain Lee, "Machine Learning, Template Matching, and the International Tracing Service Digital Archive: Automating the Retrieval of Death Certificate Reference Cards from 40 Million Document Scans," *Digital Scholarship in the Humanities* 34, no. 3 (September 1, 2019): 513–35, accessed December 15, 2020, https://doi.org/10.1093/llc/fqy063.
**35** Lee, "Compounded Mediation."

texts with a high enough degree of fidelity to facilitate reliable keyword search and textual analysis. This effective erasure of Ladino texts from search and discovery platforms is the result of a confluence of factors, from the availability of training data to the monetary value of preferentially selecting widely studied languages for inclusion in proprietary OCR engines. Significantly, this linguistic erasure is not limited to Ladino: a similar systemic problem has been documented for Yiddish and indigenous languages, which speaks to a specific form of *algorithmic bias*, in which human decisions surrounding which languages should be prioritized when training OCR algorithms have a profound impact on resulting scholarship.[36]

In this chapter, I seek not only to foreground the algorithmic marginalization of Sephardic history but also to offer an alternative approach to recover the voices that have been lost through digitization. Certainly, the utilization of machine learning to excavate visual content in Ladino newspapers is not without its own challenges. The *Newspaper Navigator* visual content recognition model performs better on pages that more closely resemble the training data, and the extracted dataset presented in this chapter suffers from a nontrivial number of false positives and false negatives as a result.[37] These false positives and false negatives motivate methodological improvements, such as training a visual content recognition model specifically for the Sephardic press in order to better capture the nuances of Sephardic visual culture. Moreover, image recognition algorithms used to evaluate image similarity have been shown to perpetuate their own forms of bias and marginalization.[38] Because these algorithms have been trained by machine learning practitioners with specific objectives and categories in mind, a fundamental question is raised as to whether the groupings identified by the algorithms capture the relationships most valuable to scholars. While these methods have the capacity to expose new groupings, they inevitably distort the viewer's perceptions of what constitutes similarity. I therefore offer these methodological approaches with such considerations in mind, a reminder of the importance of canonical historiographic approaches that can be used in concert with machine learning.

And yet, this chapter has provided the first macroscopic view of the Ladino press via the excavated visual content and thus serves as a corrective to the algo-

---

**36** Hannah Alpert-Abrams, "Machine Reading the Primeros Libros," *Digital Humanities Quarterly* 10, no. 4 (October 4, 2016), accessed December 15, 2021, http://www.digitalhumanities. org/dhq/vol/10/4/000268/000268.html; "Yiddish OCR Is Live! | Yiddish Book Center," accessed March 31, 2021, https://www.yiddishbookcenter.org/about/news/yiddish-ocr-live.

**37** Lee et al., "The Newspaper Navigator Dataset."

**38** Lee, "Compounded Mediation."

rithmic marginalization of Sephardic Studies. I therefore offer this work in pursuit of a digital humanities and a Jewish Studies that foreground Sephardic history and culture.

# References

Alpert-Abrams, Hannah. "Machine Reading the Primeros Libros." *Digital Humanities Quarterly* 10, no. 4 (October 4, 2016). Accessed December 15, 2020 , http://www.digitalhumanities. org/dhq/vol/10/4/000268/000268.html.

Arnold, Taylor, Lauren Tilton, and Annie Berke. "Visual Style in Two Network Era Sitcoms." *Journal of Cultural Analytics*, July 19, 2019, 11045. Accessed December 15, 2020. https://doi.org/10.22148/16.043.

Barnhurst, Kevin G., and John Nerone. *The Form of News: A History*. New York: *The Guilford Press*, 2002.

Chronicling America. *The Ogden Standard*, April 16, 1918. Library of Congress. Accessed December 10, 2020. https://chroniclingamerica.loc.gov/lccn/sn85058396/1918-04-16/ ed-1/seq-6/.Cole, Jean Lee. *How the Other Half Laughs: The Comic Sensibility in American Culture, 1895–1920*. Jackson: University Press of Mississippi, 2020.

Cordell, Ryan. "Machine Learning + Libraries: A Report on the State of the Field." 2020. Accessed July 22, 2020. https://labs.loc.gov/static/labs/work/reports/Cordell-LOC- ML-report.pdf.

Deng, J., W. Dong, R. Socher, L. Li, Kai Li, and Li Fei-Fei. "ImageNet: A Large-Scale Hierarchical Image Database." In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 248–55. 2009. Accessed December 10, 2020. https://doi.org/10.1109/ CVPR.2009.5206848.

DiJeSt Team. "Our Text Recognition Ground Truth and Model – DiJeSt." Accessed March 31, 2021. https://dijest.net/gtmodel/.

Fyfe, Paul, and Qian Ge. "Image Analytics and the Nineteenth-Century Illustrated Newspaper." *Journal of Cultural Analytics*, October 25, 2018, 11032. Accessed December 15, 2020. https://doi.org/10.22148/16.026.

Gomez, Joshua, Kristian Allen, Mark Matney, Tinuola Awopetu, and Sharon Shafer. "Experimenting with a Machine Generated Annotations Pipeline." *The Code4Lib Journal*, no. 48 (May 11, 2020). Accessed December 10, 2020. https://journal.code4lib.org/ articles/15209.

Griffin, Michael. "The Great War Photographs: Constructing Myths of History and Photojournalism." In *Picturing the Past: Media, History and Photography*, edited by Bonnie Brennan and Hanno Hardt, 122–57. Urbana: University of Illinois Press, 1999.

He, K., X. Zhang, S. Ren, and J. Sun. "Deep Residual Learning for Image Recognition." In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–78. 2016. Accessed December 10, 2020. https://doi.org/10.1109/CVPR.2016.90.

Jakeway, Eileen, Lauren Algee, Laurie Allen, Meghan Ferriter, Jaime Mears, Abigail Potter, and Kate Zwaard. "Machine Learning + Libraries Summit Event Summary." 2020. Accessed February 12, 2020. https://labs.loc.gov/static/labs/meta/ML-Event- Summary-Final-2020-02-13.pdf.

LC Labs. "Beyond Words." Accessed March 31, 2021. http://beyondwords.labs.loc.gov/#/.

Lee, Benjamin Charles Germain. "Compounded Mediation: A Data Archaeology of the Newspaper Navigator Dataset." *Digital Humanities Quarterly* 15, no. 4 (2021). Accessed January 10, 2022. http://digitalhumanities.org/dhq/vol/15/4/000578/000578.html.

Lee, Benjamin Charles Germain. *LibraryOfCongress/Newspaper-Navigator*. GitHub Repository. Library of Congress. 2020. Accessed December 10, 2020. https://github.com/LibraryOfCongress/newspaper-navigator.

Lee, Benjamin Charles Germain. "Machine Learning, Template Matching, and the International Tracing Service Digital Archive: Automating the Retrieval of Death Certificate Reference Cards from 40 Million Document Scans." *Digital Scholarship in the Humanities* 34, no. 3 (September 1, 2019): 513–35. Accessed December 15, 2020. https://doi.org/10.1093/llc/fqy063.

Lee, Benjamin Charles Germain, Jaime Mears, Eileen Jakeway, Meghan Ferriter, Chris Adams, Nathan Yarasavage, Deborah Thomas, Kate Zwaard, and Daniel S. Weld. "The Newspaper Navigator Dataset: Extracting Headlines and Visual Content from 16 Million Historic Newspaper Pages in Chronicling America." In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, 3055–62. CIKM '20. New York: Association for Computing Machinery, 2020. Accessed December 15, 2020. https://doi.org/10.1145/3340531.3412767.

Lee, Benjamin Charles Germain and Daniel S. Weld. "Newspaper Navigator: Open Faceted Search for 1.5 Million Images." In *Adjunct Publication of the 33rd Annual ACM Symposium on User Interface Software and Technology*, 120–22. UIST '20 Adjunct. New York: Association for Computing Machinery, 2020. Accessed December 15, 2020. https://doi.org/10.1145/3379350.3416143.

Lin, Tsung-Yi, Michael Maire, Serge Belongie, Lubomir Bourdev, Ross Girshick, James Hays, Pietro Perona, Deva Ramanan, C. Lawrence Zitnick, and Piotr Dollár. "Microsoft COCO: Common Objects in Context." *ArXiv:1405.0312 [Cs]*, February 20, 2015. Accessed December 10, 2020. http://arxiv.org/abs/1405.0312.

Lincoln, Matthew, Julia Corrin, Emily Davis, and Scott B. Weingart. "CAMPI: Computer-Aided Metadata Generation for Photo Archives Initiative." October 8, 2020. Accessed December 15, 2020. https://doi.org/10.1184/R1/12791807.v1.

Lorang, Elizabeth, Leen-Kiat Soh, Yi Liu, and Chulwoo Pack. "Digital Libraries, Intelligent Data Analytics, and Augmented Description: A Demonstration Project." 2020. Accessed December 15, 2020. https://labs.loc.gov/static/labs/work/experiments/final-report-revised_june-2020.pdf.

Maaten, Laurens van der and Geoffrey Hinton. "Visualizing Data Using T-SNE." *Journal of Machine Learning Research* 9, no. 86 (2008): 2579–605.

Manovich, Lev. "Data Science and Digital Art History." *International Journal for Digital Art History*, no. 1 (June 26, 2015). Accessed December 15, 2020. https://doi.org/10.11588/dah.2015.1.21631.

Manovich, Lev. "How to Compare One Million Images?" In *Understanding Digital Humanities*, edited by David M. Berry, 249–78. London: Palgrave Macmillan, 2012. Accessed December 15, 2020. https://doi.org/10.1057/9780230371934_14.

Mezistrano, Makena. "Why Are These Passover Ads Different from All Other Ads?" Stroum Center for Jewish Studies. 2021. Accessed May 19, 2021. https://jewishstudies.washington.edu/sephardic-studies/why-are-these-passover-ads-different-from-all-other-ads/.

Novaes, André Reyes. *Maps in Newspapers: Approaches of Study and Practices in Portraying War since 19th Century*. Leiden: Brill, 2019. Accessed June 10, 2021, https://brill.com/view/title/54806.

Padilla, Thomas. "Responsible Operations: Data Science, Machine Learning, and AI in Libraries." OCLC, August 26, 2020. Accessed December 15, 2021. https://www.oclc.org/research/publications/2019/oclcresearch-responsible-operations-data-science-machine-learning-ai.html.

Piper, Andrew, Chad Wellmon, and Mohamed Cheriet. "The Page Image: Towards a Visual History of Digital Documents." *Book History* 23, no. 1 (2020): 365–97. Accessed December 15, 2021. https://doi.org/10.1353/bh.2020.0010.

Presner, Todd. "The Ethics of the Algorithm: Close and Distant Listening to the Shoah Foundation Visual History Archive." In *Probing the Ethics of Holocaust Culture*, edited by Fogu Claudio, Wolf Kansteiner, and Todd Presner, 175–202. Cambridge, MA: Harvard University Press, 2020.

Safka, Christian. *Christiansafka/Img2vec*. Python, 2021. Accessed December 10, 2020. https://github.com/christiansafka/img2vec.

Sarın, Pınar and Necla Uluğtekin. "Analyzing Newspaper Maps for Earthquake News through Cartographic Approach." *ISPRS International Journal of Geo-Information* 8, no. 5 (May 2019): 235. Accessed June 10, 2021. https://doi.org/10.3390/ijgi8050235.

Shen, Zejiang, Ruochen Zhang, Melissa Dell, Benjamin Charles Germain Lee, Jacob Carlson, and Weining Li. "LayoutParser: A Unified Toolkit for Deep Learning Based Document Image Analysis." *ArXiv:2103.15348 [Cs]*, March 29, 2021. Accessed June 10, 2021. http://arxiv.org/abs/2103.15348.

Stein, Sarah Abrevaya. *Making Jews Modern: The Yiddish and Ladino Press in the Russian and Ottoman Empires*. Bloomington: Indiana University Press, 2004.

Wattenberg, Martin, Fernanda Viégas, and Ian Johnson. "How to Use T-SNE Effectively." *Distill* 1, no. 10 (October 13, 2016): e2. Accessed December 10, 2020. https://doi.org/10.23915/distill.00002.

Wevers, Melvin, and Thomas Smits. "The Visual Digital Turn: Using Neural Networks to Study Historical Images." *Digital Scholarship in the Humanities* 35, no. 1 (April 1, 2020): 194–207. Accessed December 15, 2020. https://doi.org/10.1093/llc/fqy085.

Williams, Andrea N. "Cultivating Black Visuality: The Controversy over Cartoons in the Indianapolis *Freeman*." *American Periodicals* 25, no. 2 (2015): 124–38. Accessed June 10, 2021. http://www.jstor.org/stable/24589083.

Winfield, Betty H. and Doyle Yoon. "Historical Images at a Glance: North Korea in American Editorial Cartoons." *Newspaper Research Journal* 23, no. 4 (September 1, 2002): 97–100. Accessed June 10, 2021. https://doi.org/10.1177/073953290202300411.

"Yiddish OCR Is Live! | Yiddish Book Center." Accessed March 31, 2021. https://www.yiddishbookcenter.org/about/news/yiddish-ocr-live.

Abby Gondek

# Using Nodegoat to Track Gendered Political Networks: Henrietta Klotz's Influence on Henry Morgenthau Jr.'s Advocacy for Jewish Refugees and the State of Israel

**Abstract:** In histories of how the U.S. government responded to the Holocaust, women's roles are often invisible or under-explored. This is because political influence and power are typically defined as being the domain of men, and because there were very few women in roles that were (or are) deemed as "powerful." Women were more likely to be in secretarial roles; secretaries are not usually perceived as being influential.[1] However, secretaries could exert political influence.[2] This chapter presents a case study of the impact that Henry Morgenthau Jr.'s assistant for 37 years, Henrietta Stein Klotz, had on the Secretary of the Treasury's positions in response to the Holocaust and his post-war fundraising for the State of Israel.

A web-based, data management, network analysis, and visualization environment, nodegoat, can be used to track the intersection of what is typically gendered female – the micro, private, social, and cultural – with what is typically gendered

---

**1** Rachel Century, "Dictating the Holocaust: Female Administrators of the Third Reich" (Royal Holloway, University of London, 2012), 182, 185, 190, 259; Leisa D. Meyer, *Creating GI Jane: Sexuality and Power in the Women's Army Corps during World War II* (New York: Columbia University Press, 1996), 71–72, 76, 80–82. Century contends that the female SS auxiliaries ("helferinnen") were subordinate to their male bosses; they were "helpers, working *for*, not with the men" (259, emphasis in original). Though they may have been close or intimate (sexual) with the men who employed them, these secretaries did not have access to their bosses' confidential papers, with rare exceptions (182, 185, 190). Meyer articulates how women's World War II employment was temporary and they were unable to change their second-class status (71–72). Seventy percent of women in the Women's Army Corp worked in fields traditionally defined as "women's work," such as communication or clerical employment (72, 76). Male officers expected women to serve them through this gendered work, as nannies, maids, waitresses, and cooks (80–82).

**2** Elisabeth Krimmer, *German Women's Life Writing and the Holocaust: Complicity and Gender in the Second World War* (Cambridge: Cambridge University Press, 2018), 10, 34, 36, 38; Judith Tydor Baumel, "Women's Agency and Survival Strategies during the Holocaust," *Women's Studies International Forum* 22, no. 3 (1999): 334–35. Krimmer argues that female secretaries within the Nazi government were able to wield more power than the term "auxiliary" typically implies. Baumel maintains that female secretaries were able to influence the male leadership of the Jewish councils, the ghetto police, and underground networks.

masculine – the macro, public, institutional, and structural. Nodegoat enables the creation of multi-layered networks in which various types of nodes ("objects") and relationships can be displayed simultaneously, demonstrating how interpersonal networks (typically gendered female) are simultaneously structural and integrate broader social phenomena such as ethnicity, religion, and politics (typically gendered male). Thus, nodegoat can demonstrate how women in secretarial roles, like Henrietta Klotz, could exert political influence through their ethnic, religious, and interpersonal networks.

**Keywords:** Henrietta Klotz, Henry Morgenthau Jr., historical social network analysis, Holocaust, Israel, Jewish women, nodegoat, politics, secretaries, U.S. government

# 1  Introduction

Henry Morgenthau Jr. (HMJ) was an assimilated German Jew whose father (Henry Morgenthau Sr.) preferred his son steer clear of "Jewish affairs."[3] Morgenthau Sr. (HM Sr.) was the U.S. Ambassador to the Ottoman Empire (Turkey) during the Armenian genocide and advocated for Armenians who were victims of "race extermination."[4] HM Sr. directly linked the plight of Armenians to that of the Jews, writing to his wife Josephine in 1913 that, like Jews, Armenians demonstrated a "stubborn adherence to their religion and a very strong race pride."[5] HMJ served as the Secretary of the Treasury in the Roosevelt and Truman administrations from January 1, 1934 to July 22, 1945.[6] In 1943, Morgenthau Jr. reluctantly became involved in the conflict between Treasury and State regarding what to do with Jewish refugees.[7] According to John Pehle, who was the director of the War Refugee Board, Morgenthau Jr. "didn't want to stand out as a Jew." Henry

---

**3** Michael Beschloss, *The Conquerors: Roosevelt, Truman, and the Destruction of Hitler's Germany, 1941–1945* (New York: Simon & Schuster, 2002), 45–46.
**4** Rouben Paul Adalian, "Morgenthau, Ambassador Henry, Sr.," Armenian National Institute, 2019, accessed January 10, 2022, https://www.armenian-genocide.org/morgenthau.html.
**5** Henry Morgenthau III, *Mostly Morgenthaus: A Family History (e-Book)*, 2nd ed. (Lexington, MA: Plunkett Lake Press, 2019), 109, accessed January 10, 2022, http://plunkettlakepress.com/mmo.html.
**6** "Henry Morgenthau Jr. Papers, 1866–1953, Finding Aid," Franklin D. Roosevelt Presidential Library & Museum, n.d., accessed January 10, 2022, http://www.fdrlibrary.marist.edu/archives/collections/franklin/index.php?p=collections/findingaid&id=159&q=&rootcontentid=72431.
**7**  Henry Morgenthau Jr. et al., "Jewish Evacuation Meeting Transcript, December 18, 1943," *Henry Morgenthau Jr. Diaries, FDR Presidential Library and Museum* 688II (1943): 85–87; Henry Morgenthau III, *Mostly Morgenthaus: A Family History* (New York: Ticknor & Fields, 1991), 323–24.

Morgenthau III (HMJ's son) explained that the Jewish advisors closest to Roosevelt (including his father) "avoided or downplayed the significance of Jewish questions."[8] After his forced resignation in 1945,[9] Morgenthau Jr. began to participate more explicitly in Jewish causes; he was the chairman of the United Jewish Appeal[10] and financially advised the State of Israel.[11] I use nodegoat to explore an understudied reason for why Henry Morgenthau Jr. took the stances he did regarding Jewish refugees at the end of World War II, and afterward.

The War Refugee Board developed through HMJ's Treasury department, and he was one of the core members of the Board (Figure 1). Morgenthau's male staff members (especially Randolph Paul, John Pehle, and Josiah DuBois) are often identified as being highly influential on his positions regarding rescue and relief during the latter part of World War II.[12] Henrietta Klotz, his assistant, has been under-examined as a central influencer for the political decisions HMJ made in

---

**8** Morgenthau III, *Mostly Morgenthaus*, 322–23.

**9** According to Henry Morgenthau III, when Roosevelt died suddenly and Truman became President, Truman did not approve of Morgenthau Jr.'s harsh stance toward post-war Germany. Truman felt HMJ tried to reach beyond his role as Secretary of the Treasury. The new president sided with his ally Henry Stimson (Secretary of War) against HMJ. HM III emphasized the antisemitism he detected in the behavior of these two men. Stimson referred to HMJ's "Jewish vengeance." HM III asserted that Truman feared that because of the order of succession, if he and Secretary of State Byrnes died, HMJ would become the new president. HM III underscored, "The possibility of having a Jewish president seems to have been one of the concerns." Morgenthau III also cited growing anti-Soviet sentiment and the association of Jews with communism as a factor in Truman's decision to replace HMJ with Fred Vinson as Secretary of the Treasury. Vinson and Byrnes, both southerners, "were often seen as a pair." See Morgenthau III, *Mostly Morgenthaus (e-Book)*, 329–30.

**10** Morgenthau III, *Mostly Morgenthaus (e-Book)*, 334. The United Jewish Appeal was a philanthropic umbrella organization established in 1939, which eventually became what is now the Jewish Federation of North America. UJA was made up of three agencies: the Joint Distribution Committee (JDC), the United Palestine Appeal (UPA), later the United Israel Appeal, and the National Coordinating Committing for Aid to Refugees (NCCR), later the National Refugee Service. The UPA represented Zionist interests.

**11** NYT Archive, "H.S. Klotz, 87, Aide to Treasury Secretary," *New York Times Online*, December 21, 1988, accessed January 10, 2022, https://www.nytimes.com/1988/12/21/obituaries/h-s-klotz-87-aide-to-treasury-secretary.html; Rebecca Erbelding, "Morgenthau Family Papers, 1860–2015, 2015.255.1 Finding Aid" (Washington DC, 2015), 2, accessed January 10, 2022, https://collections.ushmm.org/search/catalog/irn96059.

**12** Rebecca Erbelding, *Rescue Board: The Untold Story of America's Efforts to Save the Jews of Europe* (New York: Doubleday, 2018), 53; Randolph Paul, "Report to the Secretary of the Acquiescence of this Government in the Murder of the Jews, given to Henry Morgenthau Jr. by His Staff (Josiah DuBois, John Pehle, and Randolph Paul) on January 13, 1944," *Morgenthau Diaries* 693 (1944): 212–29; Morgenthau III, *Mostly Morgenthaus*, 323.

**Figure 1:** The third meeting of the War Refugee Board, March 21, 1944, in Cordell Hull's office. Pictured left to right are: Cordell Hull, Secretary of State, Henry Morgenthau Jr., Secretary of the Treasury, Henry Stimson, Secretary of War, and John Pehle, Director of the WRB. FDR Library General Photograph Collection, Folder: Conferences, Commissions and Committees, World War II, WRB, photo id: NPx82-61.

relation to the Holocaust, especially after the war, when he became increasingly involved in fundraising for the State of Israel. Henry Morgenthau III (HM III) articulated Henrietta's importance in this way: "While the Treasury lawyers [Paul, Pehle, and DuBois] were the activists in shaping and moving my father's rescue campaign, Henrietta Klotz was the catalyst."[13] However, HM III's differentiation between "activists" and "catalyst" is problematic because it implies that Henrietta only initiated HMJ's shift, while his male advisors (Paul, Pehle, and DuBois) were the real change agents regarding ongoing rescue policy. In contrast to this distinction, I maintain that Henrietta (and other women in similar "auxiliary" positions) should be recognized as executing power in political decision making during the Holocaust and immediate post-war period. "Power" can be conceptualized through Henrietta's words in her interview with HM III, "But you see, I brought things to his attention."

This chapter therefore asks the question of how, considering HMJ's assimilated Jewish upbringing and his father's non-Zionist stance, an analysis of Hen-

---

**13** Morgenthau III, *Mostly Morgenthaus*, 323.

rietta's influence can enable a clearer explanation for HMJ's eventual stances regarding relief and rescue efforts and his advocacy for Israel? It argues that digital tools, in this case the network analysis and visualization environment, nodegoat, offer new methodological possibilities to explore the interpersonal but also ethnic, religious, and political connections between Henry Morgenthau Jr. and the people closest to him, including Henrietta Klotz. This case study thus offers a topical as well as methodological contribution to Jewish Studies and the digital humanities; it demonstrates the importance of analyzing women's influence on U.S. governmental stances in response to the Holocaust, and it discusses how digital tools such as nodegoat can be used to achieve this, by visually tracking gendered political networks of impact.

This article begins with an overview of Henrietta Klotz's religious and political background (Orthodox Judaism, exposure to Zionism) and her impact on HMJ, especially in the post-war period. She furthered a cause she believed in by facilitating meetings that enabled him to become a leader in the United Jewish Appeal and the Israel Bonds program. Section 3 critiques the typical dichotomy between ego-centric (qualitative, individual) and whole network (quantitative, social structure) approaches to historical social network analysis. I utilize a case study of the networks between Mrs. Klotz and HMJ to articulate how ego-centric approaches (focusing on the lives of individuals) can be simultaneously structural and institutional, demonstrating the intersection of micro and macro levels of influence. I also clarify the benefits of using nodegoat – to visualize multiple types of nodes or "objects" at once. In Section 4 (made up of three smaller sub-sections), I explore the deeper layers of influence on HMJ and Henrietta, expanding on the simple network visualization in the second section. I ask who and what influenced the people who shaped HMJ and Henrietta, demonstrating interpersonal but also ethnic, religious, and political networks of connection. I elaborate on the relationships between Henrietta, Henry Morgenthau Sr., and Elinor Fatman Morgenthau (HMJ's first wife) in order to investigate how Henrietta managed to become more influential than both of them. Both Morgenthau Sr. and Elinor emphasized assimilation rather than affiliation with the organized Jewish community. Though HM Sr.'s belief in collaboration between minority ethnic groups inspired HMJ's dedication to saving Jewish refugees during and after World War II, Henrietta was the figure who proved to be loyal, protective, and encouraging, qualities that HMJ longed for, according to his son Morgenthau III.[14] Elinor's role was significant in the realm of democratic politics, but not in

---

**14** Morgenthau III, *Mostly Morgenthaus (e-Book)*, 342. HM III recounted that after "the three people whose counsel and affection my father most relied on had left him" (FDR in 1945, HM Sr. in

Jewish affairs. Towards the end of her life, she became slightly more involved in Jewish refugee relief efforts, but sadly, she died in 1949 before the results of this involvement could be determined.

# 2 Henrietta Klotz: "Letting Israel [and Henry Morgenthau Jr.] Grow"

In contrast to HMJ's secular and elite home environment, Henrietta Stein Klotz (Figure 2) was raised in a poor Orthodox Jewish household. Morgenthau's son, Henry Morgenthau III, argued that Henrietta was "the key to getting him involved in Jewish things."[15] Henrietta became more instrumental than HMJ's wife, Elinor Fatman.[16] Mrs. Klotz was Morgenthau's secretary for 37 years,[17] longer than his marriage to Elinor (33 years) or to his second wife, Marcelle (16 years).[18] Though Henrietta did not articulate why she was so invested in convincing Morgenthau to do more to rescue Jewish refugees, her husband, Herman, explained to HM III that the reason was in their shared familial contexts: "It is obvious … that between Henrietta's background and my own that our interests permeated Henrietta's discussions with your father [HMJ] when matters of Jewish interests were involved."[19]

According to Herman Klotz, Henrietta nudged and insisted that Morgenthau "get the President to take some action that would minimize the killing of the Jews." Herman recounted that every morning when HMJ saw Mrs. Klotz in the office, he would look at her timidly, her countenance insisting "when [will you

1946, and Elinor in 1949), "the course of his life changed as he became the devoted servant of the Jewish community. Was it partly an act of self-definition, a rebellion against his upbringing, a craving for love, a respect that he couldn't find anywhere else…? Perhaps the answer is yes to each."

**15** Herman Klotz, "Herman Klotz to Henry Morgenthau III, October 16, 1985" (Washington DC: USHMM 2015.255.1 Morgenthau Family Papers, Box 32, File 11, 1985), 7; Beschloss, *The Conquerors*, 53.

**16** Beschloss, *The Conquerors*, 55.

**17** Archive, "H.S. Klotz, 87, Aide to Treasury Secretary," 19.

**18** Edna S. Friedberg, "Elinor Morgenthau (1891–1949)," in *Jewish Women: A Comprehensive Historical Encyclopedia* (Jewish Women's Archive, 1999), accessed January 10, 2022 https://jwa.org/encyclopedia/article/morgenthau-elinor; "Mrs. Henry Morgenthau, 73 Widow of Roosevelt's Aide," *New York Times Archives*, July 19, 1972, accessed January 10, 2022, https://www.nytimes.com/1972/07/19/archives/mrs-henry-morgenthau-73-widow-of-rooseve-ts-aide.html.

**19** Klotz, "Herman Klotz to Henry Morgenthau III, October 16, 1985," 7–8.

**Figure 2:** Henrietta Klotz (uncredited) is pictured on the right of this photograph, from the Harris and Ewing collection at the Library of Congress. "Firm financeer [sic]. Sec. Henry Morgenthau, Jr., photographed at a press conference where he answered questions concerning the pending financing ... ." December 3, 1935, Library of Congress, https://www.loc.gov/resource/hec.39676/.

do something]?" If Morgenthau assumed he had done "as much as he could and absolved himself from further responsibility," Henrietta was unrelenting in "her efforts without giving up an inch." Herman attributed the creation of the War Refugee Board to his wife's persistence.[20]

In an August 1945 handwritten letter to Henrietta, HMJ proclaimed that she was the "watchdog of the Secretary of the Treasury" and asserted: "whatever credit I deserve" for rescuing Jewish refugees, "I want to share it equally with you." He underscored her persuasiveness in "Jewish affairs" where she was "particularly understanding and helpful"; in fact, he believed she "made a real contribution towards winning the war."[21]

---

**20** Klotz, "Herman Klotz to Henry Morgenthau III, October 16, 1985," 5–6. Herman does not specify the time period, but it can be assumed that he was referring to the Holocaust period because he refers to "the killing of the Jews."

**21** Henry Morgenthau Jr., "Handwritten Letter from Henry Morgenthau Jr. to Mrs. Henrietta Klotz, August 5, 1945" (Washington DC: USHMM 2015.255.1 Morgenthau Family Papers, Box 32, File 11, 1945), 5, 8–10. The quotes in this paragraph all originate in this letter.

Soon after FDR's sudden death in April 1945 (in the same month that his wife, Elinor Morgenthau, had suffered a heart attack), Morgenthau was forced to leave the Treasury department.[22] HMJ expressed to Henrietta, that if he had perceived that his time at Treasury was up:

> we both of us might have had time to plan our future together. It would have made me very happy if we could have continued working together. But to my everlasting sorrow it did not work out that way. But I will continue to hope and plan that in the not too distant future we will be working side by side once again.[23]

HMJ ensured that this possibility became a reality. He returned to New York, while Henrietta remained at Treasury in Washington, DC. HMJ called her daily and "begged her" to move back to New York to work with him. Henrietta agreed to move to New York but only if Morgenthau would work in "public service." She set up meetings with Henry Montor, Meyer Weisgal, William Rosenwald, Edward Warburg, and Rose Halpern. Through these meetings, Morgenthau became the chair of United Jewish Appeal (UJA) in 1946.[24] Montor was in charge of public relations for the United Palestine Appeal and then became the executive vice president at UJA and HMJ's "mentor in Jewish affairs."[25] They worked at UJA until 1948 and then began the Israel Bond program in 1949; Morgenthau Jr. resigned in 1954, after becoming increasingly less devoted to the program after his marriage to Marcelle Puthon Hirsch in 1951 (a Catholic, French divorced woman, who was uninterested in Israel or Jewish affairs).[26]

Henrietta's obituary made it appear that Morgenthau earned his position at UJA on his own and subsequently asked Henrietta to be his assistant.[27] However, other sources place Henrietta as the architect behind this opportunity, through

---

**22** Morgenthau III, *Mostly Morgenthaus (e-Book)*, 327–30. For a detailed explanation of the reasons for Morgenthau Jr.'s forced resignation see footnote 9.

**23** Morgenthau Jr., "Handwritten Letter from Henry Morgenthau Jr. to Mrs. Henrietta Klotz, August 5, 1945," 1–2.

**24** Unknown, "Undated Memo to Explain Henry Morgenthau Jr.'s August 5, 1945 Letter to Henrietta Klotz" (Washington DC: USHMM Accession 2015.255.1 Morgenthau Family Papers, Box 32, File 11, "Henrietta Klotz," 184–85 in file, n.d.).

**25** Morgenthau III, *Mostly Morgenthaus (e-Book)*, 334.

**26** Morgenthau III, *Mostly Morgenthaus (e-Book)*, 342, 344; Henrietta Klotz and Henry Morgenthau III, "Henry Morgenthau III, Interview with Henrietta Klotz, September 19, 1978" (Washington DC: USHMM 2015.255.1 Morgenthau Family Papers, Box 32, File 11, "Henrietta Klotz," 1978), 58, 76. The Israel Bonds program was a fundraising tool initiated in 1951 to leverage the resources of Diasporic Jewish communities, especially in the U.S. and Canada, to assist the Israeli economy following the 1948 war and the influx of Holocaust survivors.

**27** Archive, "H.S. Klotz, 87, Aide to Treasury Secretary."

her "links to the organized Jewish community."[28] Klotz "cultivated her boss's [HMJ] budding association with [Chaim] Weizmann and [Meyer] Weisgal during the war years."[29] The connection with Weizmann is especially notable considering the antipathy between Morgenthau Sr. and Weizmann (see the section on Morgenthau Sr. below). HM III underscores that his father's aims "were primarily humanitarian" since in the beginning of HMJ's association with Montor and Weisgal he "was not yet a committed Zionist."[30]

Henrietta explained that after HMJ was forced to leave the Treasury department, "Nobody wanted him … They wouldn't accept him … he was a lost sheep … it broke my heart."[31] HM III depicted the situation this way: "After the humiliating denouement of his career in government … he needed the respect and love that the Jewish community was ready to offer him."[32] When Meyer Weisgal, who was at the time the fundraiser for the Weizmann Institute, made an appointment to meet with HMJ, Morgenthau Jr. told Henrietta "that vulgarian is … arranging a meeting of the Jews and they insist that I come to this meeting."[33] Weisgal was working on relief efforts for Jewish displaced persons, a cause that Morgenthau felt deeply invested in.[34] Mrs. Klotz noted that HMJ thought Weisgal was "very, very vulgar." So, she set up a preliminary meeting with Weisgal to prepare him for the conference with her boss, so that HMJ would be favorably disposed to what Weisgal proposed. Mrs. Klotz instructed Weisgal to "speak quietly and be very gentle and use the correct word and not the vulgar word."[35] Then, after HMJ met with Weisgal, Morgenthau Jr. exclaimed to Henrietta that Weisgal was "erudite" and "gentile." Mrs. Klotz believed that the United Jewish Appeal and Israel Bonds

---

**28** Morgenthau III, *Mostly Morgenthaus (e-Book)*, 331.

**29** Monty Noam Penkower, "The Earl Harrison Report: Its Genesis and Its Significance," *The American Jewish Archives Journal* 68, no. 1 (2016): 11–12. Chaim Weizmann wrote to Henrietta in January 1944 to pass along a letter he had sent to Sam Rosenman that he hoped Henrietta would show to HMJ. The letter begins "My dear Henrietta" and ends "yours affectionately." The letter was printed on Jewish Agency for Palestine letterhead. Chaim Weizmann, "Letter from Chaim Weizmann to Henrietta Klotz, January 3, 1944," *Morgenthau Diaries* 689 (1944): 201.

**30** Morgenthau III, *Mostly Morgenthaus (e-Book)*, 334.

**31** Klotz and Morgenthau III, "Henry Morgenthau III, Interview with Henrietta Klotz, September 19, 1978," 53.

**32** Morgenthau III, *Mostly Morgenthaus (e-Book)*, 335.

**33** Klotz and Morgenthau III, "Henry Morgenthau III, Interview with Henrietta Klotz, September 19, 1978," 54–55.

**34** Morgenthau III, *Mostly Morgenthaus (e-Book)*, 332.

**35** Klotz and Morgenthau III, "Henry Morgenthau III, Interview with Henrietta Klotz, September 19, 1978," 54–55.

needed someone like HMJ "to awaken the Jews to help Israel."[36] She created the optimal conditions so that HMJ could improve his social and professional network but also so that he would further a cause that she believed in.

Henrietta and Morgenthau traveled together nationally and internationally (especially to Israel) on behalf of United Jewish Appeal and eventually launched the Israel Bonds program (1949–1954).[37] In Henrietta's words, Morgenthau Jr. "became kind of a little Jewish after a while."[38] She described how HMJ changed; at first he was not comfortable with speech-making or with shaking so many hands, but once he started seeing how much money he could raise by physically engaging with audience members, "he just curdled. He changed tremendously."[39] In reference to the urgency of HMJ's fundraising for the State of Israel, Henrietta explained: "They had to build up Israel. It was a baby. They had to let it grow."[40] Henrietta's descriptions could be applied to her impact on HMJ; she was a central figure in the growth of his Jewish identity. She built up his consciousness of the need to help the State of Israel. This is particularly striking considering his father's positions regarding the idea of an independent Jewish nation.

Rabbi Herbert Friedman, a chaplain and Haganah agent in Europe, who knew HMJ from their collaboration on the 1947 UJA fundraising campaign, told HM III that Morgenthau Jr. was "proud to be doing in his generation what his father hand done in his [Morgenthau Sr.'s] generation." Friedman believed that HMJ understood (from living through the Holocaust period) that "most of the world didn't give a damn" about Jews. Morgenthau Jr. "became more and more deeply involved … he became more convinced that Palestine was the only solution."[41] Before HMJ joined the UJA, the organization raised $35 million, but because of him, the organization raised $102 million in 1946, $124 million in 1947, and $148 million in 1948.[42]

---

**36** Klotz and Morgenthau III, "Henry Morgenthau III, Interview with Henrietta Klotz, September 19, 1978," 55–56.

**37** Archive, "H.S. Klotz, 87, Aide to Treasury Secretary"; Unknown, "Undated Memo to Explain Henry Morgenthau Jr.'s August 5, 1945 Letter to Henrietta Klotz," 1.

**38** Klotz and Morgenthau III, "Henry Morgenthau III, Interview with Henrietta Klotz, September 19, 1978," 56.

**39** Klotz and Morgenthau III, "Henry Morgenthau III, Interview with Henrietta Klotz, September 19, 1978," 57–58.

**40** Klotz and Morgenthau III, "Henry Morgenthau III, Interview with Henrietta Klotz, September 19, 1978," 60.

**41** Morgenthau III, *Mostly Morgenthaus (e-Book)*, 337–38. Haganah was the underground military organization in Palestine from 1920 to 1948.

**42** Morgenthau III, *Mostly Morgenthaus (e-Book)*, 336, 338.

During HMJ's first visit to Israel in October 1948, he cried while carrying Torah scrolls and dancing in the street with David Ben Gurion (Prime Minister of Israel) to celebrate Simchat Torah. The government named a settlement after him (Tal Shachar); at the dedication ceremony (presided over by Chaim Weizmann), HMJ called the event "one of the greatest moments of my life" and told the soldiers gathered, "You are showing the world that the Jew is a fighting man … you have raised the standard of the Jew in the eyes of the Christian world."[43]
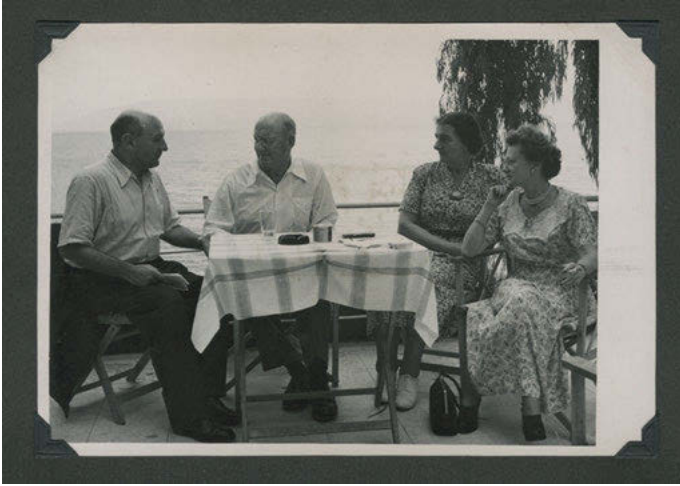


**Figure 3:** Henry Morgenthau Jr. and his secretary Mrs. Klotz meet with Israeli officials Golda Meir (Myerson) and Eliezer Kaplan. "Lake Tiberias – An important business conference with Finance Minister Kaplan, Labor Minister Myerson, and Mrs. Klotz." Photo credit: United States Holocaust Memorial Museum, courtesy of Henry Morgenthau III, October 5, 1950, photo id: 35016.

According to Henry Morgenthau III, this meeting (Figure 3) occurred during Morgenthau Jr.'s second visit to Israel in 1950. Elinor Morgenthau had died of a stroke in September 1949, only a few months before this trip. This was the meeting where the plans for the Israel Bonds drive were elaborated. HM III emphasized that Henrietta "was constantly at [my father's] side."[44]

---

**43** Morgenthau III, *Mostly Morgenthaus (e-Book)*, 339–40. Simchat Torah is a Jewish holiday marking both the end and beginning of the annual cycle of public Torah readings. Torah scrolls are taken out of the ark and celebrants dance and sing. Tal Shachar means "Valley of the Dew" and Morgenthau means "morning dew."

**44** Morgenthau III, *Mostly Morgenthaus (e-Book)*, 342.

# 3 Visualizing "Ego-Centric" Historical Social Networks in Nodegoat

My inquiry into Henrietta Klotz's impact on Henry Morgenthau Jr. is an example of an "ego-centric" approach to social network analysis. According to Franzosi and Mohr, historical social network analysis (HSNA) is necessary for historians to move beyond traditional, linear, biographical narratives which predominate when using archival materials.[45] Wetherell argues that historians have avoided social network analysis (SNA) because they are unfamiliar with social science quantitative methodologies.[46] Franzosi and Mohr contend that HSNA produces a complex non-linear representation of "broader social institutions" rather than the mindset of individuals.[47] These conceptions of HSNA maintain a dichotomy and hierarchy between the study of the mindset of individuals using traditional qualitative historical methodologies and the purportedly superior quantitative tools of SNA that provide access to societal structures. The "fact" of the *quantitative* nature of social network analysis is taken for granted. HSNA has tended towards "whole network" approaches that are more comprehensive and structural (read: quantitative) than "ego-centric network" approaches.[48] Ego-centric network analysis purportedly requires diaries and personal correspondence and has not yet been used in a systematic way to track "affection and social support."[49] "Structural" or "whole network" approaches are perceived to be quantitative (and are gendered male), while the study of individuals is perceived to be qualitative and "ego-centric" (and tends to be gendered female).

Feminist and postcolonial sociologists, anthropologists, and geographers have offered critiques of the ways that the "social" is often split from the "structural"; gender, race, and sexuality are often viewed as part of the "social," so the "structure" remains unimpacted by feminist, queer, and anti-racist critiques.[50] Gender is perceived as a variable (or an effect) rather than a complex analytical category;

---

**45** Roberto Franzosi and John W. Mohr, "New Directions in Formalization and Historical Analysis," *Theory and Society* 26, no. 2/3 (1997): 143, 145.
**46** Charles Wetherell, "Historical Social Network Analysis," *International Review of Social History* 43, supplement (1998): 125, accessed January 10, 2022, https://doi.org/10.1017/S0020859000115123.
**47** Franzosi and Mohr, "New Directions in Formalization and Historical Analysis," 143–48.
**48** Bonnie H. Erickson, "Social Networks and History: A Review Essay," *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 30, no. 3 (1997): 150–51, accessed January 10, 2022, https://doi.org/10.1080/01615449709601182.
**49** Wetherell, "Historical Social Network Analysis," 130.
**50** Gurminder K. Bhambra, "Sociology and Postcolonialism: Another 'Missing' Revolution," *Sociology* 41, no. 5, Special Issue on Sociology and Its Public Face(s) (2007): 876–77.

women are depicted as being affected by the social structure without altering the definition of this structure.[51] Feminist sociologists have challenged presumed dichotomies and hierarchies between public and private, macro and micro, institutional and cultural/emotional/embodied, material and discursive, and "the West and the rest."[52] Women are associated with the second term in each of these pairings. Institutions and "theory" (what is said to constitute valid "knowledge") are controlled by men and are perceived to be more professional and intellectual, while women's work is associated with "applied" or "popular" knowledge in the realm of "social problems."[53] As feminist geographers have noted, women are often left out of theorizing about the public realm of the workplace (coded as male); their role is perceived to be solely in the private realm where they provide emotional and sexual support to their male partners.[54] Quantitative analysis can devalue women's life histories (associated with the personal and relationships), relegating them to the margins of discussions of social structures and institutions. However, feminist sociologists have argued that life history should be defined as "doing social theory" because of the way that life history links what is typically defined as private (gender, sex, emotions) with what is often defined as public (politics, economics, education).[55]

I use this theorizing by feminist social scientists to argue that interpersonal or individual networks are simultaneously structural and integrate broader social institutions, such as national, political, and ethnic or racial affiliations. I simultaneously engage qualitative methodologies and network visualizations without depending upon the statistical analyses of SNA. To visualize the relationship between Morgenthau Jr. and Klotz (and compare it to the relationships between HMJ and his father, Morgenthau Sr., and his wife, Elinor Fatman Morgenthau), I utilize a historical social network analysis tool called nodegoat.[56]

**51** Barbara Laslett and Barrie Thorne, "Life Histories of a Movement: An Introduction," in *Feminist Sociology: Life Histories of a Movement*, ed. Barbara Laslett and Barrie Thorne (New Brunswick, NJ: Rutgers University Press, 1997), 7, 15.

**52** Barrie Thorne, "How Can Feminist Sociology Sustain Its Critical Edge?," *Social Problems* 53, no. 4 (2006): 476.

**53** Aisha Khan, "Introduction," in *Women Anthropologists: Selected Biographies*, ed. Ute Gacs et al. (Urbana: University of Illinois Press, 1989), xvii–xviii.

**54** Linda McDowell, "Space, Place and Gender Relations: Part I. Feminist Empiricism and the Geography of Social Relations," *Progress in Human Geography* 17, no. 2 (1993): 166, 170, 173, accessed January 10, 2022, https://doi.org/10.1177/030913259301700202.

**55** Laslett and Thorne, "Life Histories of a Movement," 2–4, 6–8, 10.

**56** Pim van Bree and Geert Kessels, "Nodegoat: A Web-Based Data Management, Network Analysis and Visualization Environment," LAB1100, accessed January 10, 2022, http://lab1100.com, 2013, http://nodegoat.net.

Nodegoat is a project of Lab1100 founded by Pim van Bree and Geert Kessels (https://lab1100.com/). It is an open access web-based data management, network analysis and visualization environment. I selected nodegoat because of its versatility; it enables users to design their own data models and visualize historical social networks chronologically and geographically. Another key reason is that it supports the crafting of multi-layered networks in which various types of nodes ("objects") and relationships can be displayed simultaneously or independently, promoting a "continuous process of interaction with data" leading to deeper "levels of interpretation."[57] Examples of multiple "nodes" or "objects" that I created and visualized at one time in nodegoat are: people, and organizations, or institutions. I also created "categories" that can be visualized simultaneously with these "objects": racial and religious identities (see Figure 4). However, since one's data model in nodegoat is designed by the researcher, any kind of "object" can be visualized simultaneously including (but not limited to): correspondence, events, archival collections, and publications. These are other kinds of objects that I created in my nodegoat database (but these are not visualized in the images included here). Because the nodegoat data model is designed by each researcher, the options are limitless; however, the more complicated the data model, the more complicated it becomes to decide what one wants to visualize and how to visualize it. The visualization of various types of nodes/relationships within one space is impossible within quantitative SNA, which is "ill-equipped to deal with multimodal networks" [networks with more than one kind of "node" or "object"] since when there are "three or more varieties of nodes, most algorithms used in network analysis simply do not work."[58]

Most importantly for my purposes, nodegoat supports the theoretical intersection of what is sometimes still gendered "female" – the micro, private, social, and cultural – with what is still gendered as supposedly " masculine" – the macro, public, institutional, and structural.[59] Thus, one network visualization can reveal

---

**57** Ingeborg van Vugt, "Using Multi-Layered Networks to Disclose Books in the Republic of Letters," *Journal of Historical Network Research* 1 (2017): 35, accessed January 10, 2022, https://doi.org/10.5072/jhnr.v1i1.7.

**58** Scott B. Weingart, "Demystifying Networks, Parts I & II," *Journal of Digital Humanities* 1, no. 1 (2011): 3, accessed January 10, 2022, http://journalofdigitalhumanities.org/1-1/demystifying-networks-by-scott-weingart/.

**59** Bhambra, "Sociology and Postcolonialism," 876–77; Ute Gacs et al., *Women Anthropologists: Selected Biographies* (Urbana: University of Illinois Press, 1989), xiii; Laslett and Thorne, "Life Histories of a Movement," 8, 15, 21; Linda McDowell, "Space, Place and Gender Relations: Part II. Identity, Difference, Feminist Geometries and Geographies," *Progress in Human Geography* 17, no. 3 (1993): 173, accessed January 10, 2022, https://doi.org/10.1177/030913259301700301; Thorne, "How Can Feminist Sociology Sustain Its Critical Edge?," 476.
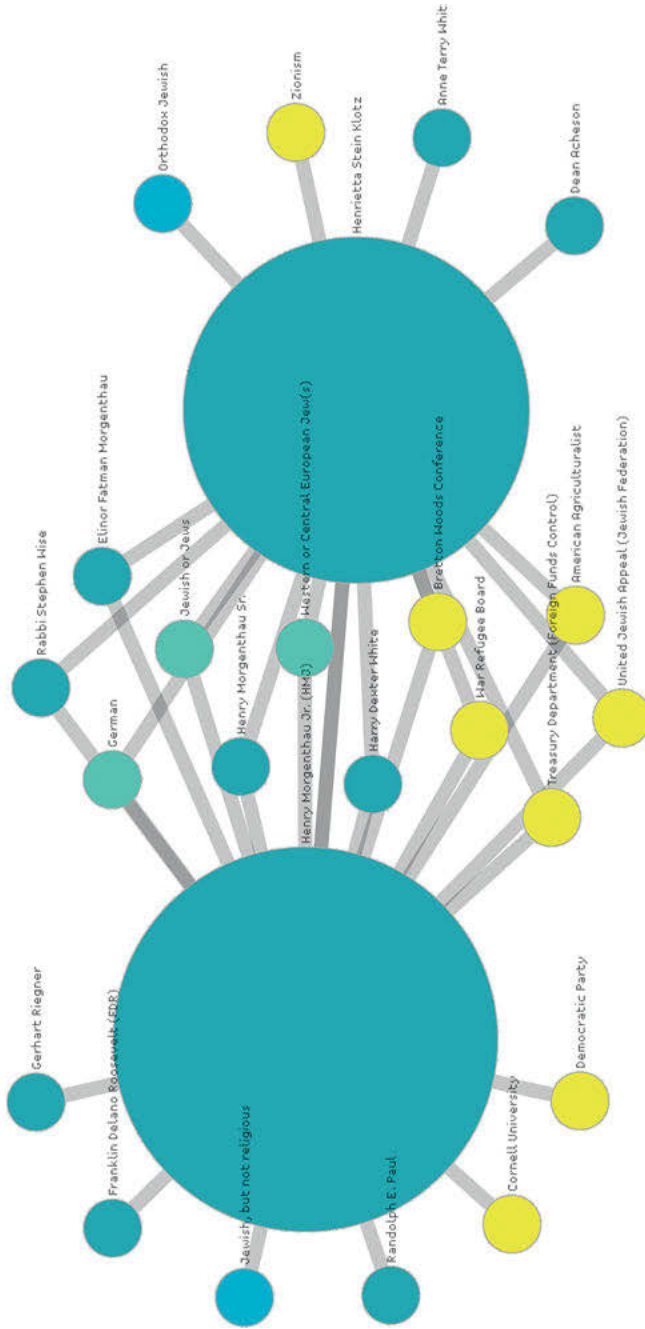
**Figure 4:** This simple social network visualization created with nodegoat displays the people (teal), institutions/organizations (yellow), racial or ethnic identities (turquoise), and religious identities (light blue) that connected Henry Morgenthau Jr. and Henrietta Klotz, based on what I entered in my nodegoat database. Color coding can be customized by the user. People and institutions are "objects" in my nodegoat database, while racial/ethnic identities and religious identities are "categories."

both macro- and micro-level connections and visually demonstrate what are typically perceived to be two different spheres (the supposed "political" sphere of men and the presumed "private" sphere of women).

The ability to visualize "broader social institutions" and the mindsets of individuals simultaneously was imperative for my analysis of how women (who were often in roles with less power) managed to influence men in leadership positions within the U.S. government to bring relief and rescue to Jewish refugees during World War II.

I found that nodegoat visualizations are clearer when only a small number of object types and categories are displayed at any one time ("scope" can be used for this purpose). Figure 4 incorporates only two object types (people and institution) and two categories (racial/ethnic and religious identities). My data model contains other categories, but they are not included here for simplicity. Because I focus on ego-centric networks, it is important to be able to see each of the "nodes" in the visualization and track who/what each is connected to, therefore displaying fewer types of "objects" at once is preferable to perform this micro-level analysis.

These types of social network visualizations serve multiple purposes; they display the connections between data that has been entered into the self-designed database, but they also allow for additional exploration of the data. By clicking on any of the "nodes" (circles) in this visualization, one can learn more about that specific person, organization, or identity (based on what one has entered into one's own database) and discover the people, organizations, and identities that are linked with the "nodes" one is interested in. For example, by clicking on "Zionism," or "Orthodox Jewish" I can learn more about the connection between Henrietta Klotz and these political and religious affiliations, but I can also discover which additional people and institutions relate to these identities in my database.

# 4 Analyzing Political Influence Networks

In this section, I demonstrate how the nodegoat network visualization above can be expanded to further explore the networks of influence on Henry Morgenthau Jr., including interpersonal but also ethnic, religious, and political connections. Especially considering HMJ's assimilated Jewish upbringing and his father's non-Zionist positions, how can an analysis of Henrietta's influence enable new explanations for HMJ's eventual stances regarding relief and rescue efforts and his advocacy for Israel?

Using a "filter" in nodegoat, a researcher can select specific people (in this case HMJ and Henrietta Klotz) and then using "scope" within the visualization settings, it is possible to visualize the people, institutions, racial, ethnic, and religious identities that influenced the key people who impacted HMJ and Mrs. Klotz. This enables a multi-layered visualization depicting micro and macro levels simultaneously. I measure "influence" qualitatively, by analyzing archival sources that describe these relationships (interviews and correspondence) or depict the relationships as they occurred (meeting transcripts in the Henry Morgenthau Jr. Diaries).

There is no way within nodegoat's social visualizations to visually compare the degree of influence or strength of a relationship between two people, except to evaluate in how many institutions they share membership or with how many people they share relationships. For example, in Figure 5, we can detect that Henrietta and HMJ were both connected to Elinor Morgenthau, Henry Morgenthau Sr., Stephen Wise, Dean Acheson (Acting Secretary of State), and Harry Dexter White (Assistant Secretary of the Treasury). We can also see that Henrietta shared her Zionist affiliation with Stephen Wise and Gerhart Riegner (World Jewish Congress representative in Geneva, Switzerland), and worked on the War Refugee Board with HMJ, John Pehle, Randolph Paul, and Josiah DuBois. However, the visualization alone does not provide more information about the content or level of influence of any of these relationships. Within each person and institution "object," it is possible to create text fields to store this qualitative data that can be tagged based on the objects and categories in one's data model. The tags within text boxes can then be visualized (if desired).

The connecting lines between "objects" (they are called "edges" within social network analysis) do not become thicker and the colors of the lines cannot be changed depending on the strength of a relationship. The "nodes" or "objects" (the circles) grow larger when they have more connected "nodes," but this does not illustrate the strength of a relationship between any two people in the network. In Figure 5, the size of the nodes reflects how many other objects and categories are connected to them. HMJ is the largest circle because he is the center of this ego-centric network and I have entered the most information into the database about the people who influenced him, his institutional affiliations, his racial, and religious identities. Since my ego-centric networks are not comprehensive and do not represent the "whole network," centrality, network, and path measures are not relevant.

This expanded and complex network visualization can be used to further explore the data. For example, how could Henrietta's connection with the institutional and religious identities "Zionism" and "Orthodox Jewish" have influenced HMJ's political stances regarding Jewish refugees and advocacy for the State of

**Figure 5:** This is a more complex network visualization, which expands upon the one above. In this visual, the advisors/influencers (teal), institutions (yellow), racial/ethnic identities (turquoise), and religious identities (light blue) of the people who influenced HMJ and Henrietta Klotz are depicted. The people and institutions in this visualization are "objects," while the racial and religious identities are "categories" in nodegoat. This enables analysis of an additional layer of influence.

Israel? How did Henrietta manage to become more influential than his father, Henry Morgenthau Sr., and his wife, Elinor Fatman Morgenthau? His father was against the creation of a Jewish nation-state and both HM Sr. and Elinor were connected to the religious identity, "Jewish, but not religious." In the next three sub-sections, I compare the Jewish identities of Henrietta, Morgenthau Sr., and Elinor and their contrasting relationships to Henry Morgenthau Jr. In the first sub-section, I examine Mrs. Klotz's Orthodox Jewish upbringing, connection to Zionism, and how she employed her Jewish identity to impact HMJ's decision making. In the second sub-section, I compare Morgenthau Sr.'s experiences of antisemitism, betrayal from his Jewish peers, and his increasingly non-Zionist positions with his son's experience of admiration from his fellow Jews through his fundraising for the State of Israel. I also argue that it was Henrietta's loyalty, protectiveness, and encouragement that endeared her to Morgenthau Jr., since he did not receive this support from his father. In the third and final sub-section, I assess Elinor's assimilationism, Christian friendship networks, involvement in democratic politics, and philanthropic causes. I emphasize HM III's evaluation that his father was closer to Henrietta than he was to Elinor.

## 4.1 Henrietta used Silence as Political Strategy: "I Am Jewish and His Cause was Justified"

Herman Klotz, Henrietta's husband, explained that Henrietta's father, Harry Stein, was a "pious, devout Jew" who was also poor.[60] Henrietta emphasized how Orthodox her parents were: "believe me, my parents were orthodox Jews, but very orthodox. I mean, really orthodox."[61] Herman was also raised Orthodox, but more flexible, and Zionism was a frequent part of his family's conversations. He also underscored how likely it was that Henrietta's and his own background influenced HMJ regarding Jewish matters.[62] Henrietta informed HM III that even though she was raised in an Orthodox Jewish home, she was not raised with Zionism. She learned about it during the second four years of her tenure at Treasury with HMJ, through her attendance at parties hosted by an unnamed Jewish woman.[63]

---

**60** Klotz, "Herman Klotz to Henry Morgenthau III, October 16, 1985," 7.

**61** Klotz and Morgenthau III, "Henry Morgenthau III, Interview with Henrietta Klotz, September 19, 1978," 48.

**62** Klotz, "Herman Klotz to Henry Morgenthau III, October 16, 1985," 7–8.

**63** Klotz and Morgenthau III, "Henry Morgenthau III, Interview with Henrietta Klotz, September 19, 1978," 48.

In one case, Henrietta hid the truth from Morgenthau Jr. so that he would unfreeze funds that were needed to help a group of rabbis escape from Poland. A rabbi (Kalmanowitz) asked for an appointment to see HMJ; he claimed he did not speak English, so he brought another rabbi to translate for him.[64] Neither of the rabbis knew that Henrietta could understand both German and Yiddish, since, as she explained to HM III, "I didn't look Jewish at that time. I was young and I was very, very blond."[65] Kalmanowitz insisted (in Yiddish) that the translator push HMJ more since "he looks like the kind you can get away with an awful lot of things." Then when HMJ explained that he would "take it under advisement," Kalmanowitz replied to his translator that once you hear that, "forget it. He's going to do nothing," and promptly pretended to faint. HMJ thought it was real and asked Henrietta to get this rabbi into HMJ's car and back to Kalmanowitz's hotel.[66] At that point, Kalmanowitz said in German to his translator, "did I cry well?" Henrietta explained to HM III: "of course I wouldn't tell your father [HMJ] because [if I did] he'd never let another rabbi in. And I am Jewish and his cause was justified."[67] The Treasury ended up releasing funds to this group of rabbis and to thank HMJ, they threw him a "big dinner" and gave him "medals."[68] Mrs. Klotz only told HMJ the truth many years later when they were working with Henry Montor on Israel Bonds.[69] This anecdote demonstrates how Henrietta's Jewish identity influenced her behavior with HMJ and in turn how these choices influenced the decisions he made in regard to rescue programs. In this situation, she chose to use silence or omission as a powerful tool of influence.

---

**64** Klotz and Morgenthau III, "Henry Morgenthau III, Interview with Henrietta Klotz, September 19, 1978," 48.
**65** Klotz and Morgenthau III, "Henry Morgenthau III, Interview with Henrietta Klotz, September 19, 1978," 49–50. Henrietta's mother had learned Yiddish upon arrival in the U.S. She taught her daughter both German and Yiddish.
**66** Klotz and Morgenthau III, "Henry Morgenthau III, Interview with Henrietta Klotz, September 19, 1978," 49.
**67** Klotz and Morgenthau III, "Henry Morgenthau III, Interview with Henrietta Klotz, September 19, 1978," 50.
**68** Klotz and Morgenthau III, "Henry Morgenthau III, Interview with Henrietta Klotz, September 19, 1978," 50.
**69** Klotz and Morgenthau III, "Henry Morgenthau III, Interview with Henrietta Klotz, September 19, 1978," 52.

## 4.2  Henry Morgenthau Sr.: "How Could a Father Say That about His Son?"

Henry Morgenthau Sr. (also pictured in the nodegoat social network visualizations, connected to both HMJ and Henrietta) was not a Zionist and had a falling out with Rabbi Stephen Wise over the issue of Zionism. HM Sr. believed that the U.S. was Zion for American Jews. This stemmed from Morgenthau Sr.'s desire to assimilate into the American upper class and wish for his son to be able to access professional and social networks which were beyond his own reach because of antisemitism.[70] His life was "centered around his business, his ambition and his son." He feared "anything that would threaten his Americanism" and discouraged his son from becoming involved in "Jewish things."[71]

HM Sr. had wanted to be Secretary of the Treasury or Commerce (or another cabinet position) when Woodrow Wilson was elected. Morgenthau Sr. had donated $20,000 to Wilson's campaign and was one of his original fundraisers. Despite the expectation among many in HM Sr.'s networks, he did not receive a cabinet post, probably because of antisemitism.[72] He was given the ambassadorship in Turkey because, as Wilson explained, this was "the point at which the interest of American Jews in the welfare of Palestine is focused, and it is almost indispensable that I have a Jew at that post."[73] Rabbi Stephen Wise convinced HM Sr. to take the ambassadorship to further Jewish interests in Palestine.[74] Morgenthau Sr. felt that his ambassadorship to Turkey was a "Jewish slot"; it was the "only diplomatic post to which a Jew can aspire," he wrote in 1922.[75]

HM Sr.'s non-Zionist stance originated in his on-the-ground experience in Palestine in 1914, when he witnessed firsthand the "instinctive hatred" that "Arabs" felt toward Morgenthau's group.[76] HM Sr. felt that a Jewish state would

---

**70** Beschloss, *The Conquerors*, 44–46; Morgenthau III, *Mostly Morgenthaus (e-Book)*, 86, 91–92, 155–56, 165.

**71** Beschloss, *The Conquerors*, 44–46.

**72** Morgenthau III, *Mostly Morgenthaus (e-Book)*, 92–93. In 1913, Stephen Wise wrote to HM Sr., "no Jew has been appointed to a single place of importance. It seems an almost deliberate slight." Charles Strauss wrote to Senator James A. O'Gorman, also in 1913, that the newspapers were criticizing President Wilson for his ingratitude to HM Sr. and were calling Wilson's neglect a form of "race prejudice."

**73** Morgenthau III, *Mostly Morgenthaus (e-Book)*, 90–93.

**74** Morgenthau III, *Mostly Morgenthaus (e-Book)*, 95–96.

**75** Klotz and Morgenthau III, "Henry Morgenthau III, Interview with Henrietta Klotz, September 19, 1978," 6; Morgenthau III, *Mostly Morgenthaus (e-Book)*, 91–93; Henry Morgenthau Sr. and French Strother, *All in a Lifetime* (Garden City, NY: Doubleday, Page & Company, 1922), 160.

**76** Morgenthau III, *Mostly Morgenthaus (e-Book)*, 119.

only cause increased antisemitism.[77] He feared that the Jewish campaign for a national homeland would trigger a genocide, like what was happening to the Armenians in the Ottoman Empire.[78] HM Sr. also became concerned when leaders of Aaron Aaronsohn's Jewish agricultural colony (Athlit, near Haifa) spoke about the need to "drive out the Arabs." Morgenthau Sr. described the profound feeling, as he prayed alongside Muslims, Christians, and Jews in the Caves of Machpelah in Hebron (near Jerusalem), that "we traced our religion back to the same source" and remarked that these ten minutes "were undoubtedly the most sacred that I have ever spent in my life."[79] He was so impressed by the farm living he observed in Petah Tikvah (near Tel Aviv) and at Aaronsohn's colony in Athlit, that he encouraged his son, HM Jr. to take up farming.[80] Morgenthau Sr. did not oppose the settlement of Eastern European Jews in Palestine, but only the creation of "a limited national state."[81]

Morgenthau Jr. echoed his father's beliefs in the importance of collaboration between minority ethnic or religious groups. In December 1943, his staff, including Henrietta, pushed him to directly approach FDR to enable the evacuation of Jews from France and Romania. HMJ was reluctant to approach the President as a "private citizen" and felt that he needed to locate his argument within his role as Secretary of the Treasury.[82] From this standpoint, he could emphasize the "question of treating minority races." Then he reaffirmed a sentiment his father had expressed in 1914 and 1921, "Just because I am a Jew, why shouldn't I look after the Jews, or the Catholics, or the Armenians?" Mrs. Klotz reassured him that taking action would be connected to his role as Secretary of the Treasury: "You got into this thing on a Treasury basis – on a financial basis. It has led into this thing, you see."[83]

In 1917, Zionists including Louis Brandeis, Felix Frankfurter, and Chaim Weizmann sabotaged Morgenthau Sr.'s mission to set up a separate peace with Turkey (initially endorsed by Woodrow Wilson but later this support was revoked). This was a consequence of a temporary alliance between anti-Semites and Zionists who wanted to establish a British-controlled Jewish state in Palestine. They depicted HM Sr. as pro-German, pro-Turkish, anti-Zionist. Morgenthau Sr. felt the ultimate betrayal when his close friend Rabbi Stephen Wise headed a delegation of Jewish

---

**77** Morgenthau III, *Mostly Morgenthaus (e-Book)*, 164.

**78** Morgenthau III, *Mostly Morgenthaus (e-Book)*, 133–34.

**79** Morgenthau III, *Mostly Morgenthaus (e-Book)*, 120, 123.

**80** Morgenthau III, *Mostly Morgenthaus (e-Book)*, 114–15, 122–23.

**81** Morgenthau III, *Mostly Morgenthaus (e-Book)*, 151, 155.

**82** Morgenthau Jr. et al., "Jewish Evacuation Meeting Transcript, December 18, 1943," 85–86.

**83** Morgenthau Jr. et al., "Jewish Evacuation Meeting Transcript, December 18, 1943," 87.

leaders (which did not include HM Sr.) who convinced Wilson to approve of the Zionist plans.[84] HM III argued that these negative incidents pushed his grandfather to become increasingly oppositional to Zionism.[85] Importantly, Rabbi Wise had officiated Morgenthau Jr.'s marriage to Elinor Fatman in 1916,[86] a fact which likely made Wise's advocacy with President Wilson even more hurtful to HM Sr. In his 1921 "Zionism a Surrender, not a Solution," HM Sr. opposed Chaim Weizmann's Zionist agenda, calling it "'the most stupendous fallacy in Jewish history,'" an "'eastern European proposal'" that would "'cost the Jews of America most [of what] they have gained.'"[87]

Henry Morgenthau III told Henrietta that his grandfather, Morgenthau Sr. used to warn his son (HMJ) "the Jews will stab you in the back." To this, Mrs. Klotz replied, "They didn't stab him [HMJ] in the back … they admired him tremendously."[88] This exchange occurred in the context of a discussion about HMJ's public speaking campaigns to fundraise for Israel. Morgenthau Sr.'s sense of betrayal from fellow Jews likely began with the treatment he suffered from Rabbi Stephen Wise, Louis D. Brandeis, and Chaim Weizmann in 1917. But he underwent another betrayal from his former friend, Samuel Untermyer, who turned against HM Sr. when Morgenthau Sr. criticized Zionism in 1921. Untermyer described HM Sr.'s essay ("Zionism a Surrender") as "personal egotism." HM III explained that after this falling out, his grandfather "quite understandably withdrew permanently from all involvement with organized Jewry."[89]

Rather than perceiving Morgenthau Sr.'s increasing antipathy toward Zionism as solely political, structural, macro, and institutional, I argue that utilizing nodegoat encourages the conceptualization of the intersections of macro and micro levels of influence. HM Sr.'s stances regarding Zionism shifted because of his personal experiences on the ground in Palestine, but also because of interpersonal duplicity he suffered from people he considered friends and colleagues. His advice to his son to avoid involvement in Jewish causes resulted from this subjective experience of disloyalty but also from the sharp acrimony he felt when antisemitism led him to be passed over for prestigious government positions.

Henrietta recounted various anecdotes illustrating how much she disliked Henry Morgenthau Sr. because of what she perceived to be his "cruel" or "crude"

---

**84** Morgenthau III, *Mostly Morgenthaus (e-Book)*, 145–56.

**85** Morgenthau III, *Mostly Morgenthaus (e-Book)*, 155.

**86** Morgenthau III, *Mostly Morgenthaus (e-Book)*, 182.

**87** Morgenthau III, *Mostly Morgenthaus (e-Book)*, 164.

**88** Klotz and Morgenthau III, "Henry Morgenthau III, Interview with Henrietta Klotz, September 19, 1978," 59.

**89** Morgenthau III, *Mostly Morgenthaus (e-Book)*, 165.

behavior toward his son and toward her. She felt that Morgenthau Sr. "dominated" his son and "was very disappointed" in him. When FDR appointed HMJ as the Secretary of the Treasury, HM Sr. apparently stated, "he's [HMJ] not up for it" but that HM Sr. would be. Henrietta exclaimed to Henry Morgenthau III: "I disliked him from that time on. How could a father say that about his son?"[90] Morgenthau Sr. likely said this because of jealousy; he had wanted that position for himself.

In contrast, Henrietta described HMJ as "fair to everybody" and "honorable."[91] Henrietta recounted another anecdote indicating HM Sr.'s snobbery, classism, and elitism toward her specifically. This incident as well as Morgenthau Sr.'s deplorable treatment of his son (from her perspective) likely contributed to her assessment of the father. She was managing a property for the Morgenthau family and HM Sr. and his wife (Josephine) returned from Europe with various gifts. Josephine asked Henrietta which gift she would like but HM Sr. snapped, "oh she'll [Henrietta] take anything she can get for nothing." This offended Henrietta who "refused to accept anything." When Josephine "went into this institution" (unclear if it was a mental institution or nursing home) both HMJ and Henrietta visited her, while HM Sr. never did.[92]

Mrs. Klotz's attitude toward HM Sr. reveals her protectiveness, devotion, and loyalty to HMJ (as the term "watchdog of the Secretary of the Treasury" connotes). Morgenthau Jr. emphasized these characteristics in his 1945 letter to her: "God help anyone who in your opinion was disloyal to me."[93] Henrietta told HM III, "if your father had asked me to jump off the roof, I think I would have done it."[94] Gabrielle Elliot Forbush (Elinor Morgenthau's friend from Vassar College), who had introduced Henrietta to Morgenthau initially, told HM III that Henrietta brought out the "'suspicious vein'" in Morgenthau: "She wanted to protect him and she was fiercely loyal. But she viewed everyone and everything in a narrowly personal way." Once Henrietta started to work at Morgenthau's magazine, *American Agriculturalist* in 1922, "'from then on they were together.'"[95] I maintain that HMJ trusted and relied upon Henrietta and followed her advice because she

---

**90** Klotz and Morgenthau III, "Henry Morgenthau III, Interview with Henrietta Klotz, September 19, 1978," 6–9.

**91** Klotz and Morgenthau III, "Henry Morgenthau III, Interview with Henrietta Klotz, September 19, 1978," 8.

**92** Klotz and Morgenthau III, "Henry Morgenthau III, Interview with Henrietta Klotz, September 19, 1978," 7–9.

**93** Morgenthau Jr., "Handwritten Letter from Henry Morgenthau Jr. to Mrs. Henrietta Klotz, August 5, 1945," 7.

**94** Klotz and Morgenthau III, "Henry Morgenthau III, Interview with Henrietta Klotz, September 19, 1978," 4.

**95** Morgenthau III, *Mostly Morgenthaus*, 251–52.

was so devoted, loyal, and protective; this was the method through which she exercised her power. She encouraged and advocated for him in a way that his father did not. Henrietta gave him the confidence in himself that allowed him to set himself apart from his father and advocate for Jewish refugees and for the State of Israel. HM III explained that after the deaths of HMJ's father in 1946 and his wife, Elinor in 1949, the two people who had tried to "steer him away" from "Jewish affairs" were gone; it was at this exact time that Morgenthau Jr. "became the devoted servant of the Jewish community." Morgenthau III reasoned that this shift was a "rebellion" as well as a "craving for love, respect …."[96] Through Henrietta's Jewish networks, she arranged for HMJ to be involved in organizations where he received that love and respect he craved.

In the December 1943 Jewish evacuation meeting discussed above, Henrietta pushed Morgenthau Jr. into action through a combination of encouragement and praise. Morgenthau wondered if he should bring other Jews (Samuel Rosenman or Herbert Lehman) with him to meet with Cordell Hull (the Secretary of State) and the President about the proposal to evacuate Jews from France and Romania. Morgenthau preferred to go as the "Treasury" because he was concerned about taking a "Jewish delegation" and thus drawing too much attention to his Jewishness. Mrs. Klotz tried to embolden Morgenthau by exclaiming: "Mr. Morgenthau, nobody would do – none of these people you mentioned, when they are put on the spot, will do what you will do."[97] This interpersonal, social, or micro-level explanation is necessary to appreciate his political or macro-level positions.

## 4.3 Elinor Fatman Morgenthau: Jewishness was "Ignored Completely"

HM III described his parents' "malaise" regarding their Jewish identity. Elinor and Morgenthau Jr. saw being Jewish as a "kind of birth defect that could not be eradicated, but with proper treatment, could be overcome, if not in this generation, then probably in the next."[98] HM III called his mother more "firmly assimilationist" than HMJ and remembered how Elinor explained to her son that he should tell his friends he was "just American" when they asked what his religion was.[99] HMJ and Elinor did not attend a synagogue or German Jewish social clubs; "almost all" of their friends were elite German Jews, but "they never talked

---

**96** Morgenthau III, *Mostly Morgenthaus (e-Book)*, 342.

**97** Morgenthau Jr. et al., "Jewish Evacuation Meeting Transcript, December 18, 1943," 89.

**98** Morgenthau III, *Mostly Morgenthaus (e-Book)*, 10.

**99** Morgenthau III, *Mostly Morgenthaus (e-Book)*, 10.

about anything Jewish" and did not keep any Jewish religious objects at home.[100] HM III explained that his parents ate pork, celebrated both Christmas and Easter rather than Jewish holidays like Passover or Yom Kippur, did not go to Jewish doctors, dentists, or lawyers, and spent summers with Elinor's Protestant Vassar friends.[101]

HM III explicitly referred to his mother's jealousy of Henrietta: "My mother had realized my father's attraction to his good-looking, intelligent, and ambitious young secretary." Henry Morgenthau III termed this the "Henry-Elinor-Henrietta triangle."[102] HM III told Henrietta during their interview that his father "was very close to you and depended [on you]." Henrietta affirmed that this was true. HM III then stated: "He was as close to you as he was to any human being."[103] HM III implied that Henrietta played a much more influential role for his father than his wife, Elinor, did. This was because as a traditional wife from an elite class, she could not be a career woman (whereas Henrietta worked from the age of 21). According to HM III, Elinor preferred that HMJ avoid "the plight of European Jewry," while Henrietta actively pushed HMJ to advocate for Jews. Also, Elinor became ill during World War II.[104] Edna Friedberg maintains that toward the end of the war, Elinor "supported her husband's increasing Jewish involvement … including campaigning on behalf of the new State of Israel."[105] Elinor visited the Emergency Refugee Shelter at Fort Ontario in Oswego, NY in 1944 with Eleanor Roosevelt and advocated for the continuing education of medical students who were refugees living at the camp.[106]

Elinor Morgenthau studied theater at Vassar College (graduated in 1913), often acting as the male lead in the all-female productions, and spoke multiple languages including German, French, Spanish, and Russian. Her closest friends at Vassar were Protestant girls of "modest means"; as Gabrielle Forbush recalled,

---

**100** Beschloss, *The Conquerors*, loc 883–889, n. 1; Morgenthau III, *Mostly Morgenthaus (e-Book)*, 11–13.

**101** Morgenthau III, *Mostly Morgenthaus (e-Book)*, 11–12.

**102** Morgenthau III, *Mostly Morgenthaus*, 305–6.

**103** Klotz and Morgenthau III, "Henry Morgenthau III, Interview with Henrietta Klotz, September 19, 1978," 14–15.

**104** Morgenthau III, *Mostly Morgenthaus*, 222, 303, 305.

**105** Friedberg, "Elinor Morgenthau (1891–1949)."

**106** Friedberg, "Elinor Morgenthau (1891–1949)"; John Pehle and Henry Morgenthau Jr., "Untitled Meeting Transcript, September 19, 1944," *Morgenthau Diaries* 772 (1944): 157, accessed January 10, 2022, http://www.fdrlibrary.marist.edu/_resources/images/morg/md1066.pdf; Eleanor Roosevelt, "Letter from Eleanor Roosevelt to John Pehle, November 2, 1944," *Morgenthau Diaries* 792 (1944): 112–13, accessed January 10, 2022, http://www.fdrlibrary.marist.edu/_resources/images/morg/md1087.pdf.

there were very few Jewish girls at Vassar and Elinor's Jewishness was "ignored completely." Going to college was "very much the exception" among women of Elinor's generation.[107] Maintaining feminine comportment was important to her.[108]

After college, Elinor Fatman taught theater at the Neighborhood Playhouse, part of the Henry Street Settlement House on Manhattan's Lower East Side. Henry Street provided social services and encouraged Americanization for Eastern European Jewish immigrants. The House was run by Lillian Wald, a nurse of German Jewish descent, and was a "kind of social service finishing school" for the children of elite families. Morgenthau Jr. also volunteered at Henry Street.[109] HM Sr. had helped to create the Bronx House (an extension of the Henry Street House), and its music school was supported by Josephine (HMJ's mother) who was a talented musician.[110] Elinor's work at Henry Street was in line with the Morgenthau family's commitment to philanthropy, focusing on Jewish assimilation into white Christian cultural norms.

Though Elinor could not continue her career as an actress because she had to focus on being a mother and "farmer's wife," she used her theater skills when she became a speaker for the state Democratic Committee, Women's Division.[111] Elinor trained in public speaking to promote Morgenthau Jr.'s political career and fill in for him when he could not attend speaking engagements. In 1941, she became Eleanor Roosevelt's assistant in the Office of Civilian Defense, but she soon resigned (because of administrative struggles and "Eleanor-haters"). After this, "her private overseeing" and "careful advice" to her husband "declined." Henrietta began to have an increasing influence on HMJ as evidenced in his stance regarding the "plight of European Jewry."[112] Elinor's influence on Morgenthau Jr. was more in the realm of democratic politics than in Jewish affairs, though it seems that she became slightly more engaged in Jewish causes near the end of the war. Unfortunately, she had a heart attack in April 1945 and then a stroke in September 1949 that led to her death.[113]

---

107 Morgenthau III, *Mostly Morgenthaus (e-Book)*, 175.

108 Morgenthau III, *Mostly Morgenthaus (e-Book)*, 173–74, 176.

109 Morgenthau III, *Mostly Morgenthaus (e-Book)*, 170.

110 Morgenthau III, *Mostly Morgenthaus (e-Book)*, 400, n. 410.

111 Morgenthau III, *Mostly Morgenthaus (e-Book)*, 185, 206–7.

112 Friedberg, "Elinor Morgenthau (1891–1949)"; Morgenthau III, *Mostly Morgenthaus (e-Book)*, 199–200, 206–7, 251–53.

113 Friedberg, "Elinor Morgenthau (1891–1949)."

# 5 Conclusion

A case study exploring Henrietta Stein Klotz's impact on Henry Morgenthau Jr., the Secretary of the Treasury from 1934 to 1945, highlights the power that women invoked in political decision making, in the position of secretary, a role not often correlated with the exercise of political authority. Specifically, exploring Mrs. Klotz's affiliations and actions can clarify why Morgenthau Jr. became more explicitly involved in Jewish causes, despite his father's and wife's avoidance of participation in "Jewish affairs." Henrietta used strategies like encouragement, loyalty, protectiveness, advocacy, and even sometimes silence or omission to push Henry Morgenthau Jr. to take political positions that were aligned with her own. Digital tools, like nodegoat (a database and network visualization platform), can make visible the people who are often either invisible or understudied in histories of the U.S. government's responses to the Holocaust: women, and more specifically, secretaries. Because nodegoat permits the visualization and exploration of multiple types of "objects" and "categories" simultaneously (whereas other network visualization software only reveals one type of object at a time), nodegoat supports the intersection of interpersonal and structural/institutional webs of influence and reveals how what is traditionally defined as feminine, micro, private, social, or cultural, is concurrently macro, public, and structural. This digital approach opens new ways to conceptualize women's and secretaries' influence on political decision making related to rescue and relief of refugees and fundraising after the war for the State of Israel. More broadly, historical social network analysis, as shown here, can help bring hitherto unheard voices and ignored agencies to the fore, offering an important contribution to Jewish Studies in the Digital Age.

# References

Adalian, Rouben Paul. "Morgenthau, Ambassador Henry, Sr." Armenian National Institute, 2019. Accessed January 10, 2022. https://www.armenian-genocide.org/morgenthau.html.

Archive, NYT. "H.S. Klotz, 87, Aide to Treasury Secretary." *New York Times Online*. December 21, 1988. Accessed January 10, 2022. https://www.nytimes.com/1988/12/21/obituaries/h-s-klotz-87-aide-to-treasury-secretary.html.

Baumel, Judith Tydor. "Women's Agency and Survival Strategies during the Holocaust." *Women's Studies International Forum* 22, no. 3 (1999): 329–47.

Beschloss, Michael. *The Conquerors: Roosevelt, Truman, and the Destruction of Hitler's Germany, 1941–1945*. New York: Simon & Schuster, 2002.

Bhambra, Gurminder K. "Sociology and Postcolonialism: Another 'Missing' Revolution." *Sociology* 41, no. 5, Special Issue on Sociology and Its Public Face(s) (2007): 871–84.

Bree, Pim van and Geert Kessels. "Nodegoat: A Web-Based Data Management, Network Analysis and Visualization Environment." LAB1100. Accessed January 10, 2022. http://lab1100.com, 2013. http://nodegoat.net.

Century, Rachel. "Dictating the Holocaust: Female Administrators of the Third Reich." Royal Holloway, University of London, 2012.

Erbelding, Rebecca. "Morgenthau Family Papers, 1860–2015, 2015.255.1 Finding Aid." Washington, DC, 2015. Accessed January 10, 2022. https://collections.ushmm.org/search/catalog/irn96059.

Erbelding, Rebecca. *Rescue Board: The Untold Story of America's Efforts to Save the Jews of Europe*. New York: Doubleday, 2018.

Erickson, Bonnie H. "Social Networks and History: A Review Essay." *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 30, no. 3 (1997): 149–57. Accessed January 10, 2022 https://doi.org/10.1080/01615449709601182.

Franzosi, Roberto and John W. Mohr. "New Directions in Formalization and Historical Analysis." *Theory and Society* 26, no. 2/3 (1997): 133–60.

Friedberg, Edna S. "Elinor Morgenthau (1891–1949)." In *Jewish Women: A Comprehensive Historical Encyclopedia*. Jewish Women's Archive, 1999. Accessed January 10, 2022. https://jwa.org/encyclopedia/article/morgenthau-elinor.

Gacs, Ute, Aisha Khan, Jerrie McIntyre, and Ruth Weinberg. *Women Anthropologists: Selected Biographies*. Urbana: University of Illinois Press, 1989.

"Henry Morgenthau Jr. Papers, 1866–1953, Finding Aid." Franklin D. Roosevelt Presidential Library and Museum, n.d. Accessed January 10, 2022. http://www.fdrlibrary.marist.edu/archives/collections/franklin/index.php?p=collections/findingaid&id=159&q=&rootcontentid=72431.

Khan, Aisha. "Introduction." In *Women Anthropologists: Selected Biographies*, edited by Ute Gacs, Aisha Khan, Jerrie McIntyre, and Ruth Weinberg, xiii–xviii. Urbana: University of Illinois Press, 1989.

Klotz, Henrietta and Henry Morgenthau III. "Henry Morgenthau III, Interview with Henrietta Klotz, September 19, 1978." Washington DC: USHMM 2015.255.1 Morgenthau Family Papers, Box 32, File 11 "Henrietta Klotz," 1978.

Klotz, Herman. "Herman Klotz to Henry Morgenthau III, October 16, 1985." Washington DC: USHMM 2015.255.1 Morgenthau Family Papers, Box 32, File 11, 1985.

Krimmer, Elisabeth. *German Women's Life Writing and the Holocaust: Complicity and Gender in the Second World War*. Cambridge: Cambridge University Press, 2018.

Laslett, Barbara and Barrie Thorne. "Life Histories of a Movement: An Introduction." In *Feminist Sociology: Life Histories of a Movement*, edited by Barbara Laslett and Barrie Thorne, 1–27. New Brunswick, NJ: Rutgers University Press, 1997.

McDowell, Linda. "Space, Place and Gender Relations: Part I. Feminist Empiricism and the Geography of Social Relations." *Progress in Human Geography* 17, no. 2 (1993): 157–79. Accessed January 10, 2022. https://doi.org/10.1177/030913259301700202.

McDowell, Linda. "Space, Place and Gender Relations: Part II. Identity, Difference, Feminist Geometries and Geographies." *Progress in Human Geography* 17, no. 3 (1993): 305–18. Accessed January 10, 2022. https://doi.org/10.1177/030913259301700301.

Meyer, Leisa D. *Creating GI Jane: Sexuality and Power in the Women's Army Corps during World War II*. New York: Columbia University Press, 1996.

Morgenthau III, Henry. *Mostly Morgenthaus: A Family History*. New York: Ticknor & Fields, 1991.

Morgenthau III, Henry. *Mostly Morgenthaus: A Family History (e-Book)*. 2nd ed. Lexington, MA: Plunkett Lake Press, 2019. Accessed January 10, 2022. http://plunkettlakepress.com/mmo.html.

Morgenthau Jr., Henry. "Handwritten Letter from Henry Morgenthau Jr. to Mrs. Henrietta Klotz, August 5, 1945." Washington DC: USHMM 2015.255.1 Morgenthau Family Papers, Box 32, File 11, 1945.

Morgenthau Jr., Henry, Randolph Paul, Josiah DuBois, Ansel Luxford, John Pehle, and Henrietta Klotz. "Jewish Evacuation Meeting Transcript, December 18, 1943." *Henry Morgenthau Jr. Diaries, FDR Presidential Library and Museum* 688II (1943): 82–94.

Morgenthau Sr., Henry, and French Strother. *All in a Lifetime*. Garden City, NY: Doubleday, Page & Company, 1922.

"Mrs. Henry Morgenthau, 73 Widow of Roosevelt's Aide." *New York Times Archives*. July 19, 1972. Accessed January 10, 2022. https://www.nytimes.com/1972/07/19/archives/mrs-henry-morgenthau-73-widow-of-rooseve-ts-aide.html.

Paul, Randolph. "Report to the Secretary of the Acquiescence of this Government in the Murder of the Jews, given to Henry Morgenthau Jr. by His Staff (Josiah DuBois, John Pehle, and Randolph Paul) on January 13, 1944." *Morgenthau Diaries* 693 (1944): 212–29.

Pehle, John and Henry Morgenthau Jr. "Untitled Meeting Transcript, September 19, 1944." *Morgenthau Diaries* 772 (1944): 157. Accessed January 10, 2022. http://www.fdrlibrary.marist.edu/_resources/images/morg/md1066.pdf.

Penkower, Monty Noam. "The Earl Harrison Report: Its Genesis and Its Significance." *The American Jewish Archives Journal* 68, no. 1 (2016): 1–75.

Roosevelt, Eleanor. "Letter from Eleanor Roosevelt to John Pehle, November 2, 1944." *Morgenthau Diaries* 792 (1944): 112–13. Accessed January 10, 2022. http://www.fdrlibrary.marist.edu/_resources/images/morg/md1087.pdf.

Thorne, Barrie. "How Can Feminist Sociology Sustain Its Critical Edge?" *Social Problems* 53, no. 4 (2006): 473–78.

Unknown. "Undated Memo to Explain Henry Morgenthau Jr.'s August 5, 1945 Letter to Henrietta Klotz." Washington DC: USHMM Accession 2015.255.1 Morgenthau Family Papers, Box 32, File 11 "Henrietta Klotz," 184–85 in file, n.d.

Vugt, Ingeborg van. "Using Multi-Layered Networks to Disclose Books in the Republic of Letters." *Journal of Historical Network Research* 1 (2017): 25–51. Accessed January 10, 2022. https://doi.org/10.5072/jhnr.v1i1.7.

Weingart, Scott B. "Demystifying Networks, Parts I & II." *Journal of Digital Humanities* 1, no. 1 (2011): 15–33. Accessed January 10, 2022. http://journalofdigitalhumanities.org/1-1/articles/demystifying-networks-by-scott-weingart/

Weizmann, Chaim. "Letter from Chaim Weizmann to Henrietta Klotz, January 3, 1944." *Morgenthau Diaries* 689 (1944): 201.

Wetherell, Charles. "Historical Social Network Analysis." *International Review of Social History* 43, supplement (1998): 125–44. Accessed January 10, 2022. https://doi.org/10.1017/S0020859000115123.

Zef M. Segal

# Constructing the Modern Jewish "Present": Time and Time Cycles in *HaTzfira*

**Abstract:** The modern periodical is an important medium in the construction of time. Its appearance and cycles of production turn artificial time cycles into seemingly natural and accepted social rhythms. Most importantly, periodicals play an important role in the construction of the "present" as a time frame of occurrences that happen "now". However, the reproduced "present" shouldn't be understood independently of the production cycle of the periodical.

Accordingly, this study characterizes the differences resulting from the shift in time cycles of the nineteenth-century Hebrew periodical *HaTzfira*. This periodical started in 1862 as a weekly and was transformed in 1886 into a daily. In order to explore the change, this chapter compares the discourse in the three years prior to the conversion of this weekly into daily (1883–1885) with the discourse in the three years following this conversion (1886–1888).

Through the use of computational tools, and in particular topic modeling algorithms, which offer a general overview of large-scale textual corpora, this chapter compares discursive patterns before and after 1886. This comparison is based, on the one hand, on a nuanced qualitative analysis of the resultant topics, and on the other hand, on an original mathematical analysis of the resultant vector space. On a theoretical level, this comparison helps characterize the differences between the discursive rhythms of weeklies and dailies. It also contributes to the introduction of computational tools into the study of Hebrew historical journalism.

**Keywords:** temporality, topic modeling, Hebrew Journalism, dailies, Nineteenth century, print culture, digital humanities, history of time

> Yet if the present were always present, it would not pass into the past: it would not be time but eternity. If then, in order to be time at all, the present is so made that passes into the past, how can we say that this present also "is"? The cause of its being is that it will cease

to be. So indeed we cannot truly say that time exists except in the sense that it tends toward non-existence.[1]

History is the study of time: locating and understanding events, people, and ideas that existed in a certain time, along a certain timeline. But the concept of time itself is usually taken for granted. In his *Confessions*, St. Augustine says "What then is time? Provided that no one asks me, I know. If I want to explain it to an enquirer, I do not know."[2] Unfortunately, the work of historians is to do exactly that, ask questions: what is time? who's time? and most importantly, "how is it that everyone, in making a choice, constructs their own personal time while still remaining subject to the restraints of social and natural time?"[3] Norbert Elias begins to answer these questions by stating that "timing thus is based on people's capacity for connecting with each other two or more different sequences of continuous changes, one of which serves as a timing standard for the other (or others)."[4] Taking Elias's definition as a guideline, this article explores the history of "timing standards," as they are expressed within a 19th-century Jewish journal.

The modern periodical is an important medium in the construction of time. Its appearance and cycles of production turn artificial time cycles into seemingly natural and accepted social rhythms. Most importantly, periodicals play an important role in the construction of the "present" as a time frame of occurrences that happen "now." As a result of the public simultaneity of newspaper production and newspaper consumption, periodicals create simultaneous social and national times "marked by temporal coincidence, and measured by clock and calendar."[5] However, the reproduced "present" shouldn't be understood independently of the production cycle of the periodical. Time is not "a smooth and standard sequence of ticks of a clock to be lived through, but rather a sequence of whole blocks of time which contain partially predictable and broadly recurring sets of meaningful events. It is around these blocks of time that we construct the cycles which organize social time into smaller and larger temporally embedded structures."[6] Accordingly, this study intends to characterize the differences resulting from the shift in time cycles of the 19th-century Hebrew periodical *HaTzfira*. This periodical started in 1862 as a weekly and was transformed in 1886 into a

---

**1** Saint Augustine, *Confessions*, trans. Henry Chadwick (Oxford: Oxford University Press, 2008), 231.
**2** Augustine, *Confessions*, 230.
**3** Simonetta Tabboni, "The Idea of Social Time in Norbert Elias," *Time & Society* 10, no. 1 (2001): 6.
**4** Norbert Elias, *Time: An Essay* (Oxford: Blackwell, 1992), 72.
**5** Benedict Anderson. *Imagined Communities: Reflections on the Origin and Spread of Nationalism* (London: Verso, 1991), 32–35.
**6** J. David Lewis and Andrew J. Weigert, "The Structures and Meanings of Social Time," *Social Forces* 60, no. 2 (1981): 439.

daily. In order to explore the change, I compare the discourse in the three years prior to the conversion of this weekly into a daily (1883–1885) with the discourse in the three years following this conversion (1886–1888).

Previous academic analysis of journalistic news production in general, and news temporality in particular has relied primarily on human-driven techniques like traditional content analysis.[7] However, as Ryan Cordell argues, the digitization of periodical archives has resulted in a shift in their study.[8] This shift is not just in the increasing accessibility of these journals; it is an actual methodological shift that allows researchers to deal with questions that were previously too difficult to ask. The growing academic field of periodical studies is a direct result of new innovative computational techniques and widespread digitization in the last two decades.[9] Instead of viewing journals as mere conveyors of discrete bits of information, computational tools transformed them to autonomous objects of analysis. They offer a new perspective on large-scale textual corpora which can reveal patterns that traditional modes of analysis typically cannot.

In an attempt to map and characterize the "present" in *HaTzfira*'s discourse, I apply algorithmic topic-modeling analysis, which enables identification of latent themes in the corpus. While applying computational tools to the study of historical journalism has become relatively known and accepted, it is still absent in the study of historical Hebrew journalism.[10] Nevertheless, such application is possible because of recent breakthroughs in the processing of previously digitized historical Hebrew periodicals.[11] This allows us to dramatically upgrade the Optical Character Recognition (OCR) identification rate of these periodicals and, consequently, use digital tools to analyze them.

---

**7** Nikki Usher, *Making News at the New York Times* (Ann Arbor: University of Michigan Press, 2014); M. Neiger, and K. Tenenboim-Weinblatt, "Understanding Journalism through a Nuanced Deconstruction of Temporal Layers in News Narratives," *Journal of Communication* 66, no. 1 (2016): 139–60; B. Zelizer, "Epilogue: Timing the Study of News Temporality," *Journalism* 19, no. 1 (2018): 111–21.
**8** Ryan Cordell, "What Has the Digital Meant to American Periodicals Scholarship?," *American Periodicals: A Journal of History and Criticism* 26 (2016): 2–7.
**9** Dean Latham and Robert Scholes, "The Rise of Periodical Studies," *PMLA* 121 (2006): 517–31; Maria DiCenzo, "Remediating the Past: Doing 'Periodical Studies' in the Digital Era," *ESC: English Studies in Canada* 41 (2015): 19–39.
**10** Zef Segal and Oren Soffer, "One Journal, One Decade, 3,797,592 Words: Computational Analysis of HaTzfira's Discourse (1874–1883)," *Journal of Jewish Studies* 72, no. 2 (2021): 369–96; Zef Segal, "'From One End of the Earth to the Other End of the Earth': Changing Perceptions of the World in Late-Nineteenth-Century Hebrew Journalism," *Jewish Studies Quarterly*, forthcoming.
**11** Oren Soffer et al., "Computational Analysis of Historical Hebrew Newspapers: Proof of Concept," *Zutot – Perspectives on Jewish Culture* 17 (2020): 97–110.

Before exploring the changing concept of "present" in *HaTzfira*, I will provide historical and methodological background. I begin by discussing the circumstances that brought about the specific shift of *HaTzfira* from a weekly publication to a daily one. This is followed by a discussion on our chosen methodology of topic modeling. The second part of the article discusses the topic-modeling analysis of *HaTzfira*'s discourse.

# 1 The Acceleration of the Hebrew Press

*HaTzfira* was founded and published in Warsaw in 1862 by Haim Zelig Slonimski, who also served as its editor. This newspaper and other Hebrew periodicals, such as *HaMaggid* and *HaMelitz*, were part of an extensive network of Jewish journals published not only in Hebrew but in Yiddish and other European languages.[12] Publication of *HaTzfira* ceased after six months and was restarted 12 years later, first in Berlin and later in Warsaw.

The original intent of *HaTzfira*'s founder was to promote scientific and technological knowledge among the observant Jews in Eastern Europe.[13] However, the style and subject matter of *HaTzfira* soon changed. In the renewed *HaTzfira* (from July 1874) an attempt was made to satisfy the readers' interest in Jewish polemics and world politics. From 1874, world news rather than scientific innovations filled the pages of the newspaper.[14] The second, and more significant, change occurred in 1881, following the eruption of pogroms in southwest Russia: world politics lost their dominance in favor of Jewish-related matters, especially those concerned with antisemitism.[15]

Another important milestone in *HaTzfira* occurred in 1886, when it became a daily newspaper. The decision of the editors to accelerate the frequency of publication was a response to similar transformations in other Hebrew periodicals. The first Hebrew daily was *HaYom*, first published in Saint Petersburg on January 31, 1886, which was soon followed by *HaTzfira* (April 13, 1886) and *HaMelitz* (July 12,

---

**12** Israel Bartal, "'Mevaser U-Modi'a Le-Ish Yehudi': Ha-Itonut Ha-Yehudit Be-Afik shel Hidush," *Katedra* 71 (1994): 154–64 [Hebrew]; Israel Bartal, "Mi-'Kahal' Le-Kehilat Kor'im," in *Ein Le-Falpel! Iton Ha-Tzfira ve-Ha-Modernizatzia shel Ha-Si'ach He-Hevrati Ha-Politi*, ed. Oren Soffer (Jerusalem: Mosad Bialik, 2007), ix–xiii [Hebrew]; Oren Soffer, "'Paper Territory': Early Hebrew Journalism and Its Political Roles," *Journalism History* 30 (2004): 31–39.
**13** Oren Soffer, *Ein Le-Falpel! Iton Ha-Tzfira ve-Ha-Modernizatzia shel Ha-Si'ach He-Hevrati Ha-Politi* (Jerusalem: Mosad Bialik, 2007) [Hebrew].
**14** Segal and Soffer, "One Journal, One Decade."
**15** Soffer et al., "Computational Analysis."

1886). The old generation of editors, who established the Hebrew weeklies in the 1850s and 1860s, warned that Hebrew journalism had not reached a stage that could support production of content suitable for a daily publication.[16] Moreover, they assumed that most general news was irrelevant for a Jewish audience and therefore did not see any merit in accelerating the frequency of news production. In contrast, a younger generation of journalists and editors, who began taking a leading role in the 1880s, saw the shift from weekly to daily as an inevitable step toward the modernization and popularization of the Hebrew press. They recognized the rapid growth in Hebrew readership of the 1880s and understood the need of readers for a daily newspaper. This clash between generations was also evident within *HaTzfira*'s editorial board. The 76-year-old establishing editor of the periodical, Slonimski, opposed the attempt to transform *HaTzfira* into a daily, while the 26-year-old Nahum Sokolov saw this transformation as unavoidable. This clash ended in Sokolov's victory. Starting from 1886, Sokolov became an equal partner to Slonimski as owner and editor of the newspaper.

The use of computational tools, and more specifically topic-modeling analysis, helps detect the changes that occurred in *HaTzfira*'s discourse following its shift to daily publication, in terms of both the pattern of the topics discussed and their content. The following section briefly discusses the methodology of topic modeling.

## 2  Topic Modeling

The problem with periodical studies is that "their seriality, abundance, ephemerality, diversity, heterogeneity – posed problems for those who wanted to access their contents."[17] As a result, distant reading approaches, such as topic modeling, that extrapolate backwards from a collection of documents to infer the discourses that could have generated them, have been used by many scholars over the past decade.[18]

Developed in 2003, this generative statistical technique identifies groups of words that tend to occur together in a large collection of documents. It assigns

---

**16** Gideon Kouts, *News and History* (Jerusalem: The Zionist Library, 2013), 81 [Hebrew].

**17** James Mussel, *The Nineteenth-Century Press in the Digital Age* (London: Palgrave Macmillan, 2012), 2.

**18** C. Jacobi, W. Van Atteveldt, and K. Welbers, "Quantitative Analysis of Large Amounts of Journalistic Texts Using Topic Modelling," *Digital Journalism* 4, no. 1 (2016): 89–106; Segal and Soffer, "One Journal, One Decade."

each appearance of a word within a single document to one of a given number of topics. The number of topics is one of the parameters required for the execution of the algorithm; however, there are no generalized guidelines for the optimal number of topics. This choice is usually made following experimentation with various numbers of topics.[19] These are not necessarily "topics" in the sense of a theme, since the algorithm has no knowledge of the actual content or context of the words. A "topic" is merely a pattern of co-occurring words. However, these clusters of co-occurring words are rarely random, and they allow us to infer the latent structure behind a collection of documents.

Once topics are generated and assigned to every appearance of each word, the corpus, its documents, and words used in the corpus become vectors that reflect the distribution of topics within each object. For example, if document A includes 100 words, of which 55 words are affiliated with "topic 1," 45 words with "topic 2," and none with "topic 3," then I could define document A by its distribution vector (55, 45, 0) and compare it with the vectors of other documents. Similarly, if the word "dog" appears 100 times throughout the whole corpus, and 50 of these appearances are affiliated with "topic 1," 30 appearances are affiliated with "topic 2," and 20 appearances are affiliated with "topic 3," then the word "dog" could be defined by its distribution vector (50, 30, 20). Thus, topic modeling can be used to detect changes within the corpus or similarities between documents or words. The technique's main advantage is its ability to switch between different levels of association within a corpus: words, documents, and topics.

As stated previously, the generated topics have no pre-defined meaning. In this study, the meanings of the unsupervised generated topics were identified by examining the semantic relations between the most frequent terms in each topic, as well as reading the journal issues, which were statistically identified as the most reflective of the topic. This allowed a critical evaluation of the computational output in order to make sure the results of the algorithm were not meaningless or arbitrary. In addition, the results were evaluated against existing qualitative scholarship.

The corpus consists of six years of *HaTzfira*'s discourse (1883–1888), which includes three years prior to the change from weekly to daily and three years after that change. On the one hand, this corpus is broad enough to show long-term changes and pursue meaningful computational analysis; on the other hand, it is focused enough to avoid discursive influences of significant historical events, such as the 1881 pogroms or the 1897 First Zionist Congress. I initiated three different

---

**19** H.M. Wallach, D.M. Mimno, and A. McCallum, "Rethinking LDA: Why Priors Matter," *Advances in Neural Information Processing Systems* 22 (2009): 1973–81.

topic models on the corpus, each designed to produce 20 topics and in which the units of analysis were individual issues of the journal. The first was conducted on the whole corpus (860 issues), thus enabling a broader overview and a detection of changes and patterns over the full period (see Section 3). The second and third models were conducted on sub-corpuses: 150 issues published 1883–1885, and 710 issues published 1886–1888 (see Section 4). These two models allow us to characterize the discourse in each period and compare patterns and topics before and after the change from daily to weekly.

# 3  The "Longue Durée": 1883–1888

Six years cannot be considered what the French historian Fernard Braudel defines as a longue durée, but the concept suits the aim of the first topic model. Braudel describes a history that opposes daily events, that is not episodic in nature, one which "requires getting to know slower temporalities, almost immobile ones."[20] Accordingly, before identifying and exploring the characteristics of the weekly "present" and the daily "present," I look at a larger picture, a corpus consisting of both periodicities at once. At first, I use vector analysis to compare topic distribution in each six-month period during the time span. The choice to group together issues published within six months ensures the readability of the data. Each six months is understood as a 20-dimensional vector, in which each coordinate denotes the portion of the relative topic within that period's discussions. Principal component analysis (PCA) reduces the 20-dimensional space into a two-dimensional visualization, in which the choice of axes optimizes the variation between the topics.[21] Figure 1 shows the differences between the periods in the relevant time span.

Although a longue durée approach concerns itself with continuity and gradual evolution, the distribution along the x-axis in this figure clearly differentiates between two groups of points. The left group consists of all the periods after the transformation, when the journal was published as a daily, and the right consists of all the periods before the conversion, when the journal was published as a weekly. As well as being far apart from each other, the distribution along the y-axis of both groups reveals entirely different patterns. The weekly group of points is

---

**20** Fernand Braudel, "Histoire et sciences sociales: La longue durée," *Annales* 13, no. 4 (1958): 725–53.

**21** Christof Schöch, "Principal Component Analysis for Literary Genre Stylistics," *The Dragonfly's Gaze*, last updated September 8, 2016, accessed January 23, 2020, http:// dragonfly.hypotheses.org/472.
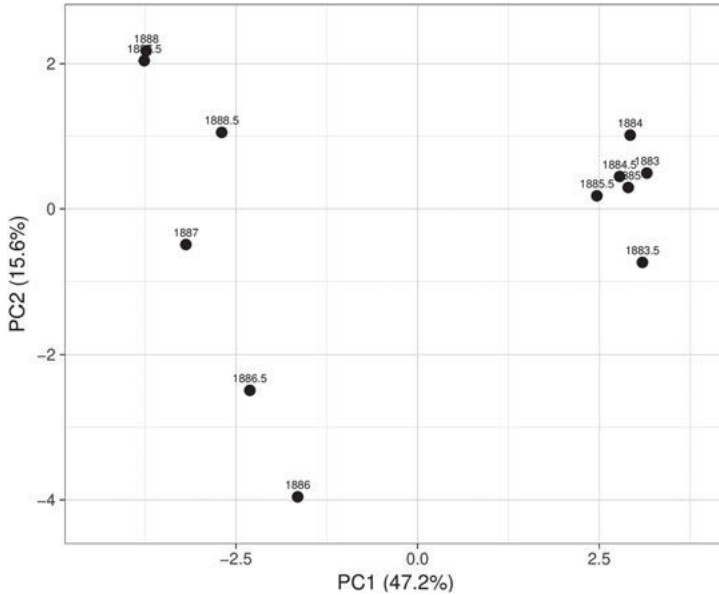
**Figure 1:** Principal component analysis (PCA) of topic distribution, 1883–1888. Each axis reflects a single dimension of the vectors as identified by the algorithm. The percentage listed on each axis reflects the proportion of variance that the specific axis reveals. The distinctions between left and right, between up and down, as well as the values listed along each axis have no significance in themselves. They show the differences between vectors. Visualization is done with ClustVis.

heavily clustered with seemingly no chronological rationale behind its distribution. In contrast, the daily group is aligned chronologically from bottom to top with a relatively uniform distribution of points, with the exception of the second half of 1887. The different characteristics of these two groups reflect a rupture, almost as if the analysis had shifted to a different newspaper. Considering the relative continuity in the internal and external circumstances surrounding the journal, we can assume that the change from weekly to daily caused this rupture.

However, some form of continuity reveals itself when examining the temporal distributions of each topic. In general, we can distinguish between two types of distributions: (i) temporary topics that vary significantly over time and (ii) constant topics. Figure 2 shows the ten constant topics during the whole time span. Although these topics are characterized by continuity, we can still see that 1886 was a turning point. For some it meant a steep decline and for others a sharp incline.
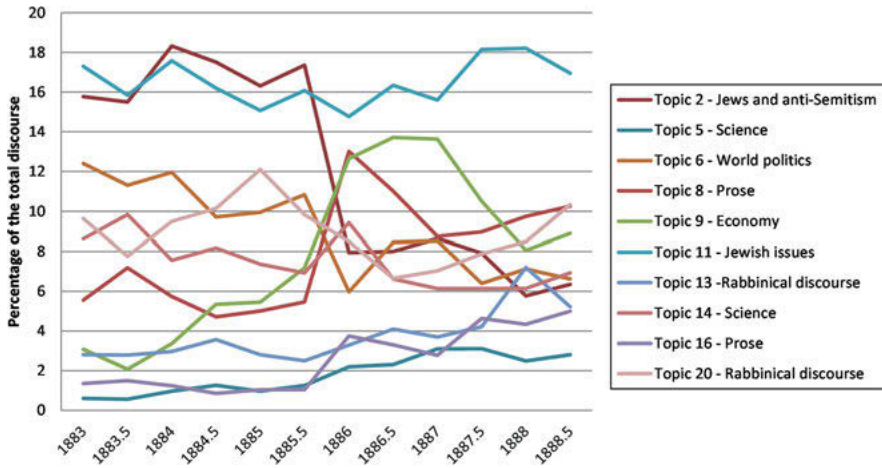
**Figure 2:** The temporal shift in the distribution of constant topics, 1883–1888. Each line signifies a single topic, listed in the legend. The value at each point marks the percentage of all the content during the relevant period.

Four topics, making up 36% of all the discussions between 1883 and 1888, remained relatively constant throughout the entire time span: topics 11, 13, and 20 are related to Jewish and rabbinical discourse and topic 14 is related to natural sciences. The consistency of these topics despite the change in publication frequency can be explained by the fact that they formed the backbone of *HaTzfira* – as a 19th-century Jewish periodical structured on a heritage of Jewish Enlightenment. Beetham discusses three different times that appear in a periodical: "monumental," "masculine," and "feminine" times.[22] "Monumental" time includes large historical forces; "masculine" time is "linear time, or the time of history and politics, the world of production associated with men"; and "feminine" time is "the time of reproduction, characterised as repetitive and circular rather than linear and progressive." The four topics can be considered the "monumental time," existing beyond the realms of frequency of publication.

The next section provides an in-depth analysis of each period by creating two different topic models, one for the period 1883–1885 (the weekly era), and another for the period 1886–1888 (the daily era). Rather than looking for similarities, this section will identify the changes.

---

**22** Margaret Beetham, "Time: Periodicals and the Time of the Now," *Victorian Periodicals Review* 48, no. 3 (2015): 332.

# 4 Changing Temporalities

As mentioned previously, some of the topics were constant while others reflect the temporal and changing nature of journalistic discourse. Figure 3 depicts the constant topics in the first period of time, during which the journal was published as a weekly, and Figure 4 depicts the constant topics in the daily era (for further information and an explanation of the topic numbers see Tables 1 and 2). What stands out in the comparison between both graphs is the rise in number of constant topics, from eight in the weekly era to eleven in the daily era, and their share of all the discussions in the journal, from 70% in the weekly era to 80% in the daily era.

The constant topics mainly relate to the recurring themes of a Hebrew journal of the late 19th century: the general paratext of the journal is affiliated with topic 14 in the weekly era and topic 8 in the daily era; recurring news regarding German politics and Jewish affairs are affiliated with topic 1 in the weekly era and topics 1 and 4 in the daily era; religious discussions appear as part of topic 6 in the weekly era and topics 13 and 19 in the daily era; and the heated debates on the fate of Jews in Europe appears in topics 5, 16, and 17 in the weekly era and topics 8 and 16 in the daily era. Topic 20 in the weekly era and 17 in the daily era, with
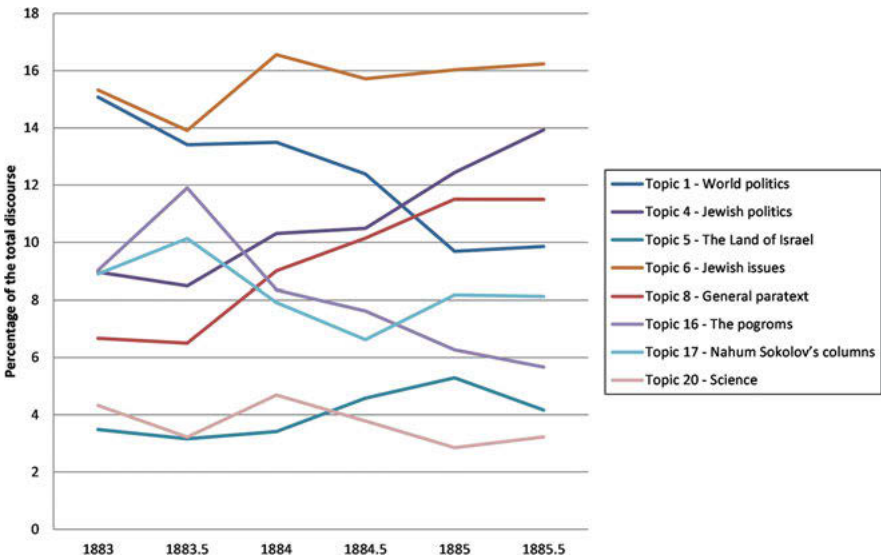


**Figure 3:** The temporal shift in the distribution of constant topics, 1883–1885. Each line signifies a single topic, listed in the legend. The value at each point marks the percentage of all the content during the relevant period.
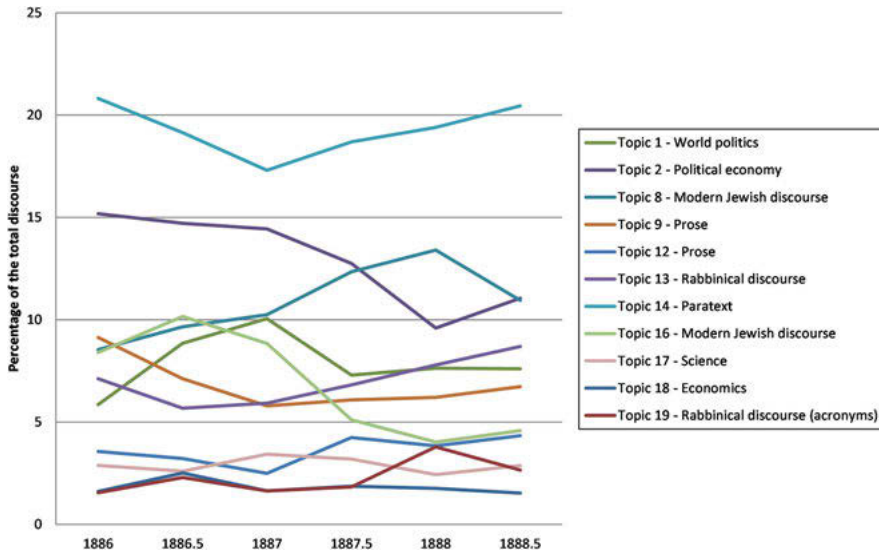
**Figure 4:** The temporal shift in the distribution of constant topics, 1886–1888. Each line signifies a single topic, listed in the legend. The value at each point marks the percentage of all the content during the relevant period.

terms such as "air," "blood," and "medicine," are unique to *HaTzfira* and reflect the initial scientific motivation for the publication of the journal. Religion and science were mentioned in the previous section as monumental themes, but the other six topics would define the journal's "feminine time": its repetitive, constant, and reproduced items.

The aforementioned themes were similarly relevant in both eras, but the daily introduced two new themes: literary content (topics 9 and 12) with terms such as "my heart," "his eyes," and "my soul"; and economics (topics 2 and 18), with terms such as "ruble," "liter," "the price," and "the trade."

Literary content entered as a separate section of the journal in the third installment of the daily journal on April 15, 1886, as Slonimski had earlier forbidden publishing such content.[23] The first story entitled "Yerushalima" (To Jerusalem), described the longing of a dying European Jew for the land of Israel.[24] The topics related to the literary theme reflect two types of content. Topic 9 is related to published prose[25] while topic 12, with its emphasis on verbs in the first person,

---

23 *HaTzfira*, February 4, 1862, 8.
24 Dober Rabinowitz, "Yerushalima," *HaTzfira*, April 15, 1886, 2–3.
25 Abraham Zuckerman, "Eshet Khayil," *HaTzfira*, June 21, 1886, 2–4.

reflects personal diaries and poems.[26] The integration of the feuilleton and literary content on a daily basis helped meet the challenge of producing enough content for a daily, as it could be prepared with no connection to current events.

The major change in content was not literary material but rather economics-oriented texts. While the economy was rarely discussed during the weekly era, it occupied 50% of the journal's content from 1886 onwards. As we will see in the analysis of the changing topics, the economy became the main feature of the daily journal. Two topics related to economic discourse appeared constantly throughout the daily era. Topic 2 connects financial terms such as "money," "trade," and different currencies with political terms such as "government." This topic relates to generalized issues of political economy. For example, a particular issue of *HaTzfira*,[27] of which 33% was affiliated with topic 2, included an article on military expenses as its main international news segment, an article on taxes as its main internal news segment, and an article on finances of Jewish philanthropy as its main segment on Jewish affairs. Topic 18, on the other hand, includes commercial terms such as "price" and "rate" and different commodities such as "sugar" and "flour." This topic reflects various commercial indices that were published on the back page of the newspaper.

In contrast to the relative similarity between the content of the constant topics in both eras, the temporary topics signify a real change in discourse. Figure 5 depicts the temporary topics in the weekly era, and Figure 6 depicts the temporary topics in the weekly era.

In general, the weekly era can be characterized by an ongoing discussion of European Jewish affairs in connection with world politics. In fact, there is no topic that is not connected in some ways to the Jewish people.[28] During the first year (1883), topics 3, 7, 11, and 18 consist of various issues around libels and rising antisemitism. This follows the pogroms of 1881–1883 in south Russia, which had a tremendous effect on Eastern European Jewry. During the second year (1884), topics 9, 10, 13, and 19 partially reflect a search for a solution for the Jewish people; thus, for example, topic 10 connects Moses Montefiore, the Jewish philanthropist, with the British colonial regime. During the third year (1885), topics 2, 12, 14, and 15 are not particularly identifiable but are still connected to general Jewish politics and issues.

---

**26** Israel Saba, "Amarti yesh li Tikva," *HaTzfira*, September 29, 1886, 2–3; "Berosh Homiyot," *HaTzfira*, September 13, 1888, 2.
**27** *HaTzfira*, September 27, 1886.
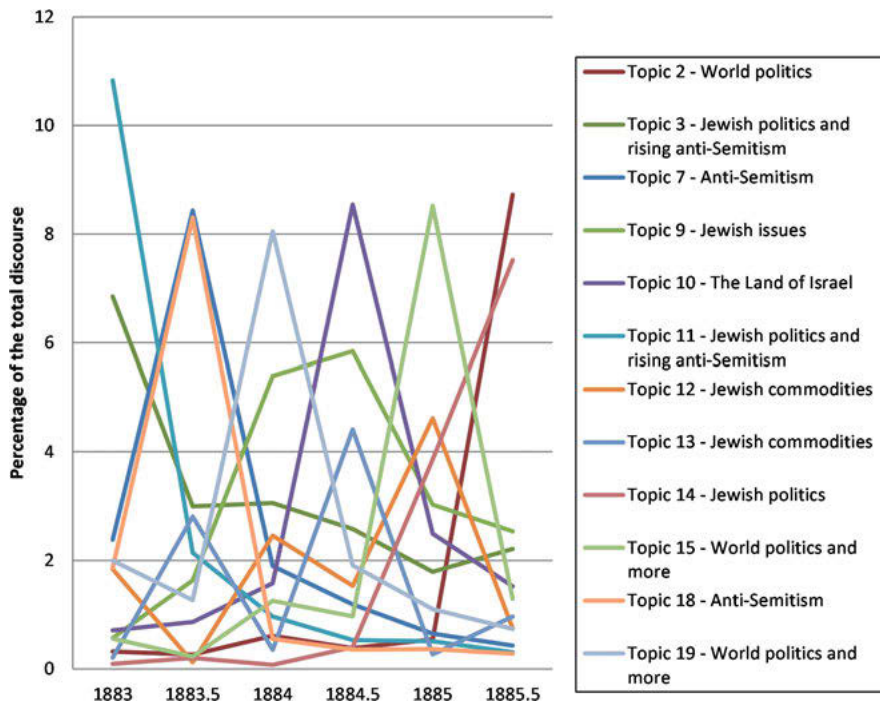**28** Segal and Soffer, "One Journal, One Decade."

**Figure 5:** The temporal shift in the distribution of temporary topics, 1883–1885. The value at each point marks the percentage of all the content during the relevant period.

What is even more striking is the wave-like pattern of the graph in Figure 5, as the rise of each journalistic theme or topic heralds the decline of another. Furthermore, the main transitional topics of *HaTzfira* reach similar peaks at approximately 8% of the journal's discourse during a single half-year period before being superseded by another topic. This temporal pattern echoes Franco Moretti's conceptualization of shifts in literary genres. In *Graphs, Maps, Trees*, Moretti refers to the temporal and ephemeral nature of genres: "the new form makes its appearance to replace an old form that has outlived its artistic usefulness …, and the decline of a ruling genre seems indeed here to be the necessary precondition for its successor's takeoff."[29] Moretti describes the life cycles of genres as waves, in which "a rather regular changing of the guard takes place, where half a dozen genres quickly leave the scene, as many move in."[30] In terms

---

**29** Franco Moretti, *Graphs, Maps, Trees* (London: Verso, 2005), 14.
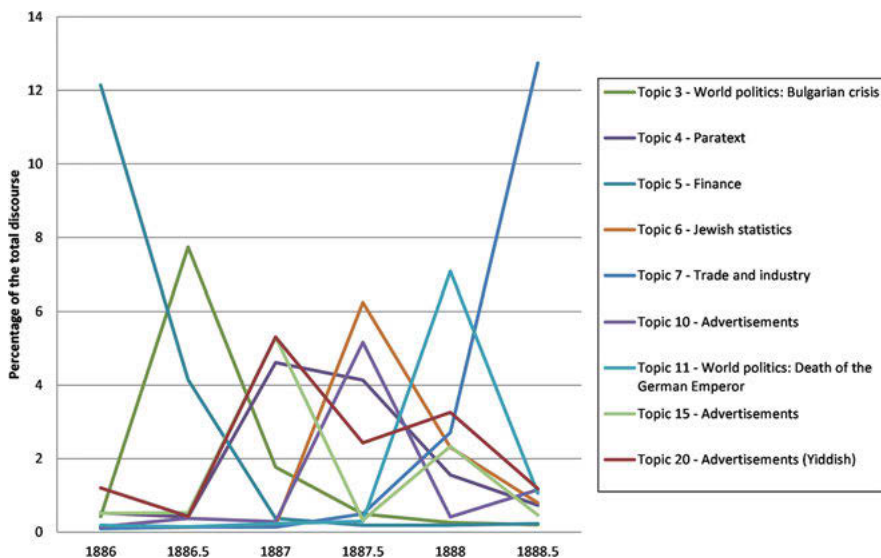**30** Moretti, *Graphs, Maps, Trees*, 18.

**Figure 6:** The temporal shift in the distribution of temporary topics, 1886–1888. Each line signifies a single topic, listed in the legend. The value at each point marks the percentage of all the content during the relevant period.

of time, the "present" of the weekly tended to operate like a metronome: some topics were continuous and constant, while others consistently appeared and disappeared.

Unlike the wave-like and uniform pattern of news cycles during the weekly era, Figure 6 reflects a chaotic change of news items in the daily era. Some topics have several peaks, others have relatively long decay periods, and many of them overlap. In addition, the peak levels are different for each topic.

The change from weekly to daily is not only apparent in the pace and rhythm of news production but also in the actual content. Of the nine changing topics, depicted in Figure 6, only two are not related to the economy (topics 3 and 11).[31] This recurring theme reflects a growing commercialization of the Hebrew journal, which was part of the capitalistic commodification of time within the commercial project of periodical print. As stated by Sommerville, "the adoption of weekly and then daily schedules for marketing information began developments, the implications of which are still being worked out. Periodicity allowed information to become a business, where it had once been a part of personal

---

[31] Topic 3 discusses the Bulgarian crisis of 1885–1888 and topic 11 is related to the reports from Berlin regarding the death of the German emperor Wilhelm I on March 11, 1888.

relations."[32] This process of commodification escalated in the daily newspaper. As the frequency of production increased, more and more space was devoted to advertisements. The first daily newspaper, the *Daily Courant* (1702–1735), for example, devoted about one-half of its space to advertisements.[33] Similarly, many of the topics in *HaTzfira*'s daily discourse relate to straightforward journalistic finances. Topic 4, for example, includes financial terms from the paratext of *HaTzfira*'s front page, emphasizing the commodification of news and the newspaper as the agent providing these news items. Topics 10, 15, and 20 relate to the advertisements in *HaTzfira*. Topics 10 and 15 are connected to commodities for Jewish holidays, such as fruit from Egypt, Passover flowers, and wine.[34] Topic 20 is connected to advertisements in Yiddish, which appeared sporadically throughout the period.[35]

As can be seen, financial issues soon entered into the journalistic texts (topics 5, 6 and 7). Topic 5, peaking at the beginning of 1886 – with terms such as "guarantee," "loan," and "departments," together with mention of various European currencies – identifies an early trend immediately following the change into daily publication.[36] During this stage, economic texts were separated from other parts of the newspaper, located in the back pages. The full integration of financial discourse within other parts of the newspaper occurred in 1887. Topic 6, for example, reflects the influence of financial reasoning on Jewish polemics through the introduction of statistics and primarily economic statistics.[37] The paper did this by encouraging readers to send figures and data concerning the finances of their own communities in answer to antisemitic accusations of the economic inefficiency of the Jewish people.[38] Topic 7 reflects a broader and more integrated influence of economic discourse on the journal. Articles related to this topic combine industrial terms such as "factory," financial terms such as "guarantee," and political terms such as "emperor" and "Jews."

---

**32** Charles J. Sommerville, *The News Revolution in England: Cultural Dynamics of Daily Information* (Oxford: Oxford University Press, 1996), 161.
**33** Will Slauter, "The Rise of the Newspaper," in *Making News: The Political Economy of Journalism in Britain and America from the Glorious Revolution to the Internet*, ed. Richard R. John and Jonathan Silberstein-Loeb (Oxford: Oxford University Press, 2015), 31.
**34** *HaTzfira*, August 19, 1887, 4; *HaTzfira*, January 23, 1888, 4.
**35** *HaTzfira*, April 1, 1887, 6–7.
**36** *HaTzfira*, May 13, 1886, 4.
**37** "Ma-Yif'al Israel," *HaTzfira*, November 28, 1887, 2–3.
**38** Soffer, "'Paper Territory.'"

**Table 1:** The ten prominent terms in each topic after removing stop words, and a defining title based on the full bank of terms, 1883–1885.

| | Title | |
|---|---|---|
| 1 | World politics | the Jews, Germany, government, Bismarck, England, [written] by, newspapers, Germans, in Germany, France |
| 2 | World politics | newspapers, Bulgarians, natives, Alexander, students, the sick, iron, Kopeks, Jews, Austria |
| 3 | Jewish politics and rising antisemitism | the emperor, Israel, the minister, France, the French, Russia, the war, the prayer, ruble, in honor of |
| 4 | Jewish politics | the Jews, the government, Jews, Russia, Jews [of], newspapers, the Jew, to the Jews, government [of], the Christians |
| 5 | The Land of Israel | God, Jerusalem, Zion, our brothers, settlement, in Jerusalem, the Land of Israel, the Ga'on, to God, houses |
| 6 | Jewish issues | Israel, God, trade, in our Land, religion, the charity, our brothers, the people of Israel, in the lands of, 20, the money |
| 7 | Antisemitism | Israel, Stoecker, the minister, Montefiore, China, France, the king, the freedom, libel, army |
| 8 | General paratext | God, ruble, rabbi, aforementioned, 10, books, in the language of, Torah, franc, of blessed memory |
| 9 | Jewish issues | bad, the Talmud, the republic, million, kopeks, doctor, the Jews, the rabbis, the order, the movement |
| 10 | The Land of Israel | the minister, Montefiore, the parliament, the colonization, the English, the council, the Land of Israel, Solomon, Germany, synagogue |
| 11 | Jewish politics and rising antisemitism | Israel, aforementioned, the government, Austria, the Christians, the elected, France, the Talmud, the bitter enemy, The Jew |
| 12 | Jewish commodities | Passover, the kosher, flour, the Passover, the enlightenment, Italy, the year, Gordon, the nature |
| 13 | Jewish commodities | citrons, myrtles, France, China, families, equal, the French, trade, purchase, Egypt |
| 14 | Jewish politics | the minister, Stoecker, to us, Salisbury, the ministry, Rothschild, the lord, in England, citrons, the Holy Land |
| 15 | World politics and more | England, the elected, Russia, 25, mit [Yiddish], 50, Gladstone, 12, 16, 10 |
| 16 | The pogroms | the Jews, the government, the army, newspapers, the riots, the mob, fire, the trade, Europe, the end |
| 17 | Nahum Sokolov's columns | the English, Russia, war, the address, army, the truth, sea, letter, the holy, and us |
| 18 | Antisemitism | the law, the trial, the judges, the blood, the boy, the libel, the witnesses, Esther, libel, Moritz |
| 19 | World politics and more | Israel, the parliament, Bismarck, the laws, cult, Lasker, the community, Egypt, the ministry, the elected [of] |
| 20 | Science | [written] by, the nature, the sun, ruble, the price [of], water, between them, the sky, 50, Austria |

**Table 2:** The ten prominent terms in each topic after removing stop words, and a defining title based on the full bank of terms, 1886–1888.

| | Title | |
|---|---|---|
| 1 | World politics | Bismarck, the army, newspaper, France, Russia, the government, government of, Ashkenaz, Europe, God |
| 2 | Political economy | the government, ruble, the trade, the value, kopek, the money, franc, trade, Warsaw, in our land |
| 3 | World politics | Bulgaria, England, December, temporal, the army, October, newspaper, government, Russia, the government |
| 4 | Paratext | 100, June, the Jews, May, 50, publication, 10, the loan, demanded, abroad |
| 5 | Finance | 100, May, 40, journey, July, to God, June, guarantee, Greece, one hundred |
| 6 | Jewish statistics | providers, 10, November, December, October, Israel, Berlin, Paris, traders, 15 |
| 7 | Trade and industry | 10, the emperor, July, 100, Wilhelm, June, August, God, 50, Shimon, Factory |
| 8 | Modern Jewish discourse | Israel, our brothers, the readers, the authors, see, in *HaTzfira*, we are, the enlightenment, the religion, articles |
| 9 | Prose | his eyes, the woman, loud, answered, called, woman, rabbi, will say, stand, sit |
| 10 | Advertisements | August, September, Israel, July, p. (acronym), 10, Bulgaria, pomegranates [of], pomegranates, number, the Jews |
| 11 | World politics | the emperor, April, 10, Berlin, May, Wilhelm, Bismarck, Friedrich, Boulanger, newspaper |
| 12 | Prose | my heart, I knew, my soul, my master, in me, I said, I saw, my brother, and I, I will be able to |
| 13 | Rabbinical discourse | God, Israel, rabbi, the congregation, the Torah, our brothers, ruble, to God, the deceased, Yitzhak |
| 14 | Paratext | kopek, 15, g' (acronym), ruble, 24, florin, saw, [written] by, fire, the time |
| 15 | Advertisements | God, February, January, Passover, rabbi, 10, Berlin, Warsaw, the Passover, flour |
| 16 | Modern Jewish discourse | the Jews, God, Israel, our brothers, the government, newspapers, about, doctor, the Christians, Jew |
| 17 | Science | by, the body, the air, the disease, the blood, the medicine, occasionally, physicians, flesh, disease [of] |
| 18 | Economics | ruble, liter, the price, the sugar, kopek, this week, 50, the traders, pood, the rate |
| 19 | Rabbinical discourse (acronyms) | vcu' [et cetera], z"l [in blessed memory], s', b', but, g', a', issue, that is, c"a [each and every one] |
| 20 | Advertisements (Yiddish) | mit, auch, fuer, im, dem, ein, werden, Fabrik, wie, Freisen |

# 5 The Editors' Perspective on Periodical Time

At the core of the previous analysis lies the question of periodical time, and more specifically its defining role in conceptualizing the present. The editors of *HaTzfira* were well aware that the shift from weekly production to daily production resulted in the need to accelerate news presentation to the readers, which they did express in short quarterly bulletins to the readers, published on the front page. These bulletins acted as policy declarations on behalf of the editors, expressing their obligation to shortening the time of news production and newspaper delivery. The summarizing bulletin of 1886,[39] for example, states that the paper will provide "current news and perceptions, brought by well-known writers that expediate all their reports to our readers *as soon as they can*" (emphasis in the original). In addition, it acknowledges the need to fill the pages of the newspapers, but at the same time expresses the obligation to provide "valuable content in each issue."[40]

The need for a more rapid cycle of news did not allow the editors to maintain the same themes discussed in the weekly era. The most dynamic and fluctuating content available to the editorial board was economic information. Data on commercial prices, global currencies, and stock market indices changed on a daily basis. Although it is hard to believe that most Jewish readers had any need for such information, it structured their time. The transformation from Jewish politics (the main theme in the weekly era) to economics (the main theme in the daily era) was initially a structural change, due to the introduction of more advertisements and specialized economic and financial columns in the back pages. Over time, economic discourse found its way into the general journalistic discussion.

While topic modeling identifies the growing role of economics within the discourse of *HaTzfira*, the quarterly bulletins show very little sign of this change. The bulletins, which reflect the editors' overview of their work, emphasize continuity from earlier periods of the journal, stressing its original scientific and political orientation. The "dailiness" of the journal is seen mainly in the editors' success in maintaining its traditional nature on a daily basis. However, the editors do reflect on a particular change in content, while never mentioning its novelty – that is, the introduction of literary columns within the daily *HaTzfira*. By September 1887,[41] the editors define the journal's mission as being "a political, scientific

---

**39** *HaTzfira*, December 28, 1886, 1.
**40** *HaTzfira*, July 1, 1887, 1.
**41** *HaTzfira*, September 30, 1887, 1.

and *literary* newspaper" (emphasis not in the original). They seem unaware of the new economic hegemony.

The single exception is the quarterly bulletin of March 30, 1887.[42] In it, the editors explicitly acknowledge the important role of "reports on the world of commerce and news related to daily work of the people of Israel." In addition, they emphasize the value of advertisements, "especially for those dealing with the people of this great metropolis [Warsaw]." This unique expression of the commercialization of the newspaper and its dependence on urbanized readership corresponds with our analysis based on computational distant reading. Although this was not repeated in previous or later bulletins, it reveals the editors' underlying awareness of these new circumstances and reality.

# 6 From "Masculine Time" to "Feminine Time"

January 5, 1886, was the date of publication of the last issue of the weekly *HaTzfira*. Three months later, on 13 April, 1886, the new daily *HaTzfira* was published, with very little apparent change. The subtitle of the newspaper changed from "published weekly" to "published daily," and the length of the newspaper was shortened from eight pages to four. Due to the fact that the editorial board, consisting of Slonimski and Sokolov, never changed, and that nothing much occurred in world affairs during this time of transition, we could assume that the journals would remain almost identical. However, this study shows that the transformation from weekly to daily had far reaching consequences on the rhythm and content of periodical times. The findings of the topic-modeling algorithm, as can be seen in the vector analysis (Section 3), provide evidence that the shift in the cycle of publication had a substantial effect on the social discourse presented in this periodical. The computational analysis shows that the difference between the two discourses was almost as great as that between two different newspapers.

The differences were related not only to the content of the topics but also to their structure, cycles, and duration. While in the weekly era we see a wavelike pattern of similarly important topics replacing each other, in the daily era we see a much more chaotic nature of changing topics as well as a larger number of unvarying topics.

---

**42** *HaTzfira*, March 30, 1887, 1.

As Beetham argues, there is always a dependency between time and money within the journalistic discourse.[43] In the daily, this dependency was radicalized, as information became a business.[44] In *HaTzfira* we see several layers of this commodification. The first layer relates to the new social function of the newspaper as the provider of information rather than reading material. In order for the daily to be relevant and commercially valuable to the readers, this information had to be printed before the readers received it through other means. The second layer is an outcome of the first layer because the most available, updated, and dynamic type of information the journal can publish is commercial and economic information. As a result, the orientation of the journal's discourse gradually shifted to economics. The third layer of the commodification is the growing proportion of advertisements in the newspaper. This was an outcome of the combination of several factors: the rising costs of daily publication required new sources of funding; the daily appearance of the newspaper created an added value for the advertisers, as the newspaper provided ongoing exposure to the readership; and the shift of the discourse to economics created a suitable environment for the commercialization of the printed space.

Frequency affected not only news cycles and journalistic content, but also the genre of journalism. The new social and commercial function of the daily newspaper was different from the distant and interpretive role of the weekly. While the Jewish weekly was often perceived as a periodical book that could be read years after the publication, the daily was designed to affect the immediate, everyday environment of the readers. The dailies were designed to grasp the attention of a daily commute to work.[45] Thomas P. O'Connor, a 19th-century journalist, wrote in 1889 that "we live in an age of hurry and of multitudinous newspapers. The newspaper is not read in the secrecy and silence of the closet as is the book. It is picked up at a railway station, hurried over in a railway carriage, dropped incontinently when read."[46] Rather than the summary and evaluation of distant events provided in the weekly era, the daily dealt with the everyday life and needs of the readers. This was manifested in a shift from politics and Jewish discourse to economics, as well as the dominant role of advertisements.

Another perspective of the fulfillment of everyday requirements was the introduction of literary columns. Similar to economic discourse, literary dis-

---

**43** Margaret Beetham, "Towards a Theory of the Periodical as a Publishing Genre," in *Investigating Victorian Journalism*, ed. Laurel Brake, Aled Jones, and Lionel Madden (London: Palgrave Macmillan, 1990), 19–32.

**44** Sommerville, *The News Revolution in England*.

**45** Beetham, "Time."

**46** Thomas P. O'Connor, "The New Journalism," *The New Review* 1 (October 1889): 434.

course was introduced partly as a result of institutional circumstances, since they were available and could be prepared in advance to fill the newspaper. However, such content related to the new personal perspective of the daily newspaper. As reflected in the prominent words of the relevant topics (9 and 12), the vocabulary was personal and related to mundane personal domains. Such style was entirely different from the dominant scientific style of the weekly era. This could reflect a much-needed balance within the daily newspaper between an instrumentalized economic style and a personal literary one.

As Turner observes, varying frequencies of journal publication create divergent time cycles that relate differently to everyday lives.[47] As we have seen, the weekly era was characterized by a metronome-like change of topics. Relatively few topics were constant, while others superseded each other in a uniform wave-like pattern. In contrast, the daily era was characterized by a larger number of constant topics while the dynamic topics reflect chaotic change. This can be explained by the need to create a sense of continuity within a daily cycle of journalistic production. The daily cycle's lack of journalistic perspective forced the creation of stabilized indexing within an unpredictable daily atmosphere. This indexing is manifested in the constant topics, which reflect Beetham's "feminine time."[48] This finding is not intuitive and stands in contradiction to Sommerville's assumption that constant change is reflected within daily production of journals.[49] While the weekly perspective enabled the production of uniform change, the rapid information provided in the daily was structured as if it was uniform continuity. This structural uniformity allowed the daily newspaper to contain relatively few changing topics, which reflect Beetham's "masculine time."

The daily *HaTzfira* was not restricted to a single sense of periodical time; it encompassed monumental, feminine, and masculine times simultaneously. The need to discipline the rapid change of information created structured journalistic formats with identified spaces within the printed topography, which elevated "repetitive and circular" representation of "linear and progressive" information. In this way, it challenged the binary distinction between feminine and masculine times. Periodicals in general, and dailies in particular, were significant in the construction of the "present." However, the meaning of this "present" was rarely, if ever, constant and coherent.

The case-study of *HaTzfira* manifests the advantages of computational tools and methods in juxtaposing two areas of interest in current Jewish Studies, the

---

**47** Mark W. Turner, "Periodical Time in the Nineteenth Century," *Media History* 8, no. 2 (2002): 183–96.
**48** Beetham, "Time."
**49** Sommerville, *The News Revolution in England*.

reliance of Jewish societies on communication networks and the significance of time in Jewish tradition.

Due to the diasporic nature of the Jewish society, communication networks assumed an exceptional significance: they were the means not only of creating an "imagined community" but a central element in maintaining the actual, real, community. "The dispersed, centerless Jewish world remained connected and in many ways intact into the modern era, based on innovative and effective media arrangements and communication strategies," claims Blondheim.[50] The importance of communication has resulted in extensive research on Jewish media in the Modern Age. Those same circumstances caused Abraham Joshua Heschel to distinguish between the "space-minded man" and the Jew and claim that Judaism is a "religion of time aiming at the sanctification of time."[51] Consequently, Jewish Studies in the last two decades has seen a rising interest in time-related research, and in particular diverse forms of temporality in Jewish culture.[52]

However, despite the fact that a lot of the work in both areas of interest has dealt with new media and digital technology, their analysis remains mostly qualitative and non-digital. Computational analysis provides new perspectives on each of these fields individually and together.

# References

Anderson, Benedict. *Imagined Communities: Reflections on the Origin and Spread of Nationalism*. London: Verso, 1991.

Augustine, Saint. *Confessions*. Translated by Henry Chadwick. Oxford: Oxford University Press, 2008.

Bartal, Israel. "'Mevaser U-Modi'a Le-Ish Yehudi': Ha-Itonut Ha-Yehudit Be-Afik shel Hidush." *Katedra* 71 (1994): 154–64. [Hebrew]

Bartal, Israel. "Mi-'Kahal' Le-Kehilat Kor'im." In *Ein Le-Falpel! Iton Ha-Tzfira ve-Ha-Modernizatzia shel Ha-Si'ach He-Hevrati Ha-Politi*, edited by Oren Soffer, ix–xiii. Jerusalem: Mosad Bialik, 2007. [Hebrew]

Beetham, Margaret. "Time: Periodicals and the Time of the Now." *Victorian Periodicals Review* 48, no. 3 (2015): 323–42.

---

**50** Menahem Blondheim, "The Jewish Communication Tradition and Its Encounters with (the) New Media," in *Digital Judaism: Jewish Negotiations with Digital Media and Culture*, ed. Heidi Campbell (New York: Routledge, 2015), 28.

**51** Abraham J. Heschel, *The Sabbath: Its Meaning for Modern Man* (New York: Farrar, Straus and Giroux, 1951), 7.

**52** Sarit K. Gribetz and Lynn Kaye, "The Temporal Turn in Ancient Judaism and Jewish Studies," *Currents in Biblical Research* 17, no. 3 (2019): 332–95.

Beetham, Margaret. "Towards a Theory of the Periodical as a Publishing Genre."
    In *Investigating Victorian Journalism*, edited by Laurel Brake, Aled Jones, and Lionel
    Madden, 19–32. London: Palgrave Macmillan, 1990.
"Berosh Homiyot." *HaTzfira*, September 13, 1888, 2. [Hebrew]
Blondheim, Menahem. "The Jewish Communication Tradition and Its Encounters with (the) New
    Media." In *Digital Judaism: Jewish Negotiations with Digital Media and Culture*, edited by
    Heidi Campbell, 24–47. New York: Routledge, 2015.
Braudel, Fernand. "Histoire et sciences sociales: La longue durée." *Annales* 13, no. 4 (1958):
    725–53.
Cordell, Ryan. "What Has the Digital Meant to American Periodicals Scholarship?" *American
    Periodicals: A Journal of History and Criticism* 26 (2016): 2–7.
DiCenzo, Maria. "Remediating the Past: Doing 'Periodical Studies' in the Digital Era." *ESC:
    English Studies in Canada* 41 (2015): 19–39.
Elias, Norbert. *Time: An Essay*. Oxford: Blackwell, 1992.
Gribetz, Sarit K. and Lynn Kaye. "The Temporal Turn in Ancient Judaism and Jewish Studies."
    *Currents in Biblical Research* 17, no. 3 (2019): 332–95.
*HaTzfira*. February 4, 1862, 8. [Hebrew]
*HaTzfira*. May 13, 1886, 4. [Hebrew]
*HaTzfira*. September 27, 1886. [Hebrew]
*HaTzfira*. December 28, 1886, 1. [Hebrew]
*HaTzfira*. March 30, 1887, 1. [Hebrew]
*HaTzfira*. April 1, 1887, 6–7. [Hebrew]
*HaTzfira*. July 1, 1887, 1. [Hebrew]
*HaTzfira*. August 19, 1887, 4. [Hebrew]
*HaTzfira*. September 30, 1887, 1. [Hebrew]
*HaTzfira*. January 23, 1888, 4. [Hebrew]
Heschel, Abraham J. *The Sabbath: Its Meaning for Modern Man*. New York: Farrar, Straus and
    Giroux, 1951.
Jacobi, C., W. Van Atteveldt, and K. Welbers. "Quantitative Analysis of Large Amounts of
    Journalistic Texts Using Topic Modelling." *Digital Journalism* 4, no. 1 (2016): 89–106.
Kouts, Gideon. *News and History*. Jerusalem: The Zionist Library, 2013. [Hebrew]
Latham, Dean and Robert Scholes. "The Rise of Periodical Studies." *PMLA* 121 (2006): 517–31.
Lewis, J. David and Andrew J. Weigert. "The Structures and Meanings of Social Time." *Social
    Forces* 60, no. 2 (1981): 432–62.
"Ma-Yif'al Israel." *HaTzfira*, November 28, 1887, 2–3. [Hebrew]
Moretti, Franco. *Graphs, Maps, Trees*. London: Verso, 2005.
Mussel, James. *The Nineteenth-Century Press in the Digital Age*. London: Palgrave Macmillan,
    2012.
Neiger, M. and K. Tenenboim-Weinblatt. "Understanding Journalism through a Nuanced
    Deconstruction of Temporal Layers in News Narratives." *Journal of Communication* 66, no.
    1 (2016): 139–60.
O'Connor, Thomas P. "The New Journalism." *The New Review* 1 (October 1889): 423–34.
Rabinowitz, Dober. "Yerushalima." *HaTzfira*, April 15, 1886, 2–3. [Hebrew]
Saba, Israel. "Amarti yesh li Tikva." *HaTzfira*, September 29, 1886, 2–3. [Hebrew]
Schöch, Christof. "Principal Component Analysis for Literary Genre Stylistics." *The Dragonfly's
    Gaze*. Last updated September 8, 2016. Accessed January 23, 2020. http:// dragonfly.
    hypotheses.org/472.

Segal, Zef. "'From One End of the Earth to the Other End of the Earth': Changing Perceptions of the World in Late-Nineteenth-Century Hebrew Journalism." *Jewish Studies Quarterly*. Forthcoming.

Segal, Zef and Oren Soffer. "One Journal, One Decade, 3,797,592 Words: Computational Analysis of HaTzfira's Discourse (1874–1883)." *Journal of Jewish Studies* 72, no. 2 (2021): 369–96.

Slauter, Will. "The Rise of the Newspaper." In *Making News: The Political Economy of Journalism in Britain and America from the Glorious Revolution to the Internet*, edited by Richard R. John and Jonathan Silberstein-Loeb, 19–46. Oxford: Oxford University Press, 2015.

Soffer, Oren. *Ein Le-Falpel! Iton Ha-Tzfira ve-Ha-Modernizatzia shel Ha-Si'ach He-Hevrati Ha-Politi*. Jerusalem: Mosad Bialik, 2007. [Hebrew]

Soffer, Oren. "'Paper Territory': Early Hebrew Journalism and Its Political Roles." *Journalism History* 30 (2004): 31–39.

Soffer, Oren, Zef Segal, Nurit Greidinger, Sinai Rusinek, and Vered Silber-Varod. "Computational Analysis of Historical Hebrew Newspapers: Proof of Concept." *Zutot – Perspectives on Jewish Culture* 17 (2020): 97–110.

Sommerville, Charles J. *The News Revolution in England: Cultural Dynamics of Daily Information*. Oxford: Oxford University Press, 1996.

Tabboni, Simonetta. "The Idea of Social Time in Norbert Elias." *Time & Society* 10, no. 1 (2001): 5–27.

Turner, Mark W. "Periodical Time in the Nineteenth Century." *Media History* 8, no. 2 (2002): 183–196.

Usher, Nikki. *Making News at the New York Times*. Ann Arbor: University of Michigan Press, 2014.

Wallach, H.M., D.M. Mimno, and A. McCallum. "Rethinking LDA: Why Priors Matter." *Advances in Neural Information Processing Systems* 22 (2009): 1973–81.

Zelizer, B. "Epilogue: Timing the Study of News Temporality." *Journalism* 19, no. 1 (2018): 111–21.

Zuckerman, Abraham. "Eshet Khayil." *HaTzfira*, June 21, 1886, 2–4. [Hebrew]

Tatsiana Astrouskaya

# "Not a Day Without a Line": Studying the Petitions of Soviet Jewish Refuseniks with the Visualization Tools in R

**Abstract:** The article offers an example of an argument-driven data analysis of 120 petitions issued in 1971–1972 by a Soviet Jewish emigrant from Minsk by the name of Ernst Levin (1934–2016). In Levin's case petitions became a principal instrument of his struggle for emigration from the Soviet Union, which lasted altogether 582 days. Probing ggmap and ggplot2 packages available for R programming language on historical data, this study intends to visualize and consequently reconstruct the way and intensity of Levin's communication with Soviet authorities and international organizations. In doing so, it approaches the emigration as a process, scaling and visualizing its durability. Displaying and highlighting the changes in the tactics of petitioning over time, the presented in the article graphs and visualizations allow considering Ernst Levin's emigration efforts from a multi-dimensional perspective of political and public actors, places, organizations, individual decisions and collective actions.

**Keywords:** Refuseniks, Soviet Jews, petitions, R, visualization, migration

# 1 Introduction

In 1968, a Soviet engineer by the name of Ernst Markovich Levin (b. 1934) resolved to abandon his USSR citizenship and to immigrate to Israel together with his wife Asia and their son Gosha. Just a few years earlier, his life resembled that of the typical Soviet *intelligent* of the time. He was born into and brought up in a family of devoted communists.[1] A graduate of the Belarusian Polytechnic Institute, one of the leading technical higher education institutions in the USSR, Ernst Levin made a successful professional career in the building construction industry.[2] Together with his family he occupied a separate apartment in the very center of Minsk, then the capital of the Belarusian Soviet Socialist Republic (BSSR), and lived a relatively prosperous life.[3] Notwithstanding appearances, throughout the 1960s, Ernst Levin developed a critical attitude towards the idea of socialism, became fascinated by the Jewish culture and learned Hebrew so well that he intended to become a Hebrew teacher.[4]

Once he decided to emigrate, Levin eagerly invested his skills and abilities in the realization of this aim, as emigration from the USSR was a lengthy and tedious process. Potential emigrants had to apply for an exit visa and frequently they faced multiple refusals before permission was finally granted, hence they became known as "refuseniks."[5] While struggling for the right to emigrate, Levin adopted

---

**1** A discussion on the relationship between the first and second generations of Soviet Jews and their attitude to the "Jewish" and the "Soviet" see Yuri Slezkine, *The Jewish Century* (Princeton, NJ: Princeton University Press, 2019), especially parts 3 and 4.
**2** For the history of the institute (now university) see K.I. Balandin et al., *Istoriia Belorusskogo natsional'nogo tekhnicheskogo universiteta* (History of the Belarusian National Technic University) (Minsk: BNTU, 2010).
**3** Ernst Levin, *I posokh v ruke vashei. Dokumental'nyi memuar 2002 goda (k tridtsatiletiiu iskhoda iz SSSR)*[And the Staff is in Your Hand. A documented Memoir of 2002 (To the 30th Anniversary of Exodus from the USSR)]. (Jerusalem: n.p., 2007), 11–13.
**4** Levin, *I posokh*, 89–93. FSO, F. 30.45 (Levin), 2–85. Compare to: Ann Komaromi, "Between Two Worlds: Late Soviet Jews in Leningrad," *East European Jewish Affairs* 48 (2018): 23–40, accessed January 13, 2022, https://doi.org/10.1080/13501674.2018.1442046. On the teaching and learning of the Hebrew language in the USSR see Mark Drachinsky, "A Brief Survey of the History of Hebrew Teaching in USSR," in *Jewish Culture and Identity in the Soviet Union*, ed. Ya'acov Ro'i and Avi Beker (New York: NYU Press, 1991), 246–54.
**5** Like many other Soviet migrants, the Levin family experienced multiple refusals, hatred, suspicions, various bureaucratic hurdles, and financial difficulties before procuring an exit visa. Larissa Remennick called this condition an "economic, legal and political" limbo. See, Larissa Remennick, *Russian Jews on Three Continents: Identity, Integration, and Conflict* (New Brunswick, NJ: Transaction Publishers, 2012), 39.

the credo "Not a day without a line" (Ni dnia bez strochki),[6] and decided to concentrate his efforts on writing and making formal requests (known as petitions) to the Soviet authorities and international organizations, thereby hoping to attract attention to his case.

Simultaneously, Ernst Levin meticulously documented the whole process of emigration (See Figure 1 for an example of his documentation). He collected an extensive archive of his and his family's departure, including the correspondence, newspaper clippings, address lists, and personal notes, which he was able to smuggle out while leaving the USSR. Owing to this careful documentation, the main bulk of which is made up of individual and collective petitions, it is now possible to reconstruct the whole process in detail.[7]

In the existing scholarship, the petitions of Soviet citizens have been studied mostly from the perspective of social history, which to a significant extent allowed historians to revise Soviet history from the standpoint of the ordinary citizens, bringing to light their agency, even if this was strongly restricted in the totalitarian (authoritarian) state.[8] Yet, a combined qualitative *and* quantitative analysis of Soviet petitions based upon their structural characteristics is still rare. Possible reasons for this are the extremely high number and diversity of petitions, which made a more precise statistical analysis difficult to handle for historians. Secondly, the sources are scattered in different post-soviet archives, some of which are still classified or difficult to access. The alternative solution, as this article suggests, is to apply data analysis to a small-scale but coherent dataset, where the materials are distributed over a specific period of time.

In what follows, I offer an example of argument-driven digital historical analysis of one particular emigration case, combining a close computational reading of petitions with a traditional study of sources. I intend to show how the tactics of petitioning has changed over time, and how the practices of both resistance *and* adaptation were adopted as instruments in the struggle for emigration.

My computational analysis is based upon a small dataset derived from Levin's petitions and uses the R programming environment, in particular the *ggplot2*

---

**6** I define petitions as all petition-like written and oral appeals to the authorities. Compare to Hale Yilmaz, "Petitions as a Source in Women's History of the Republican Period," in *Women's Memory: The Problem of Source*, ed. Fatma Türe and Birsen Talay Keşoğlu (Newcastle upon Tyne: Cambridge Scholars Publishing, 2011), 81.

**7** FSO, F. 30.45 (Levin). The other significant source has been Ernst Levin's recollections. Levin, *I posokh*.

**8** Yilmaz, "Petitions as a Source," 81–83. Compare also to the recent study on petitioning practices in late-Salazar Portugal: Duncan Simpson, "Approaching the PIDE 'From Below': Petitions, Spontaneous Applications and Denunciation Letters to Salazar's Secret Police in 1964," *Contemporary European History* 30, no. 3 (2021): 398–413, doi:10.1017/S0960777320000612.

package, to visualize its characteristics.[9] By applying R to a relatively small historical dataset we can better understand emigration as a process, and scale and visualize its durability. Moreover, it allows putting an individual case into a multi-dimension framework of political and public actors, places, organizations, and individual decisions and collective actions. It thus helps to arrive at a more nuanced understanding of the practice of petitioning and simultaneously opens an opportunity for cautious generalization. The latter is especially rewarding when dealing with unevenly represented sources as is the case with regard to the history of the Jewish refuseniks in Minsk.[10]

Though petitioning was distinctive for the relationship between the citizens and the state throughout the whole Soviet period,[11] starting from the late 1960s, petitions also mirrored the emerging political participation in the USSR.[12] By applying visual and data analysis in this study I will explore the dynamics of protest and interaction that arose in Levin's appellations. The position of a potential emigrant in Soviet society, who simultaneously was treated as a member and an outsider, was an additional factor that greatly affected the character of this dynamic.

The article begins with a brief note on emigration policies in the Soviet Union and the BSSR and discusses in some detail Ernst Levin's emigration story so as

---

**9** The use of R for these purposes is the author's choice, based on her understanding and skills. Arguably, a similar result can be achieved by other means. The purpose of this article is not to discuss the advantages of the R programming language for historical research but to combine methods of traditional source study and data analysis visualization to achieve a more nuanced understanding of the history of Jewish emigration from the USSR.

**10** The utility of digital methods for the study of Soviet history, when some archives are still closed and the sources are difficult to access, has been discussed in Susan Grunewald, "A Push for Digital History in Soviet and Post-Soviet Studies," NYU Jordan Center for the Advanced Study of Russia, accessed July 15, 2021, https://jordanrussiacenter.org/news/a-push-for-digital-history-for-soviet-and-post-soviet-studies/#.YSjUGogzaUm.

**11** Sheila Fitzpatrick, "Editor's Introduction: Petitions and Denunciations in Russian and Soviet History," *Russian History* 24, no. 1–2 (1997): 5–6. See also, in the same issue, Golfo Alexopoulos, "The Ritual Lament: A Narrative of Appeal in the 1920s and the 1930s," 117–29. It is worth noting that writing petitions had a long tradition in Russian and subsequently Soviet history, yet its premise had changed in the last decades of Soviet rule.

**12** On April 12, 1968, the Presidium of the USSR Supreme Soviet signed the "Decree on the Procedure for the Consideration of Proposals, Declarations, and Complaints of Citizens." Ukaz N 2534-VII Prezidiuma Verkhovnogo Soveta SSSR o Poriadke rassmotrenia predlozhenii, zaiavlenii i zhalob grazhdan, accessed February 13, 2021, http://docs.cntd.ru/document/9012207. On political communication see also Stephen White, "Political Communications in the USSR: Letters to Party, State and Press," *Political Studies* 31 (1983): 43–60 and Margareta Mommsen, *Hilf mir, mein Recht zu finden. Russische Bittschriften von Iwan dem Schreiklichen bis Gorbatschow* (Berlin: Propyläyen, 1987), 217–57.

to relate his petitioning activity to a broader context. After that, it considers petitions as the main weapon of Ernst Levin's struggle for emigration, and explains how the petitions could be turned into a dataset. I then explore and analyze various groups of visualizations created in R to probe the distribution of Levin's petitions over time, their different addressees and types, as well as the correlations between them.

## 2  Emigration Policies in the Soviet Union and the BSSR

In the 1970s–1980s, 240,000 Jews left the Soviet Union for Israel, the USA, and elsewhere.[13] Although the USSR signed, already on December 10, 1948, the Universal Declaration of Human Rights affirming the freedom of movement,[14] this right was not explicitly enshrined in the Soviet Constitution.[15] Moreover, even in the post-Stalin period, emigration from the USSR, if no longer a subject of criminal prosecution, was openly disapproved and condemned. Only with the onset of perestroika policies in 1987–1989 did free emigration become possible.

The success of emigration depended on multiple factors, such as international pressure and the goodwill of the local authorities. As a rule, emigration was a lengthy and challenging procedure, and the number of those allowed to emigrate fluctuated significantly.[16] To leave, one had to apply for an exit visa at the local Department for Visas and Registration (OVIR). A substantial percentage of first-time applications was refused, based on different formal reasons. The applicants who were not granted permission to emigrate were colloquially labeled "refuseniks" (*otkazniki* in Russian).[17] Refuseniks stayed in the Soviet Union for, sometimes, several years, struggling for the opportunity to emigrate. Often, they

---

**13** The figures are from Zvi Gitelman, *A Century of Ambivalence: The Jews of Russia and the Soviet Union, 1981 to the Present*, 2nd ed. (Bloomington: Indiana University Press, 2001), 185.

**14** The Universal Declaration of Human Rights, 13.1 and 13.2.

**15** Konstitutsia (Osnovnoi zakon) Soiuza Sovetskikh Sotsialisticheskikh Respublik 5 dekabria 1936 goda (The Constitution of the Union of Soviet Socialist Republics adopted on December 5, 1936), accessed March 2, 2021, http://www.hist.msu.ru/ER/Etext/cnst1936.htm. The next Constitution in the USSR was adopted on October 7, 1977.

**16** See, for instance, the numbers of emigrants by years: "Total Immigration to Israel from the Former Soviet Union (1948–present)," accessed February 5, 2021, https://www.jewishvirtuallibrary.org/total-immigration-to-israel-from-former-soviet-union.

**17** On the definition of *refuseniki* as a group see, Mordechai Altshuler, "Who Are the 'Refuseniks'? A Statistical and Demographic Analysis," *Soviet Jewish Affairs* 18, no. 1 (1988): 3–15. Vladimir (Ze'ev)

were deprived of occupation, income, and education and faced condemnations in public life and on the pages of Soviet newspapers.[18]

At the same time, from the end of the 1960s, protest activity to allow emigration started to gain force and attracted attention in the Soviet Union and internationally. As a result of the latter, the so-called Jackson-Vanick Amendment, which was designed to support human rights in the USSR and put pressure on Soviet economic interests, was adopted by the US Congress in early 1975.[19] Another significant initiative was the signing of the Final Helsinki Act in August 1975, which by many (at least in the initial stage) was perceived as a step towards democracy and civic freedoms in the Soviet Union.

# 3  Setting the Context: The Emigration Story of the Levin Family

As was already noted, Ernst Levin was a true Soviet citizen in his youth and expressed no particular interest either in Yiddish, the traditional language of the local Jewish population, or in Jewish culture.[20] The change in his *Weltanschauung* was a result of both the political liberalization of the late 1950s–1960s in the USSR[21] and personal circumstances. As a typical representative of the 1960s generation, Ernst Levin experienced the hardship of World War II and Stalinism, yet later also the loosening of ideological constraints in the Soviet Union in its prime.[22] He was also affected by the rise of common anti-Semitism in Soviet society and recalled

Khanin, "The Refusenik Community in Moscow: Social Networks and Models of Identification," *East European Jewish Affairs* 41, no. 1–2 (2011): 75–88, doi:10.1080/13501674.2011.591661.

**18** "Timeline of the Jewish Movement in the Soviet Union," University of Toronto Libraries, accessed October 17, 2020, https://samizdatcollections.library.utoronto.ca/content/timeline-jewish-movement-soviet-union.

**19** Barbara Martin, "The Sakharov-Medvedev Debate on Détente and Human Rights: From the Jackson-Vanik Amendment to the Helsinki Accords," *Journal of Cold War Studies* 23, no. 3 (2021): 138–74, accessed January 13, 2022, https://doi.org/10.1162/jcws_a_01009.

**20** Levin, *I posokh*, 17.

**21** Partial liberalization, the so-called Khrushchev Thaw, was initiated and held by the First Secretary of the Central Committee of the Tsk CPSU Nikita Khrushchev (1894–1971).

**22** On the "generation of the sixties" in the USSR see: Petr Vail and Aleksandr Genis, *60-e. Mir sovetskogo cheloveka* (Moscow: Corpus, 2013); Vladislav Zubok, *Zhivago's Children: The Last Russian Intelligentsia* (Cambridge, MA: Belknap Press, 2009); for an overview, see also Georgii Kas'ianov, *Nezgodny: Ukrainskaia inteligentsia v rusi oporu 1960-kh–1980-x rokyv* (Kyiv: Lybid', 2005), 12–31.

in his memoirs cases of anti-Semitic behavior and anti-Semitic expressions attacking him or his wife.[23] The BSSR was no exception in this regard; its daily periodicals, such as *Zviazda* (Star), regularly condemned Zionism, strongly discouraging Belarusian Jews from emigration.[24]

It was Israel's victory in the Six-Day Arab-Israeli War in June 1967 that made the idea of emigration palpable for Ernst Levin (as well as for many Soviet Jews).[25] Additional incentives were the domestic political situation in the USSR, Leonid Brezhnev's attempted de-liberalization, and increasing pressure on the dissenting and intelligentsia.[26]

Levin's wife, Asia Levina (née Rudshtein, b. 1939), and their friends' circle influenced his views greatly too. Asia was born in the small Belarusian town of Lyuban,[27] into a family in which Jewish traditions were cherished and skepticism toward Soviet authorities prevailed.[28] Among the friends of the Levin family were the Minsk theater designer Tsfania-Gedalia Kipnis, a student from Riga by the name of Il'ia (later: Eliahu) Valk, and the son of the artist and prisoner of Stalin Gulag Mark Zhitnitskii, Isaak – all persuaded Zionists.[29] The last of these, Isaak Zhitnitskii, one of the first Jewish emigrants from Minsk, provided the Levin

**23** Levin, *I posokh*, 5 and passim. On late-Soviet nationalities policies toward Jews see, Evgenii Kazakov, "V poiskakh 'sovetskogo evreiskogo': Poznesovetskaia natsional'naia politika," *Neprikosnovennyi zapas* 6 (2018): 190–215.

**24** See, for instance "Han'ba siianistskim pravakataram" (Shame on Zionist instigators), *Zviazda* 17 (January 21, 1972), 3; Safiia Kantar, "Chamu ia ne paedu ŭ Izrail" (That is why I am not going to Israel), *Zviazda* 33 (February 9, 1972), 3, etc. See also Alexander Friedman, "Antizionismus und Anti-Masonismus in der Sowjetunion nach dem israelisch-arabischen Sechstagekrieg (1967). Der Verschwörungstheoretiker Vladimir Ja. Begun (1929–1989)," in *Juden und Geheimnis. Interdisziplinäre Annäherungen*, ed. Claus Oberhauser (Innsbruck: Innsbruck University Press, 2015), 137–51.

**25** Gitelman, *A Century of Ambivalence*, 176–77. The migration intensified not only in the Soviet Union but also beyond. See, Dariusz Stola, *Kraj bez wyjścia? Migracje z Polski 1949–1989* (Warsaw: ISP PAN, 2010), 177–218; Jannis Panagiotidis, *The Unchosen Ones: Diaspora, Nation, and Migration in Israel and Germany* (Bloomington: Indiana University Press, 2019), 193–237.

**26** Levin, *I posokh*, 15.

**27** With the pre-World War II Jewish population amounting to more than one-third of all the town's inhabitants or 35.33 per cent. Source: *Vsesoiuznaia perepis'vnaseleniaia 1939 goda* (The all-Union population census of 1939), accessed March 1, 2021, http://www.demoscope.ru/weekly/ssp/ussr_nac_39_ra.php?reg=625.

**28** Interview by the author with Asia Levina, Munich, May 21, 2018, in the personal archive of the author.

**29** Eliahu (Eli) Valk had left USSR for Israel in 1971. After the dissolution of the Soviet Union, between 1993 and 1996 he became the first ambassador of Israel to the independent Republic of Belarus. For the self-published recollections of Isaak Zhitnitskii on his emigration from Minsk see, Isaak Zhitnitskii (Itskhak Bar-Zait), "Iskhod iz Minka, 1971 god," Haifa (2014), Evreiskaia
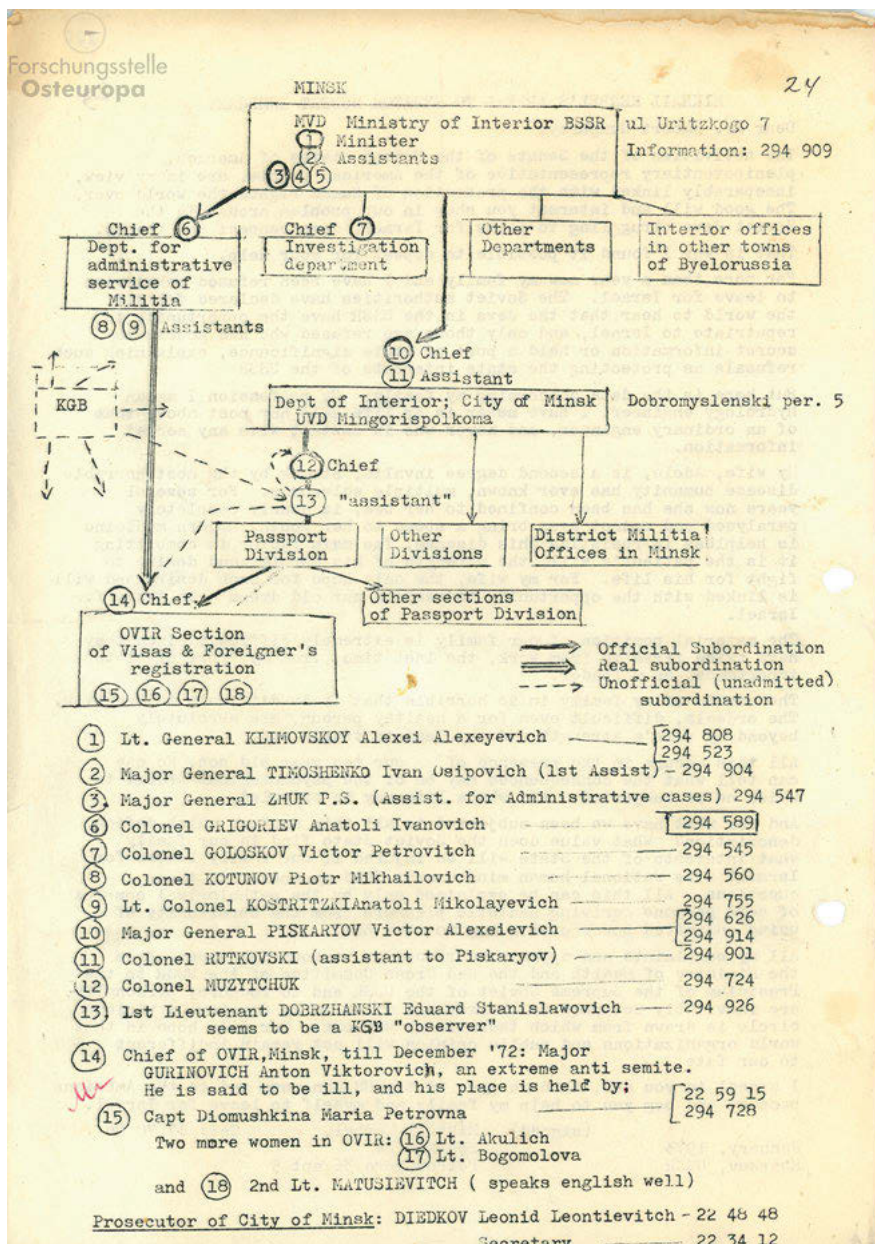
**Figure 1:** A structural organization of BSSR administrative organs dealing with Jewish emigration. A scheme drafted by Ernst Levin and translated into English.
*Source*: Forschungsstelle Osteuropa (afterward: FSO), F. 30.45 (Levin), 1–24.

family with the so-called "*vyzov*" (invitation or call) necessary for the start of the emigration process.[30]

As Minsk Jews began to form a more unified group, they also strove to make the memory of the Holocaust part of public discourse.[31] On March 5, 1972, the commemoration of the thirtieth anniversary of the Minsk Ghetto massacre, in which 5,000 people were slaughtered, took place. This gathering became one of the first collective manifestations organized by the Minsk Jewish Community and played a role also in Ernst Levin's emigration story.[32]

## 4  Petitions as a Weapon and as a Dataset

It took Ernst Levin and his family 582 days, or 83 weeks and 1 day, to emigrate from the Soviet Union.[33] The invitation arranged by Zhitnitskii arrived by post on April 28, 1971. On August 29, 1972, approval for the Levins' visa application was granted, though it required three more months to collect and pay the so-called "diploma tax," introduced by the authorities to limit the number of applications.[34] Eventually, the family crossed the border with the Polish People's Republic near the Belarusian city of Brest on November 30, 1972.

Writing petitions and appeals became a powerful weapon, which Soviet refuseniks used in their struggle for emigration. It simultaneously served the integration of the movement and helped to publicize the cause of Jewish emigration from the USSR outside its borders. Ernst Levin was one of the first Jewish refuseniks in the BSSR who recognized the power of petitioning. The study of Levin's archive together with his published recollections and other auxiliary sources

---

Wiki-Entsiklopedia/Everyday Jewish Wiki, accessed November 22, 2021, http://www.ejwiki.org/wiki/Welcome_to_Everyday_Jewish_Wiki.

**30**  Levin, *I posokh*, 15; FSO, F. 30.45/ (Levin), fols. 1–3.

**31**  Levin, *I posokh*, 89–93.

**32**  Levin, *I posokh*, 100–101. The massacre took place on March 2, 1942. The history of Jewish refuseniks in Minsk was briefly referred to in: Ludmilla Alexeyeva, *Soviet Dissent: Contemporary Movements for National, Religious, and Human Rights* (Middletown, CT: Wesleyan University Press, 1984), 175–76; Leonard Shroetter, *The Last Exodus* (Jerusalem: Weidenfeld and Nicolson, 1974), 272–85.

**33**  Author's calculations.

**34**  In Levin's case, it amounted to 15,000 rubles, while Ernst Levin's average monthly wage at the time was about 200 rubles. Levin, *I posokh*, 164–65. For more see also Gitelman, *A Century of Ambivalence*, 183–84; Viktor Dennighaus and Andrei Savin, "'Kak by Ukaz o evreiiakh ne otmeniat', a de faktom ne primeniat' L.I. Brezhnev, razriadka i evreiskaia emigratsia iz SSSR v 1972–1973 gg.," *Rossia XXI vek* 1 (2013): 130–59.

such as newspaper articles of the time allowed the reconstruction of 121 unique appeals, which the aspiring emigrant, alone and together with other refuseniks, made between May 1971 and November 1972. These petitions served as a basis for a small dataset, predominantly designed to enrich our understanding of the ways and methods in which the refuseniks communicated with the Soviet authorities.[35]

The petitions of Soviet citizens were usually written according to a specific pattern and included obligatory information such as the full name and address of sender and recipient, a date, and typically a reference to a concrete problem which should be solved with the assistance of the authorities. They thus represented a rather standardized form of communication with the authorities. Nonetheless, they varied in their intonation and contained their own unique stylistic features, depending on the personality and the position of the supplicant, leaving space for the "subjective universe" of the author, as Juliane Fürst calls it.[36] Along with some other categories of Soviet citizens – such as participants in the human rights movement[37] and religious dissidents[38] – Jewish activists for emigration operated from the margins of Soviet society. Not seen as loyal Soviet citizens on account of their activities, they practiced civil obedience – that is, they intended to follow the letter of the law in their protest actions – and dared to demand respect for their rights.[39] Simultaneously, as will be shown below through the example of Ernst Levin's petitions, their appeals and the message they contained were far more varied, as it is often presented.

---

**35** Despite Levin's thorough documentation and the author's careful examination of other available sources, as well as his recollections, there could be a small number of other petitions that he signed. This could potentially affect the correlations found between the different variables. It can be argued with a certain degree of confidence, however, that the generalizations made on the basis of the collected data are correct, as the analyzed petitions represent the bulk of Levin's appeals.

**36** Juliane Fürst, "In Search of Soviet Salvation: Young People Write to the Stalinist Authorities," *Contemporary European History* 15, no. 3 (2006): 327.

**37** The human rights movement started to form in the mid-1960s in the USSR, its principal demand to the Soviet state being that the provisions of its own law would be observed. For the history of the human rights movement in the USSR, see, Alexeyeva, *Soviet Dissent*, 267–398.

**38** On religions dissidents in the USSR, see, Alexeyeva, *Soviet Dissent*, 201–43.

**39** On the idea and practice of radical civic obedience in the milieu of Russian dissidents and its "founding father" Aleksandr Esenin-Vol'pin, see, Benjamin Nathans, "The Dictatorship of Reason: Aleksandr Volpin and the Idea of Rights under 'Developed Socialism,'" *Slavic Review* 66, no. 4 (Winter 2007): 630–63.

# 5 The First Round of Data Evaluation and Limitations

Levin's archive, as well as every single petition he signed, contains more information than could be taken into consideration in this study. During a first data evaluation round more detailed information on the recipients of Levin's appeals (addresses), which was available in some cases and included in the initial version of the dataset, was left out.

Moreover, responses from the authorities were also traceable in some cases, yet their number and the information they contained were insufficient and thus did not merit inclusion in the dataset which, as a result, focuses only on the standpoint of the sender.

Similarly, the names, ranks, and positions of the addressees (as three separate variables) were initially collected. Thus, it became apparent that in the OVIR (of the Department of Internal Affairs of the Minsk City Executive Committee) Ernst Levin mostly communicated with the head of its administration, major Anton Gurevich, whom he described as an "extreme anti-Semite."[40] The question of why refuseniks appealed to a specific functionary, and with what kind of message and at what moment, requires further consideration and could add to the history of Jewish emigration from the BSSR. Yet this question is beyond the scope of the present article.

In order to contextualize the analysis of Levin's petitions, an auxiliary dataset was also created with altogether 81 addresses of Jewish activists for emigration in Minsk in 1971–1972. As was already mentioned, in order to sign the petition (which could not be sent anonymously), one had to indicate his or her full address including the number of the apartment. These addresses were plotted on a map, using the *ggmap* package, which extends the *ggplot2* package for maps (see Figure 2).

It is worth mentioning that unlike many other post-Soviet cities, which underwent a massive renaming after the collapse of the USSR, the names of Minsk streets and the city structure have not changed greatly since the 1960s, which allows using contemporary maps for historical research.[41] Still, some changes did occur. In the 1970s, the post-war "socialist" reconstruction of the city of Minsk was still ongoing.[42]

---

**40** F. 30.45 (Levin), 1-24.

**41** Compare to: Andrei Savin, "'U Gorbacheva i El'tsina ne bylo politicheskoi voli.' Pochemu Rossia ne izbavilas' ot sovetskikh nazvanii gorodov i ulits," *Lenta.ru*, June 7, 2020, accessed November 29, 2021, https://lenta.ru/articles/2020/07/07/savin_2/.

**42** On the post-World War II socialist building of the city of Minsk see Thomas Bohn, *Minsk – die Musterstadt des Sozialismus. Stadtplanung und Urbanisierung in der Sowjetunion nach 1945* (Cologne: Böhlau, 2008).
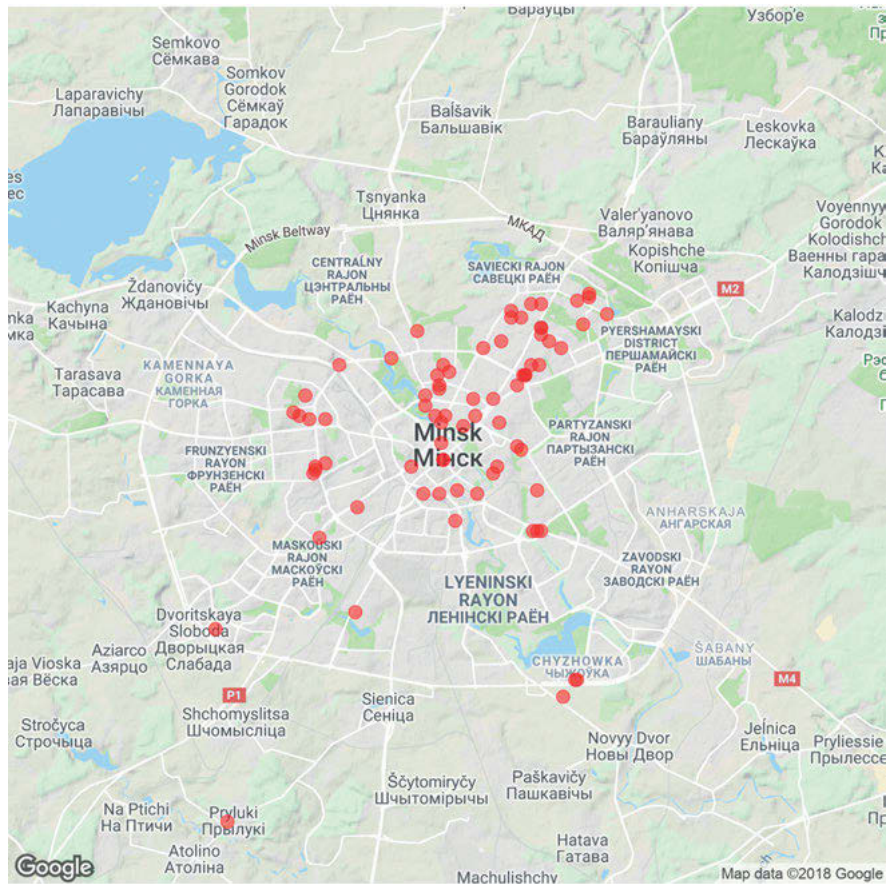
**Figure 2:** A distribution of petitioners across the city of Minsk in 1971–1972. The bright-red dots indicate that more than one appellant was residing at one address, so families signed the petitions together.

The construction of new streets and city districts led to the disappearance or renaming of old street names, which in some cases necessitated a reconstruction of the contemporary street names.[43]

––––––––––

**43** For instance the MOPR street was renamed Kalinina street in 1946, and in 1961 it was again renamed as Kommunisticheskaia street. Yet the MOPR lane on which one of the appellants resided (a side street of the former MOPR street) existed until the 1980s. A few old houses and an old biscuit factory stood on the lane but disappeared when the buildings were ruined. In this case I had to define an approximate location for the former lane. Google Maps still does not cover every street and object in Minsk, so it has to be double-checked.

As becomes apparent from the map (Figure 2), most of the aspiring emigrants resided in the center of the city and were located in clusters. The former can be explained by the distribution of housing recourses, which often depended on the individual's position in the Soviet power hierarchy, characteristic of Soviet cities.[44] As in the case of Ernst Levin, who inherited his centrally located apartment from his communist parents, some of the supplicants stemmed from the families of party functionaries or occupied a high-ranked position. Technical intelligentsia and professionals settled in newly built apartment houses which arose along the main Lenin avenue at a certain distance from the center.[45] Among the signatures were those of engineers, doctors, artists, and former Soviet army officers.[46] The location in clusters shows that the strengthening of the Jewish community, which was imperative for collective petitioning, had a certain correlation with the place of residence.

# 6  Second Round of Data Evaluation

After the first round of data evaluation, it was decided to limit the inquiry to the following questions:

1. What was the temporal distribution of Ernst Levin's petitions between April 28, 1971 and November 30, 1972?
2. What were the main addressees of Levin's appellations and where were they located? Or, more precisely: was he mostly communicating with the republican (BSSR) authorities in Minsk or central (all-Union) authorities in Moscow?
3. Was he mostly acting alone or did he collaborate with other refuseniks? And is there a correlation between individually or collectively signed petitions and their recipients?
4. What were the main channels and the main ways of communication?

The initial dataset was manually revised with a view to the data's relevance for these research questions, and the following seven variables were retained: (1) ID, (2) Date, (3) Mode of Communication (with the subsequent sub-categories: *personal*

---

**44** Bohn, "Minsk," 6.

**45** For this information I am grateful to Artur Klinau. See also his book devoted to the (post-)socialist landscape of Minsk: Artur Klinau, *Minsk – Sonnenstadt der Träume* (Berlin: Suhrkamp, 2011).

**46** Thus, active refuseniks of the time included decorated retired Red Army officers Colonel Lev Ovsischer and Lieutenant Colonel Naum Alschansky, wife of professor of history Nikolai Poletika Tamara, medical doctor Iakov Shulz, and artist Tsfania Kipnis. FSO, F. 30.45, Bl. 110, 130, 148.

*visit*, *telephone call*, or *written letter*), (4) Addressee/Organization, (5) Address/ Destination, (6) Type (*individual* or *collective*), and (7) Form (with sub-categories: *inquiry*, *appeal*, *request*, *and complaint*). (See also Dataset Summary, Figure 3). Each variable corresponds to a unique unit of information, appropriate for descriptive data analysis. Each appeal received a unique ID and was related to the date. Since several appeals could be made on one day, the same date can reappear several times in the dataset.

If the definition for most variables is quite obvious, the seventh variable – the "Form of Communication" requires a short explanation. The accessible correspondence and conversations, including those about which Levin reported, vary in their tone and message. After careful examination of the applications as well as the context in which they arose, I have sub-categorized them as follows.[47] Appeals have been classified as the applications which imply a call, a clear message "let us go home" (otpustite nas domoi) or "help us go home." Requests normally contain expressions like "we ask" (my prosim) or "we require" (my trebuem). Inquiries are formal requests for information, usually concerning the status of a visa application. Complaints (zhaloba) are typically applications to the higher authorities or prosecutorial bodies to attract attention to the injustice happening "on the ground," objecting to it, and demanding justice.

The corresponding information was loaded into an MS Excel spreadsheet which then was converted into a csv-format table, compatible with R. Ernst Levin addressed 47 Soviet and international organizations (and individuals), located in ten different cities in the Soviet Union and abroad (see also the summary of the dataset, Figure 3). Most of Levin's applications appeared in a written mode and were signed individually (74 and 83 of 121 respectively), most frequently he communicated with the local authorities in Minsk (77 applications), and the main addressee of his applications was the BSSR Ministry of Internal Affairs (24 applications). As the dataset is small and the results somewhat foreseeable, such a preliminary understanding could be drawn without the help of digital means, yet even at this stage, the analysis makes it easier to comprehend some patterns.

The seven variables in the dataset formed the basis for several visualizations that were used as an exploratory tool which served to sharpen, reframe, and support the analysis of Levin's tactics of interaction with the authorities as reflected in his petitions.

---

**47** As mentioned, these categories may require reconsideration or extension during further investigation.

```
        ID                  DATE            MODE                     ADDRESSEE.ORGANISATION        DESTINATION
Min.    :  1.00  01.03.1972: 6  Personal :33  BSSR Ministry of Internal Affairs:24     Minsk     :77
1st Qu.: 30.75  02.05.1972: 5  Telephone:14  OVIR                             :11     Moscow    :27
Median : 60.50  03.08.1971: 4  Written  :74  BSSR Minister of Internal Affairs: 9     Jerusalem : 5
Mean   : 60.50  03.11.1971: 4                Minsk City Procurator            : 6     New York  : 3
3rd Qu.: 90.25  30.09.1971: 4                TsK, Belarusian Communist Party  : 5     Paris     : 2
Max.    :121.00  31.09.1971: 4                Chairman, USSR Supreme Soviet    : 4     Washington: 2
NA's    :1       (Other)   :94                (Other)                          :62     (Other)   : 5
        TYPE            FORM
Collective:38  Appeal    :39
Individual:83  Complaint:27
               Inquiry  :38
               Request  :17
```

**Figure 3:** Dataset "petitions" summary.

# 7 Visualizing Petitions in R

R is a high-level programming language created for, primarily, data analysis and visualization and contains a vast collection of libraries suitable for almost every kind of historical analysis.[48] For this article, which relies on visualizations of a small dataset, R seems to be an appropriate tool for exploring both the quantitative and the qualitative dimensions of petitioning practices. The visualizations have been created with the help of RStudio, a programming interface or IDE – Integrated Development Environment – for R, which allows a relatively simple and user-friendly application of R.[49] I used *ggplot2*, one of the available *tidyverse* packages for R, which is specifically developed for creating graphs and which primarily bases every graph on three components – a dataset, a system of coordinates, and geoms (visual marks representing data points). In the process of data analysis, four groups of visualizations were created, which allowed for the evaluation and exploration of correlations between the defined variables.

The first group of visualizations (Figure 4–7) looks into the application process and its development over time, showing the number of petitions (sub-divided

---

**48** The advantages of R for historical research have been discussed in: Sharon Howarth, "Five Reasons for Historians to Learn R," accessed July 17, 2021, https://www.dataquest.io/blog/five-reasons-for-historians-to-learn-r/, and Lincoln Mullen, "Digital History Methods in R," accessed July 17, 2021.
**49** Mullen, "Digital History Methods."

according to the corresponding variables in different graphs) per month over a period of 20 months starting from April 1971 and ending in November 1972.

Figure 4 displays the applications grouped by the mode of communication on a timeline. As it looks, the communication was the most intensive between August 1971 and April 1972, and it especially intensified in October–November 1971 and the spring of 1972, while written applications prevailed. Figure 5 demonstrates that most of the collective applications were issued in the same period (fall 1971 and spring 1972). Notably, with the approval of his exit visa application, Ernst Levin did not completely refrain from sending petitions (as he still had to clear the issue with the "diploma tax") but preferred to do it individually. Figure 6 adds the destination of the application to this representation, the most colorful part of which is the period between February and May 1972, when the applications were sent not only to Minsk and Moscow but also to Tel Aviv, Jerusalem, Washington, New York, Paris, London, and Rome. This period coincides with the rise of collective applications. Figure 7 looks closer at the forms of applications, and, somewhat surprisingly, it indicates that complaints were more distinctive for the initial phase of the emigration process (June–October 1971). They were subsequently equaled and even outweighed by appeals.
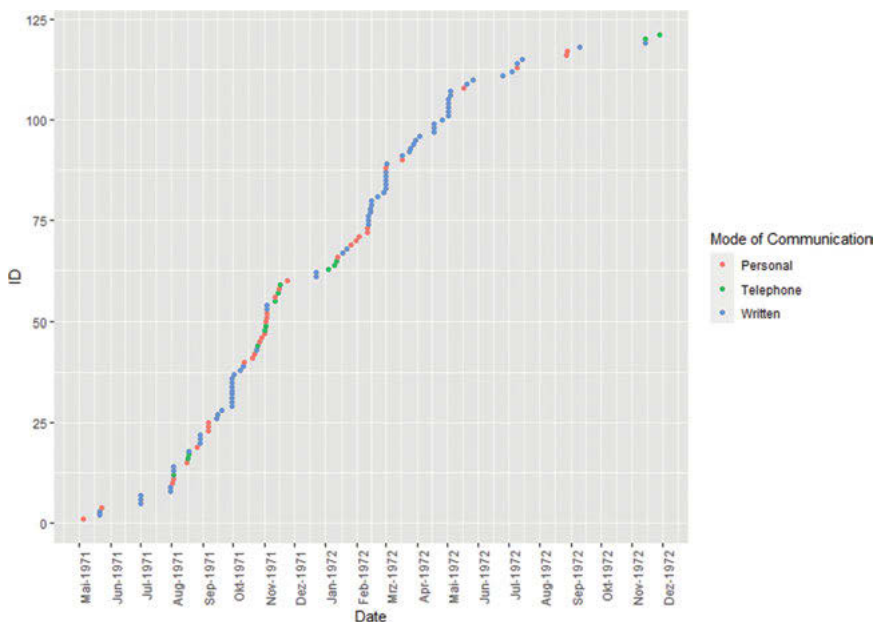


**Figure 4:** Distribution of petitions on a timeline by the mode of communication.
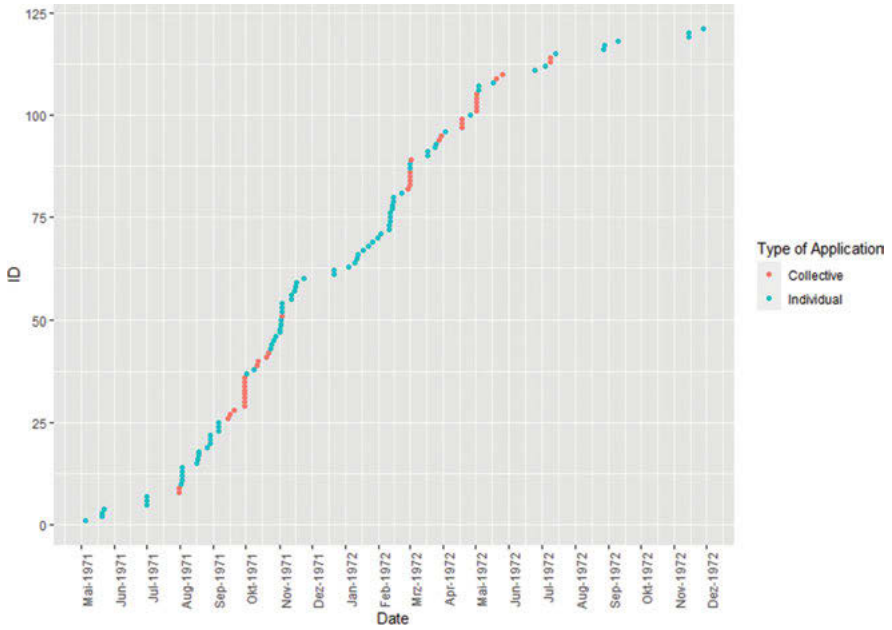
**Figure 5:** Distribution of petitions on a timeline by the type of application.
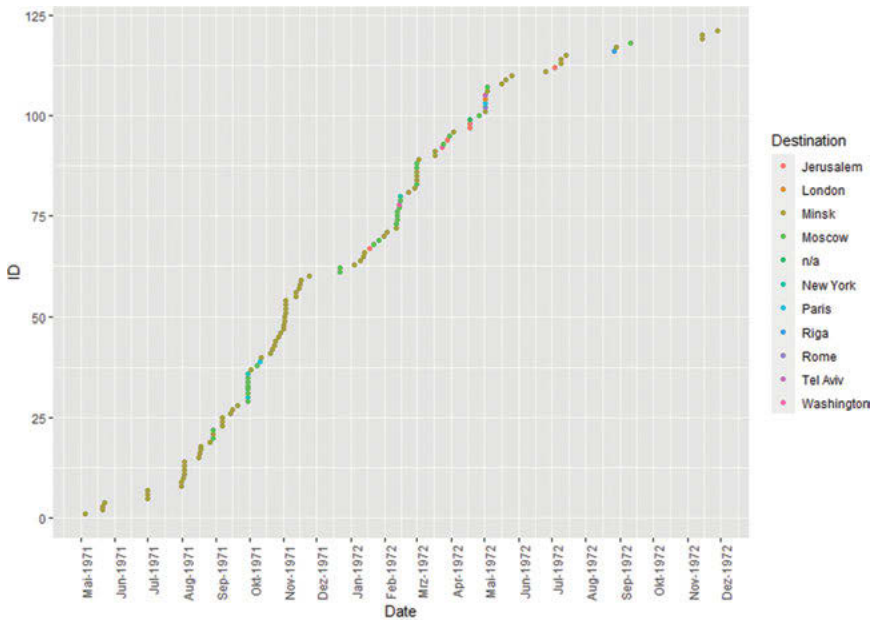


**Figure 6:** Distribution of petitions on a timeline by destination.
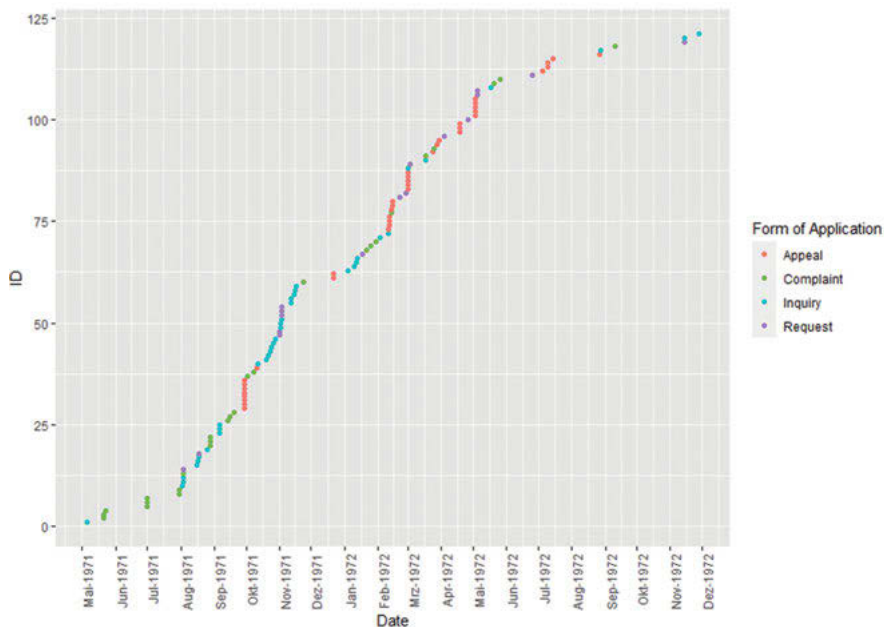
**Figure 7:** Distribution of petitions on a timeline by the form of application.

The second group of graphs (Figures 8–10) focus on the addressed organizations and officials and their correlation with the form, mode, and type of application. All three graphs make clear that the BSSR Ministry of Internal Affairs (Ministerstvo Vnutrennikh Del BSSR) was the most regularly approached organization, with second and third places taken by OVIR (the Department for Visas and Registration) and the BSSR Minister of Internal Affairs Aleksei Klimovskoi respectively. Figure 8 shows that Levin approached the Ministry and OVIR mostly with inquiries. With complaints (for instance, the cases in which he received no answer from the subordinate authorities or he believed he was treated unlawfully) he addressed the Minister himself, the Procurator General of the USSR Roman Rudenko, or the Minsk City Procurator Leonid Dedkov. Occasionally he also approached the Central Committee of the Belarusian Communist Party and the CPSU. Generally, appeals were mostly addressed to the higher authorities and even more often to the international organizations, such as the United Nations (UN).

Figure 9 demonstrates that the BSSR Ministry of Internal Affairs was approached by all modes of communication, yet mostly personally and by telephone (arguably also due to Ernst Levin's place of residence, as his apartment was located just opposite the Ministry and telephone calls to Moscow or abroad were much more expensive). Both collective and individual applications were addressed to the Min-
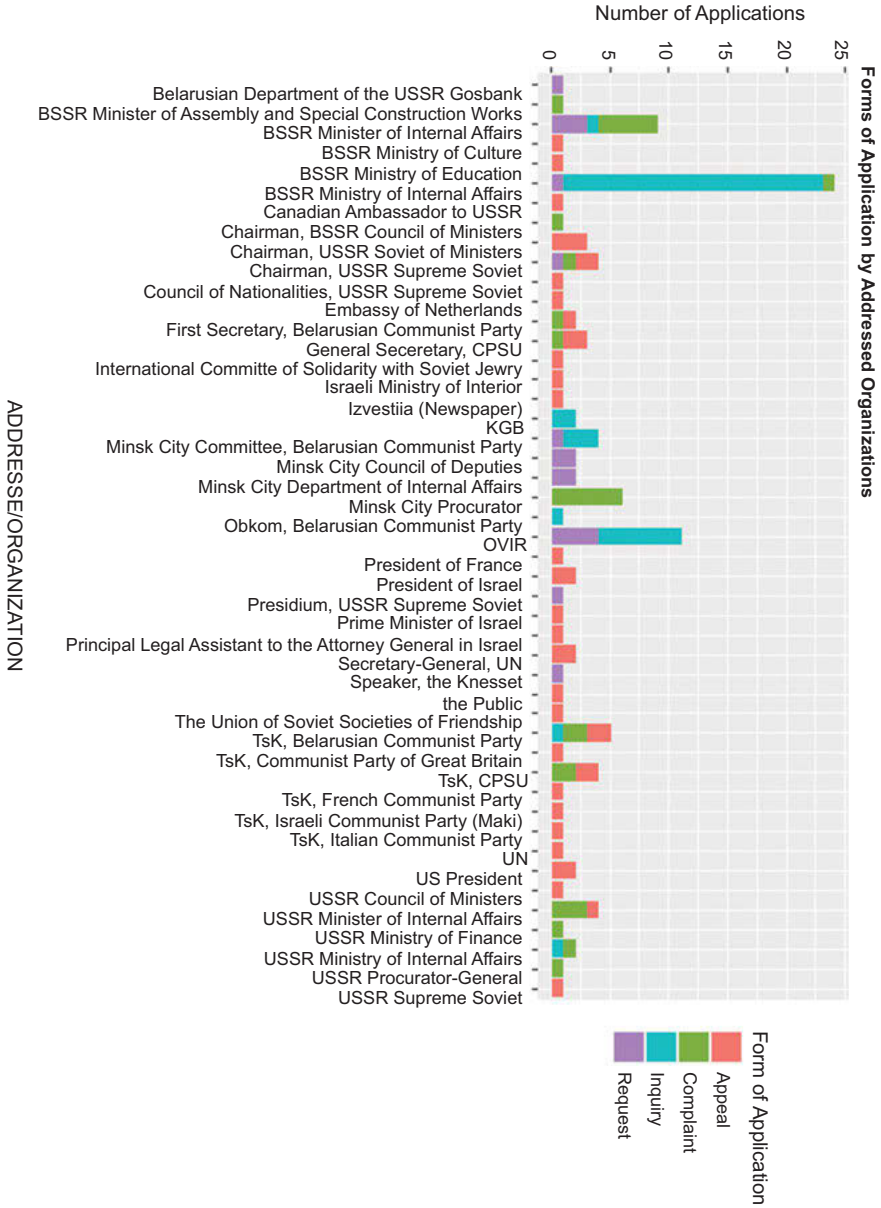
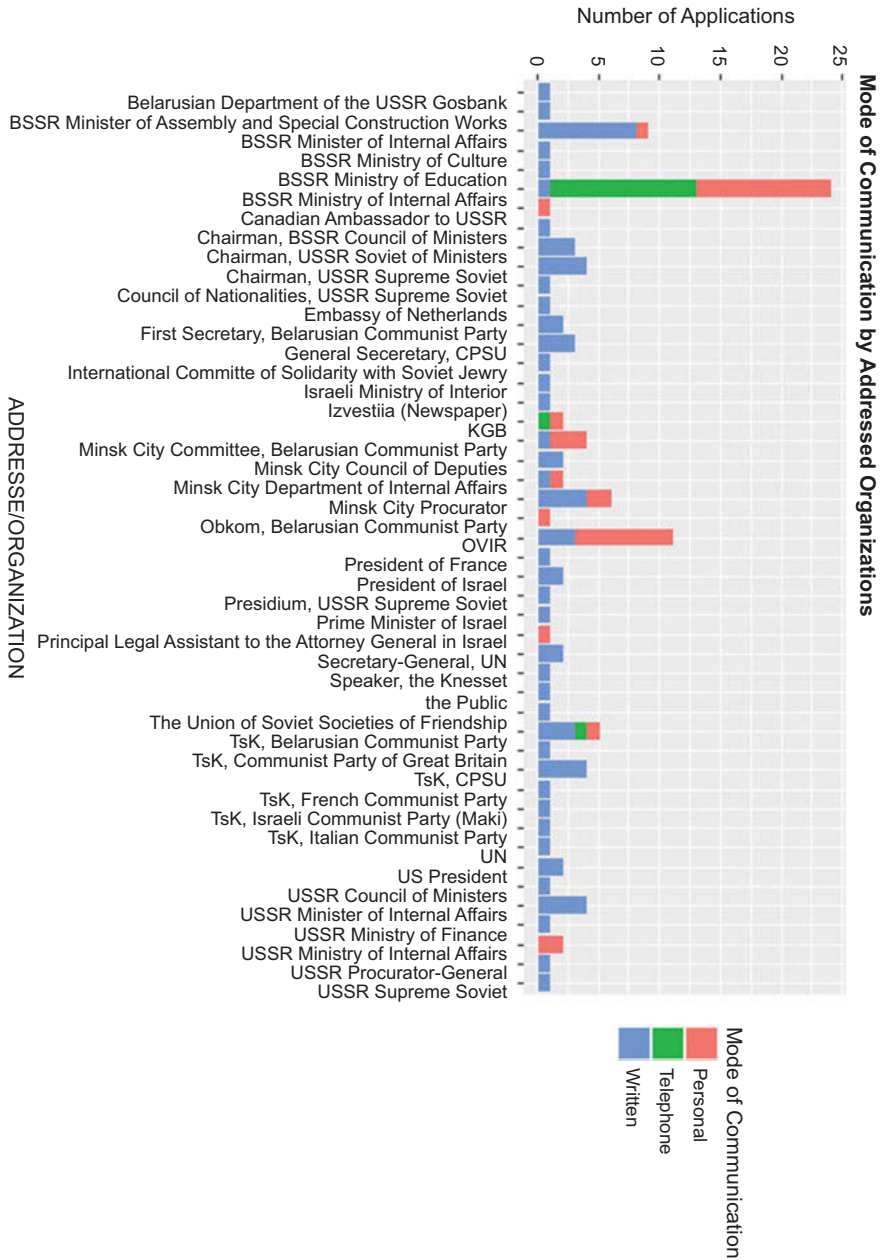**Figure 8:** Forms of application by addressed organizations.

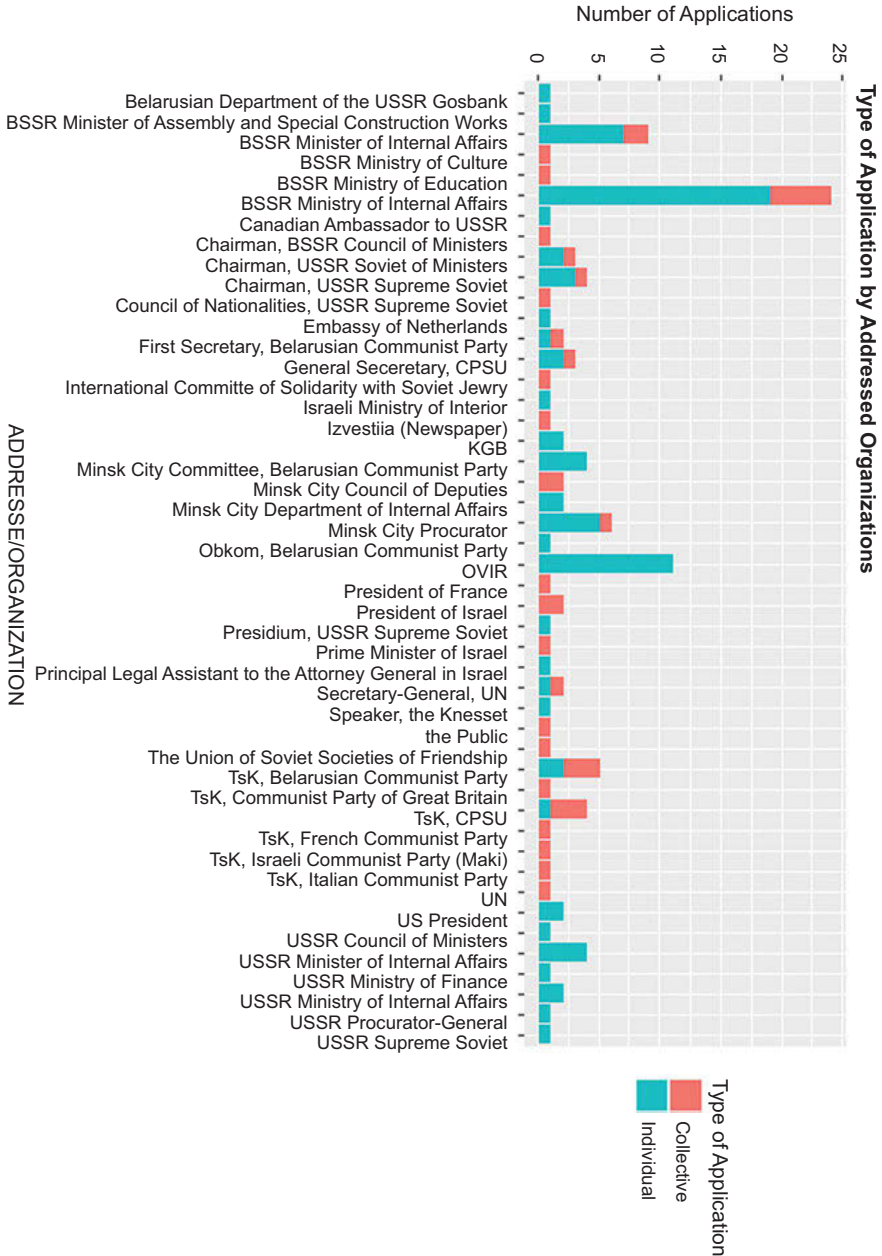**Figure 9:** Mode of communication by addressed organizations.

**Figure 10:** Type of application by addressed organizations.

istry, though individual applications prevailed. And whereas local authorities were more often approached individually (Figure 10), collective applications were predominantly directed to international organizations.

Figures 11 and 12 represent an attempt to combine (and double-check) two previous questions regarding the correlation between time and the addressees of the applications and the way they were approached. The graphs confirm that during the period between August 1971 and February 1972 Levin had been most intensively communicating with the local authorities, after which he started to summon more international attention. This shift in communication practices was very likely motivated by the developments in Soviet international politics. On May 22–30, 1972, US President Richard Nixon visited Moscow. In January and June of the same year, the USSR hosted the newly elected UN Secretary-General Kurt Waldheim. Soviet dissidents and refuseniks alike anticipated them with high hopes.[50] Thus it appears that both collaboration with other refuseniks and international attention were instrumental in advancing his case.

The final pair of graphs (Figures 13 and 14) focuses on the destination and asks about the forms and types of applications sent by Levin. Minsk (which dominated considerably, with 77 appeals) and Moscow (27 appeals), the locations of the republican and the all-Union authorities respectively, were the main destinations for Levin petitions. All forms of applications, signed both individually and collectively, were sent to these destinations. There were certainly more individual applications (approximately 69.1% of the total number), but the number of the collective ones (see Figure 5) was still significant, though it drops during the last three months, when the Levins were getting ready to depart.

## 8  Conclusion

The ultimately successful outcome of Levin's emigration attempt was, it seems, a consequence of both personal initiative and agency, as well as developments in international politics, in particular the relaxation of Soviet-US relations. The Levins were lucky to hit the first big wave of emigration and as Levin's case was among the first in Soviet Belarus, it attracted the attention of Israeli and Western media and was closely observed and supported from abroad.[51] At first glance,

---

**50** The immediate effect of these visits was rather opposite: the potential "disturbers" were threatened or even preventively detained. "Knikson," KhTS 26 (July 5, 1972).

**51** See, for instance, Nitzon, "Third Wave of Arrests: Minsk Jews Told 'No Weeping,'" *Jewish Observer and Middle East Review*, March 10, 1972; Levin, *I posokh*, 97, 148–55.
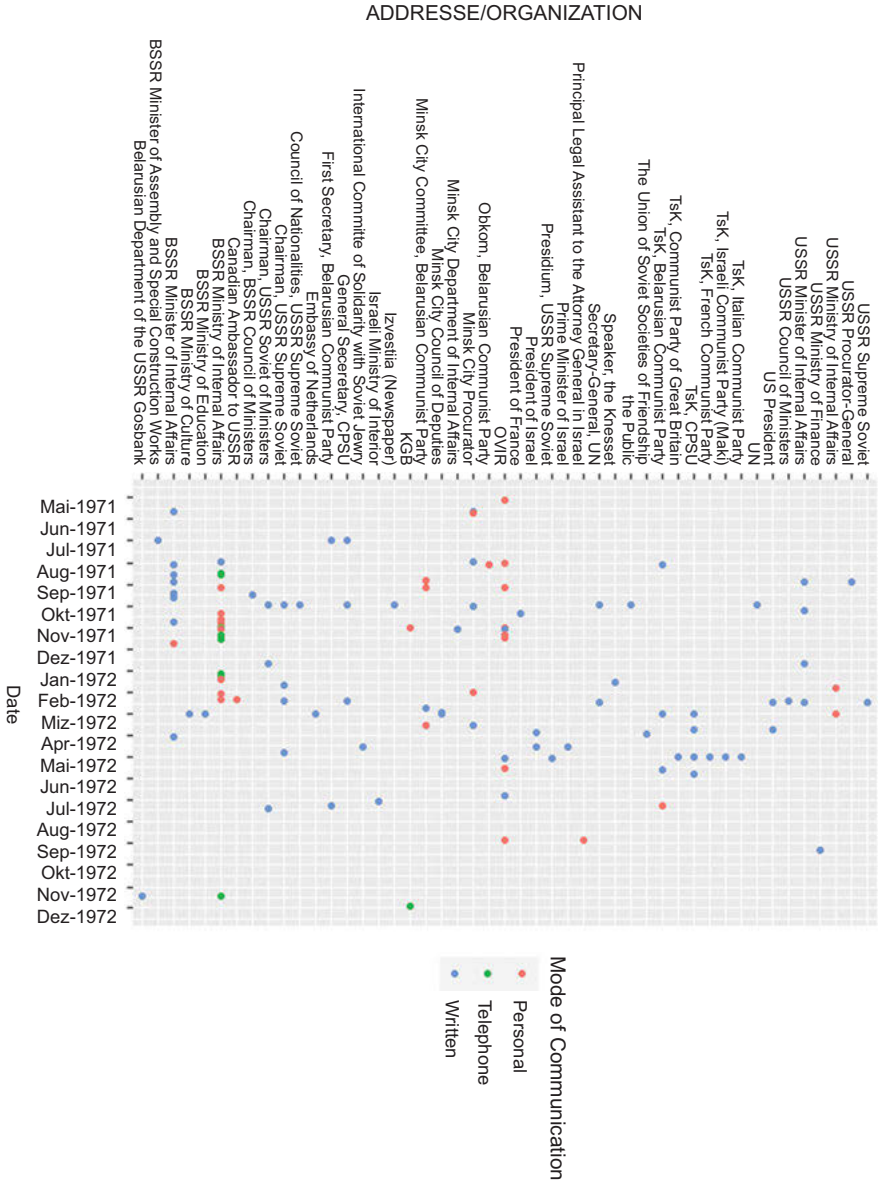
**Figure 11:** Addressed organizations by the mode of communication on a timeline.

**Figure 12:** Addressed organizations by the form of application on a timeline.
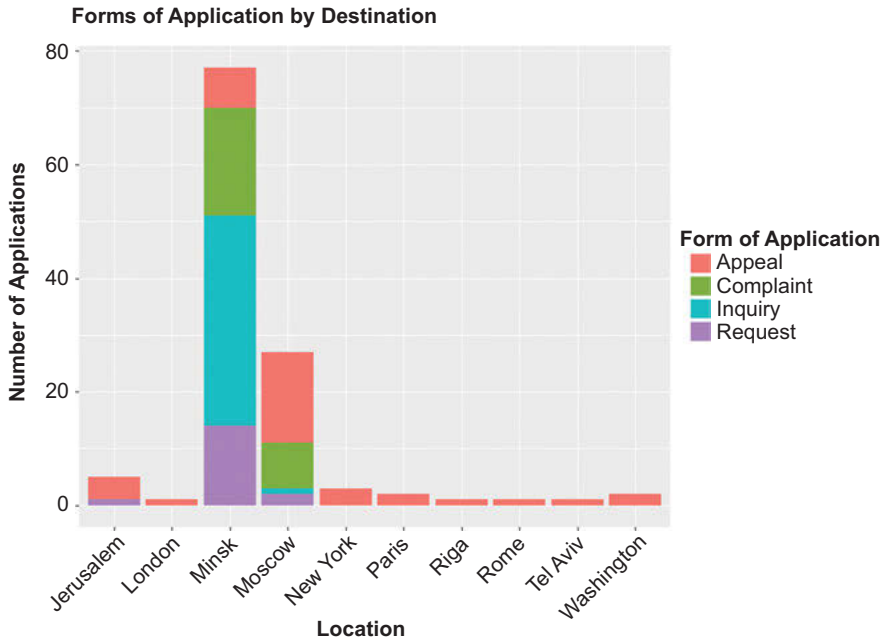
**Forms of Application by Destination**



**Figure 13:** Destination by the form of application.

then, Levin's case may be treated as typical or even not particularly remarkable within the history of Soviet Jewish emigration. His emigration was relatively quick and unproblematic; he did not face criminal charges and was not persecuted. Although time and again supporting his counterparts, Ernst Levin did not have larger political aims; his primary objective was his and his family's emigration.

And yet the emigration case of Ernst Levin is unique, because of his writing and organizational talent, his ability to formulate his petitions strictly to the letter of the relevant Soviet legislation, and his facility for adapting his appeals depending on the responses of the authorities. All this led him and his family towards successful emigration, which otherwise could have been a much more difficult and lengthy process.

Analyzing and visualizing his petitions with R reveals the specificity of Levin's case on various levels and allows for some generalizations. This becomes immediately apparent in answering the first question about the distribution of petitions over time, where visualization is instrumental in revealing patterns that are not necessarily obvious from an initial study of the petitions themselves. The distribution of petitions on a timeline serves to underline the sustainability of his effort over the period and helps explain the tactics that Levin used to optimize his emigration strategy. He switched from a more demanding line of communica-

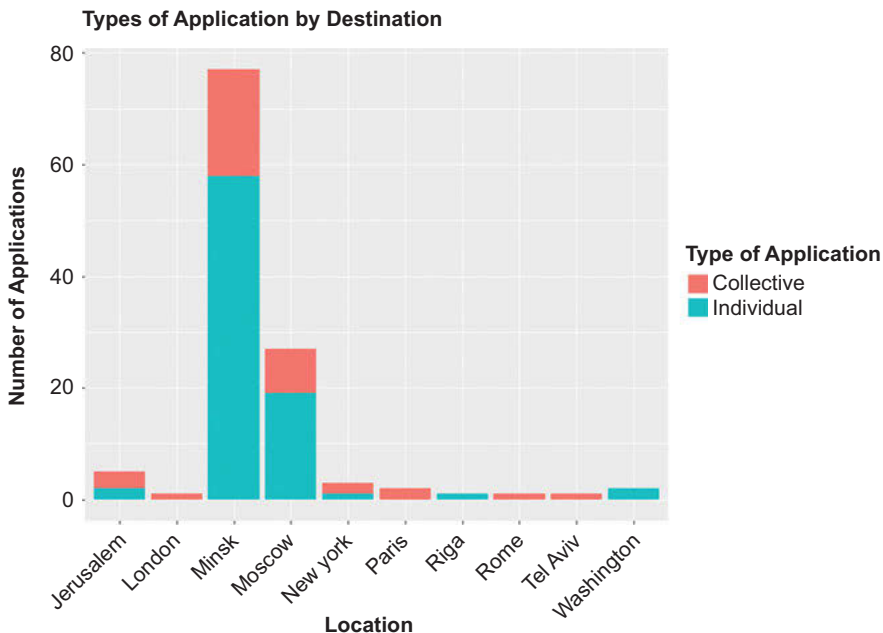**Types of Application by Destination**



Figure 14: Destination by the type of application.

tion with authorities to a more moderate one, collaborated with other refuseniks, and appealed for international support while reducing his communication to a minimum during the last three months of his stay in the USSR. Furthermore, he combined different forms of appeals depending on the context and, presumably, responses he obtained from the authorities.

The number and frequency of appeals that Levin made to the Belarusian authorities was remarkable. Even though the final decision about the status of his emigration could have been taken in Moscow, the main communication happened inside the BSSR. This may hint at the meaning and specificity of the emigration process as it took place in the Soviet periphery, though more data displaying the cases of other refuseniks is needed to confirm this assumption. Likewise, the graphic display of Ernst Levin's petitions to the Soviet authorities and international organizations brings to light the emigration efforts on the periphery of the Soviet Union during its initial phase (in the early 1970s). Even if these efforts, in this case in the Jewish community in Minsk, were most probably inspired by Jewish activity in Moscow, Leningrad, Kyiv, Kharkiv, Riga, and other big cities across the USSR, at this moment they were largely isolated.

The data visualizations highlight two important periods of intensification of Ernst Levin's struggle for emigration – fall 1971 and spring 1972, which concurred

with an increase in refusenik activity in Minsk. In the fall of 1971 it resulted in the approval of the application for several Belarusian Jewish families.[52] In the spring of 1972 the first collective public event, commemorating the victims of the Holocaust in Minsk, took place. The Jewish struggle for emigration in Minsk now reached a new level, as emigration efforts and memory work collided and created new momentum for the former.

To some extent, addressing the authorities was already an act that embodied change, as a common citizen transformed into a political subject with his or her agency becoming apparent.[53] In this way, he or she entered into relations with state institutions as an equal and full (at least outwardly) subject. Ironically enough, subjects become citizens at the moment they decide to abandon their citizenship. The visualizations display this emerging agency in detail.

By working with a small dataset we can zoom in from the general category of Soviet Jews, and more precisely of Soviet refuseniks, to the local community where the individual decision of emigration, as well as individual effort and personal connections within and outside the Jewish community, becomes visible. In this regard, a further study of other refusenik groups across the Soviet Union could yield a fruitful perspective for comparison.

The use of R in this study is rather arbitrary as it can arguably be substituted by tools available in other programming languages, especially Python. Nevertheless, R proved to be a handy tool for the analysis and visualization of the small dataset, in which time and effort expenditures are balanced and proportional to the results. Importantly, using visualizations was necessary but is not sufficient for a full understanding of the emigration process; to achieve the latter a combination of digital *and* traditional methods remains necessary, through an approach where computational and human readings complement each other.

This study demonstrates, how a relatively small amount of data can be analyzed and explored to yield new results. Especially in Jewish Studies with its great variety of languages and cultures spanning many centuries and regions, attention to "small" histories can not only enrich our existing knowledge but also help to reframe and revise existing research.[54] Thus, as the story of Ernst Levin and its digital interpretation shows, much potential remains to further explore the history of Eastern European Jewry, and the Jewish movement for emigration from the USSR in particular.

---

**52** Levin, *I Posokh*, 59.

**53** Compare to Fürst, "In Search of Salvation," 328.

**54** See also Gerben Zaagsma, "#DHJewish – Jewish Studies in the Digital Age," *Medaon – Magazin für jüdisches Leben in Forschung und Bildung* 12, no. 23 (2018): 1–11.

# References

Alexeyeva, Ludmilla. *Soviet Dissent: Contemporary Movements for National, Religious, and Human Rights*. Middletown, CT: Wesleyan University Press, 1984.

Alexopoulos, Golfo. "The Ritual Lament: A Narrative of Appeal in the 1920s and the 1930s." *Russian History* 24, no. 1–2 (1997): 117–29.

Altshuler, Mordechai. "Who Are the 'Refuseniks'? A Statistical and Demographic Analysis." *Soviet Jewish Affairs* 18, no. 1 (1988): 3–15.

Balandin, K.I. et al. *Istoriia Belorusskogo natsional'nogo tekhnicheskogo universiteta* (History of the Belarusian National Technic University). Minsk: BNTU, 2010.

Bohn, Thomas. *Minsk – die Musterstadt des Sozialismus. Stadtplanung und Urbanisierung in der Sowjetunion nach 1945*. Cologne: Böhlau, 2008.

"Decree on the Procedure for the Consideration of Proposals, Declarations, and Complaints of Citizens." Ukaz N 2534-VII Prezidiuma Verkhovnogo Soveta SSSR o Poriadke rassmotrenia predlozhenii, zaiavlenii i zhalob grazhdan. Accessed February 13, 2021. http://docs.cntd.ru/document/9012207.

Dennighaus, Viktor and Andrei Savin. "'Kak by Ukaz o evreiiakh ne otmeniat', a de faktom ne primeniat' L.I. Brezhnev, razriadka i evreiskaia emigratsia iz SSSR v 1972–1973 gg." *Rossia XXI vek* 1 (2013): 130–59.

Drachinsky, Mark. "A Brief Survey of the History of Hebrew Teaching in USSR." In *Jewish Culture and Identity in the Soviet Union*, edited by Ya'acov Ro'i and Avi Beker, 246–54. New York: NYU Press, 1991.

Fitzpatrick, Sheila. "Editor's Introduction: Petitions and Denunciations in Russian and Soviet History." *Russian History* 24, no. 1–2 (1997): 5–6.

Friedman, Alexander. "Antizionismus und Anti-Masonismus in der Sowjetunion nach dem israelisch-arabischen Sechstagekrieg (1967). Der Verschwörungstheoretiker Vladimir Ja. Begun (1929–1989)." In *Juden und Geheimnis. Interdisziplinäre Annäherungen*, edited by Claus Oberhauser, 137–51. Innsbruck: Innsbruck University Press, 2015.

Fürst, Juliane. "In Search of Soviet Salvation: Young People Write to the Stalinist Authorities." *Contemporary European History* 15, no. 3 (2006): 327–45.

Gitelman, Zvi. *A Century of Ambivalence: The Jews of Russia and the Soviet Union, 1981 to the Present*. 2nd ed. Bloomington: Indiana University Press, 2001.

Grunewald, Susan. "A Push for Digital History in Soviet and Post-Soviet Studies." NYU Jordan Center for the Advanced Study of Russia. Accessed July 15, 2021. https://jordanrussiacenter.org/news/a-push-for-digital-history-for-soviet-and-post-soviet-studies/#.YSjUGogzaUm.

"Han'ba sijanistskim pravakataram" (Shame on Zionist instigators). *Zviazda* 17 (January 21, 1972), 3.

Howarth, Sharon. "Five Reasons for Historians to Learn R." Accessed July 17, 2021. https://www.dataquest.io/blog/five-reasons-for-historians-to-learn-r/.

Kantar, Safiia. "Chamu ia ne paedu ŭ Izrail" (That is why I am not going to Israel). *Zviazda* 33 (February 9, 1972), 3.

Kas'ianov, Georgii. *Nezgodny: Ukrainskaia inteligentsia v rusi oporu 1960-kh–1980-x rokyv*. Kyiv: Lybid', 2005.

Kazakov, Evgenii. "V poiskakh 'sovetskogo evreiskogo': Poznesovetskaia natsional'naia politika." *Neprikosnovennyi zapas* 6 (2018): 190–215.

Khanin, Vladimir (Ze'ev). "The Refusenik Community in Moscow: Social Networks and Models of Identification." *East European Jewish Affairs* 41, no. 1–2 (2011): 75–88. doi:10.1080/13501 674.2011.591661.

Klinau, Artur. *Minsk – Sonnenstadt der Träume*. Berlin: Suhrkamp, 2011.

"Knikson." KhTS 26 (July 5, 1972).

Komaromi, Ann. "Between Two Worlds: Late Soviet Jews in Leningrad." *East European Jewish Affairs* 48 (2018): 23–40. Accessed December 13, 2022. https://doi.org/10.1080/1350167 4.2018.1442046.

Konstitutsia (Osnovnoi zakon) Soiuza Sovetskikh Sotsialisticheskikh Respublik 5 dekabria 1936 goda (The Constitution of the Union of Soviet Socialist Republics adopted on December 5, 1936). Accessed March 2, 2021. http://www.hist.msu.ru/ER/Etext/cnst1936.htm.

Levin, Ernst. *I posokh v ruke vashei. Dokumental'nyi memuar 2002 goda (k tridtsatiletiiu iskhoda iz SSSR)*. Jerusalem: n.p., 2007.

Martin, Barbara. "The Sakharov-Medvedev Debate on Détente and Human Rights: From the Jackson-Vanik Amendment to the Helsinki Accords." *Journal of Cold War Studies* 23, no. 3 (2021): 138–74. Accessed December 13, 2022. https://doi.org/10.1162/jcws_a_01009.

Mommsen, Margareta. *Hilf mir, mein Recht zu finden. Russische Bittschriften von Iwan dem Schreiklichen bis Gorbatschow*. Berlin: Propyläyen, 1987.

Mullen, Lincoln. "Digital History Methods in R." Accessed July 17, 2021. https://lincolnmullen. com/projects/dh-r2/introduction.html.

Nathans, Benjamin. "The Dictatorship of Reason: Aleksandr Volpin and the Idea of Rights under 'Developed Socialism.'" *Slavic Review* 66, no. 4 (Winter 2007): 630–63.

Nitzon. "Third Wave of Arrests: Minsk Jews Told 'No Weeping.'" *Jewish Observer and Middle East Review*, March 10, 1972.

Panagiotidis, Jannis. *The Unchosen Ones: Diaspora, Nation, and Migration in Israel and Germany*. Bloomington: Indiana University Press, 2019.

Remennick, Larissa. *Russian Jews on Three Continents: Identity, Integration, and Conflict*. New Brunswick, NJ: Transaction Publishers, 2012.

Savin, Andrei. "'U Gorbacheva i El'tsina ne bylo politicheskoi voli.' Pochemu Rossia ne izbavilas' ot sovetskikh nazvanii gorodov i ults." *Lenta.ru*, June 7, 2020. Accessed November 29, 2021. https://lenta.ru/articles/2020/07/07/savin_2/.

Shroetter, Leonard. *The Last Exodus*. Jerusalem: Weidenfeld and Nicolson, 1974.

Simpson, Duncan. "Approaching the PIDE 'From Below': Petitions, Spontaneous Applications and Denunciation Letters to Salazar's Secret Police in 1964." *Contemporary European History* 30, no. 3 (2021): 398–413. doi:10.1017/S0960777320000612.

Slezkine, Yuri. *The Jewish Century*. Princeton, NJ: Princeton University Press, 2019.

Stola, Dariusz. *Kraj bez wyjścia? Migracje z Polski 1949–1989*. Warsaw: ISP PAN, 2010.

"Timeline of the Jewish Movement in the Soviet Union." University of Toronto Libraries. Accessed October 17, 2020. https://samizdatcollections.library.utoronto.ca/content/ timeline-jewish-movement-soviet-union.

"Total Immigration to Israel from the Former Soviet Union (1948–present)." Accessed February 5, 2021. https://www.jewishvirtuallibrary.org/total-immigration-to-israel-from-former- soviet-union.

Vail, Petr and Aleksandr Genis. *60-e. Mir sovetskogo cheloveka*. Moscow: Corpus, 2013.

*Vsesoiuznaia perepis'vnaseleniaia 1939 goda* (The all-Union population census of 1939). Accessed March 1, 2021. http://www.demoscope.ru/weekly/ssp/ussr_nac_39_ ra.php?reg=625.

White, Stephen. "Political Communications in the USSR: Letters to Party, State and Press." *Political Studies* 31 (1983): 43–60.

Yilmaz, Hale. "Petitions as a Source in Women's History of the Republican Period." In *Women's Memory: The Problem of Source*, edited by Fatma Türe and Birsen Talay Keşoğlu, 81–93. Newcastle upon Tyne: Cambridge Scholars Publishing, 2011.

Zaagsma, Gerben. "#DHJewish – Jewish Studies in the Digital Age." *Medaon – Magazin für jüdisches Leben in Forschung und Bildung* 12, no. 23 (2018): 1–11.

Zhitnitskii, Isaak (Itskhak Bar-Zait). "Iskhod iz Minka, 1971 god." Haifa (2014). Evreiskaia Wiki-Entsiklopedia/Everyday Jewish Wiki. Accessed November 22, 2021. http://www. ejwiki.org/wiki/Welcome_to_Everyday_Jewish_Wiki.

Zubok, Vladislav. *Zhivago's Children: The Last Russian Intelligentsia*. Cambridge, MA: Belknap Press, 2009.

# Computational

Luigi Bambaci

# Digitizing Kennicott's Collation of the Hebrew Bible

## Experiences of Encoding and of Computer-assisted Stemmatic Analysis

**Abstract:** This article describes the creation of a rule-based parser for digitizing Kennicott's collation of the Hebrew Bible (Section 1). We will illustrate how it is possible to exploit the tree-like structure inherent in the critical apparatus of Kennicott's collation in order to generate XML code automatically via a context-free grammar (Section 2). Finally, we will present an experiment of computer-assisted stemmatic analysis of the manuscript tradition of the book of Qohelet, which was made possible by encoding (Section 3). Both the digitization of Kennicott's apparatus and the stemmatic analysis of the textual tradition are part of an ongoing project devoted to the preparation of a born-digital eclectic edition of the book of Qohelet.

**Keywords:** Hebrew manuscripts, textual history of the Hebrew Bible, text encoding, computer-assisted stemmatology, Digital Humanities

# 1 Introduction

The collations of Kennicott (1776, 1778) (K) and De Rossi (1788, 1798) (DR) represent to this day our main source of information about the textual history of the Hebrew Bible (HB) from the late Middle Ages to the first centuries after the invention of printing.

K and DR gathered an enormous number of variant readings of the HB consulting thousands of textual witnesses, both manuscripts and printed editions. The work of K is particularly extensive: according to Barthélemy, his critical apparatus

contains something like 1,500,000 pieces of textual information,[1] collected from approximately 600 witnesses.[2] To this corpus DR added new data, collating more than 500 additional manuscripts and 200 printed editions.[3]

Until now, the two collations have been used primarily to compile critical editions and textual commentaries. In some studies, the data provided by the two 18-century collators are exploited more systematically for investigations of various kinds, from textual history to textual criticism.[4] Managing such a large amount of material efficiently is clearly difficult and cannot in fact be accomplished except by limiting oneself to relatively small samples. Even with these, however, it is often necessary to resort to mechanical assistance to extract information relevant for research.

Thus, the digitization of critical apparatus would represent for HB philologists an important advance, since it would permit a huge variety of analyses, both qualitative and quantitative, with the aid of the computer.

Proprietary languages or software that are normally used can, however, limit data exchange and hence prevent the possibility of verifying both data and methods adopted.

The markup language promoted by the *Text Encoding Initiative* (TEI)[5] is an optimal solution in this respect. It permits us to query critical apparatuses and to extract information in an efficient way, whatever the sample size to be examined. TEI international standards also allow scholars to exchange data easily, which opens up the possibility not only to control results and methods more effectively, but also to reuse the data for further research.

In the following sections, we propose to take a stab at a concrete implementation of all of this, using as our object of study a small book of the HB, namely, Qohelet (Q).[6] We will describe the procedures which enabled us to encode the critical apparatus of Q as printed by K, and we will thereafter demonstrate how

---

**1** D. Barthélemy, *Critique textuelle de l'Ancien Testament, 1. Josué-Esther*, vol. 1 of *Orbis Biblicus et Orientalis 50* (Fribourg and Göttingen: Éditions Universitaires and Vandenhoeck & Ruprecht, 1982), 28 ff.

**2** E. Tov, *Textual Criticism of the Hebrew Bible*, 3rd ed. (Minneapolis: Fortress Press, 2012), 37.

**3** G.B. De Rossi, *Variae lectiones Veteris Testamenti* (Parma: Ex regio typographeo, 1784–1788), 1:xlvi–i.

**4** See in particular the works of M. Cohen and J. Penkower and the studies in stemmatology quoted in Section 5.

**5** Accessed February 26, 2021, https://tei-c.org/.

**6** B. Kennicott, *Vetus Testamentum Hebraicum cum variis lectionibus*, vol. 2 (Oxford: Clarendon, 1778), 549–61.

we processed the encoded data to carry out a special kind of quantitative inquiry, namely a computer-assisted stemmatic analysis.[7]

# 2  Encoding Through a Parser

Encoding a critical apparatus of printed editions or collations is a complex task. Critical apparatuses are designed to express maximum content within minimum space, so the data they contain are highly information-dense. As a consequence, such data need to be encoded in a fine-grained manner, employing a rich set of markers suitable for making the function of each apparatus component explicit for the machine. This makes manual encoding a time-consuming and error-prone enterprise.

However, when the language of a critical apparatus is sufficiently formalized, it is possible to automate the encoding operation by means of a parser, that is, a software capable of analyzing formal languages and thus of recognizing the various parts of which a critical apparatus is composed.

The method we adopted to obtain a compliant XML-TEI encoding from a copy in .pdf of K's printed apparatus (§ 2.1) consisted of four phases described in detail in this section: (1) digitization of the original source through optical character recognition technology (OCR, § 2.2); (2) creation of a context-free grammar (CFG) for describing the language of the critical apparatus (§ 2.3); (3) implementation of a general exporter to produce XML code (§ 2.4); (4) composition of XSL-T stylesheets to convert the XML into XML-TEI (§ 2.5).[8]

## 2.1  Kennicott's Critical Apparatus

K employs a rigorous annotational system to record variant readings. K's apparatus, unlike that of DR, strongly minimizes the use of natural language: each entry is organized within a structure which lists the apparatus components in a pre-established order by utilizing an unambiguous terminology and a finite set of

---

**7** The data from K's collation taken into consideration here as well as the tools used to process them – from the context-free grammar to XSL-T stylesheets – can be found on our Github repository: https://github.com/LuigiBambaci/Kennicott.

**8** For more technical details about the implementation of the parser we refer to our recently published article L. Bambaci, "Critical Apparatus as Domain Specific Languages: A Rule-Based Parser for Encoding an Eighteenth-Century Collation of Hebrew Manuscripts," *International Journal of Information Science and Technology* 5, no. 1 (2021): 22–33.

typographical conventions. This feature makes κ's apparatus highly suitable for automated processing through user-defined rules, as we are about to show.

Let us begin by examining a few examples of κ's apparatus taken from Q 1:1 (Figure 1). As can be seen, the variants are listed by verses, which are marked by numbers followed by dots ('1.', '2.' etc.). Within a verse, one finds the apparatus entries, each delimited by a dot and a long white space.[9] An apparatus entry represents a place of variation, that is, a place in the text for which variant readings are attested. For example, the following entry

76. **ירושלים** — sup. ras. 139 — 107, 109, 152 ** בירושלים**

means that there exist three readings of the word in the reference text (the lemma **בירושלם**) as printed by κ (the edition of E. van der Hooght, Amsterdam 1705): one with *scriptio plena* (**בירושלים**), one written over an erasure (*sup. ras.*), and the third with *scriptio plena* and the addition of a noun (**יהודה בירושלים**). The three groups of witnesses – expressed here as numerical *sigla* – have therefore three different readings, each separated by a long horizontal line ('—').

More complex cases can occur, as in the following entry:

**דויד** 57, 100, 260; forte 141.

Here, the annotation means that witnesses 57, 100, 260 have **דויד** instead of **דוד** of the reference text, and that the same is attested also in 141, but as a probable first-hand reading (*forte*). In this case, therefore, there is only a single group of witnesses with reading **דויד**; in the third witness, however, the reading has probably undergone correction by a second scribe, who changed **דויד** into **דוד** following the reference text. To separate witnesses containing particular readings of this sort, κ uses semicolons or, less frequently, commas.

Thus, κ's apparatus may be divided into three main units: apparatus entries, reading groups, and individual readings. The basic unit is the individual reading: here we find Hebrew words, witness *sigla* and standard Latin annotations that specify typographical details (e.g. *lit. majorib.*), the copyist's hand (*primo/forte*, *nunc*), and other phenomena.[10] Witnesses sharing readings are listed within larger units, namely the reading groups.

---

**9** For the sake of simplicity, these are encoded as tabulations in the transcription contained in the .txt file (see § 2.2).

**10** A description of κ's apparatus can be found in the introduction of the first volume of the collation, see B. Kennicott, *Vetus Testamentum Hebraicum cum variis lectionibus*, vol. 1 (Oxford: Clarendon, 1776), i–iv, 68 ff.

These three elements – apparatus entry, reading group, and individual reading – must always be present for each verse. Other elements are, by contrast, optional, for example the lemma: when the link to the reference text is rather immediate, the lemma is omitted altogether, as in the examples just illustrated; in cases, however, where there is no such immediacy, the lemma is explicitly reported at the beginning of an apparatus entry (e.g. at verse 3 in Figure 1), followed, in case of ambiguity, by the number specifying the occurrence of the word ('1º', '2º' etc.).

Other phenomena can occur, such as marginal readings, *lacunae*, transpositions, etc., each of which corresponds to a precise type of annotation in the apparatus.

ᴋ's language is, therefore, highly formalized: each element is easily definable by the 'type' it belongs to (alphabetic character, numeral, special symbol) and by the position it occupies within the overall structure. As we shall see, these two features – string class and syntax – are sufficient to instruct a machine to identify automatically the apparatus components and to assign to them a function, without recourse to direct user intervention to render such information explicit.

## 2.2  Optical Character Recognition

The first step in making the data machine-readable is digitization by means of ᴏᴄʀ software. For this purpose we acquired a `.pdf` version of ᴋ's collation (Figure 1a), which is freely available on platforms such as Google Books and Archive, and then we used the software Tesseract[11] to obtain a digitized version of the critical apparatus.[12]

ᴋ's apparatus is particularly difficult to digitize because of the presence of two different alphabets (Latin and Hebrew) and of special symbols, and because of the need to retain most typographical details of the printed source such as new-lines and tabulations, which are crucial for parsing. The output, therefore, needs to be corrected through a careful phase of pre-processing. This phase, however, can be semi-automated by using a ᴄꜰɢ: since a ᴄꜰɢ, as we will explain later (§ 2.3), permits us to check whether a given language is valid according to a set of pre-defined formation rules, it becomes possible to detect automatically certain types of errors produced by ᴏᴄʀ processing, such as those concerning misspellings of technical terms and incorrect segmentations of the apparatus structure.

---

**11**  Accessed February 26, 2021, https://tesseract-ocr.github.io/.
**12**  The reference text printed above the critical apparatus was not digitized.

This way, manual correction of the OCR results can be significally improved, as we will discuss in Section 4.

After revising the results produced by the OCR software, we obtained a `.txt` file containing the transcription of the original `.pdf` source (Figure 1b). This file was the one we used for our parsing analysis, as related in the following sections.

(a) Fragment of K's apparatus in `.pdf`

(b) Fragment of K's apparatus in `.txt`

**Figure 1:** Conversion from `.pdf` to `.txt` (example from Q 1:1–4).

## 2.3 The Context-Free Grammar

Before building the parser with a view to an automatic encoding, we opted to encode Q's apparatus manually, for two main reasons: (1) to study how the various textual phenomena are represented in K and how these could be properly expressed in a TEI-compliant encoding; and (2) to have a benchmark against which to compare the automated encoding and by which to verify the accuracy of the parsing operation (§ 2.6).

Then, we wrote a CFG by wielding the tools available in ANTLR4, a software designed to generate parsers from CFGs.[13] A CFG is a grammar consisting of a set of rules that permits one to analyze a formal language. The analysis is carried out by means of two kinds of rules: lexer rules and parser rules. The former permit the isolation of tokens from the textual flow. The latter describe the syntax, that is, how the different tokens distribute and combine in the apparatus. In ANTLR4,

---

**13** Accessed February 26, 2021, https://www.antlr.org/. See also T. Parr, *Language Implementation Patterns: Create Your Own Domain-Specific and General Programming Languages* (Dallas, TX and Raleigh, NC: Pragmatic Bookshelf, 2010) and T. Parr *The Definitive ANTLR 4 Reference* (Dallas, TX and Raleigh, NC: Pragmatic Bookshelf, 2012).

the first task is performed by the software component named lexer, the second by the parser.

An example of CFG is shown in Listing 1.[14]

**Listing 1** Context-free grammar.

```
1   grammar kennicottCFG;
2   all: listApp+;
3   listApp: loc app+;
4   app: lem? rdgGrp+ closeApp;
5   lem: w+ (lemSep|occ);
6   rdgGrp: rdg+ rdgGrpSep?;
7   rdg: term? (w+)? term? wits
    rdgSep?;
8   w: HEBW;
9   loc: verse closeLoc;
10  term: MAN_DESC;
11  wits : wit+;
12  sigl: NUM;
13  wit: sigl com?;
14  lemSep: VAR_SEP;
15  rdgGrpSep: VAR_SEP;
16  com: COMMA;
17  rdgSep: COMMA|SEMICOLON;
18  numSign: NUMEROSIGN;
19  verse: num;
20  num: NUM;
21  occ: num numSign;
22  closeMainApp: NEWLINE;
23  closeApp: END (TAB|NEWLINE);
24  closeLoc: END;

25  MAN_DESC: '\u2038' | 'forte' |
    'sup. ras.' | 'lit. majorib.' |
    'vox major, et ornata ;' | 'non
    major ;' | 'primo' | 'forte' |
    'bis';
26  NUM: [0-9]+;
27  HEBW: [\u0590-\u05ff]+;
28  NUMEROSIGN: '°';
29  VAR_SEP: '—';
30  END: '.';
31  TAB: '\t';
32  COMMA: ',';
33  SEMICOLON: ';';
34  NEWLINE: '\r\n';
35  WS: ' ' -> skip;
36  LTR: '\u200E' -> skip;
37  RTL: '\u200F' -> skip;
```

In the grammar, the parser rules are listed at the beginning, in lowercase; the lexer rules are found at the end, in uppercase. In the lexer rules we defined the tokens, which represent the minimal meaningful units of the apparatus: here we encoded, for example, the Hebrew Unicode characters (**HEBW**) necessary to detect the Hebrew words, the list of Latin annotations (**MAN_DESC**), the class of numerals (NUM) by which the witness *sigla* are expressed, and a series of separators (such as **VAR_SEP, END, TAB**), which are important because they segment the critical apparatus into its minimum components. The elements we want to ignore, such as white spaces and other special characters, are skipped (**-> skip**). As can be

---

**14** The grammar we present here is a simplified version of the one used to parse the entire apparatus of Q. It can be used to parse the example we have shown, verses 1–4. The full version of the grammar can be found on Github, see note 7.

seen, the lexer rules are defined by means of regular expressions (e.g. '`[0-9]+`' in **NUM**) or literal values (e.g. the list of technical terms and symbols in **MAN_DESC**).

Once defined, the lexer rules are passed to the parser, which checks the syntax through the parser rules. Let us look briefly now at how the parser works. In the example, it begins from the root-rule (**all**), which specifies that the language we want to analyze consists of one or more lists of apparatus entries (**listApp+**). Then it moves to **listApp**, which states that a list contains a location (**loc**) and at least one apparatus entry (**app+**). For each of these rules, the parser looks for child rules, all the way down to the lexer rules that contain the input data. In the case of the rule **loc**, for example, the parser first matches verse and closeLoc, which state that a verse consists of numbers and a final dot. Finally, it reaches **NUM** and **END**, in which the tokens in question are listed. This approach – called top-down since it proceeds from general to specific statements – is followed for the other rules of the CFG, until all the elements of the apparatus are properly recognized and assigned.

Running the CFG on the plain text of Figure 1b produces a parse tree, fragments of which are shown in Figure 2.

As can be seen, each apparatus component is attached to a label (technically corresponding to the name of the parser rule of the CFG), which acts as a mnemonic for the user, of the function it plays in the apparatus. The tokens are found at the bottom of the tree, represented as tree leaves.

## 2.4 General Exporter

The third step consisted in the creation of an XML general exporter, in order to produce XML code containing the results of the parser analysis.

We exploited a tree-walking mechanism made available in ANTLR4, named visitor. In very general terms, it can be said that the visitor traverses the tree's nodes and transforms them into XML tags. As can be seen from the code shown on the left of Listing 2, the tag names are taken from the parser rules, which in the parse trees of Figure 2 are represented as internal nodes (§ 2.3).

Although many of the tag names are borrowed from TEI terminology, and in particular from the vocabulary of Module 12 devoted to the encoding of critical apparatuses,[15] the XML code generated by the general exporter is not TEI, but is

---

**15** TEI Consortium, eds., "12 Critical Apparatus." TEI P5: Guidelines for Electronic Text Encoding and Interchange, Version 4.2.0, last updated February 25, 2021, TEI Consortium, accessed February 26, 2021, https://tei-c.org/release/doc/tei-p5-doc/en/html/TC.html.
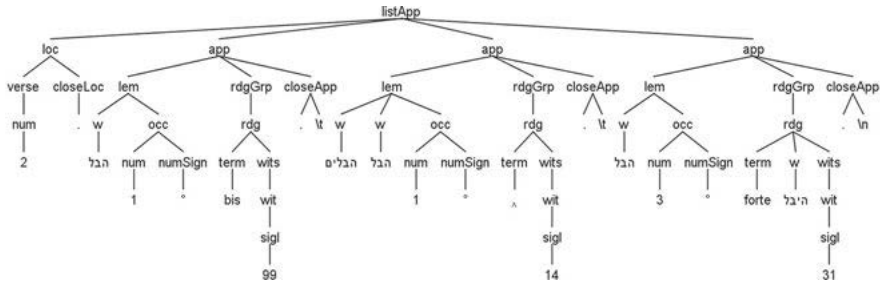
(a) Q 1:2
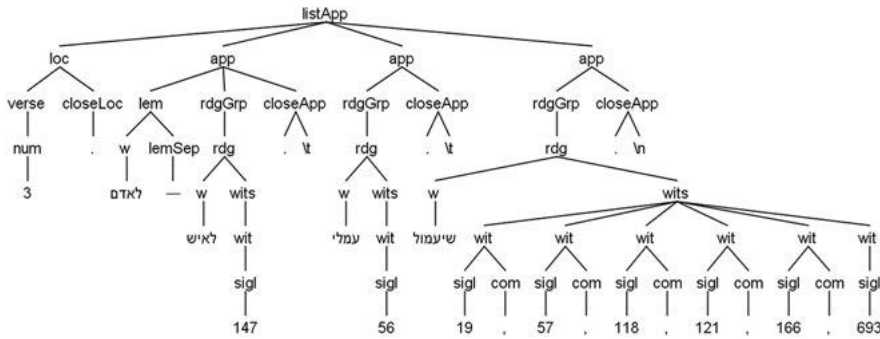
(b) Q 1:3

**Figure 2:** Examples of parse trees.

a proprietary language. In fact, it consists of a series of XML elements containing all the data of the original apparatus, including para-textual elements such as tabs and newlines. To generate a TEI critical apparatus, an XSL transformation is needed, as we describe below.

## 2.5 From XML to TEI

The fourth and last step consisted in designing an XSL-T stylesheet for converting the XML code into a compliant XML-TEI encoding (Listing 2, on the right).

**Listing 2** Conversion from XML to XML-TEI encoding (example from Q 1:1).

```
1   <listapp>
2    <loc>
3     <verse>
4      <num>1</num>
5     </verse>
```

```
29  <listapp>
30   <app loc="1 1">
31    <lem>
32     <w>דברי</w>
33    </lem>
```

```
 6    <closeLoc>.</closeLoc>        34    <rdgGrp>
 7    </loc>                        35      <rdg wit="#K136 #K139">
 8    <app>                         36        <term>vox major et ornata</term>
 9     <rdgGrp>                     37      </rdg>
10      <rdg>                       38    </rdgGrp>
11        <w>דברי</w>               39    <rdgGrp>
12        <w>קהלת</w>               40      <rdg wit="#K1 #K2 #K3 …">
13        <term>lit. majorib.</term> 41        <term>non major</term>
14        <wit>                     42      </rdg>
15          <sigl>4</sigl>          43    </rdgGrp>
16          <com>,</com>            44   </app>
17        </wit>                    45   <app loc="1 1">
18        <wit>                     46    <lem>
19          <sigl>109</sigl>        47      <w>דברי</w>
20        </wit>                    48      <w>קהלת</w>
21      </rdg>                      49    </lem>
22     </rdgGrp>                    50    <rdgGrp>
23    </app>                        51      <rdg wit="#K4 #K109">
24    <app>                         52        <term>lit. majorib.</term>
25     <lem>                        53      </rdg>
26      <w>קהלת</w>                 54    </rdgGrp>
27     </lem>                       55   </app>
      ...                                ...
28   </listapp>                     56   </listApp>
```

As can be seen, the file contains lists (`<listapp>`) of apparatus entries (`<app>`), consisting in turn of (lists of) reading groups (`<rdgGrp>`). The witnesses are encoded in the attribute `@wit` of the element `<rdg>`, while information about the status of the variant readings (e.g. the Latin annotations of the original apparatus) are encoded in the element `<term>`.

The stylesheet is customizable: here we decided to eliminate elements found in the printed source which are philologically irrelevant, such as separators and other typographical symbols.

The method chosen to link the apparatus to the text is the location-referenced method, which is the one usually recommended for digitizing printed critical editions.[16]

---

**16** TEI Consortium, eds., "12.2.1 The Location-Referenced Method." TEI P5: Guidelines for Electronic Text Encoding and Interchange, Version 4.2.0, last updated February 25, 2021, TEI Consortium, accessed February 26, 2021, https://tei-c.org/release/doc/tei-p5-doc/en/html/TC.html.

## 2.6 Evaluation

After we produced our XML-TEI code through XSL transformation, we compared the automatically encoded file with the one we had previously encoded by hand (§ 2.3), in order to assess the effectiveness of both parser analysis and XSL transformation.

The results showed that the parser as well as the XSLT stylesheet performed very well indeed: the apparatus was correctly parsed pursuant to the rules established in the CFG, and the approximately 2,600 variants[17] of Q have been correctly represented in a TEI-compliant encoding.

Such results should not, in fact, surprise us: the CFG was specifically designed to embrace all the textual phenomena found in the apparatus of Q, so that maximum accuracy in this case was rather expected.

We decided, therefore, to digitize the collation of the other Megillot,[18] and to subject it to the same procedures described above.[19] The goal was to test the robustness of the parser, that is, to verify whether the rules we set up in the CFG were valid only for the apparatus of the book of Q, in which case our model would have been overfitting and hence scarcely applicable, or whether they could be generalized, and hence used to digitize the other books of K's collation.

This time, several mismatches (syntactic errors) were reported by the parser during the analysis of the new data, meaning that it was not able to classify all the variants contained in the apparatus of these other biblical books.

Upon closer examination, however, such errors turned out to be few – less than a dozen, out of a total of about 4,000 variant readings – and of little importance: most are due to the presence of technical terms not listed in the CFG and of textual phenomena that do not occur in the book of Q. We therefore modified the grammar rules in order to encompass these other phenomena, and we ran the parser again with this latest version of the CFG. At this second running, the parser succeeded in analyzing the apparatus without syntactic errors.

As a final step, we selected several passages in the apparatus where more complex textual phenomena occur, such as double or marginal readings, and we analyzed the output produced by the parser in order to detect the presence of incorrect classifications (semantic errors). Although it cannot be completely ruled out that some entries were actually misinterpreted by the parser, the sample survey we carried out seems for the moment to exclude this possibility.

---

**17** The estimate refers to the number of `<rdgGrp>` elements.

**18** These are: Song of Songs (Kennicott, *Vetus Testamentum Hebraicum*, 2:525–33), Ruth (2:534–39), Lamentations (2:540–48), and Esther (2:562–72).

**19** Both `.txt` and `.xml` files of these books can be found on our GitHub repository, see note 7.

In conclusion, we feel confident in affirming the strong applicability of a rule-based parsing system to κ's collation, due in particular to the formal language of his apparatus. The CFG we designed, though initially developed in relation to the variant readings of a single biblical book, proved to be such a flexible and easy-to-implement tool that other books of κ's collation could be automatically encoded as well, using the identical methodology.

# 3 Computer-Assisted Stemmatic Analysis

Encoding converts textual data into machine-actionable format. This makes it possible to query the critical apparatus and to extract information from it through different kinds of techniques.

In textual criticism, one possible use of encoding of critical apparatuses is stemmatic analysis.[20] Since in an encoded critical apparatus the witnesses are aligned according to the readings they share, this information can be exploited in order to compute genealogical relationships between witnesses and to establish *stemmata codicum* following a (neo-)Lachmannian, computer-assisted methodology.[21]

In the domain of HB studies, the genealogical approach, of which Lachmann's method is one of the best known instances, is quite exceptional[22] and has never been applied, to the best of our knowledge, to medieval manuscripts. In contrast with New Testament studies, in fact, there have been few attempts to study the medieval Hebrew tradition using stemmatological criteria. The studies

---

[20] The basics of the stemmatic method are outlined in the classical work of P. Maas, "Textkritik," in *Einleitung in die Altertumswissenschaft*, 2nd ed., vol. 1.2, ed. A. Gercke and E. Norden (Leipzig: Teubner, 1927), 1–18 and P. Maas, *Textkritik*, 4th ed. (Leipzig: Teubner, 1960). For a recent introduction to stemmatology see P. Roelli, ed., *Handbook of Stemmatology* (Berlin and Boston: De Gruyter, 2020).

[21] The bibliography on computer-assisted stemmatology is extensive. Some of the classical works are P.T. van Reenen, M. van Mulken, and J. Dyk (1996). *Studies in Stemmatology*, vol. 1 (Amsterdam and Philadelphia: John Benjamins, 1996); B.J.P. Salemans, "Building Stemmas with the Computer in a Cladistic, Neo-Lachmannian, Way: The Case of Fourteen Text Versions of Lanseloet Van Denemerken" (Ph.D. thesis, Katholieke Universiteit Nijmegen, 2000); and P.T. van Reenen and M. van Mulken, *Studies in Stemmatology*, vol. 2 (Amsterdam and Philadelphia: John Benjamins, 2004).

[22] See B. Chiesa, "Textual History and Textual Criticism of the Hebrew Old Testament," in *The Madrid Qumran Congress*, ed. J.C. Trebolle Barrera (Leiden: Brill, 1992), 257–72; B. Chiesa, *Filologia storica della Bibbia Ebraica*, vol. 2 of *Studi Biblici 135* (Brescia: Paideia, 2000), 399 ff.; and Tov, *Textual Criticism of the Hebrew Bible*, 359 ff.

focused on stemmatological issues have been based on non-genealogical statistical methods,[23] and most of them have questioned even the possibility of applying the genealogical model to the transmission of the ʜʙ text in the Middle Ages.[24]

Here, we propose to investigate precisely that possibility, by exploiting phylogenetic algorithms from bioinformatics.[25] The method we employed is basically quantitative and was performed with the aid of the computer by way of the encoding. It consisted of four phases: (1) selection of variants considered to be genealogically significant (regularization, § 3.1); (2) survey of some of the manuscripts in ᴋ's collation as well as of others not included therein (§ 3.2); (3) conversion of philological data into a numerical format suitable to computational treatment (§ 3.3); and (4) stemmatic analysis through phylogenetic methods (§ 3.4).

## 3.1 Regularization

Since our task here is to recover *genealogical* relationships among the textual witnesses, it is necessary to exclude from the analysis all phenomena of textual change which are more likely due to coincident variation (polygenetic variants). To accomplish this, we have elaborated an annotation scheme which enables us to classify the variants on the basis of typological criteria, so as to select only the most significant ones from a genealogical point of view. In this way, the variants marked as irrelevant are skipped when generating the data matrix required by the phylogenetic software (§§ 3.3, 3.4).

An example of annotated code is shown in Listing 3. As can be seen, the annotation is carried out by means of tags recorded in the attribute `@ana` (short

---

**23** Such as the clustering algorithm employed in P. Sacchi, "Analisi quantitativa della tradizione medievale del testo ebraico della Bibbia secondo le collazioni del De Rossi," *Oriens Antiquus* 12 (1973): 1–13 and P.G. Borbone, *Il libro del profeta Osea. Edizione critica del testo ebraico* (Torino: Zamorani, 1990), 183–227.

**24** For a survey of such studies see D. Barthélemy, "Les manuscrits médiévaux et le texte tibérien classique," in *Critique textuelle de l'Ancien Testament, 3. Ézéchiel, Daniel et les 12 Prophètes*, vol. 3 of *Orbis Biblicus et Orientalis 50* (Fribourg and Göttingen: Éditions Universitaires and Vandenhoeck & Ruprecht, 1992), xix–xxvii.

**25** A more complete discussion of this analysis can be found in L. Bambaci, "Is a Stemma Possible for the Hebrew Bible? Towards a Genealogy of Medieval Manuscripts through Phylogenetic Analysis," *Materia Giudaica. Rivista dell'associazione italiana per lo studio del giudaismo* 26, no. 2 (2021): 3–30.

for 'analysis')[26] of the element **`<rdgGrp>`**. The tags are defined by the user[27] and can be divided into two main classes, corresponding to our distinction between accidental and substantial variants. Among the former we include, for example, variants of spelling, such as those of *scriptio plena* and *defectiva* (**`#scr_pl, #scr_def`**).

**Listing 3** Example of annotation (Q 1:1).

```
1  <listapp>
2    <app loc="1 1">
3      <lem>
4        <w>דברי</w>
5      </lem>
6      <rdgGrp ana="#typ">
7        <rdg wit="#K136 #K139">
8          <term>vox major et ornata</term>
9        </rdg>
10     </rdgGrp>
11     <rdgGrp ana="#typ">
12       <rdg wit="#K1 #K2 #K3 …">
13         <term>non major</term>
14       </rdg>
15     </rdgGrp>
16   </app>
17   <app loc="1 1">
18     <lem>
19       <w>דברי</w>
20       <w>קהלת</w>
21     </lem>
22     <rdgGrp ana="#typ">
23       <rdg wit="#K4 #K109">
24     <term>lit. majorib.</term>
25       </rdg>
26     </rdgGrp>
27   </app>
28   <app loc="1 1">
29     <lem>
30       <w>קהלת</w>
31     </lem>
32     <rdgGrp ana="#sub #sem #nm">
33       <rdg wit="#K121">
34         <w>קהוית</w>
35       </rdg>
36     </rdgGrp>
37   </app>
38   <app loc="1 1">
39     <rdgGrp ana="#scr_pl #nm">
40       <rdg wit="#K57 #K100 #K260">
41         <w>דויד</w>
42       </rdg>
43       <rdg wit="#K141">
44         <term>forte</term>
45       </rdg>
46     </rdgGrp>
       …
47   </app>
       …
48 </listapp>
```

---

Among the latter we include phenomena of addition (`#add`), deletion (`#del`), substitution (`#sub`), and transposition (`#tran`) of word(s). In the fragment of code shown in Listing 3, for example, the first three variant readings have been classified as typographical (`#typ`), since they refer to details of *mise en page* of the text; the last two, on the other hand, have been annotated, respectively, as cases of substitution and *scriptio plena* of nouns (`#sub #sem #nm, #scr_pl #nm`).

The basic assumption of the proposed distinction is that accidentals are more easily prone to random variation than substantials.[28] As a consequence, they may represent a problem for quantitative analysis, when we consider that they are the most frequent in the textual tradition.[29] Substantial variants, on the contrary, are assumed to be safer as indicators of genealogical kinship, and can offer a more solid basis for inferring descent patterns.[30]

Besides accidentals, we excluded two other categories of variants: second-hand variants (*nunc*), which may derive from external sources such as other codices or massoretic lists, and hence be the product of contamination; and variants attested in one witness only (*lectiones singulares*), which are useless in establishing relationships. Dubious variants (*videtur*, *forte*) and readings superimposed over erasures (*sup. ras.*) are excluded as well.

---

**28** This applies in particular to variants related to the use of *matres lectionis*. Other accidentals, such as some kinds of graphic variants (`#graph`), can be considered genealogically significant. An example is the alternation between סכלות/שכלות found in different passages in Q and perhaps attested in some of the ancient versions (see Q 1:17). This and other instances can be included in the analysis either by modifying the annotation (for example, from `#graph` to `#sub`, when the variant involves a possible change in meaning, as in the example just mentioned) or by applying specific tags which prevent the variant from being filtered during the XSL-T transformation phase (§ 3.3). In our annotation, the tag `#signif` (short for '[genealogically] significant [variant]') is devoted to this task.

**29** Sixty-three percent of the variants of Q are accidental according to our findings. Of these, 96% are variants of *scriptio plena* and *defectiva*.

**30** Of course, both the decision to filter accidentals, as well as to apply particular annotations to carry out the regularization, are, of necessity, the result of subjective interpretations about the logic of the copying process and the nature of textual change, and are as such questionable. The annotation scheme, on the other hand, has the advantage of making such interpretations explicit, thus allowing for the possibility of an intersubjective control. Moreover, this scheme allows for a customizing of the selection of kinship-revealing variants and permits us to automate the regularization pursuant to the criteria chosen.

## 3.2 Witnesses

The reliability of ĸ's collation has occasionally been questioned by scholars.[31] The criticisms concern in particular the lack of distinction between first and second hand,[32] and more generally errors and inaccuracies in reporting variant readings in the critical apparatus. Clearly, if criticisms of this nature were to prove true to such an extent that most variants resulted as either missing or spurious in the apparatus of Q, then the analysis would be completely unreliable. For this reason, we decided to re-examine a sample of 59 manuscripts[33] classified in ĸ as fully collated, so as to get at least a rough estimate of the accuracy of the collation as a whole.[34] Our examination proved quite satisfactory (86% accuracy) and seems to suggest that ĸ's apparatus can be considered a reliable source, at least as far as substantial variants of Q are concerned.[35]

Once we corrected the errors and integrated the missing data, we went on to collate six very ancient Oriental manuscripts not found in ĸ at all. All fragmentary manuscripts as well as manuscripts classified as partially collated were excluded.[36]

## 3.3 Data Matrix

After refining the data and selecting the most relevant witnesses, we wrote an xSL stylesheet in order to convert the variants into numerical values, and inserted them into a data matrix suitable for computational analysis (Figure 3).

---

**31** See Barthélemy, *Critique textuelle de l'Ancien Testament, 1. Josué-Esther*, 34 ff.; Chiesa, *Filologia storica della Bibbia Ebraica*, 140 ff.; R.S. Hendel, *The Text of Genesis 1–11: Textual Studies and Critical Edition* (New York and Oxford: Oxford University Press, 1998), 109 ff.

**32** A fact already known to DR, see De Rossi, *Variae lectiones Veteris Testamenti*, 1:xlvi–xvlii. See also Barthélemy, *"Les manuscrits médiévaux et le texte tibérien classique*," xxvv ff.

**33** No particular criterion was followed for sampling the manuscripts other than that of availability in digital format.

**34** The accuracy is calculated as the ratio of the variants recorded by ĸ to the total number of variants present in the manuscripts according to our surveys. The estimate refers to substantial variants only.

**35** Most inaccuracies concern the fact that ĸ does not adequately report whether a variant has been later corrected according to the received text. However, since here we take into consideration the variants of the first hand and disregard corrections, these inaccuracies are insignificant for stemmatic analysis. Cases in which substantial variants are missing or erroneous are rarer and have been corrected.

**36** The list of witnesses divided by collation degree is in Kennicott, *Vetus Testamentum Hebraicum*, 2:572.

```
                    1         2         3         4         5         6         7         8
Taxon      12345678901234567890123456789012345678901234567890123456789012345678901234567890...
- - - - - - - - - - - - - - - - - - - - - - - -
H          00000000000000000000000000000000000000000000000000000000000000000000000000000000
OOL        00000000000000000000000000000000000000000000000000000000000000000000000000000000
3IAK1      00000000000000100010000000000000000100000000100000000000000000000000100000000010
3SK2       00001000000100000010000000000000000000000000000000000000010000000000000000000000
4SK3       00000000000000000000000100000000000000000000000000000000000000000000000000000000
2AK4       00000101000000000000010000000000001000000000000000000000000001000001000000000000
4SK14      00100000000000000000000000000000000000100000000000000000000000000000010000000000
3AK17      00000100000100000001000000000000101000000000000000000000100000001000000000000000
3AK18      01000010000100010011010000000010101000001000000000000000000001100001100000000000
4SK19      00000000001000000000000000000010000000000000000000000000000000000000000000010000
2AK30      00000010001000000100010000000000000000000000001000000200000000000001000000000000
OOK31      00000000000000000010000000000000000000000000000000000000000000000000000000000000
3AK50      00001000000000000000000000000000000000000000000000000000000000000000000000000000
3AK56      00000000000000001000000100000000110100000000000000010000000001000000000000000000
4SK67      00000000000000000000000000000000000000000000000000000000000000000000000000000000
2AK76      00000001000000000000100100000000100110000000000000000000000000000000010000000000
...
```

**Figure 3:** Fragment of data matrix.

In the matrix the witnesses are represented in rows, while places of variation are represented in columns as a vector of characters. Characters contain different values, called character states, which represent different readings: '0' if the witness does not show any variant (that is, if it reads as the reference text), '1' if it has the reading of the first group (the element **<rdgGrp>** used in the encoding), '2' of the second group, and so forth. In total, the matrix consists of 116 witnesses and 371 places of variation.

## 3.4 Phylogenetic Analysis Through Maximum Parsimony

In order to compute genealogical relationships we employed phylogenetic methods proper to evolutionary biology.[37] Such methods permit the separation of witnesses into discrete groups according to inherited variants, and the representation of their relationships by the form of tree-like graphs, named phylogenetic trees.

Three main components can be distinguished in a phylogenetic tree: (1) the terminal nodes or leaves, representing the extant witnesses – in phylogenetic terms, the *taxa*; (2) the internal nodes or internodes, representing hypothetical witnesses

---

**37** We followed the methodology illustrated in A.-C. Lantin, P. Baret, and C. Macé, "Phylogenetic Analysis of Gregory of Nazianzus' Homily 27," in *Le poids des mots. Actes des 7èmes Journées Internationales d'Analyse statistique des Données Textuelles*, vol. 2 (Louvain-la-Neuve: JADT, 2004), 700–707. The appendix to the article shows the procedures applied to operate the phylogenetic software PAUP (see note 39), which we also followed.

from which the extant witnesses are presumed to descend; and (3) the tree branches or edges, which identify genealogical relationships.

When the information contained in the matrix is not conflicting, a phylogenetic tree will appear as fully dichotomous (or bifurcating): two witnesses are connected to a common ancestor, which is in turn connected to the common ancestor of other witnesses, all the way to the tree root. The similarity of phylogenetic trees to traditional *stemmata codicum*, however, must not lead us to jump to the conclusion that the first manuscript contains all the readings of the second, the second all the readings of the third, the third of the fourth, and so on, i.e., we should not automatically, as it were, read a phylogenetic tree like a proper *stemma*. Phylogenetic reconstruction of a family of manuscripts does not necessarily depend on the majority readings that are found in that given family.[38] Instead, when in a phylogenetic tree we find a group of witnesses clustered around one or more ancestors, it means that that group inherited from its respective ancestors the readings that require the least number of changes over the whole tree, as we are about to demonstrate.

Different tree-building methods exist in bioinformatics. We chose to use the Maximum Parsimony (MP) as implemented in PAUP, a widely-used software for phylogenetic analysis.[39]

According to MP, if two or more taxa share the same traits (the character states in the matrix of Figure 3), then it is far more likely that they were inherited from a common ancestor, rather than arising independently, each time in a different taxon. This last assumption, which goes in biology under the general name of *homoplasy*, is less parsimonious because it implies multiple changes at a time. MP tries to limit the amount of homoplasy as much as possible, by searching for the tree that minimizes the number of changes required to explain the actual character distribution.[40]

Given a character matrix, the basic implementation of MP must address two sub-problems: (1) generating the list of all possible trees; and (2) calculating the number of changes for each tree, or tree length.

---

**38** See M. Spencer, C.J. Howe, and K. Wachtel, "The Greek Vorlage of the Syra Harclensis: A Comparative Study on Method in Exploring Textual Genealogy," *TC: A Journal of Biblical Textual Criticism* 7 (2002), http://rosetta.reltech.org/TC/v07/SWH2002/index.html.

**39** Accessed February 26, 2021, https://paup.phylosolutions.com/.

**40** A general overview on MP analysis can be found in D.A. Baum and S.D. Smith, *Tree Thinking: An Introduction to Phylogenetic Biology* (New York: Macmillan Learning, 2013), 173–207. For a more detailed description with some PAUP running code see D.L. Swofford and J. Sullivan, "Phylogeny Inference Based on Parsimony and Other Methods Using PAUP," in *The Phylogenetic Handbook: A Practical Approach to Phylogenetic Analysis and Hypothesis Testing*, ed. P. Lemey, M. Salemi, and A.-M. Vandamme (Cambridge: Cambridge University Press, 2009), 267–312.

As far as the first task is concerned, there are two main searching methods: exhaustive and heuristic. The first method generates trees by exploring all possible combinations of taxa; the second starts by selecting an initial tree and then modifies it by using various strategies to minimize its length. However, since the number of trees rapidly increases with the number of taxa, as in our case, heuristic methods are more often employed, which, although they do not absolutely guarantee the discovery of *the* optimal tree, they have proven to yield good approximations of the most parsimonious ones. For our case study, the tree bisection-reconnection method (TBR) is used, which is set by default in PAUP.[41]

Turning now to the question of tree length, this is computed, in essence, by counting how many times a character state changes when going from one node to another. This calculation is carried out for every node on every single character in the matrix. At the end, the number of total changes (or steps) required by all the characters is summed up and this value is considered to be the length of that given tree. The tree that has the minimum length is then selected as the most parsimonious.

An example may help clarify the steps involved in the building of a phylogenetic tree through MP. Suppose we have a textual tradition with four witnesses *A B C D* and three places of variation, each with two competing readings **01**, as shown in Table 1.

**Table 1:** Example of character matrix.

| taxon | char. 1 | char. 2 | char. 3 |
|---|---|---|---|
| A | 1 | 1 | 1 |
| B | 1 | 0 | 1 |
| C | 0 | 1 | 0 |
| D | 0 | 0 | 0 |

With four witnesses, three combinations are possible: (1) *AB/CD*; (2) *AC/BD*; and (3) *AD/BC*. For each of these trees, the number of changes of every single character in the matrix is calculated as follows: first, the tree is rooted at an arbitrary point (in PAUP the rooting is done on the first taxon of the matrix); then, if two witnesses have a reading in common in a given place of variation, this is assigned to the common ancestor; otherwise, if the witnesses have two different readings, both are assigned to the ancestor and the tree length is increased by one.[42]

---

**41** An explanation of the TBR algorithm can be found in the PAUP manual, see note 39.
**42** This is the basic description of Fitch's algorithm, which presupposes that character states are unordered and that the cost of each character step is always equal to one. A more detailed

An illustration of the procedure carried out on the first three is shown in Figure 4.
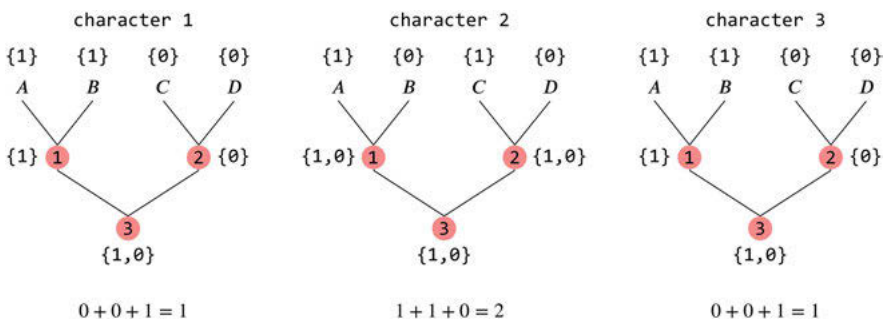


**Figure 4:** Character steps for tree no. 1.

As we can see, characters one and three change once, when going from the ancestors to the root. Character two changes twice, which means that it fits less into the tree than the other two. The total steps required by all the characters on tree one is therefore equal to four. If we carry out the analysis on the other trees, we obtain the values shown in Table 2.

**Table 2:** Length of each character state for each possible tree.

| tree | char. 1 | char. 2 | char. 3 | length |
|------|---------|---------|---------|--------|
| 1 | 1 | 2 | 1 | 4 |
| 2 | 2 | 1 | 2 | 5 |
| 3 | 2 | 2 | 2 | 6 |

The first tree is the most parsimonious, because it accounts better for the overall character distribution.

It is worth emphasizing, however, that the best tree is not optimal for all the characters: as can be seen in Table 2, character two fits better with the second tree, in which it changes just once. The first tree is better than the others, but it still contains homoplasy.

---

description of this and other algorithms available in PAUP can be found in Swofford and Sullivan, "Phylogeny Inference," 270–77.

Furthermore, different reconstructions for character two are possible on the first tree. This is illustrated in Figure 5, where the results of the ancestral reconstruction (also known as character-state reconstruction or character optimization) for the best tree are shown.
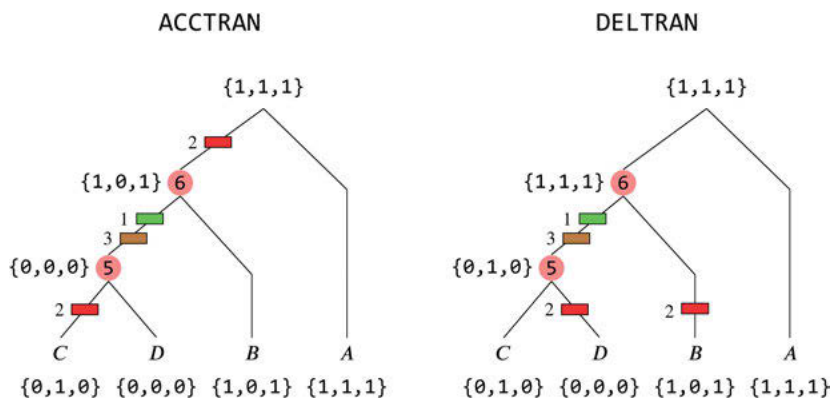


**Figure 5:** Difference between ACCTRAN and DELTRAN optimization algorithms.

In the first scenario, character two changes twice: the first time from **1** to **0** between the root and the ancestor 6; the second time from **0** back to **1** in **C**. In the second scenario, two parallel changes occur, from **1** to **0** in both **B** and **D**. These two different character reconstructions correspond to two different optimization algorithms: the first, called accelerated transformation (ACCTRAN), assumes that changes occur as soon as possible in the tree, and implies character reversal; the second, called delayed transformation (DELTRAN), assumes that changes occur as late as possible, and implies parallelism.[43] Here, we have preferred ACCTRAN, which is the default transformation in PAUP.

And so, we have seen different ancestral reconstructions based on the same tree structure (or tree topology). In many cases, however, more than one parsimonious tree is found post-analysis. For instance, imagine if we insert into the matrix shown in Table 1 a fourth place of variation with the same reading distribution as the second: in that case, we will have two places containing conflicting information. From this new matrix, two equally parsimonious trees with length 6 are generated: one with topology **AB|CD** and the other with topology **AC|BD**. In these cases a strict consensus tree can be computed, which will reveal

---

**43** Another algorithm is implemented in PAUP, called minimum-*F* value (MINF). See PAUP manual quoted at note 39.

the patterns common to all the equally parsimonious trees generated, as shown in Figure 6.[44]
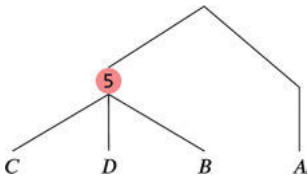


**Figure 6:** Example of strict consensus tree with polytomies.

As can be seen, the tree is not dichotomous, but presents polytomies, with more than two manuscripts descending from the same common ancestor. Such polytomies are usually interpreted as a sign of uncertainty in the data.

## 3.5 Results

The results of the analysis we carried out on κ's collation of Q are shown in the phylogenetic tree in Figure 9 at the end of this article. The tree was generated by a strict consensus of about 25,000 equally parsimonious trees, following the procedures described in the previous section.

Once we obtained the tree topology, we proceeded to examine the ancestral character reconstruction carried out by PAUP, in order to identify how many variants were assigned to the ancestors of groups of manuscripts. In particular, we tried to trace the so-called characteristic variants, i.e. ancestral variants that appear only once in a given group of manuscripts, or variants that appear in only a few groups. A list of the most relevant characteristic variants is shown in Table 3 in the Appendix. The result is that 104 out of 371 places of variation have been identified as ancestral and assigned to common ancestors. This means that most variants are classified as autapomorphies, that is, as secondary innovations introduced by the scribes of individual manuscripts. About 20 of these variants can be considered characteristic, since they appear only once, or at most, very few times, in the tree.

---

**44** Other kinds of consensus tree are implemented in PAUP in addition to the strict consensus, such as the majority rule consensus and Adam consensus, see PAUP manual quoted at note 39.

The tree presents 14 groups in total. Most of these are dichotomous, but poly-tomies are also found (such as in the fifth group of manuscripts 2AK168, 4SK144, and 0OK31).[45]

Eleven witnesses could not be classified. These witnesses, however, have very few variants from the reference text used by K (§ 2.1), which represents the received text or *textus receptus* (TR): here we find four Sephardic and four ancient Oriental manuscripts, including the *Leningradensis* (0OL). It is not incorrect, therefore, to consider them as representative of the TR group.

With regard to the other groups found by the machine, some are not certain, because they share a low number of variants or because the variants in common are weak in terms of their ability to reveal kinship. One such group is the fifth mentioned above, which is based on a single ancestral variant. Other groups, on the other hand, seem more likely, since either the quantity or the quality of inherited variants allows us to infer a genealogical kinship.

Let us take a closer look at the ancestral reconstruction of some of these groups.[46] In Figure 7 the group with three Ashkenazic manuscripts is shown (the sixth in the tree from the left).



| char. | 31AK1 | 3AK111 | 2AK185 |
|-------|-------|--------|--------|
| 79 | 1 | 1 | 1 |
| 102 | 1 | 1 | 1 |
| 187 | 1 | 1 | 1 |
| 312 | 1 | 1 | 1 |
| 147 | 1 | 1 | |
| 150 | 1 | 1 | |
| 197 | 0 | 0 | |
| 215 | 1 | 1 | |
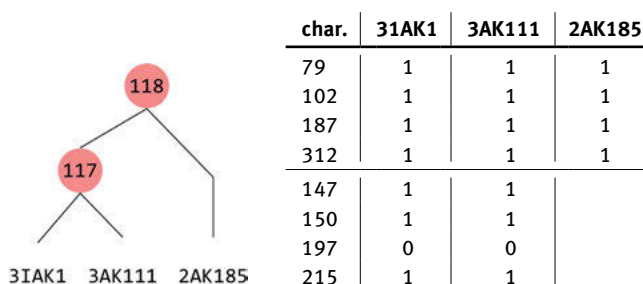
**Figure 7:** Ancestral reconstruction for 3IAK1, 3AK111, and 2AK185.

According to the ancestral reconstruction, these manuscripts inherited from their common ancestor (the internal node no. 118) a total of four readings. The manuscripts 31AK1 and 3AK111 have four other readings in common, which are not

found in 2AK185 and are assigned to another ancestor (internal node no. 117). The first reading (no. 79)[47] can be considered characteristic of this group, since it is found here and in two other manuscripts, with many variants, belonging to other branches of the tree (2AK158 and 3UK218).

As stated in Section 3.4, the best-fit algorithm underlying the ancestral reconstruction does not always assign to common ancestors the readings that are found in the majority of descendant witnesses. See for example Figure 8, in which the result of the ancestral reconstruction for three Italian manuscripts 1IK226, 2IK227, and 4IK225 is shown.

| char. | 1IK226 | 2IK227 | 4IK225 |
|---|---|---|---|
| 6 | 1 | 1 | 1 |
| 30 | 1 | 0 | 1 |
| 45 | 1 | 0 | 1 |
| 258 | 1 | 1 | 1 |
| 269 | 1 | 1 | 0 |
| 312 | 0 | 0 | 0 |
| 362 | 1 | 1 | 0 |
| 20 | | 1 | 1 |
| 124 | | 1 | 1 |
| 320 | | 0 | 0 |
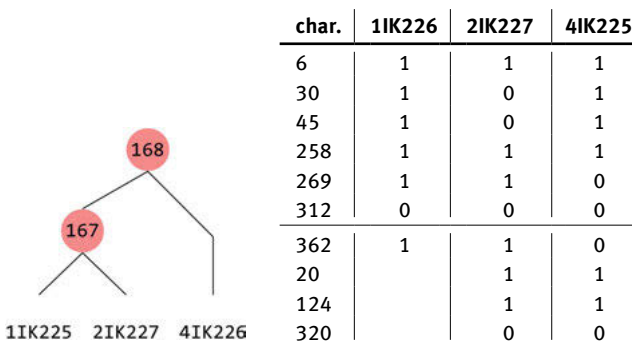
Figure 8: Ancestral reconstruction for 1IK225, 2IK227, and 4IK226.

Seven readings are assigned to the common ancestor 168. As can be seen, however, only three variants are actually common to all three manuscripts (nos. 6, 258, 312). The other readings are found in 1IK226 and 4IK225 (nos. 30, 45) or in 1IK226 and 2IK227 (nos. 269, 362). Manuscripts 2IK227 and 4IK225 share three other readings not found in 1IK226 (nos. 20, 124, 320), which are assigned to node 167. Both the distribution of readings and the fact that readings 30 and 45 are characteristic for 1IK226 and 4IK225 (see Table 3 in the Appendix) caused the three manuscripts to be lined up in this way by the algorithm. Given the common readings and script, a genealogical kinship is likely here and must be verified.

MP analysis leads to the likely conclusion that other witnesses, too, are genealogically interrelated.

As shown in Table 3, the largest number of characteristic variants are found in pairs of manuscripts of the fourteenth and last group. These manuscripts present a high number of variants with respect to the TR and seem to belong to that branch

---

47 The variant is in Q 3:11 and regards the omission of the conjunction לא.

of the textual tradition usually identified as *non-receptus* or *anti-receptus*.[48] The predominance of the Italian-Ashkenazic element in this group is remarkable.

# 4 Discussion

In Section 2 we described the creation of a parser for encoding K's collation of the HB. The parser enabled us to encode thousands of variants fully automatically, thus allowing us to avoid the complex task of manually encoding the critical apparatus. The results prove that a rule-based parsing system works well and suggest that such a system can be extended to include all the other sections of K's work.

As we mentioned in Section 2.2, the parser has proved to be a powerful tool in surmounting the hurdles presented by the first and perhaps most delicate challenge we confronted it with, namely, digitization through OCR software.

Digitizing critical apparatuses is challenging because of the specialized language that characterizes them. Errors generated by OCR are difficult for the human eye to detect, and manual correction is extremely time-consuming. Thanks to the CFG, a number of common errors, such as substitutions, omissions, and misdivisions of the apparatus, can be easily detected owing to the lexical and syntactic analysis provided by the lexer and the parser. This ensures a tight control over the structure of the critical apparatus, which allows for a correct identification of the function of its various components. The CFG also makes it possible to spell-check literal values, such as separators, technical terms, and witness *sigla*.[49]

The most important limitation of this system, as can be imagined, is represented by the impossibility of linguistically checking the Hebrew readings. In fact, Hebrew readings cannot be ratified by the CFG: as shown in the fragment of CFG in Listing 1, Hebrew tokens are described as simple sequences of Unicode alphabetic characters (`'[\u0590−\u05ff]+'`), which means that it is not possible to spellcheck with the lexer whether they are 'legal' Hebrew words or not.

---

**48** See Sacchi, "Analisi quantitativa," 8 ff.; M.H. Goshen-Gottstein, "The Rise of the Tiberian Bible Text," in *Biblical and Other Studies*, ed. A. Altmann (Cambridge, MA: Harvard University Press, 1963), 108 ff.

**49**  In the sample CFG shown in Listing 1 and in the final version available on Github (see note 7), the witness *sigla* are encoded as simple alpha-numeric characters. However, it is possible to create a rule in which the witness *sigla* used by K are encoded as literal values: in this way, the lexer will throw an error if a *siglum* does not match the corresponding value in the list of witnesses. For the Megillot, the list of witnesses can be acquired from Kennicott, *Vetus Testamentum Hebraicum*, 2:572.

At this stage, this task entirely depends on the language model built into OCR software, and the results must all be verified by direct human oversight. The implementation of a spell-checker in post-processing would be required if we want to improve the accuracy of the transcription up to this level.

In Section 3, we illustrated how data from encoded critical apparatuses have been used in computer-assisted stemmatic analysis. The task of classifying Hebrew biblical manuscripts is enormous and poses a number of problems to the investigator.

The first such problem concerns the paucity of genealogically significant variants, a fact common to traditions of authoritative texts characterized by a process of controlled transmission. As we have seen, the great majority of variants regard *minutiae*, such as the alternation of *scriptio plena* and *defectiva*, which can hardly be considered as indicators of genealogical kinship. As to the remaining variants that have been taken into account in the present study, not all can be considered as equally kinship-revealing: most are probably too weak, since they concern particles or other parts of speech that are frequently prone to random variation.

Another problem is represented by the phenomenon of horizontal transmission or contamination typical of open traditions, which cannot be properly handled with the genealogical approach adopted here.

Finally, one has to take into account the possibility of collation errors and, more generally, all the limits inherent in using works that are now, to say the least, dated, namely the collations of K and DR (§ 3.2).

In light of all these considerations, one can understand why previous attempts to discover groups or families of biblical manuscripts have been declared mostly unsuccessful, and why scholars are generally of the opinion that such classification is impossible.[50]

The problems and limitations mentioned above also emerge from the results of the phylogenetic experiment illustrated in the previous section. The tree of Figure 9 is not the best tree that can be achieved with MP analysis: as stated in Section 3.4, the most parsimonious tree cannot be arrived at, due to the size of the matrix and the nature of the data, as well as to the limits in computational power of modern computers. Most variants are useless for identifying families of manuscripts, and only a tiny minority can be considered as characteristic.

---

**50** See for example J.W. Wevers, "A Study in the Hebrew Variants in the Books of Kings," *Zeitschrift für die Alttestamentliche Wissenschaft* 61, no. 1 (1948): 43–76 and Barthélemy, "Les manuscrits médiévaux." Cf. on the other hand Sacchi, "Analisi quantitativa."

This does not mean, however, that all attempts at stemmatic analysis of the medieval biblical tradition are to be judged as doomed *a priori* to failure, nor "that certain individual manuscripts could not be judged as stemmatically connected to one another," as pointed out by Goshen-Gottstein.[51] The method we pursued did return groups of manuscripts divided by their ancestral variants, and sometimes even by their characteristic variants: some groupings rest on fragile foundations and are likely specious; others, however, seem more than plausible and need further study. The phylogenetic reconstruction which we have proposed is thus to be considered as a working hypothesis, to be verified by qualitative analysis.

Further research is needed if we want to reconstruct the genealogy of the medieval tradition of the HB. First of all, other manuscripts, more ancient or relevant from a text-critical standpoint, need to be added, as we ourselves have done to a certain extent. Other categories of potentially monogenetic variants need to be included, such as variants of punctuation and the Massora.[52]

As to methodology, we envision two strategies to improve results. One may contemplate, for example, the development of a system of weighting factors, so that readings judged as having greater kinship-revealing power (e.g. root substitutions) are assigned greater weight in building the tree.

Another strategy would be to establish an order in variation that would better reflect the logic of the copying process. Such a strategy could prevent some readings from being misclassified as ancestral to others (as in the case of readings due to *homoteleuton* being classified as ancestral to readings with intact text) and could thus help decrease the number of equally parsimonious trees that can be generated.

# 5 Conclusion

The collations of Kennicott and De Rossi represent our primary authorities on transmission of the Hebrew Bible from the Middle Ages onward. Since no other large-scale collations are available, we are obliged, still today, to rely on the efforts made by these two 18th-century scholars. Yet the obvious difficulty of dealing with the enormous mass of material they gathered has intimidated contemporary scholars from exploiting them as a direct source of text-critical and linguistic evidence.

---

**51** M.H. Goshen-Gottstein, "The Textual Criticism of the Old Testament: Rise, Decline, Rebirth," *Journal of Biblical Literature* 102, no. 3 (1983): 394.

**52** As suggested by Barthélemy, "Les manuscrits médiévaux," xxvii.

Digital technologies and methods offer the scholar of the Hebrew Bible an obvious route to pursue, since they are designed precisely to render large amounts of data far more easily searchable and processable. In fact, our research was carried out on a simple personal computer: it would have been impossible with 'traditional' methods. Further inquiries into textual history of the Hebrew Bible are scarcely imaginable without the use of the same digital tools which have become the norm in other fields of textual research, first and foremost New Testament studies. The employment of quantitative, computer-assisted methodologies, in sum, are indispensable if we want to establish a *stemma codicum* of the medieval tradition and to reconstruct the history of the biblical Hebrew text in the Middle Ages.

In the present project, we have used natural-language processing tools to demonstrate how a critical apparatus can be encoded automatically through a rule-based parser; then, we processed the resulting encoded variant readings by means of phylogenetic analysis in order to explore the medieval textual tradition of a biblical book.

Through this inquiry, we have, we hope, demonstrated that stemmatic relationships among Hebrew witnesses do exist and can be traced, and that the skeptical, indeed renunciatory, attitude dominant in this area in the scientific literature is not always justified. We believe we have shown that, at the very least, the question needs to be reopened and discussed.

A digital corpus of variant readings of the kind we created here would benefit not only stemmatology, but other research areas of the Hebrew Bible as well.

In the area of textual history, for example, variants of Hebrew codices can be used to study the formation of the *textus receptus* as reflected in most printed editions.[53]

---

**53** This line of research was carried out in particular by the Israeli school, see above all the works of M. Cohen: "מגבשי כתיב במצחפי מסורה עתיקים ומשמעם לתולדות נוסח המקרא המקובל" (Ph.D. thesis, Hebrew University, Jerusalem, 1973) [Unpublished]; "האידיאה בדבר קדושת הנוסח לאותיותיו וביקורת" "קווי יסוד לדמותו העיצורית, המקרא ואנחנו in הטכסט", vol. 1, ed. S. Uriel (Tel Aviv: דביר, 1979), 42–69; "לדמותם הקונסונאנטית" (1980): 123–82; "1 עיוני מקרא ופרשנות in של הטקסט בכתבי יד מקראיים מימי הביניים ספר השנה in של דפוסי המקרא הראשונים: המהדורה הראשונה של התנ"ך השלם – דפוס סונצ'ינו משנת 1488 "מהו 'נוסח של אוניברסיטת בר־אילן, vol. 18–19 (Ramat Gan: Bar Ilan University Press, 1981), 47–67; "עיוני מקרא ופרשנות in המסורה' ומה היקף אחיזתו בתולדות המסירה של ימה"ב Gan: Bar Ilan University Press, 1981), 229–56, and J.S. Penkower: יעקב בן חיים וצמיחת מהדורת" "המקראות הגדולות (Ph.D thesis, Hebrew University, Jerusalem, 1982) [Unpublished]; "Rabbinic Bible," in Dictionary of Biblical Interpretation, vol. 2, ed. J.H. Hayes (Nashville: Abingdon Press, 1999), 361–64; "The Development of the Masoretic Bible," in The Jewish Study Bible, ed. A. Berlin and M.Z. Brettler (Oxford and New York: Oxford University Press, 2014), 2077–84.

For textual criticism, on the other hand, quantitative methods could explore the possibility of a pre-Masoretic origin of the medieval variants, by counting and weighing the common readings between Hebrew codices and ancient versions of the Hebrew Bible.[54]

Moreover, as suggested by Goshen-Gottstein, data from medieval manuscripts can be used to study the copying process in the Middle Ages, such as scribal habits and the genesis of common copying errors.[55]

In the area of codicology and paleography, a distributional study of variants would prove useful: given that the various ethno-geographic families (Ashkenazic, Italian, Sephardic, etc.) are often characterized in terms of the extent to which they differ from the *textus receptus*, quantitative criteria would serve as an auxiliary tool to classify manuscripts whose character is dubious or unknown.[56]

Finally, a digital corpus containing thousands of variants would of necessity benefit the field of linguistics as well, as it would permit a far more extensive investigation, for example, of variant spellings and orthography.

Clearly, the data found in the 18th-century collations, and in particular in Kennicott, cannot be considered definitive. Kennicott's apparatus contains inaccuracies and *lacunae* which need correcting and integrating if our desire is to access a corpus of variant readings that is complete and truly reliable. But in the absence of an entirely new collation of Hebrew witnesses that could overcome such deficiencies, the digitization of the data in the classic collations makes an excellent starting point for the gathering of new material from which to acquire new insights and to foster new research perspectives in Hebrew Bible studies.

---

**54** A survey on these studies can be found in Barthélemy, "Les manuscrits médiévaux," xix–xxvii. These are: J. Hempel, "Chronik," *Zeitschrift für die Alttestamentliche Wissenschaft* 48 (1930): 187–206; J. Hempel, "Innermasoretische Bestätigungen des Samaritanus," *Zeitschrift für die Alttestamentliche Wissenschaft* 52, no. 1 (1934): 254–74; Wevers, "Hebrew Variants in the Books of Kings"; M.H. Goshen-Gottstein, "Die Jesaiah-Rolle und das Problem der hebräischen Bibelhandschriften," *Biblica* 35, no. 4 (1954): 429–42; H. Gese, "Die hebräischen Bibelhandschriften zum Dodekapropheton nach der Variantensammlung des Kennicott," *Zeitschrift für die Alttestamentliche Wissenschaft* 69, no. 1–4 (1957): 55; and Sacchi, "Analisi quantitativa."

**55** M.H. Goshen-Gottstein, "Hebrew Biblical Manuscripts: Their History and Their Place in the HUBP [Hebrew University Bible Project] Edition," *Biblica* 48 (1967): 49.

**56** As has been done, for example, by Penkower, see J. Penkower, כתב-יד ירושלמי של התורה מן המאה" ה-(3כתב-יד ק) מישאל בן עוזיאל שהגיהו העשירית, תרביב 58, no. 1 (1988): 47–74, and "A Sheet of Parchment from a 10th or 11th Century Torah Scroll: Determining Its Type among Four Traditions (Oriental, Sefardi, Ashkenazi, Yemenite)," *Textus* 21, no. 1 (2002): 235–64.
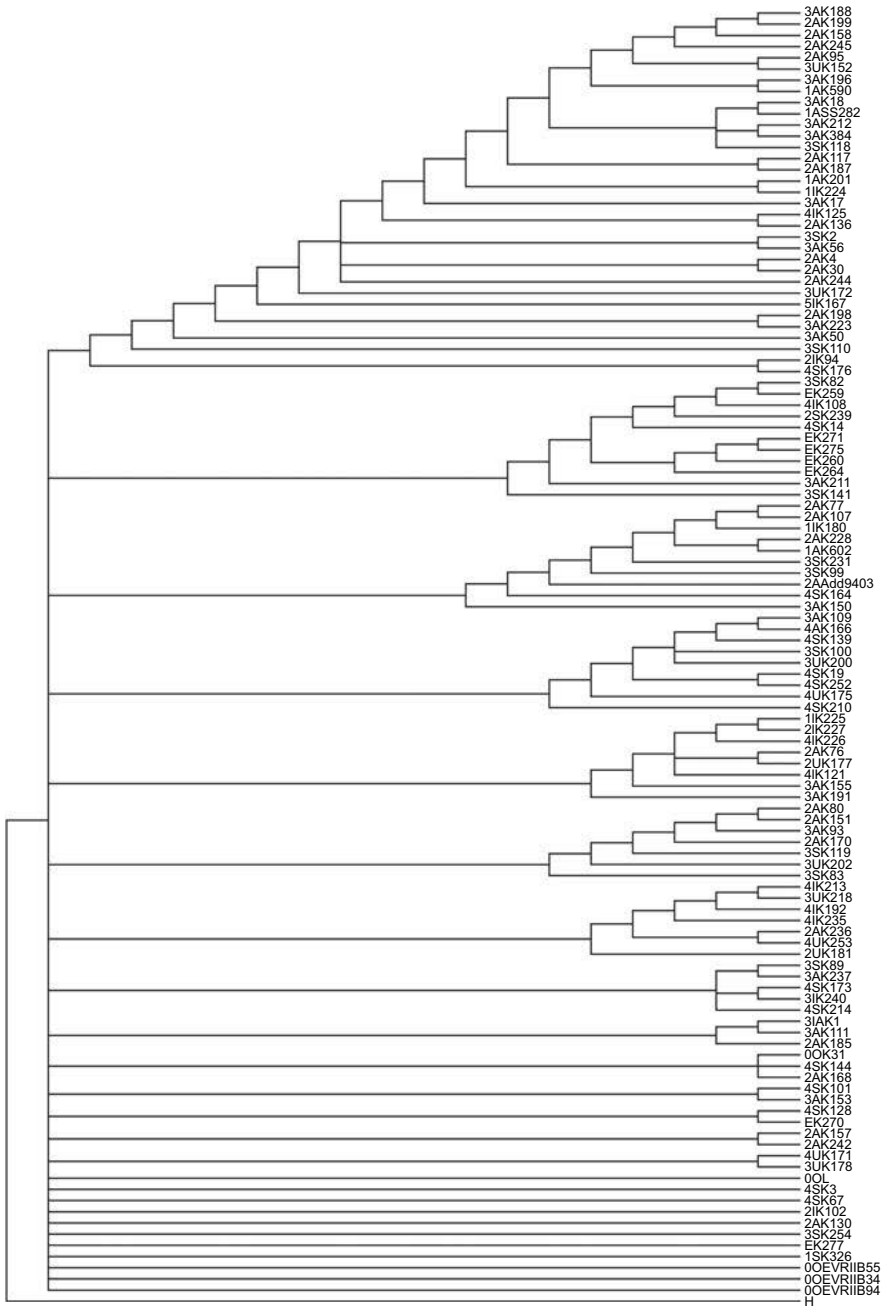
**Figure 9:** Strict consensus tree of the textual tradition of Qohelet according to Kennicott's collation.

# Appendix

In the following table the characteristic variants are listed. We used the consistency index (CI) in order to establish whether a variant can be considered characteristic and kinship-revealing. In phylogenetic analysis the CI is the measure of how much a given character fits a given tree topology: the maximum value is one and means that the character is perfectly consistent with the tree, because it changes only once; values less than one mean that more changes occur.[57] Applied to variants, a CI equal to one basically means that a variant occurs in one position only in the tree, otherwise it is found in other branches of the textual tradition.

In the table the variants are sorted by CI in descending order. Only witnesses with variants based on CI between 1 and 0.500 are shown.

**Table 3:** List of characteristic variants.

| witnesses | place | variant | CI |
|---|---|---|---|
| 2AK4 2AK30 | 6:3 | בו ← לו | 1.000 |
| | 6:5 | השמש ← שמש | 1.000 |
| | 8:9 | המעשה ← נעשה | 0.500 |
| E271 E275 | 12:1 | מי ← ימי | 1.000 |
| 2AK76 2UK177 | 2:26 | – ← ודעת | 1.000 |
| 2AK77 2AK107 | 4:4 | אני ← – | 1.000 |
| | 2:14 | ידעתי ← וידעתי | 0.400 |
| 2AK77 (2AK107) 1IK180 | 5:6 | כי ← – | 0.500 |
| | 5:5 | אל ← ואל | 0.333 |
| 2AK80 2AK151 | 7:21 | כל ← לכל | 1.000 |
| 2AK95 3UK152 | 2:15 | גם אני ← אני | 0.667 |
| | 3:19 | – ← כן מות זה | 0.500 |
| | 6:9 | מהלוך ← מהלך | 0.500 |
| | 8:4 | ושלטון ← שלטון | 0.333 |
| | 9:9 | חיי הבלך ← הבלך | 0.333 |
| 2IK225 2IK226 | 2:4 | ונטעתי ← נטעתי | 0.500 |
| | 2:10 | בכל ← מכל | 0.333 |
| 3AK18 1ASS882 | 12:9 | ותקן ← תקן | 0.500 |
| | 2:5 | נטעתי ← ונטעתי | 0.333 |
| | 9:9 | – ← כל ימי הבלך | 0.333 |
| | 12:9 | ותקן ← תקן | 0.333 |
| 3AK212 3AK384 | 8:9 | המעשה ← מעשה | 0.500 |
| | 5:18 | ונכסים וכבוד ← ונכסים | 0.333 |

---

**57** See Baum and Smith, *Tree Thinking*, 93 ff.

**Table 3** (continued)

| witnesses | place | variant | CI |
|-----------|-------|---------|-----|
| 2AK117 2AK187 | 6:10 | לא ← ולא | 0.500 |
| 1AK201 1IK224 | 5:1 | המעשה ← מעשה | 0.500 |
| | 8:11 | – ← בני | 0.333 |
| 4SK173 3IK240 | 12:8 | הבל הכל הבל הבלים ← הבל הכל | 0.500 |
| 4IK213 3UK218 | 7:12 | את בעליה ← בעליה | 0.500 |

# References

Bambaci, L. "Critical Apparatus as Domain Specific Languages: A Rule-Based Parser for Encoding an Eighteenth-Century Collation of Hebrew Manuscripts." *International Journal of Information Science and Technology* 5, no. 1 (2021): 22–33.

Bambaci, L. "Is a Stemma Possible for the Hebrew Bible? Towards a Genealogy of Medieval Manuscripts through Phylogenetic Analysis." *Materia Giudaica. Rivista dell'associazione italiana per lo studio del giudaismo* 26, no. 2 (2021): 3–30.

Barthélemy, D. *Critique textuelle de l'Ancien Testament, 1. Josué-Esther*. Volume 1 of *Orbis Biblicus et Orientalis 50*. Fribourg and Göttingen: Éditions Universitaires and Vandenhoeck & Ruprecht, 1982.

Barthélemy, D. "Les manuscrits médiévaux et le texte tibérien classique." In *Critique textuelle de l'Ancien Testament, 3. Ézéchiel, Daniel et les 12 Prophètes*. Volume 3 of *Orbis Biblicus et Orientalis 50*, xix–xcvi. Fribourg and Göttingen: Éditions Universitaires and Vandenhoeck & Ruprecht, 1992.

Baum, D.A. and S.D. Smith. *Tree Thinking: An Introduction to Phylogenetic Biology*. New York: Macmillan Learning, 2013.

Borbone, P.G. *Il libro del profeta Osea. Edizione critica del testo ebraico*. Torino: Zamorani, 1990.

Chiesa, B. "Textual History and Textual Criticism of the Hebrew Old Testament." In *The Madrid Qumran Congress*, edited by J.C. Trebolle Barrera, 257–72. Leiden: Brill, 1992.

Chiesa, B. *Filologia storica della Bibbia Ebraica*. Volume 2 of *Studi Biblici 135*. Brescia: Paideia, 2000.

Cohen, M. "מגבשי כתיב במצחפי מסורה עתיקים ומשמעם לתולדות נוסח המקרא המקובל" Ph.D. thesis, Hebrew University, Jerusalem, 1973. [Unpublished].

Cohen, M. "המקרא ואנחנו in האידיאה בדבר קדושת הנוסח לאותיותיו וביקורת הטקסט", volume 1, edited by S. Uriel, 42–69. Tel Aviv: דביר, 1979.

Cohen, M. "קווי יסוד לדמותו העיצורית של הטקסט בכתבי יד מקראיים מימי הביניים in עיוני מקרא ופרשנות 1 (1980): 123–82.

Cohen, M. "לדמותם הקונסונאנטית של דפוסי המקרא הראשונים: המהדורה הראשונה של התנ"ך השלם" in "ספר השנה של אוניברסיטת בר-אילן – דפוס סונצ'ינו משנת 1488", volume 18–19, 47–67. Ramat Gan: Bar-Ilan University Press, 1981.

Cohen, M. "מהו 'נוסח המסורה' ומה היקף אחיזתו בתולדות המסירה של ימה"ב in עיוני מקרא ופרשנות", volume 2, edited by S. Uriel, 229–56. Ramat Gan: Bar-Ilan University Press, 1986.

De Rossi, G.B. *Variae lectiones Veteris Testamenti*. Parma: Ex regio typographeo, 1784–1788.

De Rossi, G.B. *Scholia critica in V.T. libros, seu supplementa ad varias sacri textus lectiones*. Parma: Ex regio typographeo, 1798.

Gese, H. "Die hebräischen Bibelhandschriften zum Dodekapropheton nach der Variantensammlung des Kennicott." *Zeitschrift für die Alttestamentliche Wissenschaft* 69, no. 1–4 (1957): 55.

Goshen-Gottstein, M.H. "Die Jesaiah-Rolle und das Problem der hebräischen Bibelhandschriften." *Biblica* 35, no. 4 (1954): 429–42.

Goshen-Gottstein, M.H. "The Rise of the Tiberian Bible Text." In *Biblical and Other Studies*, edited by A. Altmann, 79–122. Cambridge, MA: Harvard University Press, 1963.

Goshen-Gottstein, M.H. "Hebrew Biblical Manuscripts: Their History and Their Place in the HUBP [Hebrew University Bible Project] Edition." *Biblica* 48 (1967): 243–90.

Goshen-Gottstein, M.H. "The Textual Criticism of the Old Testament: Rise, Decline, Rebirth." *Journal of Biblical Literature* 102, no. 3 (1983): 365–99.

Hempel, J. "Chronik." *Zeitschrift für die Alttestamentliche Wissenschaft* 48 (1930): 187–206.

Hempel, J. (1934). "Innermasoretische Bestätigungen des Samaritanus." *Zeitschrift für die Alttestamentliche Wissenschaft* 52, no. 1 (1934): 254–74.

Hendel, R.S. *The Text of Genesis 1–11: Textual Studies and Critical Edition*. New York and Oxford: Oxford University Press, 1998.

Kennicott, B. *Vetus Testamentum Hebraicum cum variis lectionibus*, volume 1. Oxford: Clarendon, 1776.

Kennicott, B. *Vetus Testamentum Hebraicum cum variis lectionibus*, volume 2. Oxford: Clarendon, 1778.

Lantin, A.-C., P. Baret, and C. Macé. "Phylogenetic Analysis of Gregory of Nazianzus' Homily 27." In *Le poids des mots. Actes des 7èmes Journées Internationales d'Analyse statistique des Donées Textuelles*, volume 2, 700–707. Louvain-la-Neuve: JADT, 2004.

Maas, P. "Textkritik." In *Einleitung in die Altertumswissenschaft*, 2nd ed., volume 1.2, edited by A. Gercke and E. Norden, 1–18. Leipzig: Teubner, 1927.

Maas, P. *Textkritik*, 4th ed. Leipzig: Teubner, 1960.

Parr, T. *Language Implementation Patterns: Create Your Own Domain-Specific and General Programming Languages*. Dallas, TX and Raleigh, NC: Pragmatic Bookshelf, 2010.

Parr, T. *The Definitive ANTLR 4 Reference*. Dallas, TX and Raleigh, NC: Pragmatic Bookshelf, 2012.

Penkower, J.S. "יעקב בן חיים וצמיחת מהדורת המקראות הגדולות" Ph.D. thesis, Hebrew University, Jerusalem, 1982. [Unpublished].

Penkower, J.S. "כתב-יד ירושלמי של התורה מן המאה העשירית שהגיהו מישאל בן עוזיאל (כתב-יד קק 3)" תרביץ 58, no. 1 (1988): 49–74.

Reenen, P.T. van and M. van Mulken. *Studies in Stemmatology*, volume 2. Amsterdam and Philadelphia: John Benjamins, 2004.

Reenen, P.T. van, M. van Mulken, and J. Dyk. *Studies in Stemmatology*, volume 1. Amsterdam and Philadelphia: John Benjamins, 1996.

Roelli, P., ed. *Handbook of Stemmatology*. Berlin and Boston: De Gruyter, 2020.

Sacchi, P. "Analisi quantitativa della tradizione medievale del testo ebraico della Bibbia secondo le collazioni del De Rossi." *Oriens Antiquus* 12 (1973): 1–13.

Salemans, B.J.P. "Building Stemmas with the Computer in a Cladistic, Neo-Lachmannian, Way: The Case of Fourteen Text Versions of Lanseloet Van Denemerken." Ph.D. thesis, Katholieke Universiteit Nijmegen, 2000.

Spencer, M., C.J. Howe, and K. Wachtel. "The Greek Vorlage of the Syra Harclensis: A Comparative Study on Method in Exploring Textual Genealogy." *TC: A Journal of Biblical Textual Criticism* 7 (2002). http://rosetta.reltech.org/TC/v07/SWH2002/index.html.

Swofford, D.L. and J. Sullivan. "Phylogeny Inference Based on Parsimony and Other Methods Using PAUP." In *The Phylogenetic Handbook: A Practical Approach to Phylogenetic Analysis and Hypothesis Testing*, edited by P. Lemey, M. Salemi, and A.-M. Vandamme, 267–312. Cambridge: Cambridge University Press, 2009.

Tov, E. *Textual Criticism of the Hebrew Bible*, 3rd ed. Minneapolis: Fortress Press, 2012.

Wevers, J.W. "A Study in the Hebrew Variants in the Books of Kings." *Zeitschrift für die Alttestamentliche Wissenschaft* 61, no. 1 (1948): 43–76.

Avi Shmidman

# Automatic Identification of Biblical Citations and Allusions in Hebrew Texts

**Abstract:** The Hebrew text of scripture forms a foundational piece of shared culture for Hebrew writers. Hebrew texts through the ages, whether medieval or modern, virtually all draw upon the Hebrew Biblical text in one way or another. The identification of these Biblical citations and allusions is generally a primary prerequisite for the analysis or publication of the text. The automation of this process has long been a desideratum among scholars; however, for many years, an effective solution proved elusive.

To be sure, exact citations of Biblical prooftexts are trivial to identify: they are generally introduced with a stock phrase such as "as it is written", and the exact reproduction of the Biblical texts allows for a quick and easy lookup to determine the source of the quote. However, the overwhelming majority of Biblical text reuse is not in the form of exact citations, but is rather found as a reworking and adaptation of the Biblical text, adjusting the text to fit the new context, while still maintaining a set of linguistic cues in order to keep the connection to the original verse. For instance, a case of Biblical allusion will often be based upon the reuse of a set of lexemes from a given verse, even as those lexemes are adjusted in terms of their tense, person and gender. Additionally, prefixes and suffixes may be added, dropped or altered; additional words may be interpolated; and the order of the lexemes may be altered as well.

In this chapter, I present a new algorithm for the identification of Biblical allusions, designed to address all of these challenges.

**Keywords:** Hebrew scripture, natural language processing, approximate matching, intertextuality

---

# 1 The Challenge

Hebrew Scripture forms a foundational piece of shared culture for Hebrew writers. Hebrew texts through the ages, both medieval and modern, have often drawn upon the biblical text in one way or another. The identification of these biblical citations and allusions is a primary prerequisite for the analysis or publication of the texts.

To be sure, exact citations of biblical verses are fairly trivial to identify, using standard string-searching algorithms.[1] However, the overwhelming majority of biblical text reuse is not in the form of exact citations. Rather, when reusing biblical material, authors of Hebrew texts will generally rework the biblical text, adjusting it to fit the new context. The lexemes of the original verse may be altered in terms of their tense, person and gender; individual words may be added, dropped, or altered; prefixes and suffixes may be altered; and the order of the words may be inverted as well. To take one example, consider the following line from S.Y. Agnon's novella, "In the Prime of Her Life": ותתחזק אמי ותשב על המטה ("my mother gathered her strength and sat up on the bed").[2] This line is patterned after Genesis 48:2: וַיִּתְחַזֵּק יִשְׂרָאֵל וַיֵּשֶׁב עַל הַמִּטָּה ("Israel gathered his strength and sat up on the bed"). However, the first and third words have been altered from masculine to feminine (ותתחזק <- ויתחזק, and ותשב <- וישב), and the second word, "Israel," has been swapped out for "my mother."

Furthermore, biblical allusions can be very short, sometimes consisting of no more than two words.[3] For example, Agnon writes: וכחה ברח כצל ("and her strength fled like a shadow").[4] The simile here derives from Job 14:2: וַיִּבְרַח כַּצֵּל וְלֹא

---

**1** For an overview of exact-match search algorithms, see: Dan Gusfield, *Algorithms on Strings, Trees, and Sequences* (Cambridge: Cambridge University Press, 1997), 5–69. To be sure, when it comes to the biblical text, there is the additional complication of non-normative defective and plene orthography. Citations of verses in later texts often adopt normalized spellings, and thus, in order to locate all exact citations, these orthographic variants must be taken into account. For more regarding this point, see below, note 9.

**2** S.Y. Agnon, *Al Kapot Ha-man'ul* (Jerusalem and Tel-Aviv: Schocken, 1998), 7 [Hebrew].

**3** Theoretically, one might also consider single words to constitute biblical allusions, especially in the case of hapax legomena. However, as Shulamit Elizur has compellingly argued, individual biblical words should never be considered biblical allusions in and of themselves. (Shulamit Elizur, *Hebrew Poetry in Spain in the Middle Ages*, vol. 3 [Tel-Aviv: The Open University, 2004], 355–57 [Hebrew]). As she explains, individual words are the lexicographical building blocks that a Hebrew writer uses when formulating a sentence. If a single word could constitute a biblical allusion, then writers would never be able to use the word without invoking the said allusion. We would thus be left with the absurd situation in which a word would be effectively blocked from use, except in the case where a writer desires to forge an allusion to the source verse.

**4** Agnon, *Al Kapot Ha-man'ul*, 5.

יַעֲמֹד ("he fled like a shadow and could not endure"). This allusion consists of no more than two solitary words; and, further, the first word is altered in its morphological form (the biblical vuv-inversive form ויברח is replaced by the normative modern verb ברח).

We now turn to the question of how to design an algorithm to automatically identify allusions of this nature. Specifically, we consider how to break down the biblical text into minimal representational units which can then be matched to corresponding units in other Hebrew texts, such that these matches will capture the biblical allusions therein. These minimal representational units are termed "hashes," and the process of breaking down a text into these units is called "hashing." The non-biblical Hebrew text in which we wish to identify the allusions is termed the "target text."

Given the examples of biblical allusions presented above, we might consider hashing every two-word lexeme pair in the biblical text. This would ensure, first of all, that we catch matches as short as two words. Secondly, because the hashes are based on lexemes and not exact wordforms, they would allow us to capture any case in which the two lexemes appear together, regardless of how the lexemes might have been altered grammatically. Further, in order to allow for allusions which delete, add, or change a word in the middle, we could hash every two-out-of-three sequence. Thus, given a sequence of three words A-B-C from the biblical text, we would prepare three hashes: one for A-B, one for B-C, and one for A-C, each time storing the lexemes underlying the two words. We would run the same two-out-of-three hash procedure on the target text, and then look for all cases of hashes in the target text which have a match in the biblical text.

Such an approach would certainly capture virtually all of the relevant cases of biblical allusions in the target text (high recall). The problem, however, is that it would capture an overwhelming amount of irrelevant material as well (low precision); for a mere match of a pair of lexemes does not always constitute sufficient basis for establishing a biblical allusion. Indeed, using this method, almost every modern Hebrew sentence matches up with multiple biblical verses, based upon lexemes which coincidentally happen to co-occur, and the number of false positives quickly becomes unmanageable. To illustrate, we take the first three paragraphs of Agnon's aforementioned novella "In the Prime of Her Life," consisting of 198 words. My manual inspection of these paragraphs reveals 12 bona fide biblical allusions. Yet, running the same paragraphs through the proposed lexeme-pair hashing method turns up no less than 4,937 matches, connecting those 198 words to over 3,400 different biblical verses. This ratio, consisting of several hundred false positives for every true positive, clearly renders the algorithm useless, with a precision of less than 1%. Furthermore, considering that the Bible only contains some 23,000 verses, it emerges that this algorithm thus

finds "allusions" to over 10% of the Bible's verses within these three solitary paragraphs – clearly an absurd result.

One might suggest that this overflow can be obviated by defining a set of "stop words" – very frequent words which should be ignored when hashing both the biblical text and the target text. We thus assembled a list of the 25 most common biblical words; this list includes words such as אֶת (accusative marker), אֲשֶׁר ("that"), לֹא ("not"), and יִשְׂרָאֵל ("Israel"). We defined these 25 words as stop words and reran the hashing procedure as above on the Agnon text. Although this did help somewhat, the number of false positives was still overwhelming: 2,624 matches in over 1,000 distinct verses. Furthermore, the eliminated stop words are often crucial in identifying allusions; for instance, without the word לֹא ("not"), core biblical phrases such as לֹא תַחְמֹד ("thou shalt not covet") or לֹא תִגְנֹב ("thou shall not steal") would go unnoticed. Thus, the use of stop words does not provide an adequate solution.

How then can we allow for a sufficiently high degree of matching flexibility, without incurring a flood of meaningless results in return?

## 2  What Makes an Allusion?

The key intuition in solving this challenge is that our goal is not to find all instances of parallel phrases between the Bible and the target text, but rather to find allusions within the target text to *specific* biblical verses. In order for a given phrase to be considered an allusion to a biblical verse, it is not sufficient for it to incorporate a biblical phrase; rather, it must include textual material which connects it to a particular biblical verse.[5] Thus, when it comes to Agnon's ותתחזק אמי ותשב על המטה phrase, referenced above, a direct connection is formed to the verse in Genesis, because there are no other verses in the entire Bible which demonstrate that sequence. In contrast, were a target text to contain phrases such as ערי יהודה ("cities of Judah"), קדוש ישראל ("Holy One of Israel") or חי אני ("as I live"), no such connection would be formed. These three phrases are certainly biblical; each one of them appears over 20 times in Hebrew Scripture. However, precisely because of their high frequency of occurrence, their reuse in later Hebrew texts would not point the reader to any specific biblical verse, and thus they would not constitute instances of biblical allusion.

In this sense, we may posit that the best predictor as to whether a match constitutes a significant biblical allusion or not is the rarity of the phrase. If a

---

**5** This claim is underscored and explicated by Elizur, *Hebrew Poetry*, 3:358.

sequence that is unique within the biblical text recurs in the target text, then that phrase has a high likelihood of representing a biblical allusion.

Given this premise, we can severely cut down the number of hashes that we include in our search infrastructure, by discarding any hash which occurs in more than one verse. To be sure, in practice, the criterion of absolute uniqueness would be too restrictive. In many cases, a given verse or set of verses is requoted within Scripture itself. For instance, the decalogue appears in Exodus 20 and repeats again almost verbatim in Deuteronomy 5. Similarly, regarding 2 Samuel 22 and Psalm 18 (David's song); Psalms 14 and 53; and more. In such cases, the repetition does not indicate that the phrases contained therein are generic; rather, the phrases of the decalogue would still point the reader to specific biblical content. Therefore, we would not wish to block allusions to this material just because of the recurrence. In other cases, a given phrase appears a few times in the space of one pericope, such that a reuse of the phrase could be said to allude to that pericope overall. For instance, in Genesis 23 the words קְבֹר ("bury") and מֵתֶךָ ("your deceased") appear together three times within the chapter; nevertheless, because they are all used within the same biblical story in the same biblical chapter, a reuse of that phrase could be said to point the reader specifically to this section. In order to make allowance for such cases, we will slightly relax the uniqueness requirement, allowing all hashes which occur no more than three times within the biblical text (henceforth: "semi-unique hashes"). Nevertheless, as we will see in the next section, a further complication arises.

# 3 Beyond Lexemes

Based on the foregoing, it might seem that it would be sufficient to simply hash all semi-unique lexeme pairs within the Bible. However, the reality turns out to be more complex. Our assertion that biblical allusions must be based on hashes which occur no more than three times in the Bible gives rise to a further question: what exactly are we counting? Is it the pair of lexemes in their precise order of appearance which must be semi-unique? Or in any order? Or is it the exact pair of words that must be semi-unique?

In truth, the semi-uniqueness requirement can be satisfied in any one of these ways (and many more), and it is critical to consider them all. To illustrate the point, let's examine the first half of Psalms 89:41: פָּרַצְתָּ כָל גְּדֵרֹתָיו ("You have breached all his walls"). The combination of the words פָּרַצְתָּ and גְּדֵרֹתָיו is unique in the Bible, and the reappearance of those specific words together in a target text would likely constitute allusion to this verse, even without the word כָל in

the middle. In contrast, the lexeme pair of פרץ ("to breach") and גדר ("wall") occurs several times in the Bible in various other conjugations, such as פֹּרֵץ גָּדֵר ("one who breaches a wall," Ecclesiastes 10:8) and פָּרֹץ גְּדֵרוֹ ("breaching his wall," Isaiah 5:5). Therefore, if a target text were to use some other conjugation of the lexemes פרץ and גדר (e.g., פרצתי גדרותיך), there would be no reason for that word pair to trigger an allusion specifically to the Psalms verse. For the pair פָּרַצְתָּ גְדֵרֹתָיו, then, uniqueness is achieved by focusing upon the particular word forms, rather than the lexemes. On the other hand, if we look at the full phrase פָּרַצְתָּ כָל גְּדֵרֹתָיו, and we examine the sequence of underlying lexemes (פרץ, כל, גדר), we find that it is completely unique across the Bible. Therefore, if a target text were to use those three lexemes together, it would likely constitute an allusion to the Psalms verse, even if the words were morphologically altered (e.g., פרצתי כל גדרותיך). The subtle addition of the lexeme כל provides the extra hint necessary to link a new usage of פרץ+גדר to the verse in Psalms.

What emerges, then, is that it is not sufficient to simply hash the lexemes, because often the semi-uniqueness which creates a connection to a specific biblical verse is achieved only through the exact word forms. On the other hand, if the lexemes are sufficiently unique to induce an allusion, then we would not want to suffice with a hash of the word forms; rather, we would want to use a hash of the lexemes, in order to capture matches in which the words are morphologically altered. Furthermore, it is not always a matter of choosing either lexemes or word forms; in many cases, the requisite semi-uniqueness is found specifically in a combination of the two – e.g., a pair consisting of one lexeme and one word form.

The key principle is that, for any given biblical phrase, we wish to find the most generic representation which also satisfies the semi-uniqueness requirement. In addition to lexemes and exact word forms, we also consider the word forms with prefixes removed, which we will term "base words." Thus, if the lexemes underlying a given phrase are not sufficiently unique, we don't necessarily need to limit matches to the verbatim phrase; rather, we can first fall back to the base words. For example, consider the phrase בְּשִׂמְחֹת וָגִיל ("with joy and gladness," Psalms 45:16). Although the full phrase בְּשִׂמְחֹת וָגִיל is unique in the Bible, the lexeme pair is not; the combination of the lexemes שמחה ("joy") and גיל ("gladness") is quite frequent in the Bible, and their co-occurrence would not constitute an allusion to this verse. However, this does not mean that we need to limit matches to the exact phrase בְּשִׂמְחֹת וָגִיל. Both of the words in this phrase contain prefixes (ב and ו). An examination of the base words (שמחות, גיל) demonstrates that this is the only verse where they co-occur, regardless of prefixes. We can thus define the hash as: "base word שמחות plus base word גיל," in order to capture reuse cases such as השמחות שבגיל; והשמחות והגיל; ושמחות וגיל, and so on. All of these would constitute valid allusions to the verse.

We also consider whether the semi-uniqueness holds only given the specific order, or regardless of order. If the latter is true, then we catalog the hash as an order-neutral hash, enabling it to match a wider set of allusions. For instance, the set of lexemes underlying the three-word phrase לְחָיָו כַּעֲרוּגַת הַבֹּשֶׂם (Song of Songs 5:13) does not appear anywhere else in the Bible, whether in that order, or in any other order. In such a case, there is no need to store the order-sensitive hash, because the order-neutral hash is sufficient to identify all relevant instances in the target text.

Finally, we apply the same procedure with regard to hash length as well. A given sequence of two or three lexemes may not be sufficiently unique, but if we extend the sequence and add one or two subsequent lexemes, we may then arrive at a semi-unique hash. As above, we aim to hash it as the shortest possible sequence, for maximum applicability; but when the short sequence is not sufficient to secure the connection to the verse, we fall back to longer sequences.

Naturally, applying stricter definitions in any one of these areas will allow more flexible definitions in a different area. For instance, a short sequence is more likely to satisfy the semi-uniqueness requirement if its components are defined strictly in terms of their full forms in their specific order, while longer sequences are more likely to accommodate order-neutral lexeme-based hashes. We wish to store all such hashes, because each one provides flexibility from a different angle.

# 4 Implementation

We now formalize the aforementioned ideas into a step-by-step algorithm, with specific implementation details. We start with the procedure for hashing the biblical text:

- For every word in the biblical text, we store three separate representations: the full vocalized word; the vocalized word without prefix; and the lexeme underlying the word.[6] All such unique word representations are assembled in a single linear list, and an identifier is assigned to each one. In total, this list numbers under 100,000 items; thus, we can store any word representation within 17 bits.

---

**6** In order to determine the precise lexeme and prefix segmentation of each word in the biblical text, we use the ETCBC4B tagged database of Hebrew Scripture, generously provided by the Eep Talstra Centre for Bible and Computing, VU University Amsterdam. See: W.T. van Peursen, C. Sikkel, and D. Roorda, *Hebrew Text Database ETCBC4b*, DANS, 2015, accessed January 25, 2022, https://github.com/ETCBC/bhsa/tree/master/source/4b.

- We now iterate across all words of the biblical text. For each word, we consider a set of hashes starting at that word position, with sequence length ranging in length from two to five (without crossing verse boundaries).
  - For sequences of length two to three, we allow one word skip (two out of three, or three out of four), and for each configuration we compute all permutations of the three aforementioned word representations of each word within the sequence. Thus, for each starting word, we compute 18 two-word hashes, and 81 three-word hashes. To illustrate the process, Figure 1 rolls out all two-word hashes computed for a sample three-word biblical phrase.
  - For sequences of length four to five, we allow two word skips (four out of six, or five out of seven). For these longer sequences, we don't compute hashes for all possible permutations of the three word representations. Rather, for each configuration, we compute three hashes: one based solely on the lexemes, one based on the base words; and one based on the exact words.[7]

Hashing Isaiah 24:17: פַּחַד וָפַחַת וָפָח
All 2-word hashes



**Figure 1:** Demonstration of our hashing procedure. Quotation marks indicate an exact word; an underscore indicates a base word regardless of prefix; and square brackets indicate a lexeme. When considering words 1, 2 (right column) and words 1, 3 (middle column), all hashes satisfy the semi-uniqueness criterion, and the lexeme-based hash (9) is sufficiently generic to cover all of the other cases. Thus, we need only store hash (9). In contrast, when considering words 2, 3 (left column), we discard hashes (5), (6), (8), and (9), because they are not sufficiently unique. Of the remaining hashes, (3) and (7) are sufficiently generic to cover all of the remaining cases. Hash (3) pairs the exact form of word 2 with the lexeme of word 3, while hash (7) pairs the lexeme of word 2 with the exact form of word 3.

---

7 The less flexible hashing for longer hashes is motivated as follows. When it comes to sets of two to three words, it is often the case that the lexemes or words will coincidentally co-occur

- – Regardless of length, all lexeme-based hashes are additionally stored in an order-neutral manner in a separate hash collection.[8]
  - – Because each hash covers a maximum of five words, and because each word representation entails only 17 bits, we can handily store any hash in a single 128-bit integer.
- – At this point, having built up our two hash collections of order-sensitive hashes and order-neutral hashes, we proceed to prune the collections as follows:
  - – First we apply our criterion of semi-uniqueness. For each collection, any hash with a count of more than three is removed from the collection.[9] At this point, we are left with some 27 million semi-unique hashes.
  - – Next, we remove all hashes which are already covered by a more generic hash. For instance, if a given sequence is covered both by an exact-word hash and also a lexeme-based hash, we can discard the former. Further, if the same lexeme-based hash is covered by both an order-sensitive hash and an order-neutral hash, we can discard the order-sensitive hash. In such cases, we can of course also discard any hashes of the same sequence that are based on a combination of word forms and lexemes. This process eliminates over 90% of the semi-unique hashes; at the end of

---

in multiple biblical verses. Biblical allusions of two or three words are thus often quite subtle, picking up on a slight uniqueness in one biblical form or another in order to forge the connection to the verse. In order to ensure that we capture these particularly subtle references, we consider all permutations of all the word representations; this enables us to determine the precise set of terms which can form a maximally flexible semi-unique connection to the verse. In contrast, sequences of four or more words tend to be fairly unique from the start, and allusions of four or more words are naturally more explicit and easier to identify. Thus, in these cases, the precise determination of the maximally flexible word-representation premutation is less critical. Skipping the extra permutation calculations for the longer sequences provides significant computation savings, due to the exponential growth in number of permutations (for each starting word position, there would be 3,645 five-word permutations, given all permutations of the three word representations for each five-out-of-seven configuration).

**8** Our working assumption is that when it comes to allusions with altered word order, the exact word forms are less relevant, because the phrase is undergoing reformulation in any case. We therefore store order-neutral hashes only for the lexeme-based hashes.

**9** Of course, in computing the number of occurrences of vocalized words, it is imperative to count orthographic variants as the same word. From the perspective of a biblical allusion, it makes no difference whether text is *plene* or defective (e.g., תּוֹלְדֹת vs. תּוֹלְדוֹת, or כָּמֹכָה vs. כְּמוֹךָ), or whether an alternate *matres lectionis* is chosen (e.g., יְמַלֵּא vs. יְמַלֵּה). Thus, when encoding the word representations for the hashes, we group together orthographic variants which are pronounced identically and which fill the same morphological function. We assign a single numeric identifier to all variants in each group, so that their hashes will match one another.

this stage, we are left with some 2.5 million hashes (2.1 million order-neutral hashes, and 400,000 order-sensitive hashes).
–   As noted, each of the 2.5 million hashes can be represented in 128 bits. We store each hash along with a list of biblical verses (and word positions) identifying where the hash occurs. As per our semi-uniqueness criterion described above, a given hash will always be linked to a maximum of three verses, which fits within an additional 128-bit integer, so that the entirety of the hashes requires only 80 megabytes. Thus, we can easily keep the entire hash collection in memory, allowing for near-instant processing of the data.

Given this hash collection, we now present the algorithm for identifying these allusions within a target text:
–   We hash the target text in an analogous manner to that of the biblical text: for each word within the text, we consider all sequences of length two to five from that starting word position, allowing for word-skips and hash permutations as above.
–   Because the input text will generally be unvocalized, a given word can admit to multiple possible prefix-segmentation possibilities, and to multiple possible lexemes. For instance, the word מדברים can be a single word מִדְבָּרִים ("deserts," from the lexeme מִדְבָּר), or מְדַבְּרִים ("speaking," from the verbal lexeme דבר), or מִדְּבָרִים ("from items," from the lexeme דָּבָר). Therefore, for each word, we consider word representations for all such possible analyses, and for each sequence, we compute hashes for all permutations of the word analyses.
–   If a given sequence contains a lexeme or wordform which is not attested in the Bible, then we can discard it immediately, because it will certainly not match any of the biblical hashes.
–   For all remaining sequences, we compute the corresponding 128-bit hash, and we check to see whether that same hash appears in our collection of biblical hashes. A match indicates a basis for identifying a biblical allusion at that point in the text.

# 5 Interactive Online Implementation

A full interactive implementation of the algorithm presented here is freely available on the Dicta website: https://citation.dicta.org.il/. The site (pictured in Figure 2) allows the user to paste in any Hebrew text, and to immediately receive a visual

display of the results, mapping each biblical allusion in the input text to the relevant biblical verse.[10]
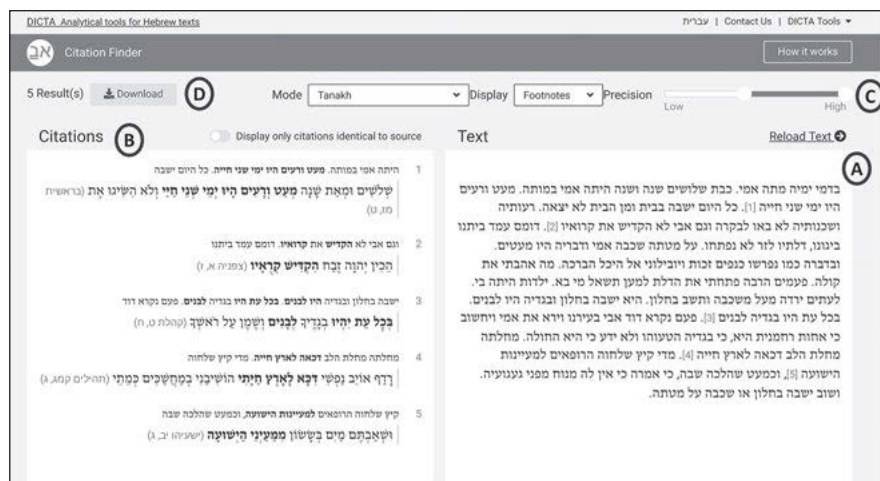
**Figure 2:** The Dicta Citation Finder interface, https://citation.dicta.org.il/, showing the input target text on the right (A), and identified biblical allusions on the left (B). The user can use a slider (C) to adjust the threshold, setting the precision/recall balance as desired. The user can request a download of the results in Microsoft Word format (D); in the resulting file, all identified allusions are integrated as footnotes. The relevant biblical verses are quoted in full in the text of the footnotes, and the relevant words within each verse are highlighted in bold.

Significantly, the site includes a slider to allow the user to set a threshold indicating how closely the text must match a verse in order to be considered a biblical allusion. As described above, every hash match indicates a possible basis for a biblical allusion, because it represents a case in which the target text uses a set of words or lexemes which is semi-unique within the Bible. Nevertheless, some matches are more significant than others, and some may seem to be more of a coincidence than an intended allusion.[11] In order to address this issue, we apply a scoring system to

---

**10** For the design and implementation of the interactive front-end, I wish to credit the following members of the Dicta team: Joshua Guedalia, Dvora Bloch, Dovid Lipman, Aryeh Sanders, and Rivka Sharfman.

**11** One example is the phrase כי אמרה כי ("because she said that"), which appears on the first page of Agnon's "In the Prime of Her Life" (Agnon, *Al Kapot Ha-man'ul*, 5). This phrase appears only a single time in all of Hebrew Scripture (Genesis 29:32), and thus prima facie its appearance in Agnon might be said to allude to that verse. However, because the phrase comprises two frequent function words and one very common verb, it may be considered a coincidental juxta-

rate the matches returned by the algorithm. Matches are awarded points for length and for rareness of the individual words within the match, and they are penalized for interpolated or deleted words, for morphologically altered words, and for deviation from the order of the biblical text. The slider allows the user to filter the matches based upon this score, effectively adjusting the precision/recall balance of the results. A higher threshold will ensure high precision – that is, the matches displayed will virtually all comprise cases of compelling and significant biblical allusions – but it will come at the cost of lower recall, wherein meaningful allusions might be omitted. On the flipside, a lower threshold ensures higher recall, but with more false positives.

Different use cases will entail different requirements in this regard. A user who is focusing on a short text – perhaps a single stanza of a poem – will wish to aim for maximum recall, even if it means wading through many less compelling matches. Thus, such a user will wish to set the slider to the minimum setting. In contrast, a user who is working on a lengthy text will likely prefer a higher threshold, to immediately identify as many substantial allusions as possible, even if it means sacrificing a certain number of less evident cases.

# 6 Evaluation

In order to evaluate the performance of our algorithm, we examine the first chapter of the book *The Fathers and the Sons* by S.J. Abramowitch (also known as "Mendele Mocher Sforim").[12] A domain expert evaluated the chapter (1,929 words in length) and identified 77 phrases as valid biblical allusions, alluding to 91 verses. Our algorithm successfully identified 76 of these 91 verses (recall = 83.5%), with a precision of 60%.[13] An inspection of the false positives indicates that over half of them are extra support verses for the same valid allusions.[14] We thus also consider how well

---

position of common words, rather than a meaningful allusion. Accordingly, the scoring system on the site rates the phrase rather low, such that the phrase will not appear at most thresholds. However, for users who wish to examine every reasonable possibility to ensure maximum recall, setting the threshold to a low setting will nevertheless turn up this match.

**12** Accessed January 25, 2022, https://benyehuda.org/read/5766.

**13** When evaluating the test set, we used a threshold of 18, equivalent to placing the slider one third of the way from the minimum position. We have found that this threshold generally provides optimal results.

**14** For instance, regarding the phrase בושת וכלמה, the domain expert noted Psalms 35:26, which contains those two words in succession. Our algorithm identified the Psalms verse, plus another verse with those same two lexemes, albeit with an intervening word (Isaiah 61:7).

our algorithm succeeds in identifying the phrases within the target text which comprise valid biblical allusions. Regarding this evaluation, we achieve a recall of 86% (66 out of 77), with a precision of 81% (15 false positives).

We release our test set as a public domain Zenodo dataset, so that it can be used to benchmark future research.[15]

# 7  Limitations and Future Directions

The algorithm discussed herein assumes that at the basis of any biblical allusion lies a set of shared lexemes with the biblical text, which may be morphologically altered or reordered in the target text. However, there is another class of biblical allusions, which connect to the text not with lexemes, but with paronomasia. For instance, consider the phrase שבטו יתן ("he shall apply his staff") used in a consolation poem penned by Moroccan sages during the mid-19th century.[16] This phrase clearly alludes to Exodus 21:19, שִׁבְתּוֹ יִתֵּן ("he must pay for his idleness"). The connection is cleverly forged by words which sound identical (שבטו, שבתו), but which are actually derived from two different lexemes (the first from שבט = "staff," and the second from שֶׁבֶת = "inaction"). The present algorithm would fail to find such allusions. In the future we hope to extend our algorithm to capture cases of soundplay as well.

# 8  Conclusion

The automation of biblical allusion identification presents a formidable challenge, because it entails capturing extremely flexible and subtle matches, while at the same time demanding discernment to prevent a flood of false positives. This chapter presented an algorithm to efficiently address this challenge. The algorithm allows instant processing of target texts while utilizing only a minimal amount of computer memory, capturing a wide range of biblical allusions both explicit and subtle, while limiting false positives to a manageable quantity.

The significance of this algorithm within the overall framework of Jewish Studies in the digital age is two-fold. First, this algorithm allows efficient automa-

---

**15** Avi Shmidman, "Biblical Allusions Test Set," *Zenodo*, accessed January 25, 2022, http://doi.org/10.5281/zenodo.5059159.

**16** David Ovaḍia, *La communaute de Sefrou*, vol. 2 (Jerusalem: self-pub., 1975), 84.

tion of processes previously performed laboriously by hand. Jewish Studies texts make extensive use of biblical allusions, regardless of genre: whether we are dealing with *halakhic* responsa, liturgical poetry, *haskalah* literature, or otherwise, biblical allusions abound, and those who prepare critical editions of such texts aim to document as many of those allusions as possible in the accompanying critical commentaries. Until now, the identification of such allusions throughout a text entailed substantial time and effort on the part of the editors and publishers of such editions. Fortunately, the shift to the digital age has brought along with it algorithms such as the one presented herein, which can greatly expedite the process of allusion identification. However, the impact of this algorithm upon Jewish Studies research is not limited to speeding up existing research patterns. Rather, the ability to automatically identify biblical allusions within a corpus of unlimited size paves the path for new "big data" analyses of Jewish Studies texts. For instance, it is now practical to generate statistical characterizations of texts in terms of the density of biblical allusions within them, and in terms of the particular biblical books that they tend to quote. Similarly, we can now generate statistical profiles for any given author of Hebrew texts, based upon the author's predilections in terms of biblical allusions. In this regard, the dawn of the digital age opens up new directions for comparing, contrasting, and analyzing Jewish Studies texts.

# References

Abramowitch, S.J. *The Fathers and the Sons*. Accessed January 25, 2022, https://benyehuda.org/read/5766. [Hebrew]

Agnon, S.Y. *Al Kapot Ha-man'ul*. Jerusalem and Tel-Aviv: Schocken, 1998. [Hebrew]

Elizur, Shulamit. *Hebrew Poetry in Spain in the Middle Ages*, vol. 3. Tel-Aviv: The Open University, 2004. [Hebrew]

Gusfield, Dan. *Algorithms on Strings, Trees, and Sequences*. Cambridge: Cambridge University Press, 1997.

Ovạdia, David. *La communaute de Sefrou*, vol. 2. Jerusalem: self-pub., 1975. [Hebrew]

Peursen, W.T. van, C. Sikkel, and D. Roorda. *Hebrew Text Database ETCBC4b*. DANS. 2015. Accessed January 25, 2022. https://github.com/ETCBC/bhsa/tree/master/source/4b.

Shmidman, Avi. "Biblical Allusions Test Set." *Zenodo*. Accessed January 25, 2022. http://doi.org/10.5281/zenodo.5059159.

Daria Vasyutinsky Shapira, Irina Rabaev, Ahmad Droby,
Berat Kurar Barakat, and Jihad El-Sana

# Is a Deep Learning Algorithm Effective for the Classification of Medieval Hebrew Scripts?

**Abstract:** In this research, we apply deep-learning techniques to Hebrew paleography to automatically classify and process medieval Hebrew manuscripts. Our work is based on contemporary Hebrew paleography (Malachi Beit-Arié, Colette Sirat, Norman Golb, Ada Yardeni, Benjamin Richler) that recognizes fifteen sub-types of medieval Hebrew script. Automatic recognition of these scripts allows to determine the approximate origin and date of writing for not-dated, fragmentary, and damaged manuscripts. To train the deep neural network, we compile a Visual Media Lab – Hebrew Paleography (VML-HP) dataset that contains 537 high-resolution manuscript page images. The images were hand-picked from the SfarData (http:/sfardata.nli.org.il/) dataset; in some rare cases, we also included pages from other manuscripts' collections. For testing the model, we define a notion of typical and blind test sets. The typical test set consists of the unseen pages of the manuscripts used in training. The blind test set, on the contrary, consists of pages from unseen manuscripts, thus, providing us with a real-life scenario. To train the model, we used patches extracted from the documents' pages. To filter irrelevant patches (empty patches or patches that contain decorations), we developed a clean patch generation algorithm that can generate patches containing pure text regions (for the VML-HP dataset, we generated 150K train patches). In all the experiments, we trained the network on the training set and tested it on both test sets, typical and blind. The objective training function was cross-entropy loss and was minimized using the Adam optimizer algorithm. The training was performed until there was no improvement in validation loss with five epochs' patience. The model with the least validation loss was used for testing.

**Keywords:** deep learning, digital Hebrew paleography

In this chapter we are presenting an interdisciplinary project that applies deep learning models to classify script types and sub-types in medieval Hebrew manuscripts. It incorporates the techniques and databases of Hebrew paleography and (with reservations) Hebrew codicology. This research project is part of our ongoing effort to develop algorithmic tools for processing historical documents

within the Visual Media Lab at the Department of Computer Science at Ben-Gurion University of the Negev, Israel.[1]

The ongoing digitization of manuscripts' collections kept in different libraries worldwide leads to the increasing availability of more and more volumes of manuscripts that once could have been studied only *in situ*. We have all reasons to believe that, within a few years, thousands more manuscripts around the globe would be properly digitized and available online. In the case of Hebrew manuscripts, this process is already very advanced, with the Institute for Microfilmed Hebrew Manuscripts at the National Library of Israel that already hosts more than 70,000 microfilms and thousands of digital images. These digitized documents constitute more than 90% of the known Hebrew manuscripts. Thus, automatic processing, or at least the primary computerized categorization of manuscripts, has become the most urgent task of modern Hebrew paleography.

Hebrew paleography emerged in the mid-20th century, side by side with modern Latin paleography, and with the same basic principles. The theoretical basis of Hebrew paleography is formulated in the works of Malachi Beit-Arié,[2] Norman Golb,[3] Benjamin Richler,[4] Colette Sirat,[5] and Ada Yardeni.[6] Contemporary Hebrew paleography identifies six main-types of scripts: Ashkenazi, Italian, Sephardic, Oriental, Byzantine, and Yemenite. Each main script type may contain up to three sub-types of scripts: square, semi-square, and cursive. In total, there are 15 Hebrew script sub-types. The paleographical classification of the ground truth for our project comes from the SfarData dataset,[7] which includes full codico-

---

**1** The participation of Dr. Vasyutinsky Shapira in this project is funded by Israeli Ministry of Science, Technology and Space, Yuval Ne'eman scholarship n. 3-16784.
**2** Malachi Beit-Arié, *Hebrew Codicology* (Jerusalem: Israel Academy of Sciences and Humanities, 1981); Malachi Beit-Arié and Edna Engel, *Specimens of Mediaeval Hebrew Scripts*, 3 vols. (Jerusalem: Israel Academy of Sciences and Humanities, 1987, 2002, 2017). See now the complete *opus magnum* Malachi Beit-Arié, *Hebrew Codicology – קודיקולוגיה עברית* (2021). Accessed January 26, 2022. http://doi.org/10.25592/uhhfdm.8849.
**3** Norman Golb and Omeljan Pritsak, *Khazarian Hebrew Documents of the Tenth Century* (Ithaca, NY: Cornell University Press, 1982).
**4** Binyamin Richler and Malachi Beit-Arié, eds., *Hebrew Manuscripts in the Biblioteca Palatina in Parma: Catalogue* (Jerusalem: Hebrew University of Jerusalem, Jewish National and University Library, 2001); Benjamin Richler, Malachi Beit-Airé, and Nurit Pasternak, "Hebrew Manuscripts in the Vatican Library," *Catalogue. Compiled by the Staff of the Institute of the Microfilmed Hebrew Manuscripts, Jewish National and University Library (Città del Vaticano)* (2008).
**5** Colette Sirat, *Hebrew Manuscripts of the Middle Ages* (Cambridge: Cambridge University Press, 2002).
**6** Ada Yardeni, *The Book of Hebrew Script: History, Palaeography, Script Styles, Calligraphy and Design* (Jerusalem: Carta, 1997).
**7** Accessed January 26, 2022, http://sfardata.nli.org.il/.

logical descriptions and paleographical definitions of all dated medieval Hebrew manuscripts until the year 1540 (this constitutes about 95% of the known dated medieval Hebrew manuscripts). The SfarData project was initiated by Malachi Beit-Arié in the 1970s and it is currently hosted at the site of the National Library of Israel.

Our project is an ongoing research. Our current goal is to develop algorithms to recognize Hebrew scripts and their sub-types. The practical applications at this stage would include:

– Determining the date and the area of writing. The paleographical classification of verified manuscripts enables machine learning models to learn the features common to each type and sub-type. The trained models can determine the sub-type of a query manuscript, which enables estimating the date of an undated manuscript or the place of copying. Thus, the application of this technology to fragmentary and faked text has the potential to roughly estimate where and when it was written. Today this task poses serious challenges and often only an experienced librarian or paleographer is capable of a plausible guess. There are many forged and incorrectly dated manuscripts, on the basis of which historical theories and histories of entire peoples are built. The use of a well-trained algorithm will allow us to objectively resolve such issues.

– Already at this stage, we expect the algorithm to be capable of producing a rough catalog of a collection of manuscripts where no trained human paleographer is available. Alongside the effort of the Institute for Microfilmed Hebrew Manuscripts to assemble the digital images of all the known Hebrew manuscripts, there still are important collections that have not been digitized and properly cataloged, such as the big collection of Hebrew manuscripts in the Vernadsky Library in Kyiv, currently in the most alarming state of preservation. Even the basic catalog made by the algorithm could attract to such collections the much-needed attention of the researchers.

– Identifying important parts of a manuscript, such as colophons, owner's notes. These additions to a manuscript are often written in a different script sub-type. Identifying them allows a researcher to recognize the date, place of copying, name of the scribe, etc.

– Tracking the movement of scribes, scholars, and communities over time through script and/or hand similarities.

– When the algorithm is further trained to recognize specific words, we would apply it to the biggest manuscript collections, such as the Firkowicz collections kept in St. Petersburg, collections of the Bibliothèque nationale de France, and others. This will allow us to take a closer look at some intriguing and fascinating but extremely complicated objects of research, when pieces of information

about them are scattered in the libraries around the globe. To take just one example, we could learn more about the Jews of Magna Graecia, with their physical and social mobility and intricate history. The relevant manuscripts from different libraries' collections can be identified, brought together, connected, and sorted out with the help of machine learning.

– Another possible application at this stage includes research of little-known, challenging, and often mysterious marginal Jewish communities, such as Georgian, Bukharan, Mountain Jews, about whom little is known today and whose history remains to a great extent legendary. History and works of the Jews of the Kingdom of the Two Sicilies and the Jews of Malta (to whom belonged the famous kabbalist Abraham Abulafia) before the expulsion by the king of Aragon, is another example of a potential application of the algorithm.

There are several ongoing projects in the research of the Hebrew manuscripts that complement ours; the most important among them are the Friedberg Genizah Project with its Cairo Genizah site[8] and the Judeo-Arabic corpus,[9] the eScriptorium,[10] and the Haifa Project for Research on the Dead Sea Scrolls.[11] There is also a very promising project at the Bar-Ilan university that works on building Hebrew manuscript metadata records and is focused on the manuscripts dated after 1540 – i.e., later than the classical Hebrew paleography.[12] Similar efforts to train an algorithm to recognize script types and build a web database exist in Latin paleography,[13] with its database.[14] A recent deep learning method[15] studies the impact of varying patch sizes on the performance of writer identification

---

**8** Accessed January 26, 2022, https://fjms.genizah.org/.

**9** Accessed January 26, 2022, https://ja.genizah.org/Home.aspx?isDoubleLogin=False.

**10** Accessed January 26, 2022, https://www.escriptorium.uk/.

**11** Accessed January 26, 2022, http://megillot.haifa.ac.il/index.php/en/.

**12** Gila Prebor, Maayan Zhitomirsky-Geffet, and Yitzchak Miller, "A New Analytic Framework for Prediction of Migration Patterns and Locations of Historical Manuscripts Based on Their Script Types," *Digital Scholarship in the Humanities* 35, no. 2 (2020): 441–58.

**13** Florence Cloppet et al., "Icdar2017 Competition on the Classification of Medieval Handwritings in Latin Script," *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, vol. 1 (IEEE, 2017), 1371–76; Linda Studer et al., "A Comprehensive Study of Imagenet Pre-training for Historical Document Image Analysis," *2019 International Conference on Document Analysis and Recognition (ICDAR)* (IEEE, 2019), 720–25.

**14** Accessed January 26, 2022, http://www.digipal.eu/.

**15** Akshai Punjabi et al., "Writer Identification Using Deep Neural Networks: Impact of Patch Size and Number of Patches," *2020 International Conference on Pattern Recognition* (IEEE, 2020), 3065–68.

for modern handwritten documents. Their results expose that the performance depends on the patch size and for each dataset a different patch size gives the best performance. Arabadjis et al.[16] classify the hands who wrote a given set of historical Byzantine Codices using manually designed features for matching a similarity score.

In our project, we built a medieval Hebrew manuscripts dataset, Visual Media Lab – Hebrew Paleography (VML-HP). The VML-HP dataset includes 500 pages labeled with 15 script types. To our best knowledge, this is the first publicly available Hebrew paleographic dataset. Currently, the dataset can be downloaded from https://www.cs.bgu.ac.il/~berat/. To provide a common baseline for algorithm assessment and comparison, we supply the partition of the VML-HP. The dataset is split into training and two test sets. The first test set, the typical test set, consists of unseen pages of documents present in the training set. The second, blind test set contains unseen manuscripts and imitates a real-life scenario. We present a case study for script type classification on the introduced dataset. We introduce a homogeneous style patch extraction method, where each patch contains a fixed number of lines. We also compare several established deep learning classification models and preprocessing methods. The obtained results show that there is big room for improvement on the blind test set, whereas the typical test set is an easier problem. Currently, we are working on exploring more advanced deep learning architectures that can capture fine-grained features of the Hebrew manuscripts. The fine-grained features are the features that aim to differentiate between hard-to-distinguish object classes, such as subtle differences in letter forms in different script sub-types.

# 1 Method

We propose to develop a computational tool that can recognize the script sub-type of a given Hebrew manuscript. Conventional recognition methods utilize handcrafted features, which mainly depend on careful design and expert knowledge. More advanced recognition methods are based on deep learning and can acquire effective feature representations from training data. Deep learning algorithms are backboned by neural networks inspired by human brain architecture, consisting of neurons and synapses among them. A deep learning algorithm is organized as a stack of layers, each of which is a collection of feature extrac-

---

**16** Dimitrios Arabadjis et al., "A General Methodology for Identifying the Writer of Codices: Application to the Celebrated 'Twins,'" *Journal of Cultural Heritage* 39 (2019): 186–201.

tors, so-called filters (Figure 1). Raster pixel values of a document image patch are fed into the network and are transformed into feature maps as they pass forward through the layers. Each layer extracts features at a different abstraction level. Initial layers detect primitive features such as dots, lines, and curves. Final layers combine these features into complex features such as corners, circles, and letters. At the final layer, the document image patch is classified into one of the script sub-types. The filters are updated according to a measure of the difference between the target and the predicted labels. The major drawback of a deep learning network is the necessity of a large amount of labeled train data, for example, 1,000 samples per class.
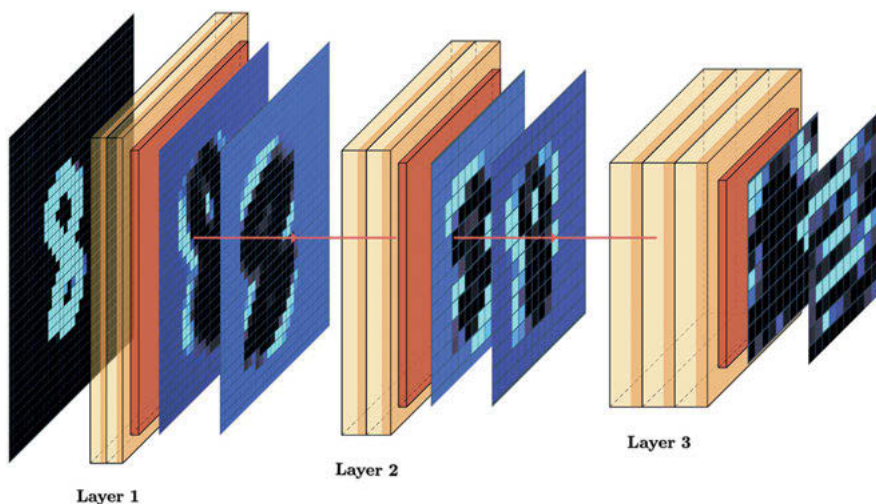


**Figure 1:** Illustration of a deep learning network It consists of stacked feature extraction layers. The early layers extract primitive features, and the later layers extract more complex features. Figure prepared by authors.

## 2 Building the dataset

SfarData, the database of Hebrew paleography and codicology, completed by Malachi Beit-Arié and his team, contains descriptions and classification of almost all known dated medieval Hebrew manuscripts; all the manuscripts in the database were studied in the libraries where they were kept. Malachi Beit-Arié and his team met with our team, discussed our project, gave us their full support, and allowed us to use their database in its entirety. Our team's paleographer, who is

herself a student of Malachi Beit-Arié, handpicked digitized pages from the manuscripts described in the SfarData as the raw material for our project. When for certain script sub-types we had to add manuscripts not described in SfarData, our paleographer picked them in accordance with the classification of medieval Hebrew manuscripts as described in SfarData.

Pages in the VML-HP dataset were extracted from high-quality digitized manuscripts, and we gave first preference to those kept in the National Library of Israel. We also used manuscripts from other libraries, first and foremost the British Library and the Bibliothèque nationale de France, with their significant collections of digitized manuscripts available for download. The dataset includes 500 pages in total. Table 1 details the distributions of the number of pages per main-type and sub-type scripts in the train set and two test sets.

**Table 1:** Summary of the VML-HP dataset.

| Main-Type | Sub-Type | Train | Typical Test | Blind Test | Total |
|---|---|---|---|---|---|
| Ashkenazi | Square | 16 | 4 | 10 | 30 |
| | Semi-Square | 16 | 3 | 10 | 29 |
| | Cursive | 16 | 4 | 10 | 30 |
| | **Total** | 48 | 11 | 30 | 89 |
| Byzantine | Square | 16 | 4 | 10 | 30 |
| | Semi-Square | 16 | 4 | 10 | 30 |
| | **Total** | 32 | 8 | 20 | 60 |
| Italian | Square | 16 | 4 | 10 | 30 |
| | Semi-Square | 16 | 4 | 10 | 30 |
| | Cursive | 16 | 4 | 10 | 30 |
| | **Total** | 48 | 12 | 30 | 90 |
| Oriental | Square | 64 | 14 | 10 | 98 |
| | Semi-Square | 16 | 4 | 10 | 30 |
| | **Total** | 80 | 18 | 20 | 118 |
| Sephardic | Square | 16 | 4 | 10 | 30 |
| | Semi-Square | 24 | 6 | 10 | 40 |
| | Cursive | 16 | 4 | 10 | 30 |
| | **Total** | 56 | 14 | 30 | 100 |
| Yemenite | Square | 24 | 6 | 10 | 40 |
| | Semi-Square | 24 | 6 | 10 | 40 |
| | **Total** | 48 | 12 | 20 | 80 |
| **Total** | | 312 | 75 | 150 | 537 |

# 3 Clean Patch Generation Algorithm

The VML-HP dataset contains 500 pages that represent 15 script sub-types. The ideal solution is to feed whole pages into the network because with larger input images, the network can capture more fine-grained features.[17] However, the input image cannot be greater than the size that fits the memory requirements. Therefore, we balance this tradeoff by cropping image patches that contain approximately five text lines, which is a sufficient size for human paleographers to classify the script type. Some parts of the pages contain irrelevant information, such as decorations, marginal drawings, or noisy background, as illustrated in Figure 2. Therefore, we developed a clean patch generation algorithm (https://www.cs.bgu.ac.il/~berat/data/hp_dataset.zip) that generates patches containing pure text regions and an approximately equal number of text lines.

To achieve this, we first calculate a square patch size for each page s×s that will include five lines. Then, we extract random patches of size s×s. The size of the extracted patches – i.e., the value of s – varies across manuscripts. Therefore, to remain consistent with our previous experiments, the patches are resized to 350 × 350. Examples of such clean patches are shown in Figure 3.

Calculating the patch size s×s for each page is done by first, extracting $k$ random patches of the size equal to one-tenth of the page height, as a patch of this size usually includes several text lines. Then the number of lines in a given



**Figure 2:** Example output patches from a naive patch generation algorithm. Some patches contain irrelevant features, some contain only a few characters, and others do not contain any text.



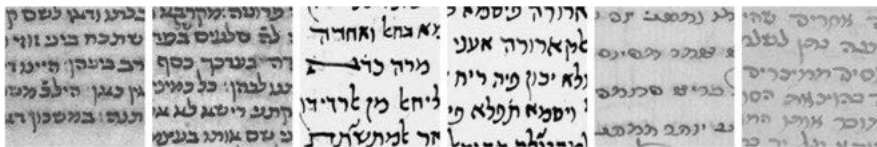**Figure 3:** Example output patches from the clean patch generation algorithm.

---

**17** Mingxing Tan and Quoc Le, "Efficientnet: Rethinking Model Scaling for Convolutional Neural Networks," *International Conference on Machine Learning* (PMLR, 2019).

patch is computed by counting the peaks of the *y* profile using Savitzky-Golay filter. Finally, the desired patch size is given by s = $\frac{h}{10} \times \frac{n}{m}$, where *h* is the height of the page, *n* is the average targeted number of lines, and *m* is the actual average number of lines in the *k* extracted patches. We used *n* = 5 and *k* = 20.

Furthermore, each extracted patch is validated according to the following conditions:

- The foreground area should be at least 20% of the total patch area and not exceed 70% of the total patch area. This condition eliminates almost empty patches and patches with large spots, stains, or decorations.
- The patch should contain at least 30 connected components. This condition eliminates patches with few foreground elements.
- The variance of the x and y profiles denoted by $\sigma_x$ and $\sigma_y$, respectively, should satisfy the conditions $\sigma_x \leq T_x$, $\sigma_y \geq T_y$. Assuming horizontal text lines, the variance of the *x* profile should be relatively low. During our experiments we set $T_x$ = 1500 and $T_y$ = 500.
- The following inequality should be satisfied:

$$0.5 \leq \frac{\sum_{i=0}^{\frac{v}{2}} P_x(i)}{\sum_{i=\frac{v}{2}}^{v} P_x(i)} \leq 1.5$$

- Where *v* is the number of values in the *x* profile and $P_x(i)$ is the *i*-th value. This condition eliminates the patches with text lines that occupy only a fraction of a patch.

## 4 Results

We experimented with several convolutional network architectures. In all the experiments, we train the network on the training set and test it on both test sets, the typical and the blind. We generated 150,000 patches from the training set, 10,000 patches from the typical test set, and 10,000 patches from the blind test set. The patches are generated using the clean patch extraction algorithm described in the previous section and are resized to the size of 350×350 pixels. The generated patches are equally distributed amongst all of the script types. The classification results are evaluated by the patch level accuracy and the page level accuracy. For the page level accuracy, the label of a page is computed by taking the majority vote of the predictions of 15 patches from the page.

## 4.1 Classifying Into 15 Script Types

Table 2 shows the accuracy results for classifying 15 script sub-types using different convolutional networks and compares the results on the typical and blind test sets at patch and page levels. As we can see from the results, the typical test set patches and pages are easier to classify. The gap in results on typical and blind test sets shows that the models are overfitting. The models have seen the pages from the typical test set during the training; however, the blind test set contains pages from unseen manuscripts. The difference in results shows that the models' learned features are specific to the manuscripts and not to the script type, like background texture. At nearly all levels and sets, the performance of the ResNet50 classifier is consistently higher; however, it does not surpass 40% accuracy on the blind test set. The random guess accuracy of 15 classes is 7.6%, indicating that the network can extract some script type features and improves the random classification accuracy. We can argue that script type classification is an expressible function, but the network needs more data to learn this function.

**Table 2:** Patch and page level accuracies on typical test set and blind test set using different network architectures for classifying 15 script sub-types.

|  | Patch level | | Page level | |
| --- | --- | --- | --- | --- |
|  | **Typical** | **Blind** | **Typical** | **Blind** |
| DenseNet | 97.97 | 32.95 | 98.63 | 38.36 |
| AlexNet | 91.99 | 27.03 | 93.15 | 28.28 |
| VGG11 | 99.16 | 35.55 | **100.00** | 35.63 |
| SqueezeNet | 98.03 | 30.38 | 98.63 | 29.45 |
| ResNet18 | 97.07 | 30.95 | 98.63 | 34.25 |
| ResNet50 | **99.55** | **36.15** | 98.63 | **39.73** |
| InceptionV3 | 94.94 | 26.41 | 95.89 | 26.71 |

## 4.2 Classifying Square and Cursive Script Types

Table 3 shows the accuracy results for classifying only two script types, square and cursive, using different convolutional networks. From a human paleographer's point of view, it is almost impossible to make a mistake and mix up square and cursive script (while the boundaries between square and semi-square, and semi-square and cursive can be blurry). Thus, a good result obtained by the algorithm in this case indicates that the algorithm learns the correct features in the manuscript, which represent the script itself. We can note that the typical test

set accuracy is fully saturated, whereas there is still little room for improvement at blind test set accuracy. This result strengthens the above argument that more samples should be used in the training phase; we see that decreasing the number of classes from 15 to two (which increased the number of samples per class), leads to higher accuracy.

**Table 3:** Patch and page level accuracies on typical test set and blind test set using different network architectures for classifying square and cursive script sub-types.

| | Patch level | | Page level | |
|---|---|---|---|---|
| | Typical | Blind | Typical | Blind |
| DenseNet | 99.85 | 87.06 | 100.00 | 83.72 |
| AlexNet | 99.48 | 88.01 | 100.00 | 91.86 |
| VGG11 | 99.93 | 86.45 | 100.00 | 88.37 |
| ResNet18 | 96.65 | 86.85 | 100.00 | 87.21 |
| ResNet50 | 99.99 | 90.58 | 100.00 | 94.19 |
| SqueezeNet | 98.03 | 82.45 | 100.00 | 86.05 |
| InceptionV3 | 99.16 | 82.06 | 100.00 | 20.23 |

# 5 Discussion

The classification accuracy of the best performing model – i.e., ResNet50 – is around 35%. When comparing class accuracies (Figure 4) we found that particular classes have an accuracy over 50% – i.e., Yemenite semi-square, Sephardic square, Sephardic semi-square, Oriental square, and Ashkenazi semi-square. These results indicate that a classification system designed only for these classes will have higher page level accuracy, since the page level accuracy is computed by the majority vote over the patches from the same page.

The Byzantine square sub-type has a very low accuracy because it was confused with Byzantine semi-square (Figure 5). Interestingly, this confusion is not mutual because the Byzantine semi-square sub-type was confused with Italian. In contrast, the confusion among the Italian, Oriental, Sephardic, and Yemenite semi-square sub-types are mutual. The mutual confusions can be due to the paleographers' ambiguity in the ground truth of semi-square types or to insufficient ground truth.
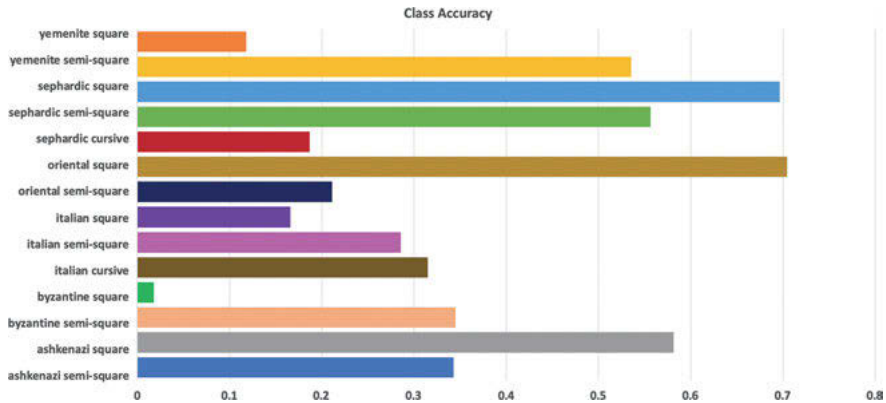
**Figure 4:** Patch level class accuracies on blind test set using ResNet50 network, prepared by authors.
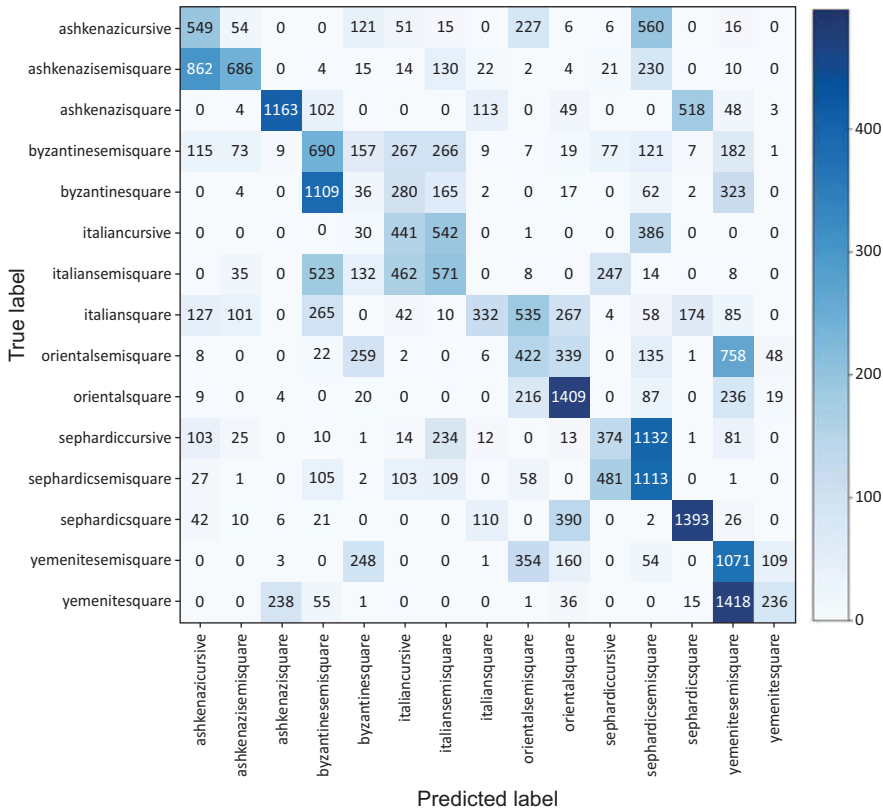


**Figure 5:** Patch level confusion matrix on blind test set using ResNet50 network; prepared by authors using https://www.cs.ryerson.ca/~aharley/vis/conv/flat.html (accessed 26 January, 2022).

# 6 Conclusions

From the paleographic point of view, it would be beneficial to gain insight into the features that underlie the class decisions. We are developing a fine-grained classification model that can spot the regions taken into account for script type decisions. In addition, we are collecting and labeling more document page images, as the machine learning for a 15-class problem requires around 15,000 samples in total. Our algorithm significantly surpasses the random guess accuracy of 15 classes (7.6%) and this indicates that the network can extract some script type features. We can argue that script type classification is an expressible function, but the network needs more data to learn this function. When more material is brought for comparison and the size of the test, train, and blind sets increases, the accuracy of the algorithm will improve.

# 7 Our Work and its Place in the Overall Theme of Jewish Studies in the Digital Age

Our project belongs to the field of digital research of manuscripts and historical documents.

The quantity of digitally available manuscripts and documents in different libraries and archives is constantly growing and already the amount of material available is often more than an individual researcher could process manually. In all likelihood, in the foreseeable future, a human researcher will formulate a problem and the processing of large amounts of data will be assigned to an algorithm. For this to be possible, algorithms must recognize, classify, and ultimately search through large quantities of unrecognized manuscripts and documents. The integration of computer-based techniques can now bring to the manuscripts' research the often-missing quality of objectivity, possibility of objective verification of results. It also brings with it the possibility of solving problems that are beyond the physical capacities of an individual researcher.

We use the theoretical framework of Hebrew paleography to train deep learning neural networks to classify Hebrew script types and sub-types and our project works side by side and complements such ongoing projects as eScriptorium, the Friedberg Genizah Project, and more.

# References

Arabadjis, Dimitrios et al. "A General Methodology for Identifying the Writer of Codices: Application to the Celebrated 'Twins.'" *Journal of Cultural Heritage* 39 (2019): 186–201.

Beit-Arié, Malachi. *Hebrew Codicology* – קודיקולוגיה עברית. 2021. Accessed 26 January, 2022. http://doi.org/10.25592/uhhfdm.8849

Beit-Arié, Malachi. *Hebrew Codicology*. Jerusalem: Israel Academy of Sciences and Humanities, 1981.

Beit-Arié, Malachi and Edna Engel. *Specimens of Mediaeval Hebrew Scripts*, 3 vols. Jerusalem: Israel Academy of Sciences and Humanities, 1987, 2002, 2017.

Cloppet, Florence et al. "Icdar2017 Competition on the Classification of Medieval Handwritings in Latin Script." *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, vol. 1, 1371–76. IEEE, 2017.

Golb, Norman and Omeljan Pritsak. *Khazarian Hebrew Documents of the Tenth Century*. Ithaca, NY: Cornell University Press, 1982.

Prebor, Gila, Maayan Zhitomirsky-Geffet, and Yitzchak Miller. "A New Analytic Framework for Prediction of Migration Patterns and Locations of Historical Manuscripts Based on Their Script Types." *Digital Scholarship in the Humanities* 35, no. 2 (2020): 441–58.

Punjabi, Akshai et al., "Writer Identification Using Deep Neural Networks: Impact of Patch Size and Number of Patches." *2020 International Conference on Pattern Recognition*, 3065–68. IEEE, 2020.

Richler, Binyamin and Malachi Beit-Arié, eds. *Hebrew Manuscripts in the Biblioteca Palatina in Parma: Catalogue*. Jerusalem: Hebrew University of Jerusalem, Jewish National and University Library, 2001.

Richler, Benjamin, Malachi Beit-Arié, and Nurit Pasternak. "Hebrew Manuscripts in the Vatican Library." *Catalogue. Compiled by the Staff of the Institute of the Microfilmed Hebrew Manuscripts, Jewish National and University Library (Città del Vaticano)*. 2008.

Sirat, Colette. *Hebrew Manuscripts of the Middle Ages*. Cambridge: Cambridge University Press, 2002.

Studer, Linda et al., "A Comprehensive Study of Imagenet Pre-training for Historical Document Image Analysis." *2019 International Conference on Document Analysis and Recognition (ICDAR)*, 720–25. IEEE, 2019.

Tan, Mingxing and Quoc Le. "Efficientnet: Rethinking Model Scaling for Convolutional Neural Networks." *International Conference on Machine Learning*. PMLR, 2019.

Yardeni, Ada. *The Book of Hebrew Script: History, Palaeography, Script Styles, Calligraphy and Design*. Jerusalem: Carta, 1997.

Joshua Waxman

# Projecting Punctuation From an Interpolated Translation and Commentary

**Abstract:** Many classical Jewish texts were composed without punctuation, and modern punctuated versions are desired. We describe an algorithm we employed to create a punctuated digital edition of the Babylonian Talmud, a text composed in a mixture of Hebrew and Aramaic. Our approach is to first word-align the original Talmudic text with an interpolated translation and commentary composed by Rabbi Adin Steinsaltz. We apply heuristics to identify the governing punctuation in the commentary text and then project that punctuation along with other lexical features. Our results are quite good overall, with both recall and precision often in the 90%–95% range, when compared with an edited punctuated text.

**Keywords:** digital editions, projection, alignment, automatic punctuation

## 1 Introduction

Traditional historical Hebrew texts lack punctuation, such as commas and periods, which could inform the reader where a phrase or sentence begins and ends. Even when readers are able to determine sentence boundaries, they may be unsure of the tenor of a sentence – whether it is a question, answer, exclamation, or simple statement – something that exclamation marks, question marks, and periods can help disambiguate. These texts are often unvocalized as well, introducing an entirely different set of ambiguities, which couple with the punctuation ambiguities. This presents a huge obstacle and learning curve to the novice first approaching these texts.

An example of such an unpunctuated text is the Babylonian Talmud. Indeed, most of the Talmudic tractates that appear on Sefaria's website lack punctuation.[1] While the Koren printed version of this Talmud has punctuation, the digital version does not. We therefore set out to generate a freely available punctuated

---

[1] See www.sefaria.org. After implementing this first version of this algorithm, I contacted Sefaria to offer them the tool and generated corpus, prior to the beginning of the Daf Yomi cycle. It turns out that they independently had developed a similar process to extract punctuation from the Steinsaltz Hebrew commentary. Following the Daf Yomi cycle, they release each punctuated tractate, with human editing to correct egregious errors.

digital edition of the Talmud, which can be useful both for readers and for downstream computational analysis.

## 1.1 The Talmud's Concern with Punctuation

An ambiguity in the opening statement of the Babylonian Talmud illustrates the difficulties of ambiguous sentence segmentation and how punctuation can help. The first Mishnah on Berakhot 2a begins:

> מאימתי קורין את שמע בערבין משעה שהכהנים נכנסים לאכול בתרומתן עד סוף האשמורה הראשונה דברי ר' אליעזר וחכמים אומרים עד חצות

> (From when does one recite *Shema* in the evening? From the time when the priests enter to partake of their *teruma*. Until the end of the first watch. That is the statement of Rabbi Eliezer. The Rabbis say: until midnight.)

The Mishnah is concerned with a start time for recitation of the evening Shema, and a start time is given. Then, an end time is given, followed by an attribution to Rabbi Eliezer ben Hyrcanus. Finally, the Sages disagree with the end time. Due to the lack of punctuation, it is unclear whether the answer as to the start time is part of Rabbi Eliezer's statement, or stands apart as an anonymous first Tanna (the *Tanna Kamma*).

A *brayta* on Berakhot 2b records Rabbi Eliezer as giving a different starting time, "from the time when the day becomes sanctified on the eve of Shabbat," but the Talmud (Berakhot 3a) addresses the contradiction with the Mishnah:

> קשיא דר' אליעזר אדר' אליעזר תרי תנאי אליבא דר' אליעזר ואב"א רישא לאו ר' אליעזר היא:

> (The opinion of Rabbi Eliezer contradicts the opinion of Rabbi Eliezer! Either these are two *tannaim* in accordance with Rabbi Eliezer, or alternatively, the first part of the statement is not from Rabbi Eliezer.)

To put it into our own terms, there might be a comma after the word בתרומתן, in which case the attribution of the full statement belongs to Rabbi Eliezer (with a resultant contradiction). Alternatively, there is a period after the word בתרומתן. If so, Rabbi Eliezer did not author the statement in the Mishnah about the starting time, eliminating the contradiction.

This example demonstrates that punctuation can play an important role in understanding the Talmudic discourse, and that resolving such ambiguities sometimes even appears in the Talmudic discourse. Further, it shows possible limitations of automatic linguistic approaches when compared to manual punctuation by a knowledgeable human. Simple syntactic and lexical knowledge are insuffi-

cient here – a deep contextual knowledge, in this case, of Rabbi Eliezer's opinion as stated in the *brayta* – is required.

While usually punctuation aids readers by making the parsing of a text more straightforward, this is an interesting case because the Talmud's question relies on the ambiguity. Which of the two parses offered by the Talmud should be used in a punctuated text? This is an issue shared by translations and commentaries, which translate or explain ambiguous terms or phrases based on the ultimate conclusions of the Talmud's analysis, where readers must reinject the ambiguity. Automated punctuated would always select one option. Here, Steinsaltz took pains to punctuate in a special way that preserves the ambiguity.[2] Clearly, this punctuating decision is something best performed by a thoughtful human.

## 1.2 The Importance of Punctuation

People who try to read the Talmud are faced with several obstacles and ambiguities to overcome. Those for whom Hebrew is not their first language must mentally translate the text, and possess an extensive foreign-language vocabulary. Talmudic statements are often terse, and many details need to be filled in. (See Figure 1 and compare the amount of literal and gloss text.) Further, the words in the standard Vilna edition of the Talmud lack vowel points, and several point combinations are possible. Is the word שקמתי to be read *shiqmati* ("my sycamore") or *she-qamti* ("that I have arisen")?[3] Hebrew is morphologically rich, and one word may be subdivided into several constituting morphemes, with the mental division again based on the aforementioned vowels points. Thus, הקפה could be ה + קפה ("orbit"), ("the coffee"), הקף שלה ("her perimeter") and so on, to take an example from.[4] One would select the best meaning based on context. Meanwhile, the Talmud is com-

---

**2** He wrote:

מאימתי קורין את שמע בערבין? משעה שהכהנים נכנסים לאכול בתרומתן. עד סוף האשמורה הראשונה. דברי ר' אליעזר. וחכמים אומרים: עד חצות.

By placing a period after both statements that might be ascribed to Rabbi Eliezer, including immediately before דברי ר' אליעזר, "These are the words of Rabbi Eliezer," he deliberately leaves the parsing ambiguous and anticipates the Talmud's question. In contrast, his regular punctuation would be to place a comma before דברי, such as on Berakhot 2b, preceding both דברי רבי מאיר and דברי רבי אליעזר; and on Berakhot 31b, preceding דברי רבי ישמעאל.

**3** This example is drawn from the Song of Deborah in Judges 5:7, where it is vocalized *she-qamti*.

**4** Reut Tsarfaty, Amit Seker, Shoval Sadde, and Stav Klein, "What's Wrong with Hebrew NLP? And How to Make It Right," in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, EMNLP, Hong Kong, 2019.

posed in a mix of Biblical Hebrew, Middle Hebrew, and Babylonian Aramaic, with different vocabulary and syntax, and with frequent code switching between them, often at phrase or sentence boundaries. The presence of multiple languages introduces new ambiguities, so דבר no longer could mean only "speak" or "thing" but also "of the son of." People who study the Talmud read a phrase at a time and then resolve its meaning, and without punctuation, it can be difficult to know where a phrase or sentence ends. Knowledge of phrase boundaries could limit the preceding ambiguities. Finally, the Talmud is a record of intensive debate, and it can be ambiguous whether a statement is a question or an answer. Punctuation can make the reading of the Talmud somewhat easier.

These ambiguities present a problem as well for computer programs. Tsarfaty and colleagues present a joint morpho-syntactic parsing framework for Modern Hebrew (YAP), as opposed to the typical pipeline model.[5] Each ambiguity (vowel points, lemmatization, part of speech tagging) is not resolvable in sequence, but is only apparent in context. For the aforementioned reasons, the Talmudic text is even more difficult to process. Our ongoing work in named entity recognition, relation extraction, and discourse analysis of the Talmud, described in part by Waxman would benefit from a punctuated text.[6] The process described in this article will become an upstream process for future work, enabling processing of the actual Talmudic text.

# 2 Method

## 2.1 Resources

The William Davidson Talmud is a digital edition of the Babylonian Talmud made available at Sefaria.[7] Among its features are three parallel and aligned texts: the actual Hebrew and Aramaic Talmudic text (henceforth called "Original Text"); Rabbi Adin Steinsaltz's Modern Hebrew translation and commentary ("Hebrew Commentary"), and his English translation and commentary ("English Commentary"). Some of these texts were previously available in printed form – the

---

**5** Tsarfaty et al., "What's Wrong with Hebrew NLP?"
**6** Joshua Waxman, "A Graph Database of Scholastic Relationships in the Babylonian Talmud," in *Proceedings of the Digital Humanities Conference 2019*, Utrecht, 2019; Joshua Waxman, "A Graph Database of Scholastic Relationships in the Babylonian Talmud," *Digital Scholarship in the Humanities* 36, no. 2 (2021): 277–89. Accessed March 1, 2022, https://doi.org/10.1093/llc/fqab015.
**7** Accessed March 1, 2022, https://www.sefaria.org/william-davidson-talmud.

**Figure 1:** A selection of the Hebrew Steinsaltz Talmud, Berakhot 2a.

Hebrew Commentary in the original Hebrew Steinsaltz Talmud, 1968–2010, and the English Commentary in the Noé Edition of the Koren Talmud Bavli.

The Original Text differs slightly from the standard Vilna Edition Talmud. For instance, abbreviations ('ר for רבי, כ"א for אם כן) are expanded, and censored passages have been restored. It also differs from the Talmudic text in the printed Steinsaltz editions of the Talmud. Those editions contained elaborate punctuation as well as Hebrew vowel points, which the Original Text lacks. Also, it draws from a variant text, so many words and phrases are different.[8] The center column of Figure 1 contains an example of the printed vocalized, punctuated Talmudic text.

The Hebrew Commentary is an interpolated translation and commentary, as shown in the leftmost column of Figure 1. Since Mishnaic / Talmudic Hebrew, Babylonian Aramaic, and Modern Hebrew are all Semitic languages, there is a significant overlap in their vocabulary. The original and concise Talmudic words are included in the commentary as bold text, with smaller nonbold gloss text in between providing commentary and elaboration. This gloss is sometimes brief and provides a smoother flow. For instance, the original Talmudic text is

---

**8** See, for instance, Kiddushin 31a, where the Original Text has נכרי due to censorship, while Steinsaltz's printed text and Hebrew Commentary restore the word גוי. Similarly, see Shabbat 135b where the Original Text has an entire repeated sentence from the *brayta* omitted by the printed Steinsaltz Talmud and Hebrew commentary.

תנא היכא קאי דקתני מאימתי. In the commentary text, the first word is expanded into התנא של משנתנו. The appended Hebrew prefix ה is the Hebrew definite article, where Aramaic does not employ it, and של משנתנו elaborates that the Tanna under discussion is the anonymous narrator of the Mishnah. This becomes flowing Modern Hebrew text, or a mixture of Modern Hebrew, Hebrew, and Aramaic text. Other short interjections include framing the discourse, as a question or answer. Other times, the commentary introduces background and Scriptural basis, or elaborates on unfamiliar concepts, in which case the nonbold commentary text is longer.

When an Aramaic word or phrase occurs which cannot pass for Hebrew, it still appears in bold, but is followed by a Modern Hebrew translation in square brackets. Within these brackets, the literal translation of words is printed in a `Courier font`, while any elaboration for the sake of flow appears in the typical gloss font. For example, in the figure, the continuation of the sentence under discussion is היכא קאי דקתני מאימתי, which is glossed in brackets as [היכן הוא עומד **"מאימתי"** שונה], בזה שהוא ממשיך, הוא דברים אלו על בהסתמך. When Mishnaic or Talmudic Hebrew appears as literal text but is sufficiently awkward in Modern Hebrew, it is similarly glossed, but in parentheses. Thus, in the Mishnah, מאימתי is glossed as (ממתי, מאיזו שעה) and בערבין is glossed as (בערבים).

Importantly for our purposes, the Hebrew Commentary contains punctuation. Further, the digital edition of the text carries over all of the punctuation and formatting of the original, so we can distinguish between bolded literal Talmudic text and nonbolded gloss, gloss affixes to literal text, translations of Hebrew and translations of Aramaic, and which translations of Aramaic are literal and which are gloss.

Our third resource, the English Commentary, is an interpolated commentary as well, alternating between bolded literal translated text and nonbolded gloss text, with transliterations in italics. Within the Sefaria digital edition, these three texts (Original Text, Hebrew Commentary, and English Commentary) are translation-unit aligned. This is not a word, phrase, sentence, or paragraph, but some logical unit of the text that an editor selected. Sometimes this translation unit is an entire paragraph, including multiple statements by multiple speakers.

Another noteworthy resource is the vocalized Talmudic text produced by Dicta's Nakdan project.[9] Their system produces Hebrew vowel points for unvocalized text. While it works best on Modern Hebrew, by interacting with a human editor, they have produced a vocalized Talmudic text for all of the Orders of *Zera'im* and *Moed*. This digital edition is freely available at http://daf-yomi.com/.

---

**9** See Avi Shmidman, Shaltiel Shmidman, Moshe Koppel, and Yoav Goldberg, "Nakdan: Professional Hebrew Diacritizer," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, 2020.

מְאִירֵימוּ

"קוֹרִין אֶת "שְׁמַע" בָּעֲרָבִין? — מִשָּׁעָה שֶׁהַכֹּהֲנִים
נִכְנָסִים לֶאֱכוֹל בִּתְרוּמָתָן עַד סוֹף הָאַשְׁמוּרָה
הָרִאשׁוֹנָה, דִּבְרֵי רַבִּי אֱלִיעֶזֶר. וַחֲכָמִים אוֹמְרִים:
עַד חֲצוֹת. רַבָּן גַּמְלִיאֵל אוֹמֵר: עַד שֶׁיַּעֲלֶה עַמּוּד
הַשַּׁחַר. מַעֲשֶׂה וּבָאוּ בָנָיו מִבֵּית הַמִּשְׁתֶּה, אָמְרוּ
לוֹ: לֹא קָרִינוּ אֶת "שְׁמַע". אָמַר לָהֶם: אִם לֹא עָלָה
עַמּוּד הַשַּׁחַר חַיָּיבִין אַתֶּם לִקְרוֹת. וְלֹא זוֹ בִּלְבַד
אָמְרוּ, אֶלָּא כָּל מַה שֶּׁאָמְרוּ חֲכָמִים עַד חֲצוֹת
מִצְוָתָן עַד שֶׁיַּעֲלֶה עַמּוּד הַשַּׁחַר; הֶקְטֵר חֲלָבִים
וְאֵבָרִים — מִצְוָתָן עַד שֶׁיַּעֲלֶה עַמּוּד הַשַּׁחַר, וְכָל
הַנֶּאֱכָלִים לְיוֹם אֶחָד — מִצְוָתָן עַד שֶׁיַּעֲלֶה עַמּוּד
הַשַּׁחַר, יְאִם כֵּן, לָמָּה אָמְרוּ חֲכָמִים עַד חֲצוֹת? —
כְּדֵי לְהַרְחִיק אָדָם מִן הָעֲבֵירָה.

ומביא רבן גמליאל כמה דוגמאות לעקרון זה, כגון: מצות **הקטר** (שריפת) **חלבים ואיברים** של הקרבנות על המזבח, אשר בשל ריבוי הקרבנות שהוקרבו לא הספיקו במהלך היום להעלות על המזבח את כל החלבים והאיברים, ועל כן המשיכו להקטירם כל הלילה, ככתוב "זאת תורת העולה, היא העולה על מוקדה על המזבח כל הלילה עד הבוקר ואש המזבח תוקד בו" (ויקרא ו, ב), וכן **כל הנאכלים ליום אחד**, אותם קרבנות, כמו החטאת והאשם, שמותר לאכלם רק ביום שחיטתם ובלילה שלאחריו, אף שאמרו חכמים שהם נאכלים "ליום ולילה עד חצות" מדין התורה מצותן עד שיעלה עמוד השחר, אבל לא מאוחר יותר, ככתוב "ביום קרבנו יאכל לא יניח ממנו עד בקר" (שם ז, טו). אם כן, למה כן אמרו חכמים שזמנם "עד חצות" — כדי להרחיק את האדם מן העבירה, עלול אדם להתרשל ולדחותה, עד אשר יעבור זמנה של המצוה.

**Figure 2:** Steinsaltz punctuated text of the Mishnah on Berakhot 2a (right) and a selection of the Hebrew commentary (left).

Sefaria has paired Dicta's vocalized Talmudic text with a punctuation algorithm similar to the one described in this chapter.[10] Before publishing this text on their website, it is manually edited to fix errors and add missing punctuation. They regularly publish each tractate in advance of the Daf Yomi cycle start of that tractate, and that vocalized punctuated text is what is presented as the Hebrew-Aramaic Talmudic text on their website. We use this edited punctuated text as a "Gold Standard" to evaluate our own system's results.

## 2.2 Approach

Within each translation unit of the digital Sefaria corpus, we select the Original Text and the aligned Hebrew Commentary, in order to project the punctuation from the commentary to the Talmud.

In our first step, we parse the HTML of the Hebrew Commentary to distinguish bolded from nonbolded text, and tokenize the text to separate words and punctuation. We thus obtain a corpus of tokens, each marked as literal translation or gloss. Some of the punctuation marks in this corpus are marked literal, since they were bolded along with their adjacent word, but these are not necessarily the

---

**10** See footnote 1.

punctuation marks we want in the final corpus. For instance, in the gloss on the Mishnah on Berakhot 2a, consider the commentary text in Figure 2. The bolded punctuation after the word השחר is a comma, but we would prefer the nonbolded punctuation mark of a period which appears at the end of the gloss text. There are punctuation marks in the gloss text that we would like to utilize as well. For example, in Figure 2, after הקטר חלבים ואברים, we would like to obtain one of the commas which follow in the gloss text.

In theory, this Hebrew Commentary, stripped of all gloss words, could serve as the entire basis for our punctuated corpus. After all, the bold, literal words are the Talmudic text, and it is just a matter of deciding which punctuation marks to interject. In practice, there are several reasons this would not suffice. Firstly, the version of the Talmudic text in the commentary will occasionally differ from our desired text. Thus, in Figure 2, the desired text of the Mishnah is as on the right, in which there is a first clause of הֶקְטֵר חֲלָבִים וְאֵבָרִים מִצְוָתָן עַד שֶׁיַּעֲלֶה עַמּוּד הַשַּׁחַר and a second clause of וְכָל הַנֶּאֱכָלִים לְיוֹם אֶחָד מִצְוָתָן עַד שֶׁיַּעֲלֶה עַמּוּד הַשַּׁחַר. The version in the commentary text collapses the statements into הֶקְטֵר חֲלָבִים וְאֵבָרִים וְכָל הַנֶּאֱכָלִים לְיוֹם אֶחָד מִצְוָתָן עַד שֶׁיַּעֲלֶה עַמּוּד הַשַּׁחַר. Other differences occur due to censorship – e.g., גוי vs נכרי for "gentile." Still others include plene vs. deficient spellings of words, corrections of errors such as דתנן for דתניא, or separation of one word into two (אי בעית vs. איבעית). Secondly, translation considerations occasionally cause omissions in the commentary text. For instance, in Figure 1, the correct Talmudic text is לִתְנֵי דְּשַׁחֲרִית בְּרֵישָׁא, but the bolded commentary text omits the leading ד, because of the gloss Hebrew word של. Thirdly, we would like to project the punctuation onto other Talmudic texts available on Sefaria, which have expanded rabbinic titles from 'ר to רב or רבי, expanded abbreviation such as כ"א to אם כן, and have introduced Hebrew vowel points.

Therefore, our second step is word alignment of the Original Text and Hebrew Commentary. These texts were already aligned by translation unit (sentence or paragraph), but this is a further alignment. Since we are aligning words that are highly similar and occur in the same order across texts, but with gaps due to gloss text or textual variants, we use the Needleman-Wunch algorithm. This algorithm aligns matches in parallel sequences, rewards matches, and penalizes mismatches and gaps (with a constant penalty regardless of gap length). To capture matches even with slight spelling differences, our award is based on the ratio of overlapping character bigrams of the letters of the two words.

We actually first select only the bolded, literal tokens of the Hebrew Commentary, and align just those tokens against the Original Text tokens. We then utilize the original word positions to apply the alignment to the full corpus. While utilizing the unfiltered Hebrew Commentary would have worked because of the gap penalty, the alignment could fail in instances such as short phrases surrounded by

gloss text. Instead, we utilize the manually produced bolded text, which is more reliable.

Our third step is projection of punctuation across these word-aligned texts. Only literal words, and not gloss text or bolded or nonbolded punctuation, have been aligned. We loop through the aligned Original Text, and any time there is a gap, we consider all of the Hebrew Commentary punctuation tokens parallel to that gap. We rank the punctuation tokens as in Table 1.

**Table 1:** Punctuation ranking.

| Punctuation | Rank |
|---|---:|
| Comma | 1 |
| Semicolon | 2 |
| Colon | 3 |
| Em-dash | 4 |
| Period | 5 |
| Exclamation Mark, Question Mark, Interrobang | 6 |

This is because, at times, the intervening text has multiple punctuation marks, which can appear in any order. Our heuristic is to select the first occurrence of the maximum ranked punctuation mark. Thus, a period would replace a comma, but a comma would not replace a period. If we have already encountered a question mark, which finishes the sentence, we would not replace it with an exclamation mark that appears later in the gloss text, because it is of equal rank.

Parentheses in the Hebrew Commentary are not projected, because they either represent citations or translations of Rabbinic Hebrew to Modern Hebrew, which is not punctuation that should appear in the Original Text. Furthermore, in searching for punctuation with the highest rank, we ignore all other punctuation marks which appear inside parentheses, since they do not refer to anything in the Talmudic text. Likewise, we ignore brackets, which represent translations from Aramaic into Hebrew, as well as most punctuation inside those brackets. We don't ignore punctuation immediately preceding the closing bracket, as often the commentator places the punctuation there instead of after the closing bracket.

We also project quotation marks, which are employed in quotations of biblical verses as well as in quotation of key phrases under discussion from Tannaitic sources mentioned earlier in the text. In the Hebrew Commentary, these are non-directional quotation marks, and we would like to project them in pairs, so as to have both open and closed quotation marks. We consider two cases. If the

quotes begin and end in the gloss text, then they are not relevant to the actual Talmudic text and are ignored. If the open quote appears in one gap and the close quote appears in another gap, then they enclose the intervening literal Talmudic text and are properly projected. We ignore other punctuation marks inside of ignored quotation marks.

We do not model level of certainty in our output corpus. Our heuristic selects and projects just one punctuation mark for each gap. In theory, we could have projected each occurring punctuation mark and labeled our judgment about each – e.g., that a comma followed by a period in gloss text is only 10% likely to be authentic and 90% likely to be the period. In future work, we may take up this problem.

The final version of our algorithm works well, but we encountered a few challenges during our implementation. The Hebrew Commentary is encoded in HTML, with markup tags for different types of text (original words, gloss, Hebrew translation). Stripping all HTML or parsing the HTML ourselves provided noisy or insufficiently detailed data. We ultimately used the BeautifulSoup4 module which allowed looping through each textual element and its (optionally) associated tag. For the unpunctuated Original Text, we tokenized by splitting on spaces (which kept parentheses and brackets, indicated textual variants) with their words, but used NLTK's WordPunctTokenizer to separate off punctuation in the Hebrew Commentary. Postprocessing the output was required to merge separated characters of interrobang and revert curly single and double end-quotes to their original form. We also found that projecting quotation marks alongside other punctuation was too complex – edge cases led to spurious punctuation before and after the quotation mark. A better process was to identify and store positions of pairs of quotation marks, then project the punctuation, and finally to project a subset of the quotation mark pairs. The ranking of punctuation marks was subjective and was repeatedly adjusted after examining the output sentences and the particular employment of the punctuation in the Hebrew Commentary.[11] Finally, an unfortunate omission of our process is that colons in the Original Text, which indicated the end of a *sugya* or a citation from a Mishnah, have been stripped out.

---

**11** For example, a statement in a *brayta* is cited and ends with a period. In the context of the discourse, the purpose of the citation is to attack an Amora's position, so the gloss text contains a question mark. Which would we project? If we select the question mark, then consider that sometimes the *brayta* is cited for discussion, and the question in the gloss text is simply to anticipate an Amora's explanation. If so, a period is better. Some of these punctuation pairs should be more carefully considered and fine-tuned, rather than ranked, such as the case of an em-dash after a period.

# 3 Results and Evaluation

We have released punctuated Talmudic text at https://github.com/joshwaxman/Punctuated/. It contains punctuation for 37 tractates. Each line in the file corresponds to a translation unit (that is, paragraph) in the Sefaria database. We perused the text and the punctuation is quite helpful.

As discussed above, we compare to a Gold Standard, which is a manually edited partial Talmudic text. Table 2 presents a confusion matrix comparing all expected punctuation according to this Gold Standard compared with the punctuation marks generated by our algorithm, tractate Berakhot. For instance, the row for "?" corresponds to each time "?" appeared in the Gold Standard, and each column indicates how many times a given punctuation mark appeared in that position in our generated corpus – e.g., 794 times as the same question mark, and two times as an interrobang. The null column indicates how many times no punctuation appeared in that position.

Overall, these results are quite good, with both recall and precision often in the 90%–95% range. If we would relax our evaluation, and consider marks on the same punctuation level as equivalent, or even the mere helpful existence of punctuation as a match, the recall and precision would be even higher. An examination of the confusion matrices reveals that the alternatives are often reasonable. Quotation marks, projected via a separate process which does not lead to confusion with other punctuation, do not appear on the table, but have a recall of 95.6% and precision of 93.1%.

While we referred to this above as recall and precision, what we are really describing is degree of overlap. While some differences are due to the human editor, many others are due to differences in our respective processes. For instance, the Gold Standard will often have an em-dash following another punctuation mark, we would only project one mark, leading to 88% recall; we also project many more commas where the Gold Standard has none, leading to 88% precision. We might also resolve ambiguities among multiple punctuation marks in the gloss text differently. Additionally, we don't project any punctuation appearing in parentheses, but relaxing that restriction increases our recall. Some of our results are indeed erroneous. For instance, the commentary often omitted punctuation marks at the end of paragraphs, especially where brackets translated the final Aramaic phrase. These account for some nulls instead of sentence terminators in Table 2.

**Table 2:** Confusion matrix. Precision and recall for tractate Berakhot.

|  |  | Gold Standard punctuation | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
|  |  | , | ; | : | — | . | ? | ?! | ! | null | precision |
| Generated punctuation | , | 4884 | 3 | 51 | 27 | 101 | 7 | 0 | 4 | 442 | 88.5% |
|  | ; | 9 | 1 | 1 | 0 | 5 | 0 | 0 | 0 | 5 | 4.8% |
|  | : | 39 | 0 | 3692 | 12 | 39 | 0 | 0 | 0 | 91 | 95.3% |
|  | — | 20 | 0 | 8 | 1816 | 3 | 1 | 0 | 0 | 76 | 94.4% |
|  | . | 90 | 3 | 66 | 12 | 4817 | 8 | 0 | 2 | 290 | 91.1% |
|  | ? | 7 | 0 | 2 | 5 | 15 | 793 | 2 | 3 | 15 | 94.2% |
|  | ?! | 0 | 0 | 0 | 0 | 3 | 2 | 119 | 0 | 1 | 95.2% |
|  | ! | 3 | 0 | 2 | 0 | 29 | 1 | 0 | 341 | 14 | 87.4% |
|  | null | 202 | 1 | 219 | 193 | 64 | 15 | 1 | 9 | . |  |
|  | recall | 92.9% | 12.5% | 91.4% | 87.9% | 94.9% | 95.9% | 97.5% | 95.0% |  |  |

# 4  Future Work

In this work, we projected punctuation from a parallel interpolated translation and commentary in a related Semitic language. Therefore, there was significant overlap in actual incorporated words and the order in which they appeared. In future work, we intend to phrase-align and project punctuation from the parallel English Commentary.

Further, this work is part of a larger project – a digital edition of the Babylonian Talmud, described in part by Waxman.[12] The punctuated corpus is useful in its own right, but its availability enables various downstream processes. For instance, while much previous work was most successful on the (punctuated) English parallel aligned Talmudic text, we anticipate better named entity recognition and relation extraction on the punctuated Hebrew / Aramaic text. Other work enabled by the punctuation includes language identification and discourse classification (based on machine learning and on a crafted semantic grammar).

Looking at recent digital humanities work such as those presented in this volume, I see a great deal of promise in the new digital age of Jewish Studies. Within my own focus of academic Talmudic study, digital approaches can be transformative. The depth and type of research that the expert in humanities but relative novice in computers can quickly and readily perform has expanded. For instance, performing stylometry for authorship identification on an unattributed rabbinic commentary is relatively straightforward.

This is due to several factors. Modern high-level programming languages (notably Python 3) are designed with powerful abstractions such as lists, sets, dictionaries, and list comprehensions so that it is easy for a programmer to accomplish a lot with very little work. There are many well-supported libraries for these programming languages (such as, for Python: Natural Language Toolkit, Stanza, spaCy, scikit-learn) which have implemented necessary algorithms, so that a programmer can simply pass texts into them and extract linguistic analyses. Texts are widely and increasingly available in digital form, on Wikitext or on Sefaria, and OCR works well on other texts.

Even for non-programmers, collections and digital tools are made available on websites, such as Hachi Garsinan, Dicta, and Shebanq. I often need to look up Talmudic variants. In the past, I would see variants in Dikdukei Soferim, or look at individual manuscripts in a library. The Hachi Garsinan project, at the Friedberg Jewish Manuscript Society website, gives the full text of several printings

**12** Waxman, "A Graph Database of Scholastic Relationships in the Babylonian Talmud" (2019); Waxman, "A Graph Database of Scholastic Relationships in the Babylonian Talmud" (2021).

and manuscripts in columnar format, highlighting differences in red, and also makes available the digitized image of the page in the manuscript. (They allow download of a spreadsheet with the variants for a page.) This allows me to see relevant changes across several manuscripts, and double-check that I agree with the reading, something that would otherwise have taken much time and effort.

An extremely useful task for many Talmud scholars is identifying commonality between texts. When studying one Babylonian Talmudic passage, it is helpful to know of parallels elsewhere in the Talmud, which can provide a fuller picture of how the ideas are discussed or allow tracing the development of a *sugya* – e.g., that there was *ha'avara*, transfer of the passage. It is helpful to find earlier sources and later commentaries quoting the passage. This cannot be performed by simple string matching, because of paraphrase or other textual variation. Dicta provides a tool to perform such analysis, and results have been integrated into other projects, such as Sefaria.

This digital age of Jewish Studies is increasingly open and collaborative, with the texts and tools open source or with public APIs, so that projects can build upon one another. Alongside advances in big data and deep learning, this collaborative field allows for the advancement of even more sophisticated work, by both novices and experts, which differ in number and kind from previous efforts. I look forward to seeing how this develops further.

# References

Shmidman, Avi, Shaltiel Shmidman, Moshe Koppel, and Yoav Goldberg. "Nakdan: Professional Hebrew Diacritizer." In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*. 2020.

Tsarfaty, Reut, Amit Seker, Shoval Sadde, and Stav Klein. "What's Wrong with Hebrew NLP? And How to Make It Right." In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*. EMNLP, Hong Kong, 2019.

Waxman, Joshua. "A Graph Database of Scholastic Relationships in the Babylonian Talmud." In *Proceedings of the Digital Humanities Conference 2019*. Utrecht, 2019.

Waxman, Joshua. "A Graph Database of Scholastic Relationships in the Babylonian Talmud." *Digital Scholarship in the Humanities* 36, no. 2 (2021): 277–89. Accessed March 1, 2022. https://doi.org/10.1093/llc/fqab015.

# List of Contributors

**Julia Afanasieva** is Head Bibliographer at the Research Library, Siberian Federal University. Her research interests are scientometrics and statistics in bibliographic data analysis.

**Tatsiana Astrouskaya** holds a PhD in East European History from the University of Greifswald (2018); she studied social theory and political philosophy (M.A.) and philosophy (B.A.) in Vilnius and Minsk. She is the author of "Cultural Dissent in Soviet Belarus (1968–1988). Intelligentsia, Samizdat and Non-Conformist Discources" (Harrassowitz, 2019). Since 2018, she works as a post-doctoral researcher in the field of digital history at the Herder Institute for Historical Research on East Central Europe in Marburg, Germany. In her current project, she looks into the motives and setting of the Jewish Emigration from the Belarusian Soviet Socialist Republic from the crossing of urban and digital history as well as migration studies.

**Luigi Bambaci** is a PhD student in Jewish Studies at the University of Bologna. He graduated in 2019 at the University of Pisa in Biblical Philology, with a thesis on the Hebrew text of Qohelet 1–3. He is member of the Laboratory of Cooperative Philology (CoPhiLab) of the Institute of Computational Linguistics "Antonio Zampolli" of the Italian National Council of Research in Pisa. His doctoral dissertation concerns the preparation of a digital scholarly edition of the Hebrew book of Qohelet. His main research interests are in textual history and textual criticism of the Hebrew Bible, in digital and computational philology. His recent research activities concern the application of formal and quantitative methods to the textual tradition of the Hebrew Bible, and in particular domain-specific language design, and computer-assisted stemmatology.

**Berat Kurar Barakat** is a PhD research student of computer science in Visual Media Lab (VML) at Ben-Gurion University of the Negev. Her latest research is on historical document image analysis using digital image processing and machine learning methods.

**Anna Bonazzi** is a PhD candidate in Germanic Languages in the department of European Languages and Transcultural Studies (ELTS), University of California, Los Angeles (UCLA). Anna works on testimony, memory, and multilingualism in the areas of Holocaust and migration testimonies from the 1940's to the present. Her research is based on large multi-media archives of oral testimonies and interviews, and she works primarily with Digital Humanities methods (corpus linguistics, natural language processing, text analysis, data visualization). Anna's working languages are Italian, English, German, Yiddish, and French.

**Michelle Margolis** is the Norman E. Alexander Librarian for Jewish Studies at Columbia University and the current President of the Association for Jewish Libraries. She co-directs a digital project called *Footprints: Jewish Books Through Time and Place*, which traces the movement of early printed Jewish books throughout the world. Michelle also edits a column in *Judaica Librarianship* on digital humanities and Jewish Studies.

**Sinja Clavadetscher** M.A., studied history and educational science at the Universities of Bern and Basel and is currently working on her PhD as part of Stefanie Mahrer's research group.

**Ahmad Droby** is a Ph.D. candidate at Ben Grunion university studying the use of weakly supervised learning to solve various visual computing problems. Previously, he received his Masters for his work on surface detection and deformation Detection in a video stream.

**Jihad El-Sana** is a professor at the Department of Computer Science, Ben Gurion University of the Negev and headed the department from 2015–2017. He is an expert in image/video processing, computer vision, and augmented reality with over 20 years of experience in these fields. He has published over 150 papers and five patents.

**Abby Gondek** served as the Morgenthau Scholar-in-Residence at the FDR Presidential Library and Museum from 2019–2021. She employed digital humanities resources (blogs, digital exhibits, databases, data visualizations) to raise awareness about Holocaust-related archival collections. Dr. Gondek tracked and shared findings about the influential roles of women in the Treasury Department and War Refugee Board on refugee rescue policy and letter campaigns from Jewish women's organizations in support of a refugee camp in Oswego, NY. She received her PhD in Global and Socio-cultural studies in 2018 and an MA in African and African Diaspora Studies from Florida International University in 2015. She also holds MAs in Women's Studies and Special Education. Her previous research used historical social network analysis to analyze the political, institutional, and personal networks between Jewish women social scientists and women scholars of color in the early – mid 20th century. Her use of nodegoat originated in this earlier research.

**Karin Hofmeisterová** is a PhD candidate, teaching assistant and coordinator of the Malach Centre for Visual History at Charles University. Her research focuses primarily on religious institutions and their engagement in politics of memory in Southeast Europe.

**Maja Hultman** is a postdoctoral researcher at Centre for European Research and Department of Historical Studies at University of Gothenburg, and a research fellow at Centre for Business History in Stockholm. She received her doctorate in History from University of Southampton in 2019 and is currently working on her first monograph where she explores the intersection of digital humanities, urban studies, and the history of emotions through the case study of Stockholm's modern Jewry. Using spatial concepts and interdisciplinary methodologies, her current research projects focus on Jewish cultural-economic networks in both urban and transnational settings.

**Benjamin Lee** is a fourth year Ph.D. candidate in the Paul G. Allen School for Computer Science & Engineering at the University of Washington. He was also a 2020 Innovator in Residence at the Library of Congress, where he created *Newspaper Navigator*. Ben served as the 2020–2021 Richard and Ina Willner Memorial Fellow in the Stroum Center for Jewish Studies at the University of Washington. Previously, he was the inaugural Digital Humanities Associate Fellow at the United States Holocaust Memorial Museum and a Visiting Fellow in Harvard's History Department. Ben is currently a National Science Foundation Graduate Research Fellow in Machine Learning.

**Amalia S. Levi** is an archivist, and the founding director of *The HeritEdge Connection*, a non-profit organization based in Barbados dedicated to forging cultural heritage collaborations, connecting people, resources, and institutions. She has completed numerous digitization and archival processing projects. Currently she is conducting doctoral research at the Bonn Center for Dependency and Slavery Studies. Amalia holds an MLS with a concentration in Archives, and an M.A. in History with a concentration in Jewish Studies, both from the University of Maryland; an M.A. in Museum Studies from Yildiz University in Istanbul, Turkey; and a B.A. in Archaeology and History of Art from the University of Athens, Greece. She is interested in augmenting and facilitating historical scholarship on diasporas and underrepresented populations by linking and enriching dispersed collections through digital tools and Linked Data.

**Inna Kizhner** is an Assistant Professor at the Department of Information Technology in Creative and Cultural Industries, Siberian Federal University. Her research interests are digitization of cultural heritage and data-let approaches to understanding bias in cultural heritage collections. She was an editor for the Russian translation of Defining Digital Humanities (Siberian Federal University Press, 2017).

**Jiří Kocián** received his PhD in Area Studies from the Charles University, Prague, where he currently holds the positions of research associate at the Institute of International Studies and Coordinator of the Malach Centre for Visual History. In his research, he focuses on comparative history and politics, with special emphasis on Jewish history, national minorities, identity building, transitions to democracy and populism in Central and Southeast Europe. Simultaneously, he strives for application of digital humanities methods in his research and promotes interdisciplinary work of this kind by all means possible.

**Prof. Dr. Stefanie Mahrer** is an SNSF-PRIMA assistant professor of Modern General and Jewish History at the Universities of Bern and Basel, Switzerland, she is the PI and group leader of the SNSF-funded project "Forced Academic Migration during the NS-Period and its Aftermath".

**Piergabriele Mancuso** received his doctoral degree in Jewish Studies from University College London in 2009. He is the director of the Eugene Grant Jewish History Program at the Medici Archive Project, Florence, and coordinator of the Ghetto Mapping Project. He is Senior Lecturer in History at Boston University (Padua), and former research fellow in Jewish studies at Insubria University, Como. His research fields include Jewish music and ethnomusicology and history of Italian Jews.

**Jakub Mlynář** is a sociologist working at Charles University (Czech Republic) and at the University of Applied Sciences of Western Switzerland (HES-SO Valais/Wallis, Switzerland). For his PhD thesis, he conducted research on Holocaust oral history interviews addressing issues of narrative, identity, and memory. At the Malach Center for Visual History at Charles University, he has worked on how to introduce these interviews in educational contexts. In his postdoctoral research at the Department of Social Sciences of the University of Fribourg, he has studied classroom interactions involving digital technologies, in Swiss and Czech schools, in the framework of a one-year project funded by the Swiss Government Excellence Scholarships. He is also interested in technology-mediated interaction and social dimensions of artificial intelligence.

**Diana Pusenkova** is a bachelor's student at the School of Space and Information Technologies, Siberian Federal University.

**Irina Rabaev** received her Ph.D. in Computer Science from Ben-Gurion University of the Negev, Israel. She is currently a faculty member at Department of Software Engineering of Shamoon Academic College of Engineering, Beer-Sheva, Israel. Her main research interests include the areas of computer vision and image processing with a focus on historical documents analysis.

**Miriam Ruerup** is Director of the Moses Mendelssohn Center for European-Jewish Studies and Professor at the University of Potsdam. Before that she was director of the Institute for the History of the German Jews in Hamburg and a postdoctoral fellow at the German Historical Institute in Washington, DC. She is a historian with a focus on German-Jewish History of the 19th and 20th Century, with a strong interest in Jewish heritage studies and memorial culture. She has been involved in the field of Jewish Digital history and was the founding editor of the online source edition "Key Documents on German-Jewish History." She also works as Review Editor for Jewish History for the Online-Forum and Mailing List H-Soz-u-Kult and is the head of the Academic Working Group of the Leo Baeck Institute in Germany.

**Stefanie Salvisberg** M.A., is a PhD student in history at the University of Bern in the SNF-Prima project "Science transnational. Switzerland and the academic forced migrants 1933 to 1950.

**Zef Segal** is a historian and digital humanist at the Open University of Israel. His research interests are mobility, space, and communication in the nineteenth century as well as the application of digital methodologies in historical research. His latest books are *The Political Fragmentation of Germany*, published in 2019 by Palgrave Macmillan, and *Motion in Maps, Maps in Motion*, published in 2020 by Amsterdam University Press. His forthcoming books are *Culture in the Age of the Enlightenment: Reshaping the Private Sphere*, and *Who's Afraid of Numbers: Math and Computation for the Humanities*, both soon to be published at the Open University of Israel.

**Jeffrey Shandler** is Distinguished Professor of Jewish Studies at Rutgers University. He is the author of *While America Watches: Televising the Holocaust* (Oxford University Press, 1999); *Jews, God, and Videotape: Religion and Media in America* (New York University Press, 2009); *Shtetl: A Vernacular Intellectual History* (Rutgers University Press, 2014); and *Holocaust Memory in the Digital Age: Survivors' Stories and New Media Practices* (Stanford University Press, 2017), among other titles. Shandler is also the editor of *Awakening Lives: Autobiographies of Jewish Youth in Poland before the Holocaust* (Yale University Press, 2002) and co-editor, with Barbara Kirshenblatt-Gimblett of *Anne Frank Unbound: Media, Imagination, Memory* (Indiana University Press, 2012). His translations of Yiddish literature include *Emil and Karl*, a Holocaust novel for young readers by Yankev Glatshteyn (Roaring Brook, 2006).

**Maria Sherer** is a bachelor's student at the School for the Humanities, Siberian Federal University.

**Dr. Avi Shmidman** is a senior lecturer in the Department of Hebrew Literature at Bar-Ilan University. He is an associate member of the Academy of the Hebrew Language, and a senior researcher at DICTA: The Israel Center for Text Analysis. His research focuses upon Hebrew poetry, Jewish liturgy, S.Y. Agnon and digital humanities.

**Daniil Skorinkin** is an Associate Professor at the Humanities Faculty of the Higher School of Economics (HSE) in Moscow. He has a Ph.D. degree in linguistics and literary studies. His thesis focused on the application of digital methods to the study of Russian literature. He teaches Digital Humanities and Python Programming at HSE and does research at the HSE Centre for Digital Humanities. His research interests include stylometry, network analysis, and the use of NLP models and tools for research in the humanities.

**Daniel Stein Kokin** is an adjunct faculty member in the School of Historical, Philosophical and Religious Studies at Arizona State University, and a lecturer in the Judaic Studies program at the University at Albany SUNY. He has previously taught Jewish and Israel Studies at Yale, UCLA, the University of Oregon, and the University of Greifswald (Germany). Stein Kokin's publications have ranged across Renaissance humanism, Christian Kabbalah, Jewish visual culture, and Israeli film and music, and he is the founder and director of the "All the Points" project.

**Melissa Terras** is the Professor of Digital Cultural Heritage at the University of Edinburgh's College of Arts, Humanities, and Social Sciences, Director of the Edinburgh Centre for Data, Culture and Society, and Director of Research in the Edinburgh Futures Institute. Her latest book is Picture-Book Professors: Academia and Children's Literature (CUP, 2018). She was an editor for the Russian translation of Defining Digital Humanities (Siberian Federal University Press, 2017).

**Daria Vasyutinsky Shapira** is a PhD in Jewish Thought, her research interests include digital Hebrew paleography, digital epigraphy, codicology, Modern Jewish intellectual history, Karaitica. She is currently a postdoctoral researcher at the CS Visual Media Lab at Ben-Gurion University of the Negev, Israel.

**Joshua Waxman** is an Assistant Professor of Computer Science at Stern College for Women, Yeshiva University. He has rabbinic ordination from RIETS and a PhD in Computer Science from CUNY Graduate Center. His research interests include natural language processing (especially of low-density languages), machine learning, social network graphs, Talmudic prosopography, and digital humanities as they intersect with academic Judaic studies. One current focus is the creation of a digital edition of the Babylonian Talmud, aspects of which include discourse classification, named entity recognition, relation extraction, and language identification.

**Gerben Zaagsma** is Assistant Professor at the Centre for Contemporary and Digital History (C2 DH) at the University of Luxembourg. His main research and teaching interests are modern Jewish history, digital history, and music history. He currently works on two related projects that investigate the politics of digitisation and the history of digital history, in addition to ongoing work in modern Jewish history. He holds a Ph.D. in modern history from the European University

Institute in Florence and was a research fellow at the Department of Hebrew and Jewish Studies at University College London and the Lichtenberg Kolleg, The Göttingen Institute of Advanced Study at University of Göttingen. As a web developer and editor, he was also involved in the creation of several online resources, including the recently launched portal #DHJewish – Jewish Studies and Digital Humanities (dhjewish.org). His first book Jewish volunteers, the International Brigades and the Spanish Civil War was published with Bloomsbury Academic in 2017. For more information, see: gerbenzaagsma.org.