

STEPHEN DÖRR , CLEMENS LIEDTKE  & HANNA LISS 

## Transkription, Edition und digitale Präsentation in den Projekten *Corpus Masoreticum* und *Bibelglossare*

**Keywords** Middle Ages; digital edition; Jewish Studies/Romance Studies; medieval biblical text research and interpretive literature; historical Lexicography

### **PROJEKT 1: *Corpus Masoreticum. Die Inkulturation der Masora in die jüdische Gelehrsamkeit Westeuropas im 11.–13. Jahrhundert. Digitale Erschließung einer vergessenen Wissenskultur***

#### **Projektbeteiligte**

Clemens Liedtke, Hanna Liss, Sebastian Seemann, Ilona Steimann, Melissa Anwar Uthman

#### **Institutionelle Anbindung**

Hochschule für Jüdische Studien, Heidelberg

#### **Förderung**

Deutsche Forschungsgemeinschaft (DFG), Langfristvorhaben

#### **Laufzeit**

2018–2030

### **PROJEKT 2: *Bibelglossare als verborgene Kulturträger. Judäo-französischer Kulturaustausch im Hochmittelalter***

#### **Projektbeteiligte**

Sabine Arndt, Stephen Dörr, Hanna Liss, Katelyn Mesler, Fabian Strobel, Isabelle Schmiederer

#### **Institutionelle Anbindung**

Forschungsstelle an der Heidelberger Akademie der Wissenschaften

#### **Förderung**

Akademienprogramm

#### **Laufzeit**

2023–2040

#### **Kurzbeschreibung**

Die in den verschiedenen Philologien entwickelten Editionswissenschaften, die mit Karl Lachmann (1793–1851) eine erste grundlegende Systematisierung erhielten (SAHLE 2013, 23–37; KRISTELLER 1981), basierten auf einer implizit gesetzten Prämisse, die vor allem die Literaturen des Westens durchgehend auszeichnete: Es handelte sich oftmals um Autorenliteratur, deren ursprüngliche Fassungen („Urtexte“) man zu rekonstruieren suchte. Bei den hebräisch-aramäischen Literaturen liegen

allerdings häufig Traditions- bzw. Kompilationsliteraturen vor. Dadurch können sich unterschiedliche und jeweils distinkte Makroformen von ein und demselben ‚Text‘ ergeben. Dieses Spezifikum antiker und mittelalterlicher hebräisch-aramäischer Texte gilt es bei einer Edition zu berücksichtigen. An der Hochschule für Jüdische Studien Heidelberg laufen derzeit zwei Langzeitprojekte. In beiden Fällen geht es um Textkulturen, die aus dem westeuropäischen Mittelalter (Ashkenas) zwischen dem 12. und 14. Jahrhundert stammen.

## 1 Das Projekt *Corpus Masoreticum*

Das Projekt *Corpus Masoreticum* läuft derzeit in der dritten von vier Phasen. Ziel ist die erstmalige philologische Auseinandersetzung mit der westeuropäischen masoretischen Tradition zwischen dem 11. und 14. Jahrhundert. Der Begriff ‚Masora‘ bezieht sich auf paratextliche Elemente, die den Konsonantentext der orientalischen Bibelcodices des 9. Jahrhunderts ergänzen und grammatikalische, syntaktische und andere Phänomene adressieren. Oftmals sind diese Beobachtungen in langen Listen zusammengefasst. Hier ist der Text linear gestaltet. Seit dem 12. Jahrhundert erscheinen in Frankreich und Deutschland Bibeln, in denen dieses masoretische Material in ornamentaler Mikrographie dargestellt wird (Abb. 1), oftmals im Stil romanischer Bibel-Illustrationen (Chimären, Drachen, Ritter) oder in geometrischen Formen (LISS et al. 2021).

Masoretische Paratexte, die transkribiert und für eine Edition aufbereitet werden, sind also

- Texte, die von sog. Masoreten in Bibelmanuskripte hineingeschrieben, abgeschrieben, manchmal aber auch neu verfasst wurden (sog. *masora parva*, *masora magna*, *masora finalis*);
- Texte, die auf unterschiedlich umfangreiche Listencorpora zurückgehen und sich sowohl als isolierte Listensammlungen als auch integriert in die Bibelmanuskripte finden, wie beispielsweise die Sammlungen *Okhla we-Okhla* oder *Sefer ha-Chilufim* (OFER 2019, 49–60; DOTAN 2007).

Wir verstehen dabei Paratexte im Genette’schen Sinne: „Man weiß nicht immer, ob sie zum Text gehören oder nicht, aber auf jeden Fall umgeben sie ihn und verlängern ihn, eben um ihn zu präsentieren, im üblichen Sinne dieses Verbs, aber auch in seiner stärksten Bedeutung: um ihn präsent zu machen, um seine Präsenz in der Welt, seine ‚Rezeption‘ und seinen Konsum, zumindest heute in Form eines Buches, zu gewährleisten“ (übersetzt aus: GENETTE/MACLEAN 1991, 261).

Das Projekt *Corpus Masoreticum* steht editionstheoretisch daher vor folgenden Herausforderungen:



Abb. 1: Ausschnitt von masoretischen Listen in illustrativer Mikrographie (MS London BL Or 2091, fol. 203r).

- Was bedeutet ‚diplomatische Edition‘?
- Was kann/soll wie ediert werden?
- Soll alles ediert werden, was man auf einer Manuskriptseite sieht?
- In welcher Weise werden Artefakte/Manuskripte in den Transkriptions- und Editionsprozess einbezogen, d. h. sollen nur ‚Texte‘ transkribiert werden oder soll die Edition auch Textebenen, *mises-en-page* und *mises-en-texte*, präsentieren (Objekt- vs. Textedition)?
- Soll eine digitale Edition masoretischer Paratexte Perspektivwechsel ermöglichen, d. h. Textpräsentation aus der Perspektive des Schreibers/Masoreten, des (ersten) Lesers/Auftraggebers, der heutigen Leser:innen?

Die bisherige Arbeit hat gezeigt, dass vor allem die Präsentation transkribierter Texte, die sich in mehreren Bibelmanuskripten befinden, z. B. eine distinkte masoretische Liste, in der Edition die verschiedenen Formgebungen oder *mises-en-page* und *mises-en-texte* berücksichtigen muss, weil diese auf eine erweiterte inhaltliche Aussage schließen lassen. So verweist der als *masora figurata* gestaltete Ritter in MS London BL Or 2091, 203r (Abb. 1) darauf, dass der Masoret mit dieser Liste bei seinen Lesern einen ‚Gegenkreuzzug‘ andeuten wollte (LISS 2020; LISS 2021, 131–175); die Präsentation der Text-Bild-Relation ist hier unerlässlich (LISS 2021). Das MS Wien ÖNB Cod.hebr.16 (vgl. Abb. 4) zeigt eine Klage des Masoreten Abersush (fol. 248v-268r),

der die Ermordung seiner Familie während der Rintfleisch-Pogrome 1291 beklagt (vgl. <https://buchkultur.ausstellung.corpusmasoreticum.de>; Stand: 30.09.2025). Abersush schrieb seinen Text aus masoretischem Material (<https://bimaz.corpusmasoreticum.de>; Stand: 30.09.2025). Die in diesem Text verarbeitete Masora ist stets zum Haupttext relationiert. Gleichzeitig zeigt hier der digitale Corpusvergleich (siehe 3.), dass das im Text verarbeitete Listenmaterial auch entsprechend der Wortlängen der Klage angeordnet wurde. Der philologische Wert für die heutige masoretische Listenforschung ist grundlegend, stand aber natürlich nicht im Fokus des traurigen Masoreten.

## **2 *Bibelglossare als verborgene Kulturträger*** **(Heidelberger Akademie der Wissenschaften)**

Im Zentrum des interdisziplinären Projektes (Jüdische Studien/Romanistik) stehen sechs hebräisch-französische (Teil-)Glossare (Basel, UB, A III 39; Leipzig, UB, Vollers 1099; Paris, BNF, hébr. 301; Paris, BNF, hébr. 302; Parma, Palatina, Parm. 2780; Parma, Palatina, Parm. 2924), die (zum größten Teil erstmals) ediert und historisch-philologisch bearbeitet und kontextualisiert werden. Die französischen Glossen sind durchgehend in hebräischer Graphie notiert und umfassen etwa ein Viertel des heute bekannten altfranzösischen Wortschatzes (LISS/DÖRR 2022; BANITT 1995–2005). Die Glossare enthalten eine Vielzahl unterschiedlicher Textgenres, die in ihrem Zueinander, aber auch in ihrer eigenen Überlieferungsgeschichte editionstechnisch transparent werden sollen. Alle Glossare umfassen (zumeist textchronologisch gelistet) bibelhebräische Termini: Verben (konjugiert), Substantive (dekliniert), Adjektive und Adverbien sowie eine altfranzösische Erklärung in hebräischer Notation.

Je nach Glossar finden sich noch Textentitäten wie eine biblische Parallelstelle, in der das Wort vorkommt, oder eine oder mehrere Zitation(en) aus Bibelkommentaren und/oder zeitgenössischen Wurzelwörterbüchern sowie Worterklärungen/Synonyme (Abb. 2).

Auch hier muss die (digitale) Edition praxeologische Ergebnisse deutlich werden lassen, denn die unterschiedlichen *mises-en-pages* und *mises-en-textes* lassen eine mögliche Rekonstruktion und plausible Szenarien der Verwendungen der Artefakte zu: Ein in Spalten gestaltetes Glossar verweist auf eine andere Verwendung als ein einspaltiges (Nachschlagefunktion gegenüber kontinuierlicher Lektüre). Auch für die Bibelglossare muss die Edition unterschiedliche Forschungsperspektiven adressieren: Für die moderne Forschung kann die Entstehungsgeschichte von Texten sowie ihre Rezeption zurückverfolgt werden. Für die Jüdischen Studien geht es dabei um die Rekonstruktion des jüdischen Erziehungs- und Bildungswesen zwischen dem 12. und 15. Jahrhundert. Die Romanistik erwartet die editorische Erschließung der Glossen nach den in der Forschung des Altfranzösischen etablierten und von der *École des Chartes* definierten Kriterien sowie die Kontextualisierung aller Lexeme (ein Glosseneintrag weist häufig mehr als ein Lexem auf) mithilfe

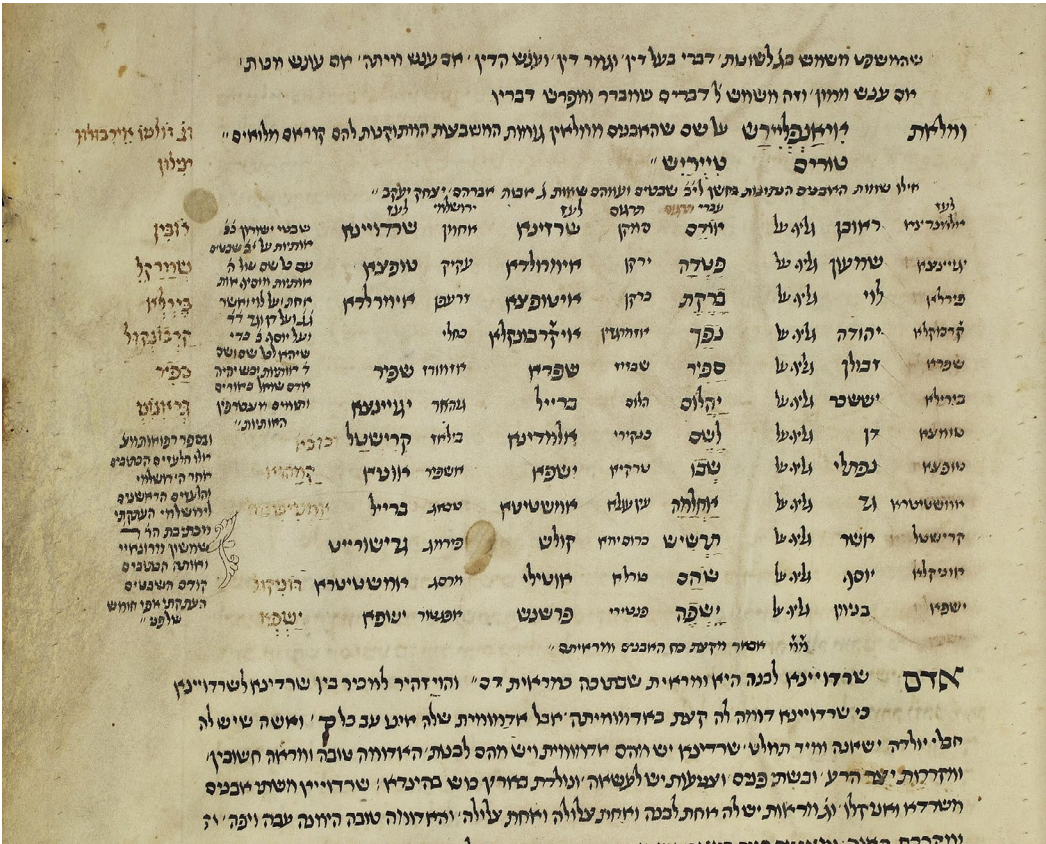


Abb. 2: MS Leipzig, UB, Vollers 1099, fol. 17r.

der wissenschaftlichen Lexikographie. Es ergeben sich diachronische, diatopische, diastratische und diaphasische Ergebnisse, sowohl für die Geschichte des französischen Wortschatzes als auch für die jüdische Interpretation der Glossen.

### 3 Die Digitale Infrastruktur BIMA 2

#### 3.1 Anforderungen an eine digitale Infrastruktur

Editionstheoretisch bedeutet die projektübergreifende Ausrichtung von Forschungsfragen an genreübergreifende Quellenlagen, dass eine Edition mehr leisten muss als nur eine digitale Lesbarmachung und Annotation eines Dokuments in Gestalt eines (weiteren) Dokuments. Einzelne transkribierte Elemente müssen sich im Sinne multipler Kontexte neu anordnen, arrangieren und verschränken lassen, so dass eine solche Edition nicht nur den Befundzustand dessen dokumentiert, was die Manuskriptseite präsentiert, sondern auch Neuarrangements von Editionselementen, die

potentiell eine Vielzahl von weiteren interpretativen Ebenen nach sich ziehen. Benötigt wird eine digitale Editions Umgebung, die Folgendes leisten muss:

- die Möglichkeit der Bereitstellung und Metadatifizierung von Manuskripten über entfernte Datenquellen hinweg mit Hilfe eines gemeinsamen Protokolls (IIIF),
- die redaktionelle Erschließung, Bearbeitung und Langzeitarchivierung des transkribierten Materials,
- die Präsentation der digitalen Transkriptionen und ihren (TEI/XML)-Export,
- die Möglichkeit eines sofortigen textkritischen Zugriffs während des Transkriptionsprozesses,
- ein automatisches Transliterationstool (für die altfranzösischen und mittelhochdeutschen Glossen),
- die Möglichkeit, Schnittstellen zu schaffen, z. B. für die Integration jüdisch-französischer Glossar-Ressourcen wie DEAFél (<https://deaf.hadw-bw.de/>) in das BIMA-2-Graphdatenbankmodell.

### 3.2 Erste Lösungsversuche

Die von Clemens Liedtke entwickelte Dateninfrastruktur BIMA 2 stellt nicht einfach ‚Texte‘ vor oder nimmt von einem Autor verfasste Texte in getippter Form auf; vielmehr werden sowohl im Projekt *Corpus Masoreticum* als auch im Projekt *Bibelglossare* ‚Texte am oder im Text‘ transkribiert und für unterschiedliche Editions- und Präsentationsformen aufbereitet. Der gesamte Systemaufbau umfasst derzeit 37 verschiedene sogenannte Servicecontainer innerhalb einer hoch skalierbaren Cloud-Infrastruktur, darunter Webservices, Datenbanken, Speicher, Anwendungen und Hochleistungs-Caches (Abb. 3). Einige zusätzliche Ressourcen wurden einer Instanz von eScriptorium zugewiesen, um diesen und anderen Projekten maschinelle Texterkennungsdienste (Handwritten Text Recognition, HTR) zur Verfügung zu stellen.

BIMA 2 beruht im Wesentlichen auf drei Säulen:

- 1) Die erste Säule ist die konsequente Verwendung von IIIF, dem *International Image Interoperability Framework*, einem Protokoll, das beschreibt, wie Bibliotheken ihre digitalen Bestände unabhängig von ihrem physischen Ort mit allen Erschließungsdaten so zur Verfügung stellen, dass sie weiter annotiert und als eigenständige Ressource nachgenutzt werden können.
- 2) Die zweite Säule ist die Verwendung einer Graphdatenbank anstelle einer traditionellen relationalen Datenbank (wir verwenden neo4j; <https://neo4j.com/>). Graphdatenbanken behandeln ihre Daten als möglichst kleinteilige Informationseinheiten, auch Knoten genannt, die mit anderen Knoten durch ‚Kanten‘ (edges) verknüpft sind. So genannte ‚labelled

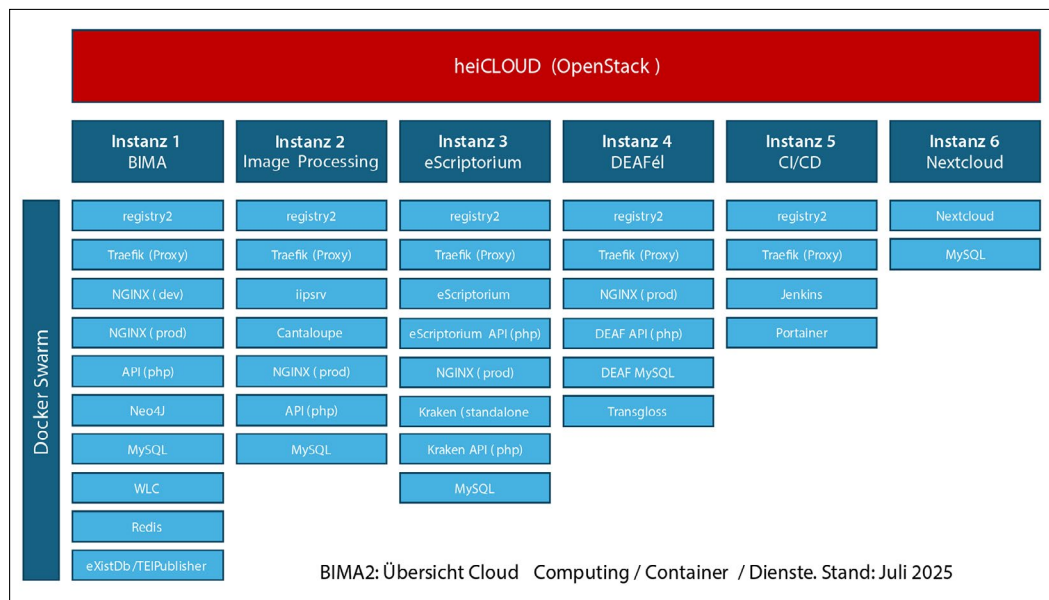


Abb. 3: Corpus Masoreticum Cloud Computing.

property graphs‘ machen es möglich, dass die Art der Verknüpfung mit eigenen Eigenschaften beschrieben werden kann. Anstelle von statischen Tabellen erhalten wir dynamische Netzwerke und Cluster von Informationen zu Manuskripten, Schreibern, Lesarten, Materialumfang („text-as-a-graph“).

- Die dritte Säule ist die Transkription mit Hilfe von SVG-Textpfaden; SVG („Scalable Vector Graphics“) ist ein Web-Standard zur Beschreibung von Vektor-Primitiven, also Linien, Kreisen, Polygonen ... Beliebiger Text kann hier entlang von gezeichneten Textpfaden aufgetragen werden; wir benutzen das, um maschinenlesbare Transkriptionen direkt auf dem Handschriftendigitalisat aufzubringen (LIEDTKE 2021, 119–121). Abbildung 4 zeigt den IIIF-basierten Textpfadeditor und rechts den Kontextstapel, von dem aus die Transkriptionen zur Weiterverarbeitung für Editionen und weiterführende Analysen verfügbar sind. Außerdem nutzen wir die Ähnlichkeiten zwischen der SVG-Textpfad-Technologie von BIMA 2 und der HTR-Software eScriptorium (Kooperation Stoekl ben Ezra, EPHE Paris; <https://gitlab.com/scripta/escriptorium>). Es ist bereits jetzt möglich, HTR-generierte Transkriptionen direkt aus BIMA 2 nach eScriptorium zu importieren sowie umgekehrt eScriptorium mit gebrauchsfertigen ‚Ground-Truth-Daten‘ zu füttern, um bessere HTR-Modelle zu trainieren.

BIMA 2 unterstützt so den editorischen Workflow, um die verschiedenen ‚Layer‘, d. h. die Art der Neuverknüpfung einzelner transkribierter Elemente, ihre Struktur und Interpretation selbst zu Elementen einer ‚offenen Edition‘ zu rekonfigurieren.

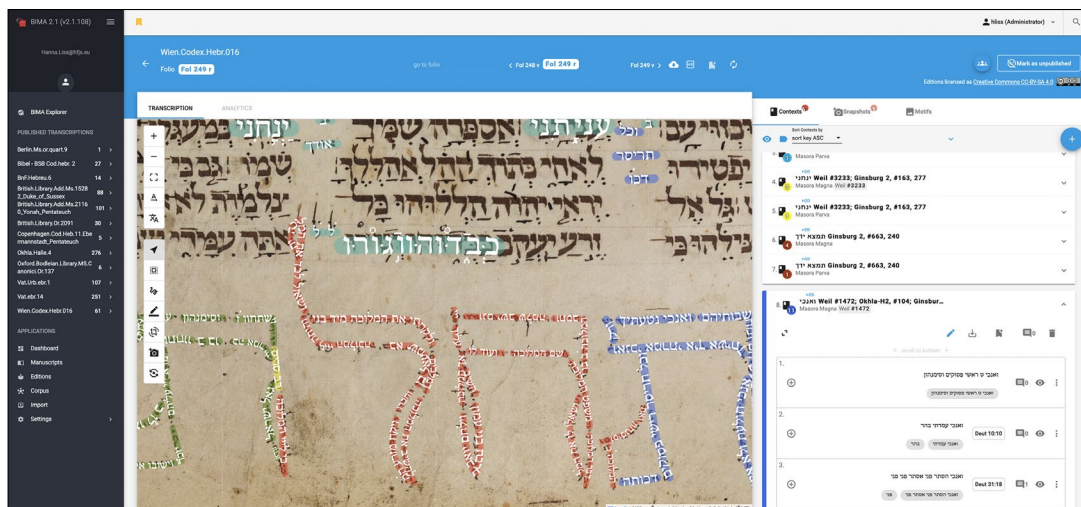



Abb. 4: Oberfläche BIMA 2 *Corpus Masoreticum*.

Dies bedeutet gleichzeitig, dass eine reine Metadatifizierung und ein enzyklopädisches ‚open data linking‘ (zu Wörterbüchern, Orten, Personendaten, anderen Texten aus demselben geokulturellen Raum u. v. m.) kein Selbstzweck ist, sondern ein Baustein unter vielen eines solchen polykontexturalen Editionsprozesses, der sich stets am Ziel der Erschließung der einzelnen zu edierenden Literaturen messen und erproben lassen muss.

Insofern sich also die oben skizzierte Edition von Quellentexten jüdischer Textkulturen des Mittelalters nicht als statisches Produkt eines kommentierten und ggf. rekonstruierten Lesetextes versteht, sondern als mehrdimensionale Abbildung verschiedener literarischer Makro- und Mikroformen, müssen auch jenseits etablierter digitaler Editionsstandards experimentell neue Wege gegangen werden. BIMA 2 als projektübergreifende lokale Forschungsdateninfrastruktur (Cloud-Computing + Datenmodellierung + Anwendungsentwicklung + Schnittstellen + Langzeitarchivierung + Publikation im OpenAccess) versteht sich daher als ‚hybrides‘ Editionsframework, innerhalb dessen sowohl innovative, aber noch nicht als ‚best practice‘ eingestufte Technologien (Graphendaten, SVG-Textpfad-Transkriptionen) als auch bewährte DH-Standards (TEI-XML, RDE, IIIF) sinnvoll miteinander kombiniert und zum Einsatz gebracht werden (dazu LIEDTKE 2020, 107–111; LIEDTKE 2021, 121–124).

## ORCID®

Stephen Dörr  <https://orcid.org/0009-0007-2325-8548>

Clemens Liedtke  <https://orcid.org/0000-0002-2181-0606>

Hanna Liss  <https://orcid.org/0000-0003-0137-3995>

## Literaturverzeichnis

- Banitt, Menahem (Hg.) (1995–2005), *Le Glossaire de Leipzig*, 4 Bde., Jerusalem.
- Dotan, Aron (2007), Art. „Masorah“, in: Michael Berenbaum u. Fred Skolnik (Hgg.), *Encyclopedia Judaica*, 2. Aufl., Bd. 13, Detroit et al, 603–656.
- Genette, Gérard/Maclean, Marie (1991), „Introduction to the Paratext“, in: *New Literary History* 22 (2), 261–272.
- Kristeller, Paul Oscar (1981), „The Lachmann Method. Merits and Limitations“, in: *TEXT – Transactions of the Society for Textual Scholarship* 1, 11–20.
- Liedtke, Clemens (2020), „DH's Next Top- Model? Digitale Editionsentwicklung zwischen Best Practice und Innovation am Beispiel des ‚Corpus Masoreticum‘“, in: *Conference Abstract Dhd2020: Spielräume. Digital Humanities zwischen Modellierung und Interpretation* (Konferenzabstracts Universität Paderborn, 02. bis 06. März 2020), 107–11, [https://zenodo.org/record/3666690/files/2020\\_DHd\\_BookOfAbstracts-web.pdf](https://zenodo.org/record/3666690/files/2020_DHd_BookOfAbstracts-web.pdf) (Stand: 07.2024).
- Liedtke, Clemens (2021), „How Am I Supposed to Read This? Challenges and Opportunities of Medieval Western Masorah as a Digital Scholarly Edition“, in: Hanna Liss u. Jonas Leipziger (Hgg.), *Philology and Aesthetics: Figurative Masorah in Western European Manuscripts*, Frankfurt am Main, 103–129.
- Liss, Hanna (2021a), „Negation oder Transformation? Illustrative Auflösung masoretischer Listen in einem mittelalterlichen hebräischen Manuskript“, in: Şirin Dadaş u. Christian Vogelin (Hgg.), *(Nicht)Wissen – Dynamiken der Negation in vormodernen Kulturen*, Wiesbaden, 313–330, <https://t1p.de/yulz3> (Stand: 30.06.2025).
- Liss, Hanna (2021b), „Masorah as Counter-Crusade? The Use of Masoretic List Material in MS London, British Library Or. 2091“, in: Hanna Liss u. Jonas Leipziger (Hgg.), *Philology and Aesthetics: Figurative Masorah in Western European Manuscripts*, Frankfurt am Main, 131–175.
- Liss, Hanna (2023), „Masorah Re-Arranged: Eight Masoretic Lists in MS London Oriental 2091, Fol. 335v“, in: *Corpus Masoreticum Working Papers* 6, 36–120.
- Liss, Hanna/Dörr, Stephen (2022), „Hebrew-French Bible Glossaries and the Question of Jewish-French Cultural Exchange in the High Middle Ages: A Reevaluation“, in: *Corpus masoreticum working papers* 2, 22–50.
- Ofer, Yosef (2019), *The Masora on Scripture and Its Methods*, Berlin et al.
- Sahle, Patrick (2013), *Digitale Editionsformen. Zum Umgang mit der Überlieferung unter den Bedingungen des Medienwandels*, Teil 1: *Das typografische Erbe*, <https://kups.ub.uni-koeln.de/5351/>; Teil 2: *Befunde, Theorie und Methodik*, <https://kups.ub.uni-koeln.de/5352/>; Teil 3: *Textbegriffe und Recodierung*, <https://kups.ub.uni-koeln.de/5353/>, Norderstedt.

## Abbildungsnachweis

- Abb. 1 MS London British Library Or 2091, fol. 203r (Public Domain).
- Abb. 2 MS Leipzig Universitätsbibliothek Leipzig 1099, fol. 17r (Public Domain).
- Abb. 3 Corpus Masoreticum Cloud Computing (© Clemens Liedtke).
- Abb. 4 Oberfläche BIMA 2 Corpus Masoreticum (CC-BY-SA 4.0).