

Varianz vermitteln

Zur editorischen Aufarbeitung abweichender Textfassungen in Redaktionen von *Der Heiligen Leben*

In der altgermanistischen Editionspraxis hat sich bekanntlich seit langem ein Wandel in der Auffassung davon vollzogen, wie mit varianten Textüberlieferungen umzugehen sei: Galten der Lachmann'schen Schule Abweichungen zwischen Handschriften vor allem als ‚Fehler‘, auf deren Grundlage die Rekonstruktion eines idealen Textes vollzogen wurde,¹ rückten die Überlieferungsvarianten unter dem Einfluss von Strömungen wie der ‚überlieferungsgeschichtlichen Schule‘² oder der ‚New Philology‘³ immer mehr in das Zentrum des Interesses.⁴ Dies zumindest in theo-

¹ Sebastiano Timpanaro: Die Entstehung der Lachmannschen Methode. 2., erweiterte und überarb. Aufl. Hamburg 1971; Hans-Gert Roloff: Karl Lachmann, seine Methode und die Folgen. In: Geschichte der Editionsverfahren vom Altertum bis zur Gegenwart im Überblick. Ringvorlesung. Hrsg. von Hans-Gert Roloff. Berlin 2003 (Berliner Beiträge zur Editionswissenschaft. 5), S. 63–81; Paolo Trovato: Everything you Always Wanted to Know about Lachmann's Method. A Non-Standard Handbook of Genealogical Textual Criticism in the Age of Post-Structuralism, Cladistics, and Copy-Text. Padova 2014 (Storie e linguaggi. 7); Oliver Primavesi, Anna Kathrin Bleuler: Einleitung: Lachmanns Programm einer historischen Textkritik und seine Wirkung. In: Lachmanns Erbe. Editionsmethoden in klassischer Philologie und germanistischer Mediävistik. Hrsg. von Anna Kathrin Bleuler und Oliver Primavesi. Berlin 2022 (Beihefte zur Zeitschrift für deutsche Philologie. 19), S. 11–107.

² Überlieferungsgeschichtliche Prosaforschung. Beiträge der Würzburger Forschergruppe zur Methode und Auswertung. Hrsg. von Kurt Ruh. Tübingen 1985 (Texte und Textgeschichte. 19); Werner Williams-Krapp: Die überlieferungsgeschichtliche Methode. Rückblick und Ausblick. In: Internationales Archiv für Sozialgeschichte der deutschen Literatur 25, 2000, S. 1–21; Freimut Löser: Überlieferungsgeschichte(n) schreiben. In: Überlieferungsgeschichte transdisziplinär. Neue Perspektiven auf ein germanistisches Forschungsparadigma. Hrsg. von Dorothea Klein, Horst Brunner und Freimut Löser. Wiesbaden 2016 (Wissensliteratur im Mittelalter. 52), S. 1–20.

³ Bernard Cerquiglini: Éloge de la variante. Histoire critique de la philologie. Paris 1989; Speculum: A journal of medieval studies 65, 1990, H. 1; Towards a Synthesis? Essays on the New Philology. Hrsg. von Keith Busby. Amsterdam 1993 (Faux Titre. 68); Alte und Neue Philologie. Hrsg. von Martin-Dietrich Gleßgen und Franz Lebsanft. Tübingen 1997 (Beihefte zu editio. 8); Peter Strohschneider: Situationen des Textes. Okkasionelle Bemerkungen zur ‚New Philology‘. In: Zeitschrift für deutsche Philologie 116, 1997, S. 62–86; Karl Stackmann: Neue Philologie? In: Modernes Mittelalter. Neue Bilder einer populären Epoche. Hrsg. von Joachim Heinzle. Frankfurt/Main, Leipzig 1999, S. 398–427; Freimut Löser: Postmodernes Mittelalter? ‚New Philology‘ und Überlieferungsgeschichte. In: Kulturen des Manuskriptzeitalters. Ergebnisse der amerikanisch-deutschen Arbeitstagung an der Georg-August-Universität Göttingen vom 17. bis 20. Oktober 2002. Hrsg. von Arthur Groos und Hans-Jochen Schiewer. Göttingen 2004, S. 215–236; Freimut Löser: Überlieferungsgeschichte und New Philology. Methodische Varianten in der Altgermanistik. Wiesbaden 2005.

⁴ Die vielleicht prominentesten Stationen dieser Entwicklung stellen die Aufsätze von Karl Stackmann: Mittelalterliche Texte als Aufgabe. In: Festschrift Jost Trier zum 70. Geburtstag. Hrsg. von William

retischer Hinsicht; in der Interpretationspraxis wurde diese Wende zur Handschriftennähe jedoch nicht immer in gleicher Weise nachvollzogen. Interpretiert wird in vielen Fällen immer noch auf der Grundlage von rekonstruierend verfahrenen Eintexteditionen, und das interpretatorische Potential von Varianz scheint erst in Ansätzen ausgeschöpft worden zu sein. Dies hängt sicher auch damit zusammen, dass sich Unterschiede zwischen Textfassungen nur in für unsere Lesegewohnheiten ungewöhnlicher Weise darstellen lassen: Wir sind es in unserer Prägung durch das Buchdruckmedium gewohnt, Texte linear zu lesen; Überlieferungsvarianten eröffnen demgegenüber sozusagen eine vertikale Dimension, die mithin geradezu quer zu unseren Bemühungen steht, Texte zu verstehen.

In Editionen der Lachmann'schen Schule sind Varianten – natürlich überspitzt gesagt – im ‚Lesartenfriedhof‘ des Apparats marginalisiert, sie stehen unter dem Haupttext und dienen eigentlich nur als Argumente zur Stützung des Ausgaben-textes, über sie kann problemlos hinweggelesen werden.⁵ Die ideelle Geringschätzung der Variante und ihre materielle Marginalisierung im Buchmedium gehen hier Hand in Hand, Editionseinrichtung und -grundlagen bedingen sich gegenseitig.⁶ Es scheint daher offensichtlich, dass mit der theoretischen Aufwertung der Abweichungen der Überlieferung auch neue Darstellungsformen zu finden sind, die das Potential der Varianz an die Nutzerinnen und Nutzer vermitteln.⁷

Vielversprechend für eine solche Revolution der Darstellungsformen und ein Durchbrechen der linearen Lesegewohnheit scheinen insbesondere digitale Editions-methoden zu sein. In digitalen Editionen ist es möglich, unterschiedliche Ansichten auf den Text nebeneinander zur Darstellung zu bringen und fast beliebig viele Kontextualisierungen vorzunehmen.⁸ Dennoch ist gerade die Frage, wie man die Darstellung von Varianz bewerkstelligt und Querlesungen zwischen Texten und

Foerste und Karl Heinz Borck. Köln, Graz 1964, S. 240–267, und Joachim Bumke: Der unfeste Text. Überlegungen zur Überlieferungsgeschichte und Textkritik der höfischen Epik im 13. Jahrhundert. In: ‚Aufführung‘ und ‚Schrift‘ in Mittelalter und Früher Neuzeit. Hrsg. von Jan-Dirk Müller. Stuttgart, Weimar 1996 (Germanistische Symposien. Berichtsbände. 17), S. 118–129, dar.

⁵ Zur Theorie des Apparats vgl. zuletzt etwa Florian Kragl: Kritik des Apparats. Anlässlich der neuen Ausgabe der *Laurine*. In: Beiträge zur Geschichte der deutschen Sprache und Literatur (PBB) 136, 2014, S. 601–623.

⁶ Dieser Eindruck dürfte letztlich auch bestehen bleiben, wenn man konzediert, dass die im Verlauf der frühen altgermanistischen Editions-geschichte ersichtliche Tendenz zur Verschiebung des Apparats aus dem Anhang unter den Text überhaupt erst die Rückbindung der Edition an die dort ersichtliche Beweisführung befördert hat; vgl. Kragl 2014 (Anm. 5), S. 602.

⁷ Vgl. Kragl 2014 (Anm. 5), S. 604; Karina van Dalen-Oskam: In Praise of the Variant Analysis Tool: A Computational Approach to Medieval Literature. In: Texts, Transmissions, Receptions: Modern Approaches to Narratives. Hrsg. von André Lardinois, Sophie Levie, Hans Hoeken und Christoph Lüthy. Leiden, Boston 2015 (Radboud Studies in Humanities. 1), S. 35–54.

⁸ Vgl. grundlegend Patrick Sahle: Digitale Editionsformen. Zum Umgang mit der Überlieferung unter den Bedingungen des Medienwandels. 3 Bde. Norderstedt 2013; Elena Pierazzo: Digital Scholarly Editing. Theories, Models and Methods. New York 2015.

auch Kontexten ermöglicht, noch keineswegs erschöpfend beantwortet.⁹ Obwohl die Möglichkeiten des digitalen Mediums bekannt sind, bleibt also die Frage, wie man Textvarianz ansprechend vermittelt, ein vordringliches Forschungsgebiet, gerade um den Einzug des Paradigmenwechsel hin zu einem handschriftennahen Textverständnis auch in der interpretatorischen Praxis endlich zu verankern.

Im Folgenden werden einige Möglichkeiten des Umgangs mit und der Darstellung von Varianz anhand eines digitalen Editionsprojekts zu einem spätmittelalterlichen Legendar präsentiert, das in der Forschung den sperrigen Namen *Der Heiligen Leben, Redaktion* erhalten hat. Das Werk bietet ein besonders gutes Beispiel dafür, wieso sich die Betrachtung von Textunterschieden lohnen kann, zugleich stellt es in Bezug auf die Vermittlung von Varianz durchaus besondere Herausforderungen.

Der Heiligen Leben, Redaktion ist im frühen 15. Jahrhundert vermutlich in der Region von Nürnberg entstanden.¹⁰ Als Legendar bietet es eine Sammlung von nach dem Jahrestag gegliederten Heiligenviten ‚per circulum anni‘. Wie der Name mehr schlecht als recht ausdrückt, stellt *Der Heiligen Leben, Redaktion* eine Bearbeitung von *Der Heiligen Leben* dar, dem erfolgreichsten Legendar seiner Zeit in Europa.¹¹ Entstanden am Ende des 14. Jahrhunderts, vereint *Der Heiligen Leben* 251 Viten,¹² die in der *Heiligen Leben, Redaktion* nochmals auf nahezu 400 erweitert wurden. Zusätzlich wurden Tageseinträge des Martyrologiums des Usuardus eingetragen.¹³ Insgesamt ist so das „mit Abstand [...] umfangreichste [Legendar] des deutschen Mittelalters“¹⁴ entstanden und damit ein voluminöses Textensemble, das sehr früh schon in zwei unterschiedlichen Versionen vorliegt. Denn kurz nach seiner Entstehung wurde in einer zweiten Phase das gesamte Legendar wiederum bearbeitet, wobei sich zwischen erster und zweiter Stufe tiefgreifende stilistische und narrative

⁹ Ein erster Überblick über bestehende Ansätze aus Sicht der Visualisierungstechniken bei Tariq Yousef, Stefan Jänicke: A Survey of Text Alignment Visualization. In: IEEE Transactions on Visualization and Computer Graphics 27, 2021, H. 2, S. 1149–1159.

¹⁰ Zum mutmaßlichen Entstehungsort Nürnberg s. Werner Williams-Krapp: Die deutschen und niederländischen Legendare des Mittelalters. Studien zu ihrer Überlieferungs-, Text- und Wirkungsgeschichte. Tübingen 1986 (Texte und Textgeschichte. 20), S. 331f. Vgl. auch Konrad Kunze: Der Heiligen Leben, Redaktion. In: Die deutsche Literatur des Mittelalters. Verfasserlexikon. Bd. 3. 2., völlig neu bearb. Aufl. Berlin, New York 1981, Sp. 625–627, der die Entstehungszeit von *Der Heiligen Leben, Redaktion* auf „vor 1434“ ansetzt (s. ebd., Sp. 627).

¹¹ Vgl. Kunze 1981 (Anm. 10), Sp. 617–625; zur umfangreichen Überlieferungs-, Text- und Wirkungsgeschichte von *Der Heiligen Leben* s. Williams-Krapp 1986 (Anm. 10), S. 188–346. Eine Edition des Legendars liegt in zwei Bänden vor: Der Heiligen Leben. Bd. I: Der Sommerteil. Hrsg. von Margit Brand, Kristina Freienhagen-Baumgardt, Ruth Meyer und Werner Williams-Krapp. Tübingen 1996 (Texte und Textgeschichte. 44); Der Heiligen Leben. Bd. II: Der Winterteil. Hrsg. von Margit Brand, Bettina Jung und Werner Williams-Krapp. Tübingen 2004 (Texte und Textgeschichte. 51).

¹² Zum Normalcorpus s. Williams-Krapp 1986 (Anm. 10), S. 292–295.

¹³ Siehe zu Vorgängern dieses Kombinationstyps in der lateinischen Legendartradition Kunze 1981 (Anm. 10), Sp. 626.

¹⁴ Williams-Krapp 1986 (Anm. 10), S. 315.

Umgestaltungen finden lassen, die mitunter auch die inhaltlichen Elemente der Vorlagentexte betreffen.¹⁵

Die textuellen Abweichungen könnten sich dabei durch die Einbindung in unterschiedliche Gebrauchssituationen, etwa in monastische Kontexte, ergeben haben oder durch anders gelagerte Erwartungen einer neuen Rezipientengruppe, an die die Texte angepasst wurden. Neben diesen externen Ursachen wären aber auch andere, sozusagen interne Faktoren als Erklärung denkbar, die mit den Besonderheiten der literarischen Gattung der Legende zu tun haben könnten. Es ist etwa in der jüngeren Forschung diskutiert worden, ob sich im rhetorischen Duktus von Legenden Spannungen manifestieren, die sich aus der Aufgabe ergeben, nüchtern über die Exorbitanz eines Heiligen zu erzählen und zugleich zur Nachahmung einer aus der Gemeinschaft herausgehobenen Persönlichkeit anzuregen.¹⁶ Zudem wird dadurch, dass in Legenden geradezu serienmäßig vom Unglaublichen erzählt wird, dieses Unglaubliche gerade hinterfragbar;¹⁷ Strategien der Historisierung des Inhalts wie auch der Emotionalisierung der Rezipienten könnten in Zusammenhang mit dem komplexen Geltungsanspruch dieser Textgattung stehen. Zudem ist die Retextualisierung auch als ein Akt der spirituellen Aneignung der Vorlagen seitens eines Bearbeiters oder Redakteurs zu bedenken.¹⁸ Es dürften also im Fall der Legenden nicht allein unterschiedliche Gebrauchssituationen, sondern auch diese inneren Spannungen, theologische wie zugleich ästhetisch-literarische, dazu geführt haben, dass Legenden in den unterschiedlichsten Ausprägungen vorliegen und eine große textuelle Varianz aufweisen.

Die große Textmasse des dreibändigen Legendaris *Der Heiligen Leben, Redaktion* ist noch nicht ediert.¹⁹ Insgesamt ist die Überlieferung auf einen relativ engen zeitlichen Rahmen (bis zum dritten Viertel des 15. Jahrhunderts) und auf den

¹⁵ Die beiden Stufen der *Redaktion* werden im Folgenden mit den Siglen HL-Red. I und HL-Red. II abgekürzt.

¹⁶ Susanne Köbele: Die Illusion der ‚einfachen Form‘. Über das ästhetische und religiöse Risiko der Legende. In: Beiträge zur Geschichte der deutschen Sprache und Literatur (PBB) 134, 2012, S. 365–404.

¹⁷ Siehe Werner Williams-Krapp: Laienbildung und volkssprachliche Hagiographie. In: Literatur und Laienbildung im Spätmittelalter und in der Reformationszeit. Symposium Wolfenbüttel 1981. Hrsg. von Ludger Grenzmann und Karl Stackmann. Stuttgart 1984 (Germanistische Symposien. Berichtsbände. 5), S. 697–709. Zum Fiktionalitätsverdacht in nachreformatorischer Zeit s. Antje Sablotny: Metalegende. Die protestantische Lügende als invektive Metagattung. In: Beiträge zur mediävistischen Erzählforschung 2, 2019, S. 148–200.

¹⁸ Vgl. Cornelia Herberichs: *ein zusammenclawber vnd ein auszsprecher gar nahent aller heiligen merterer*. Dynamiken des Sammelns als kreative Memoria im Legendar ‚Der Heiligen Leben, Redaktion‘. In: Sammelns als literarische Praxis im Mittelalter und in der Frühen Neuzeit. XXVI. Anglo-German Colloquium 26.–30. August 2019. Hrsg. von Mark Chinca, Manfred Eikelmann, Michael Stolz und Christopher Young. Tübingen 2022, S. 143–159, hier S. 153–158.

¹⁹ Die Texte der HL-Red. II sind nur vereinzelt in Editionen zugänglich; s. zu den edierten Texten Williams-Krapp 1986 (Anm. 10), S. 82. Die Texte der HL-Red. I, die aus *Der Heiligen Leben* übernommen wurden, sind nach dem *Y-Zweig ediert in *Der Heiligen Leben*, Bd. I und II (Anm. 11) (die aus *Der Heiligen Leben* übernommenen Texte der HL-Red. I sind stemmatologisch mit dem *Y-Zweig der Legendarüberlieferung verwandt; s. *Der Heiligen Leben*, Bd. II, Anm. 11, S. XIII).

begrenzten geographischen Raum des Bistums Bamberg beschränkt; ein Gesamtlegendar, das alle drei Bände umfasst, ist nicht erhalten.²⁰ Die Zahl an Parallelüberlieferungen der Einzelbände nimmt sich für die einzelnen Legendarteile jeweils unterschiedlich aus: Band 1, welcher inklusive der Prologe²¹ insgesamt 112 Einzeltexte für die Lektüre während der Tage vom 1. Januar bis zum 30. April enthält, ist in jeweils einer Vollhandschrift überliefert (HL-Red I: München, BSB, Cgm 536, und HL-Red. II: München, BSB, Cgm 535). Hingegen ist Band 2 mit seinen – ohne Mitzählen der hier ebenfalls überlieferten fünf Prologe – 133 Einzeltexten für die Zeit zwischen dem 1. Mai und dem 31. August jeweils in mehreren Vollhandschriften überliefert (HL-Red. I dreimal: Heidelberg, UB, Cpg 153; Karlsruhe LB, Donauesch. cod. 447b; Melk, StiBi., cod. 361; HL-Red. II viermal: Augsburg, UB, cod. III, 1, 20, 20; Augsburg, UB, cod. III, 1, 20, 21; Budapest, NB, Cod. Germ. 48; Prag, NB., Teplá MS. D 17, früher: MS. Teplá 42). Für Band 3, der mit 136 Einzeltexten den Zeitraum vom 1. September bis zum 31. Dezember abdeckt, ist keine einzige Vollhandschrift der HL-Red. I überliefert, während für die HL-Red. II dieses Bandes zwei Handschriften vorliegen (Augsburg, UB, cod. III, 1, 20, 2; München, BSB, Cgm 537). Darüber hinaus gibt es weitere Codices, die einzelne Legenden der beiden Bearbeitungsstufen der *Redaktion* und zum Teil zusätzliche Sondergutlegenden beinhalten.²²

Das hier skizzierte Editionsprojekt stellt sich der Aufgabe, diese Textfülle von nahezu 400 Texten, zumeist in jeweils zwei Versionen, zugänglich und so weitere Variationen im großen Korpus der Legendare sichtbar zu machen. Mit der Edition soll es also möglich sein, Varianten zwischen den Textfassungen aufzuspüren, darüber hinaus Varianten aber auch zu anderen Legendenerzählungen aufzuzeigen und Bearbeitungstendenzen nachzuvollziehen. Eine entscheidende Rolle spielen daher die Alignierung von Textversionen und die Darstellungsmöglichkeiten, die deren Vergleich unterstützen. Wie etwa Stefan Jänicke und David Wrisley herausgestellt haben, ist die Darstellung von Varianz ein Problem, dem man sich auf mehreren perspektivischen Ebenen nähern kann.²³

²⁰ Nur in einem Fall sind Band 2 und 3 der Phase HL-Red. II als ursprünglich zusammengehörende Bände überliefert: München, BSB, Cgm 535 und München, BSB, Cgm 537. Die bislang ausführlichste Beschreibung der Überlieferung stammt von Williams-Krapp 1986 (Anm. 10), S. 317–320.

²¹ Zur Reihenbildung der Prologe siehe Herberichs 2022 (Anm. 18), S. 149–153.

²² Unter anderem Wunsiedel, Stadt- und Kirchenbibliothek, Ha 7 und die Handschrift München, BSB, Cgm 539, welche zudem Sondergutlegenden enthält; zusätzlich sind Fragmente und weitere Einzeltextüberlieferungen im Rahmen von Streuüberlieferung zu verzeichnen, wobei mit weiteren Funden zu rechnen ist, da „viele [] HL-Hss. [existieren], in die einzelne Texte aus der HL-Red. zurückgeflossen sind“ (Williams-Krapp 1986, Anm. 10, S. 317).

²³ Etwa auf einer Makro-, einer Meso- oder einer Mikro-Ebene; vgl. Stefan Jänicke, David Joseph Wrisley: Interactive Visual Alignment of Medieval Text Versions. In: 2017 IEEE Conference on Visual Analytics Science and Technology (VAST). 2017, S. 127–138, DOI: 10.1109/VAST.2017.8585505 (sämtliche Links in diesem Beitrag sind zuletzt gesehen am 17.12.2022); Stefan Jänicke, David Joseph Wrisley: Visualizing Mouvance. Toward a visual analysis of variant medieval text traditions. In: Digital Scholarship in the Humanities 32, Issue suppl_2, 2017, S. ii106–ii123.

Dementsprechend sollen im Folgenden drei Aspekte der Vermittlung von Varianz differenziert werden, denen sich das Editionsprojekt besonders widmen wird und die im Folgenden kurz umrissen seien: Da ist zum ersten die Frage nach der Alignierung von Textversionen in der synoptischen Darstellung, die insbesondere bei Prosatexten eine große Herausforderung darstellt. Zum Zweiten sind Möglichkeiten der Kategorisierung von Varianz und deren Visualisierung zu bedenken. Zum Dritten ist danach zu fragen, wie das Konzept des Textvergleichs über das Legendar hinaus in den Kontext eines umfassenden Korpus von legendarischen Texten oder auch anderen medialen Formen hagiographischer Wissenskontexte verlängert werden kann.

Zum ersten Aspekt, der Alignierung von Textversionen: Im digitalen Medium ist eine solche Alignierung, also der Abgleich von Textversionen in einer meist synoptischen Darstellung, im Falle von Verstexten vergleichsweise einfach zu erreichen und dementsprechend ein durchaus gängiges Feature von Editionen. Als Beispiel kann hier etwa die digitale Edition des *Armen Heinrich* von Hartmann von Aue dienen (vgl. Abb. 1).²⁴

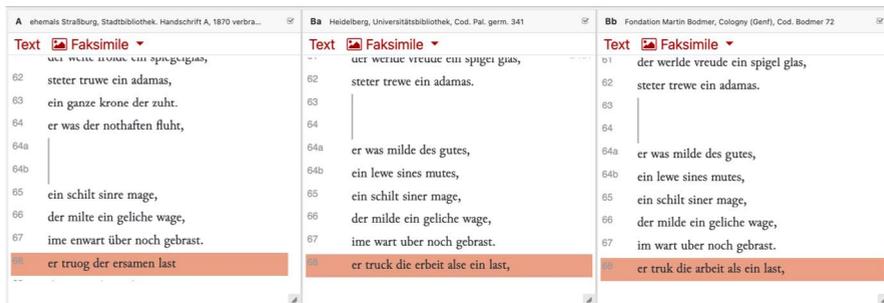


Abb. 1: Digitale Edition des *Armen Heinrich* (Ausschnitt).

Die synoptische Darstellung der Edition zeigt die drei überliefernden Handschriften, die zeilengenau miteinander synchronisiert werden können. Fehlen Verse oder kommen welche hinzu (etwa V. 63/64 und 64a/64b), dann lässt sich dies durch leere Zeilen oder Zusatzversnummern andeuten. Durch den zeilengenauen Abgleich wird es leicht möglich, Abweichungen im Wortlaut festzustellen, wie etwa in den unterschiedlichen Versionen des Verses 68 („er truog der ersamen last“ A gegen „er truck die erbeit alse ein last“ Ba bzw. „er truk die arbeit als ein last“ Bb).

²⁴ Hartmann von Aue: *Der arme Heinrich*. Textgeschichtliche elektronische Ausgabe. Hrsg. von Gustavo Fernández Riva und Victor Millet, unter Mitarbeit von Jakob Šimek, mit Übersetzungen von Dietmar Peschel. Heidelberg 2018, URL: <https://doi.org/10.11588/edition.ahd>. Die Darstellung ist abrufbar unter <https://digi.ub.uni-heidelberg.de/diglit/edition/ArmerHeinrich>.

Probleme ergeben sich bei dieser Darstellungsweise am ehesten bei Verschiebungen von Versen und Textumstellungen, die aber – in Relation etwa zur Legendenüberlieferung gesehen – in vergleichsweise stabil überlieferten Texten wie dem *Armen Heinrich* ohnedies nicht allzu häufig vorkommen.

Deutlich komplexer stellt sich die Situation jedoch bei stark veränderten Prosatexten wie unserem Legendar dar (Abb. 2).

The screenshot displays a web-based digital edition interface. At the top, the browser address bar shows a local host URL. Below it, a navigation bar contains the title 'Der Heiligen Leben, Redaktion - Ambrosius'. The main content area is split into two columns of text, each starting with 'E 2'. The left column contains German text with red arrows pointing to specific words or phrases. The right column contains a different version of the text, also with red arrows. To the right of the text columns are two sidebars, each titled 'Personen'. The first sidebar lists names like Ambrosius, Jesus Christus, Valentinian, Rufinus, Simplicianus, Isidor, and Orte, with small circular icons next to them. The second sidebar lists names like Ambrosius, Rufinus, Simplicianus, Isidor, Valentinian, Orte, Mailand, Rom, Italien, Hispalia, Lincio, and Konstantinopel, also with small circular icons.

Abb. 2: Prototyp der digitalen Edition von *Der Heiligen Leben, Redaktion*: Synoptische Gegenüberstellung des Beginns der Ambrosius-Legende nach HL-Red. I und HL-Red. II.

Abbildung 2 zeigt einen Prototyp unseres Editionsvorhabens, in welchem die beiden angesprochenen Textversionen synoptisch gegenübergestellt sind, und zwar anhand des Beispiels der Legende des Hl. Ambrosius, des Schutzpatrons der Bienen, Kirchenvater und Bischof von Mailand in der zweiten Hälfte des 4. Jahrhunderts.²⁵

Auch im Fall eines frühneuhochdeutschen Prosatextes ist es natürlich sinnvoll, den Text in kleinere Segmente zu unterteilen, diese Segmentierung ist jedoch nicht durch Verszeilen vorgegeben. In Prosa bieten sich zunächst Einzelsätze als Segmente an, wie dies in unserem Beispiel gehandhabt ist. Abbildung 2 zeigt Passagen vom Beginn der Legende, in denen zunächst eine Art Kurzzusammenfassung der Verdienste des Heiligen geboten wird.

²⁵ Die Ambrosius-Legende ist für den 4. April in Bd. I der *Heiligen Leben, Redaktion*, enthalten. Sie ist ausschließlich in zwei Handschriften der HL-Red. vollständig überliefert: in der Fassung von HL-Red. I auf den Blättern München, BSB, Cgm 536, fol. 201vb–204vb, und in der Fassung von HL-Red. II auf München, BSB, Cgm 535, fol. 326ra–330ra.

In Version 1 (HL-Red. II, Cgm 536, fol. 201vb) links lautet dieser Beginn:

6 DEr heilig fant Ambrosius was gar dimütig vnd tugenlich vnd grofzer weifhait vnd ein lerer der heiligen schrift.

7 vnd er leget allen feinen fleifz darauff, wie er das lob gotes mocht weiten aufz gepreiten.

8 Vnd er hat di heiligen criftenhait wol erleucht vnd gezired mit manigen edeln fpruchen.

9 Vnd ift der vier lerer einer.

10 Er ift auch geporen von Ram von reichem vnd edelm gefchlechte.

Die Satzgrenzen finden sich nicht in der Handschrift, die Interpunktion wurde in der Edition hinzugefügt und versteht sich als vorläufig. Was an den Beispielsätzen bereits ersichtlich wird, sind die bekannten Probleme spätmittelalterlicher deutscher Syntax.²⁶ Die Satzgrenzen sind nämlich häufig nicht genau festzulegen, da die Sätze syndetisch mit ‚und‘-Konjunktionen aneinandergereiht sind.²⁷ Nach heutigen Regeln könnte man aus Satz 6 bis 9 z.B. einen einzigen Satz machen. Wie die Handschrift jedoch selbst mit der Rubrizierung der Initialbuchstaben anzeigt,²⁸ hier etwa bei Satz 8 und 9, sind die ‚und‘-Perioden wohl durchaus auch als abgetrennt verstanden worden.

Noch schwieriger wird die Angelegenheit aber, wenn man die andere Version der Legende (Cgm 535) mitberücksichtigt, die hier in der Konkordanz neben der ersten Version steht. Dort ist das Problem noch verschärft, hier wären die ‚und‘-Syntagmen nun für sich genommen klare Ellipsen, weil das Subjekt fehlt bzw. nicht nochmals pronominal aufgegriffen wird.²⁹ Es heißt also: „vnd was ein grozer lerer der heiligen geschriff“, „vnd hatt die cristenheit wol geziret“ etc. Aber auch hier ist der Text wohl kaum als im heutigen Sinn gegliedert verstanden, was wieder die handschriftlichen Strichelungen anzeigen, die einzelne ‚und‘-Teilsätze markieren. Für den Vergleich wird dies nun besonders prekär bei Satz 10, denn hier hat Version 1 einen vollständigen Satz, während Version 2 eine solche ‚und‘-Ellipse hat. Teilsatz 10 in Version 2 gehört also eigentlich, nach heutigen Maßstäben, noch zu dem ganzen Zusammenhang, der ab Satz 6 beginnt, während er in Version 1 ein neuer, eigenständiger, aber sinnreicher Satz ist. Was soll man hier also mit was vergleichen?

Unter diesen Umständen kann es zudem leicht passieren, dass ein Satz in der einen Version zwei oder gar mehr Sätzen in der anderen Version entspricht, was dann zur Folge hat, dass es zu Verschiebungen in der synoptischen Darstellung kommt (vgl. Abb. 3).

²⁶ Siehe den Überblick in: Frühneuhochdeutsch. Eine Einführung in die deutsche Sprache des Spätmittelalters und der frühen Neuzeit. Hrsg. von Frédéric Hartweg und Klaus-Peter Wegera. Tübingen 1989 (Germanistische Arbeitshefte. 33), S. 133–140.

²⁷ Vgl. Hermann Paul: Mittelhochdeutsche Grammatik. Tübingen 2007 (Sammlung kurzer Grammatiken germanischer Dialekte. A: Hauptreihe. 2), § S 221.

²⁸ Rubrizierte Buchstaben sind in der Editionsansicht rot dargestellt.

²⁹ Zur Gängigkeit der Konstruktion vgl. Paul 2007 (Anm. 27), § S 208,1, Anm. 2, und § S 208,2.

Der Heiligen Leben, Redaktion - Ambrosius

16 → Darnach zu einen zeiten, do waz das kind fand Ambrosius mit feiner schwelster gengen in di kirchen.
 17 → Do opfert fy dem pilchoff vnd küzlet im fein hant, Als dann gewonheit waz.
 18 → Do weitaget das kint Ambrosius - Vnd sprach zu feiner schwelster:
 19 → - "du wirrt mir noch mein hant auch küzen."
 20 → Aber fy hot kain achtung auf di red, wann fy meint, er ret es in feiner kintheit.
 21 → - Aber do er her nach ein pilchoff wart, do gefchah, daz im fein schwelsterj fein hant küzlet.
 E 5
 22 → Darnach do lifz man fand Ambrosius zu der lernung.
 23 → do lertt er mit grotzer weihait di heilig-n schrifft.
 24 → Vnd hutet fich alzeit vor fünden.
 25 → Vnd nam zu an der lernung Vnd an allen tugenten, Das er kurzlichen ein grotzer mailter der heiligen schrifft wart.
 E 6
 E 6.1
 → Paa vnd dem heilf fclation, mofant den

16 → das kint wuchs auff in vil tugenden.
 E 4
 17 → Eins mols was das kint zu kirchen vnd was fein schwelster auch do.
 18 → Do fahs fant ambrosius dem pilchoff opfert nach gewonheit vnd im auff fein hant küzlet küzet.
 19 → Do weyffaget fant am-brosius als palkte vnd sprach:
 20 → - Schwester kufz mir mein hant auch wann das mütz fein."
 21 → Do sprach fie:
 22 → - habe dein gemach.
 23 → Wann fie hett im die rede vor er - kintheit.
 24 → Do ward es her-nach wars, wann do er pilchoff warde, Do küzlet fie im fein hant.
 E 5
 25 → Das liebe kint ginge fchrie) zu fchul vnd lernat mit grotzer weyffheit die heiligen gefchriift vnd hüt fich alzeit vor fünden.
 26 → Vnd ward fchir darnach ein grotzer mailter.

Personen		Personen	
Ambrosius	20	Ambrosius	13
Jesus Christus	3	Rufinus	4
Valentinian	2	Simplicianus	2
Rufinus	2	Isidor	1
Simplicianus	2	Valentinian	1
Isidor	1	Orte	
Orte		Malland	4
Malland	7	Rom	2
Italien	1	Italien	1
Hispalis	1	Hispalis	1
Rom	1	Linicio	1
lininio	1	Konstantinopel	1

Abb. 3: Prototyp der digitalen Edition von *Der Heiligen Leben, Redaktion: Synoptische Gegenüberstellung eines Ausschnitts aus der Ambrosius-Legende nach HL-Red. I und HL-Red. II.*

In der Ambrosius-Legende tritt dieser Fall bereits bei Satz 16 ein. In Version 1 lautet der Abschnitt: „Darnach zu einen zeiten, do waz das kind fand Ambrosius mit feiner schwelster gengen in di kirchen.“ In Version 2 heißt es unter Satz 16 hingegen: „das kint wuchs auff in vil tugenden.“ Blickt man aber auf Satz 17, dann sieht man, dass erst dieser der eigentlich entsprechende Satz wäre; Satz 17 lautet nämlich: „Eins mols was das kint zu kirchen vnd was fein schwelster auch do.“

Wollte man sich an einer automatischen Alignierung solcher Verschiebungen versuchen, könnte man, zumindest bei nicht stark varianten Texten, die Levenshtein-Distanz zum Einsatz bringen.³⁰ Das Prinzip dieses Algorithmus kommt, mehr oder weniger abgewandelt, in den meisten gängigen Kollationierungstools zur Anwendung.³¹ Bei der Levenshtein-Distanz wird zur Berechnung der Distanz zwischen zwei Wörtern ein Score errechnet, der sich aus der Anzahl der Umstellungen, Auslassungen oder Hinzufügungen von Buchstaben zusammensetzt, die nötig sind, um von einem Wort auf das andere zu kommen. Die Wortpaare mit der geringsten Distanz werden dann zusammensortiert. Dies erlaubt ein wenig orthographischen

³⁰ V. I. Levenshtein: Binary Codes Capable of Correcting Deletions, Insertions and Reversals. In: Soviet Physics Doklady 10, 1966, H. 8, S. 707–710.

³¹ Etwa in den Tools Juxta, TRAViz oder CollateX; Dana Wheelles, Kristin Jensen: Juxta Commons. In: Proceedings of the Digital Humanities 2013, URL: <http://dh2013.unl.edu/abstracts/ab-142.html>; Stefan Jänicke, Annette Geßner, Greta Franzini, Melissa Terras, Simon Mahony, Geric Scheuermann: TRAViz: A Visualization for Variant Graphs. In: Digital Scholarship in the Humanities 30, 2015, Issue suppl_1, S. i83–i99, URL: <https://doi.org/10.1093/lc/fqvo49>; CollateX – Software for Collating Textual Sources, URL: <https://collatex.net/>.

Spielraum, die Wörter müssen nicht genau gleich lauten, um von einem Computer gegenseitig zugeordnet zu werden.

Die textuelle Varianz im Falle von *Der Heiligen Leben*, *Redaktion* beschränkt sich allerdings nicht nur auf unterschiedliche Schreibungen, sondern betrifft zum Teil auch die Ersetzung sinnverwandter Wörter, deren Schreibweisen überhaupt nichts miteinander zu tun haben und daher für den Computer mit konventionellen Mitteln nicht als zugehörig zu erkennen sind.

Hier können nun neuere digitale Verfahrensweisen, wie sie zurzeit auch in der digitalen ‘Text-Reuse’-Forschung zum Einsatz kommen, weitergehende Möglichkeiten eröffnen.³² Gemeint sind hier Verfahrensweisen der distributionellen Semantik wie ‘Word Embeddings’, denen seit einigen Jahren in der digitalen Textanalyse gesteigerte Aufmerksamkeit zukommt.³³ Die distributionelle Semantik geht von der Grundannahme aus, dass sich die Bedeutung eines Wortes nicht bloß aus sich selbst heraus ergibt, sondern aus dem Kontext, in dem es erscheint: “You shall know a word by the company it keeps” – so hat der Linguist John Rupert Firth diesen Sachverhalt bündig zum Ausdruck gebracht.³⁴

Ihren Durchbruch haben diese distributionellen Methoden in den Digital Humanities mit dem im Umkreis der Firma Google im Jahr 2013 entwickelten Word2Vec-Algorithmus erreicht. Word2Vec ist ein Verfahren, das in effizienter Weise Kontextvektoren für jedes Wort eines Korpus berechnen kann.³⁵ Kontextvektoren sind dabei letztlich nichts anderes als Zahlenreihen, aus denen ersichtlich wird, mit welchen anderen Wörtern das Zielwort überdurchschnittlich häufig in Zusammenhang erscheint. Mit diesen Vektoren lässt sich sehr gut rechnen – und damit mit semantischen Ähnlichkeiten kalkulieren. So kann etwa mit Hilfe eines geometrischen Distanzmaßes wie der Cosinus-Distanz der Abstand zwischen Vek-

³² Vgl. grundsätzlich zum Einsatz dieser Methoden in den Digital Humanities Marco Büchler: Informationstechnische Aspekte des ‘Historical Text Re-use’. Leipzig 2013, für den hier vorliegenden Anwendungsfall auf die Detektion von Textvarianz insbesondere Christofer Meinecke, David Joseph Wrisley, Stefan Jänicke: Automated Alignment of Medieval Text Versions based on Word Embeddings. In: LEVIA 2019: Leipzig Symposium on Visualization in Applications. Leipzig 2019, o.S., DOI: 10.31219/osf.io/tah3y; Stefan Tomasek, Christian Reul, Maximilian Wehner: Handwritten Text Recognition und Word Mover’s Distance als Grundlagen der digitalen Edition *Die Kindheit Jesu* Konrads von Fußesbrunnen. In: DHd 2022. 8. Tagung des Verbands Digital Humanities im deutschsprachigen Raum, S. 134–138, URL: <https://zenodo.org/record/6304590>.

³³ Siehe Peter David Turney, Patrick Pantel: From Frequency to Meaning: Vector Space Models of Semantics. In: Journal of Artificial Intelligence Research 37, 2010, H. 1, S. 141–88; Ben Schmidt: Vector Space Models for the Digital Humanities 2015, <http://bookworm.benschmidt.org/posts/2015-10-25-Word-Embeddings.html>; Christof Schöch: Quantitative Semantik: Word Embedding Models für literaturwissenschaftliche Fragestellungen. In: Digitale Literaturwissenschaft. Hrsg. von Fotis Jannidis. Stuttgart 2023 (im Druck).

³⁴ John Rupert Firth: A synopsis of linguistic theory. In: John Rupert Firth: Studies in linguistic analysis. Oxford 1957, S. 1–31, hier S. 11.

³⁵ Vgl. Tomas Mikolov, Kai Chen, Greg Corrado, Jeffrey Dean: Efficient Estimation of Word Representations in Vector Space. In: 1st International Conference on Learning Representations, ICLR 2013, Scottsdale, Arizona, USA, May 2–4, 2013, Workshop Track Proceedings. 2013, S. 1–12, <https://arxiv.org/pdf/1301.3781.pdf>.

toren ermittelt und auf diese Weise eine Liste der ähnlichsten Wörter zu einem Ausgangswort erstellt werden (Abb. 4).³⁶

```
In [10]: model.wv.most_similar('minne', topn=20)
Out[10]: [('minnen', 0.7144430875778198),
          ('minn', 0.6799044609069824),
          ('liebe', 0.6033757925033569),
          ('herzeliebe', 0.5968233346939087),
          ('sinne', 0.5832563638687134),
          ('güete', 0.5694595575332642),
          ('süeze', 0.5338948369026184),
          ('herze', 0.5032045245170593),
          ('höchgemüete', 0.48634767532348633),
          ('reine', 0.4820777475833893),
          ('gedanke', 0.47799330949783325),
          ('triuwe', 0.4726966917514801),
          ('trüren', 0.47019490599632263),
          ('huote', 0.4656956195831299),
          ('herzen', 0.46428489685058594),
          ('fröide', 0.463878333568573),
          ('vruht', 0.46041738986968994),
          ('unminne', 0.4603417217731476),
          ('stäte', 0.4573539197444916),
          ('fröude', 0.4560955762863159)]
```

Abb. 4: Word2Vec-Liste von Wörtern mit den geringsten Cosinus-Abständen zu mhd. ‚minne‘.

Abbildung 4 zeigt eine solche Liste für die zwanzig Wörter mit den geringsten Cosinus-Abständen zum mittelhochdeutschen Wort ‚minne‘, die nach den Annahmen des Verfahrens damit als die zwanzig semantisch ähnlichsten Wörter gelten können.³⁷ Die Liste wurde mit Hilfe eines Word2Vec-Modells erstellt, das auf der Basis der mittelhochdeutschen Begriffsdatenbank trainiert wurde, die wohl das umfangreichste digital verfügbare Korpus von mittelalterlichen deutschen Texten bietet.³⁸

Einmal in Zahlen gefasst, lassen sich noch weitere Rechenoperationen mit diesen Vektoren unternehmen, beispielsweise diese addieren oder subtrahieren. Um nur ein berühmtes Beispiel anzuführen: Wenn man den Vektor ‚Mann‘ vom Vektor von ‚König‘ abzieht und den Vektor von ‚Frau‘ aufaddiert, dann kommt man in den

³⁶ Vgl. Daniel Jurafsky, James H. Martin: *Speech and Language Processing. An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. 3. Aufl. New Jersey 2020, S. 111 f.

³⁷ Diese Annahme ist freilich nur eine Approximation. Auffälligerweise begegnen in solchen Listen immer wieder auch Gegenbegriffe zum Zielwort (etwa ‚Leid‘ zu ‚Liebe‘ oder wie in den vorliegenden Beispielen ‚trüren‘ zur ‚minne‘), was darauf hindeutet, dass sich die Wörter mit ähnlichen Vektoren bis zu einem gewissen Grad auch als mögliche Kandidaten für denselben paradigmatischen ‚slot‘ verstehen lassen.

³⁸ *Mittelhochdeutsche Begriffsdatenbank (MHDDBD)*. Universität Salzburg. Interdisziplinäres Zentrum für Mittelalter und Frühneuzeit (IZMF). Koordination: Katharina Zeppezauer-Wachauer. 1992–2022 (laufend), <http://www.mhdadb.sbg.ac.at/>. Die Berechnung erfolgte mit dem Gensim-Package für Python; Radim Řehůřek, Petr Sojka: *Software Framework for Topic Modelling with Large Corpora*. In: *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*. 2010, S. 45–50, DOI: 10.13140/2.1.23931847.

meisten natürlichsprachlichen Word2vec-Modellen an eine Position, die ungefähr dem Vektor von ‚Königin‘ entspricht (Abb. 5).

```
In [13]: model_n.vw.most_similar(positive=["küninc", "vrouwe"], negative=["man"], topn=10)
Out[13]: [('küniginne', 0.6205001473426819),
          ('juncvrouwe', 0.5573724508285522),
          ('tohter', 0.5536460876464844),
          ('künigin', 0.5199984312057495),
          ('rüedar', 0.4757816195487976),
          ('gezogenliche', 0.4662765860557556),
          ('herzogin', 0.45327040553092957),
          ('maget', 0.45232900977134705),
          ('swester', 0.4399452805519104),
          ('bät', 0.43746665120124817)]
```

Abb. 5: Word2Vec-Liste von Wörtern mit den geringsten Cosinus-Abständen zum Vektor, der sich aus der Addition der Vektoren von ‚küninc‘ und ‚vrouwe‘ sowie der Subtraktion des Vektors von ‚man‘ ergibt.

Wie aus Abbildung 5 ersichtlich wird, in der das Beispiel auf die Wörter ‚küninc‘, ‚vrouwe‘, ‚man‘ und ‚küniginne‘ übertragen wurde, funktioniert eine solche Rechnung auch für das auf dem mittelhochdeutschen Korpus trainierte Modell. Das ist fast ein wenig erstaunlich, denn für das Erstellen von solchen ‘Word-Embedding’-Modellen werden in der Regel relativ große Textmengen benötigt, die in mittelhochdeutscher Sprache eigentlich gar nicht recht zur Verfügung stehen. Mit seinen ca. 10 Millionen Token ist das Korpus der mittelhochdeutschen Begriffsdatenbank schon eher an der unteren Grenze hinsichtlich der Textmenge angesiedelt, die normalerweise als Grundlage für die Erstellung von ‘Word-Embedding’-Modellen herangezogen wird. Was die Situation im Vergleich zu Korpora neuerer Sprachstufen jedoch noch besonders verschärft, ist der Umstand, dass die der Begriffsdatenbank zugrundeliegenden Texte aus mehreren Jahrhunderten stammen und damit entsprechende sprachgeschichtliche Veränderungen aufweisen sowie zum Teil aus nicht normalisierten, zum Teil aber auch aus nach ganz unterschiedlichen Kriterien vereinheitlichten Texteditionen bezogen sind. Für das Modell bedeutet dies, dass es erheblich abweichende Schreibweisen des Mittel- und auch Frühneuhochdeutschen miteinander verrechnen muss.

Das Modell lässt sich daher verbessern, indem man das Mittelhochdeutsche-Korpus automatisch vorprozessiert und normalisiert. Bis vor Kurzem war dies nicht möglich, da es entsprechende Tools noch nicht gab bzw. diese noch nicht zu befriedigenden Resultaten führten. Mittlerweile liegt aber ein von Helmut Schmid trainiertes Normalisierungs- und Lemmatisierungs-Modell für den von ihm programmierten RNN-Tagger vor, mit dem nun auch für das Mittelhochdeutsche gute Ergebnisse erzielt werden können.³⁹ Dementsprechend ist das hier zum Einsatz

³⁹ Helmut Schmid: Deep Learning-Based Morphological Taggers and Lemmatizers for Annotating Historical Texts. In: Proceedings of Digital Access to Cultural Heritage (DATeCH). New York 2019, o.S., DOI: 10.1145/3322905.3322915.

kommende Modell auf der Basis derart automatisch normalisierter Texte erstellt worden.

Mit Hilfe eines Word2Vec-Modells wird es für den Computer also möglich, semantische Ähnlichkeiten zwischen Wörtern zu erkennen, ohne dass diese Wörter auf der Oberfläche Gemeinsamkeiten aufweisen müssen.⁴⁰ Für die Erfordernisse einer Zeilensynopse muss die Methode allerdings noch über den Wort-für-Wort-Vergleich hinaus auf die Satzebene erweitert werden. Zu diesem Zweck lässt sich etwa die sogenannte ‘Word Mover’s Distance’ (WMD) einsetzen, bei der auf Grundlage der einzelnen Wort-Vektoren die Minimaldistanz berechnet wird, die zwischen zwei Texten oder Sätzen im Vektorraum liegt.⁴¹

Diese WMD-Methode lässt sich nun auch auf die beiden Versionen der Ambrosius-Legende anwenden. Dazu haben wir mit dem Computer für jeden Satz aus der ersten Version des Legendars die WMD zu allen Sätzen aus der zweiten Version berechnet. Je kleiner die WMD ist, desto ähnlicher sollten – nach den Vorannahmen des Verfahrens – die Sätze sein. Auf diese Weise kann eine Rangliste der sinngemäß am meisten übereinstimmenden Sätze zu jedem Satz aus der ersten Version erstellt werden. Derselbe Vorgang kann auch in anderer Richtung ausgeführt werden, mit der zweiten Textversion als Ausgangspunkt.⁴²

Diese Berechnung der WMD soll nun im Kontext unserer Edition dazu verwendet werden, die Varianz der Textversionen miteinander und damit an die Nutzerinnen und Nutzer zu vermitteln. Eine mögliche Implementierung zeigt

⁴⁰ Wie erwähnt, handelt es sich beim Word2Vec-Algorithmus um das erste erfolgreiche abgekürzte Verfahren zur Erstellung von ‘Word Embeddings’. Mittlerweile ist eine Reihe weiterführender Methoden entwickelt worden. Eine Alternative zu Word2Vec wären etwa das Fasttext-Modell (Piotr Bojanowski, Edouard Grave, Armand Joulin, Tomas Mikolov: Enriching word vectors with subword information. In: Transactions of the Association for Computational Linguistics 5, 2017, S. 135–146; Fasttext kommt etwa in den Studien von Tomasek/Reul/Wehner 2022, Anm. 32, und Meinecke/Wrisley/Jänicke 2019, Anm. 32, zum Einsatz) sowie Token-basierte Embeddings, erstellt mit BERT (vgl. zu deren Einsatz in den Digital Humanities aber Anton Ehrmantraut, Thora Hagen, Leonard Konle, Fotis Jannidis: Type- and Token-based Word Embeddings in the Digital Humanities. In: CHR 2021: Computational Humanities Research Conference, November 17–19. Amsterdam 2021, S. 15–38, https://ceur-ws.org/Vol-2989/long_paper35.pdf). Eine Evaluierung dieser unterschiedlichen Systeme für den vorliegenden Task steht noch aus (und soll in einer zukünftigen Studie geleistet werden), für den hier zu erbringenden ‘proof of concept’ hat sich der Rückgriff auf Word2Vec jedoch bereits als vielversprechend erwiesen.

⁴¹ Vgl. Matt J. Kusner, Yu Sun, Nicholas I. Kolkin, Kilian Q. Weinberger: From Word Embedding To Document Distances. In: ICML’15: Proceedings of the 32nd International Conference on Machine Learning 37, 2015, S. 957–966, URL: <https://dl.acm.org/doi/10.5555/3045118.3045221>.

⁴² Bei einem solchen Verfahren steigt die Berechnungsdauer exponentiell an, je länger die Texte sind, weshalb für umfangreichere Textvergleiche eine Vorselektion vonnöten wäre; vgl. etwa das Verfahren bei Meinecke/Wrisley/Jänicke 2019 (Anm. 32), S. 4. Segmentiert man das Konvolut der *Heiligen-Leben-Redaktion* jedoch nach den einzelnen Viten, entstehen ausreichend kurze Texteinheiten, um den Vergleich mit akzeptablem Aufwand zu berechnen (eine Detektion von allfälligen über die Vitengrenzen hinausgehenden Textähnlichkeiten ist damit freilich nicht mehr möglich, aber auch nicht der hier vordringlichste Anwendungsfall).

ein weiterer Screenshot unseres Editionsprototyps in Abbildung 6: Klickt man auf eines der Pfeilsymbole, die in den beiden Editionstexten jeweils neben den Zeilennummern positioniert sind, werden auf der rechten Seite des Bildschirms für den entsprechenden Satz die drei gemäß der WMD ähnlichsten Sätze der jeweils anderen Version eingblendet. Der Computer bietet also einen ‚Dreiervorschlag‘ für mögliche Textentsprechungen.⁴³ Die Abbildung zeigt daher rechts die ähnlichsten Sätze aus Version 2 zum oben angesprochenen Satz 16 aus Version 1. Tatsächlich ist der erste Vorschlag mit der geringsten WMD auch jener Vers 17, dem man wohl auch nach menschlichem Ermessen die größte Ähnlichkeit zubilligen würde.

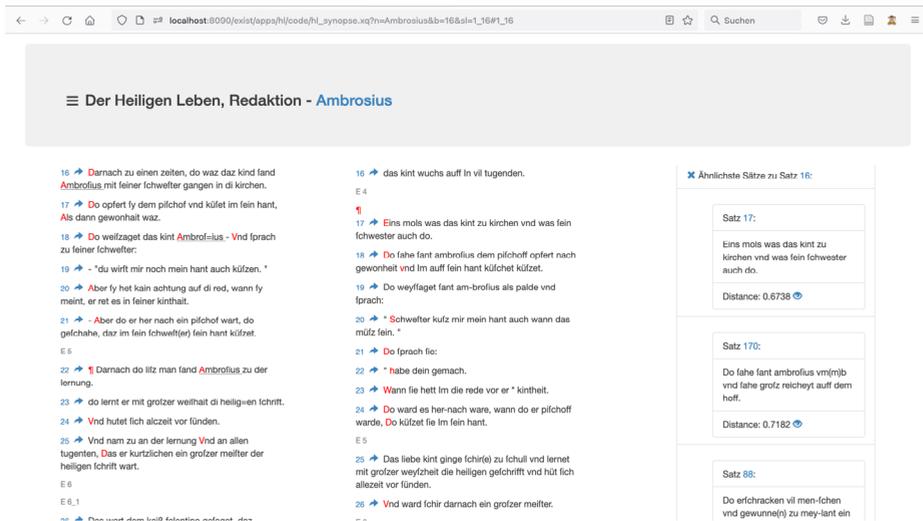


Abb. 6: Prototyp der digitalen Edition von *Der Heiligen Leben, Redaktion*: Zuordnung ähnlicher Sätze.

Auf diese Weise kann der Computer also dabei helfen, den jeweils passenden Satz in Hinblick auf die Synopse zu ermitteln, selbst wenn dieser im Wortlaut erheblich von seinem Pendant abweicht. Allfällige Unschärfen der Methode (die sich bei digitalen Verfahren immer einstellen werden), lassen sich dabei durch den interaktiven Zuschnitt des Interfaces ausgleichen: Sollte der Computer einmal nicht das Richtige getroffen haben oder gibt es gar keinen entsprechenden Satz, kann die

⁴³ Die Berechnung kann aufgrund der in Anmerkung 42 erwähnten Rechenintensität der Aufgabe nicht ‘on the fly’ vorgenommen werden, sondern beruht auf vorkalkulierten Listen, die als XML-Dateien in die Editionsdatenbank aufgenommen werden. Die Bildschirmdarstellung versteht sich als vorläufig und wird später noch optimiert werden.

Nutzerin oder der Nutzer eine Alternative aus den drei generierten Vorschlägen auswählen.⁴⁴

Die interaktive Auswahlmöglichkeit aus dem Dreivorschlag hat jedoch gleichsam als Nebeneffekt noch eine weitere hermeneutische Komponente, die sogar über den unmittelbaren Anwendungsfall der Synopse und die Vorstellung vom ‚richtigen‘ Textabgleich hinausführt, und zwar in Richtung eines allgemeineren Verständnisses von Intertextualität: Mit dieser Methode werden nämlich auch Sätze gefunden, die zwar nicht die genaue Entsprechung der beiden Textversionen im Sinne von Textvarianz darstellen, aber doch zumindest Ähnlichkeiten aufweisen. Das Verfahren ließe sich daher auch dazu einsetzen, Motive oder Floskeln aufzuspüren, die an anderen Stellen im Text auftreten, wodurch allfällige strukturelle Isotopien aufgedeckt werden können oder auch schlicht Passagen, an denen bereits bestehender Text sozusagen ‚recycelt‘ wurde. Dies lässt sich etwa an folgendem Beispiel aus unserem Text demonstrieren: Ambrosius kommt nach Mailand, wo Streit zwischen Christen und Arianern herrscht und gerade der Bischof gestorben ist, was in Version 1 mit Satz 30 „vnd den waz ir pifhof dot in der selben fstat“ ausgedrückt wird. Lässt man sich den Dreivorschlag für diesen Satz in Version 2 ausgeben, wird man zunächst auf den (in Hinblick auf die Synopse) korrekten Satz 33 verwiesen („Den was ir pifhoff tot in der fstat“), erhält aber durch den zweiten Vorschlag aus der Liste mit Satz 74 („Den was ir pifhoff tot“) einen fast identen Textbaustein. Damit kommt in den Blick, dass die Formel offensichtlich in einer weiteren Episode an späterer Stelle erneut vorkommt.

Sieht man Textvarianz als einen Sonderfall von Intertextualität an, kommen hier also durch die bewusst eingesetzte Unschärfe des Algorithmus die Grenzen dieser Sonderkategorie ins Wanken. Gerade dadurch, dass der Algorithmus nur skalierbare Ähnlichkeitswerte, aber keine binären Entscheidungen darüber vorgibt, was Textvariante ist und was nicht, wird der qualitative Befund hinsichtlich der Textunterschiede eigentlich bestätigt: Die Varianz zwischen den Versionen ist teilweise so hoch, dass man sich an manchen Stellen fragen könnte, ob es sich noch um Varianten ein und desselben Textes oder schon um neuen Text handelt.

Zum zweiten Aspekt, der Kategorisierung und Visualisierung von Varianz: Wenn man Varianten in elektronischer Form gespeichert hat, liegt es nahe, diese auch für ein ‚Distant Reading‘ zugänglich zu machen, etwa in der Form, dass man über das ganze Legendar hinweg bestimmte Tendenzen der beiden Textversionen ausfindig

⁴⁴ Zur Bedeutung von Feedback bei der Entwicklung von WMD-Routinen für den Variantenvergleich vgl. auch Meinecke/Wrisley/Jänicke 2019 (Anm. 32). In unserer Implementierung wird dieses Feedback also durch die Verbindung von quantitativem und qualitativem Zugang in das System eingebracht.

macht. Dafür reicht es aber nicht aus, bloß Varianz zu verzeichnen, sondern sie muss auch semantisch ausgezeichnet und kategorisiert werden.⁴⁵

Im Rahmen des Editionsprojekts wird derzeit eine manuelle Kategorisierung von Textvarianz erprobt, die beispielsweise Abweichungen registriert, welche Personen, Orte oder Namen betreffen, oder sich auf sprachliche und stilistische Änderungen richtet, wie beispielsweise die Hinzufügung von Adjektiven, den Wechsel vom Aktiv zum Passiv, indirekte vs. direkte Rede sowie inhaltliche Unterschiede wie die Streichung beziehungsweise Einfügung ganzer Episoden.⁴⁶ Zum Teil könnte man auch hier zur Unterstützung der Klassifikation automatisierte Verfahren einsetzen, und der Umstand, dass mit dem RNN-Tagger-Modell fürs Mittelhochdeutsche nun ein brauchbares 'Preprocessing'-Tool vorliegt, das zum Beispiel auch Wortarten bestimmen kann, eröffnet neue Möglichkeiten, die es erst auszuschöpfen gilt.

Eine besondere Herausforderung stellt aber in diesem Zusammenhang erneut die Vermittlung und Darstellung von Varianz dar. Als Vorarbeit für die Edition wurde zu diesem Zweck in Zusammenarbeit mit dem Institut für Visualisierung der Universität Stuttgart ein Visualisierungstool entwickelt, dessen Oberfläche in Abbildung 7 als Screenshot wiedergegeben ist.⁴⁷

Das Tool bewegt sich zunächst auf der Makroebene, da zwischen der synoptischen Darstellung der Texte in zwei Spalten ein Mittelbalken als grafische Abstraktion eingesetzt ist, welche die Varianzverhältnisse über den Gesamttext einer Legende visualisiert.

Die von Hand vorgenommene Kategorisierung der Varianz wird mit unterschiedlicher Farbgebung wiedergegeben. Die Nutzerin oder der Nutzer erhalten

⁴⁵ Beispiele etwa bei Amir Hazem, Béatrice Daille, Dominique Stutzmann, Jacob Currie, Christine Jacquin: Towards Automatic Variant Analysis of Ancient Devotional Texts. In: Proceedings of the 1st International Workshop on Computational Approaches to Historical Language Change. Hrsg. von Nina Tahmasebi, Lars Borin, Adam Jatowt und Yang Xu. Florenz 2019, S. 240–249; Erik Ketzan, Christof Schöch: Classifying and Contextualizing Edits in Variants with Coletto: Three Versions of Andy Weir's *The Martian*. In: Digital Humanities Quarterly 15, 2021, H. 4, o.S., URL: <http://www.digitalhumanities.org/dhq/vol/15/4/000579/000579.html>; David Lassner, Anne Baillot, Sergej Dogadov, Klaus-Robert Müller, Shinichi Nakajima: Automatic Identification of Types of Alterations in Historical Manuscripts. In: Digital Humanities Quarterly 15, 2021, H. 2, o.S., URL: <https://arxiv.org/abs/2003.09136>; Thomas Bremer, Paul Molitor, Marcus Pöckelmann, Jörg Ritter, Susanne Schütz: Zum Einsatz digitaler Methoden bei der Erstellung und Nutzung genetischer Editionen gedruckter Texte mit verschiedenen Fassungen. Das Fallbeispiel der *Histoire philosophique des deux Indes* von Guillaume-Thomas Raynal. In: editio 29, 2015, S. 29–51.

⁴⁶ Die Kategorisierung ist damit auf die spezifischen Erfordernisse der Varianzanalyse im *Heiligen Leben* zugeschnitten und setzt auf frühere Versuche zur Variantentypologisierung auf; vgl. etwa Tara Andrews, Caroline Macé: Beyond the tree of texts. Building an empirical model of scribal variation through graph analysis of texts and stemmata. In: Literary and Linguistic Computing 28, 2013, H. 4, S. 504–521.

⁴⁷ Vgl. Martin Baumann, Markus John, Hermann Pflüger, Cornelia Herberichs, Gabriel Viehhauser, Wolfgang Knopki, Thomas Ertl: An Interactive Visualization for the Analysis of Annotated Text Variance in the Legendary *Der Heiligen Leben*, Redaktion. In: OSF Preprints 2020, DOI: 10.31219/osf.io/wd9yz.

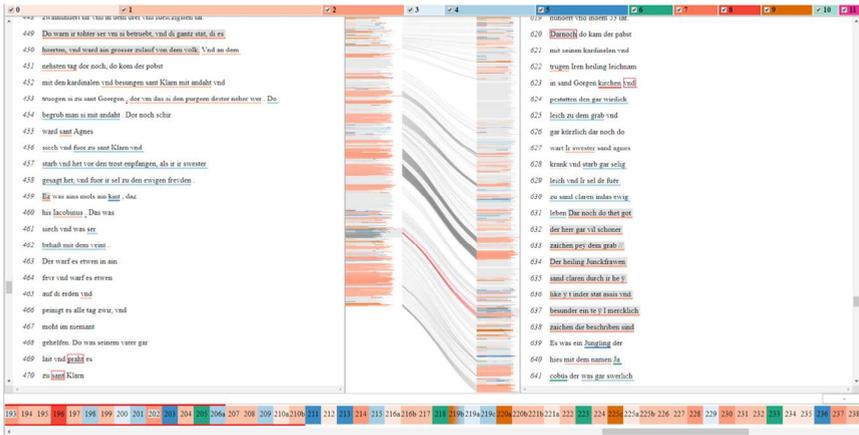


Abb. 7: Experimentelles Visualisierungstool für Fassungsvarianz am Beispiel der Clara-Legende nach HL-Red. I und HL-Red. II.

so einen schnellen Überblick, welche Arten von Varianz existieren und in welcher Quantität sie in den verschiedenen Abschnitten der Legende vorkommen.

Diese distanzierte Ansicht lässt sich dann in Hinblick auf die beiden Einzeltexte links und rechts des Mittelbalkens ins Detail verfeinern; so lassen sich die Einzelstellen der Varianz genauer unter die Lupe nehmen. Um gezielt und in übersichtlicher Form einzelne Kategorien zu untersuchen, können die Farbgebungen für einzelne Kategorien jeweils aus- und eingeschaltet werden.

Zum dritten Aspekt, der Verlängerung des Textvergleichs über das Legendar hinaus: Wie erwähnt, ist *Der Heiligen Leben, Redaktion* nur eines von vielen Legendaren des Mittelalters, zahlreiche in ihm enthaltene Heiligenviten sind überdies in anderen Versionen als Einzeltexte außerhalb von Legendaren überliefert. Nicht nur existieren folglich umfassende Parallelen zu anderen Legendaren- und Legendensammlungen, überdies weisen Legendareninhalte auch Verbindungen zu anderen Formen der Heiligenverehrung auf. Der Name Ambrosius referenziert ja auf einen spezifischen Heiligen, der noch dazu historisch eine realweltliche Referenz hat, es handelt sich bei ihm – wie oben schon erwähnt – um einen Bischof von Mailand aus dem 4. Jahrhundert.

Der Gedanke liegt daher nahe, für das Nachspüren dieser Verflechtungen die Kennzeichnungen von Entitäten in unserer Edition vorzunehmen. Das umfasst insbesondere, wie dies ja auch in anderen digitalen Editionen mittlerweile üblich ist, die Aufgabe, Personen und Ortsnamen zu annotieren und mit Normdaten zu versehen.⁴⁸ Unter diesen Normdaten sind auch Anknüpfungspunkte für Linked

⁴⁸ Vgl. Jörg Wettlaufer: Der nächste Schritt? Semantic Web und digitale Editionen. In: *Digitale Metamorphose: Digital Humanities und Editionswissenschaft*. Hrsg. von Roland S. Kamzelak und Timo

Open Data, etwa über den Wikidata-Identifizier.⁴⁹ Von hier aus kann man sich nicht nur unkompliziert Informationen zu realgeschichtlichen Daten über den Heiligen abrufen, sondern erhält auch einen Ausblick etwa auf ikonographische Zeugnisse und damit allgemein auf Praktiken der Heiligenverehrung.⁵⁰ Die Ausweitung des intertextuellen Bezugsrahmens, zu dem digitale Editionen ja tendieren, geht hier also über das bloße Medium Text hinaus.

Die gezeigten Beispiele sollten deutlich gemacht haben, dass gerade Legenden und Legendare sich nicht nur als Einzeltexte, sondern auch als Teil eines umfangreichen Textgeflechts verstehen lassen, zu dessen Erforschung dem Vergleich zwischen unterschiedlichen Textausprägungen besondere Bedeutung zukommt. Dieser Vergleich lässt sich im Extremfall sogar über den Text hinausführen, in den Interferenzbereich zwischen Texten, Bildern und Praktiken, welcher sich im digitalen Medium mit seiner Tendenz zur potentiell unbegrenzten Ausweitung der Kontexte weit besser darstellen lässt als etwa in der Buchform. Doch selbst wenn man die Ausweitung der Kontexte nicht so weit treiben möchte und die Grenzen der Betrachtung auf ein einzelnes Legendar wie etwa *Der Heiligen Leben, Redaktion* verengt, zeigt sich bald die Notwendigkeit, Varianz zwischen den Textversionen sprechend zu machen, da, wie man gesehen hat, der Text schon dieses Legendars in zwei unterschiedlichen Versionen vorliegt.

Um diese Varianz auszuwerten und damit auch interpretatorisch sowohl für die Literaturwissenschaft als auch für die Frömmigkeitsgeschichte nutzbar machen zu können, ist es nötig, sie möglichst übersichtlich in einer Edition zu vermitteln. Zu einer ansprechenden Vermittlung bedarf es unterschiedlicher Arbeitsschritte, von denen in diesem Beitrag ein Spektrum präsentiert wurde. Die Praktikabilität der angewandten Methoden wird sich gewiss im Prozess der Projektentwicklung erweisen müssen. Diese Herausforderung anzunehmen scheint im Hinblick auf die kulturelle Bedeutung einer so verbreiteten Buch- und Textform wie den Legendaren und auf das aktuelle Forschungsinteresse an dieser Gattung unerlässlich.⁵¹

Steyer. Wolfenbüttel 2018 (Sonderband der Zeitschrift für digitale Geisteswissenschaften. 2), o.S., DOI: 10.17175/sb002_007.

⁴⁹ Vgl. Georg Schelbert: ... warum nicht gleich Wikidata?! In: Digitale Nachhaltigkeit. 4. Tagung des Verbands Digital Humanities im deutschsprachigen Raum (DHD 2017), Bern 2017, o.S., URL: <https://doi.org/10.5281/zenodo.4622713>.

⁵⁰ Im Fall von Ambrosius etwa unter dem Link <https://www.wikidata.org/wiki/Q43689>, der aufgrund der Graphstruktur der Datenbank auch maschinell ausgewertet werden kann.

⁵¹ Vgl. zur aktuellen Legendarforschung u. a. Andreas Hammer: Erzählen vom Heiligen. Narrative Inszenierungsformen von Heiligkeit im *Passional*. Berlin, Boston 2015 (Literatur. Theorie. Geschichte. 10); Eva von Contzen: *The Scottish Legendary*. Towards a Poetics of Hagiographic Narration. Manchester 2016.

Abstract

Textual variance poses a particular challenge to editing medieval texts. Digital editions in particular can respond to these challenges in order to convey to the user the potential of variance and to enable an interpretation of medieval literary works that is close to their actual transmission. This paper presents some possibilities of dealing with and presenting variance through the example of the digital edition of a late-medieval legendary, the so-called *Der Heiligen Leben, Redaktion*. This collection of saints' legends offers a particularly good example of the value of considering textual variants, while also posing particular challenges for communicating variance.