

Entwicklung eines mittelhochdeutschen Sentiment-Wörterbuchs aus korpusermeneutischer Perspektive

Friedrich Michael Dimpel 

Eingegangen: 31. Oktober 2023 / Angenommen: 30. Dezember 2023
© The Author(s) 2024

Zusammenfassung Im ersten Teil werden Probleme und Lösungsansätze einer mediävistischen Korpusforschung skizziert. Ein zentrales Problem besteht immer noch in der unzureichenden Verfügbarkeit von digitalen Texten. Zwar gibt es mit der Mittelhochdeutschen Begriffsdatenbank und einigen vorbildlichen Editionsprojekten begrüßenswerte Ansätze, allerdings wäre bei öffentlich geförderten Editionsprojekten eine digitale Publikationspflicht ohne Schlupflöcher überfällig. Zudem hat die Mediävistik gegenüber der Korpusforschung zu neueren Sprachen noch weitere Nachteile, die zum Teil auf der nicht normierten mittelhochdeutschen Graphie beruhen und zum Teil darauf, dass es weniger bzw. weniger leistungsfähigere Tools und Ressourcen aus dem Bereich der automatischen Sprachverarbeitung für das Mittelhochdeutsche gibt.

Im zweiten Teil des Aufsatzes wird exemplarisch ein solches Desiderat aufgegriffen: Die Sentiment-Forschung. Auch hier hat die Mediävistik einen Rückstand aufzuholen, daher wird das erste mittelhochdeutsche Sentiment-Wörterbuch »SentiMhd« vorgestellt. Mit automatischen Verfahren zur Sentimentanalyse wird es möglich, große Korpora oder auch Textabschnitte hinsichtlich positiver oder negativer Stimmungen zu untersuchen. Diskutiert werden dabei Probleme, die sich bei mehrdeutigen, kontextabhängigen oder in Negation stehenden Lemmata ergeben, wenn verschiedene Annotator*innen den gleichen Text hinsichtlich seines Sentimentgehaltes auszeichnen. Solche Mehrdeutigkeiten führen eine automatische Analyse unter hermeneutischen Aspekten an ihre Grenzen. Neben der Evaluation von SentiMhd anhand eines manuell annotierten Korpus werden erste Analysen zu Hartmanns *Iwein* vorgestellt. Dank annotierter Figurenreferenzen können auch Sentimentwörter im Kontext von Figurenreferenzen erfasst werden. Das Sentiment-Modell deutet darauf hin, dass es im *Iwein* eher zwei Tiefpunkte gibt (Terminversäumnis und Lunetes

✉ Friedrich Michael Dimpel
GND: 1111656460, FAU Erlangen-Nürnberg, Erlangen, Deutschland
E-Mail: mail@dimpel.de

Einkerkerung), während die Forschung meist nur eine Krise zwischen dem ersten und zweiten Romanteil sieht. Nach einer Auswertung zu männlichen und weiblichen Hauptfiguren, Zofen und Opponent*innen schließt ein exemplarischer Makro-Blick auf verschiedene vier verschiedene Textsorten die Studie ab.

Schlüsselwörter Digitale Textanalyse · Sentiment-Analyse · Mediävistik · Korpusanalyse an mittelhochdeutschen Texten · Hermeneutik · Iwein

Development of a Middle High German Sentiment Dictionary from a Corpus Hermeneutic Perspective

Abstract The first part outlines problems and possible solutions for medieval corpus research. The inadequate availability of digital texts is still a central issue. Although the Mittelhochdeutsche Begriffsdatenbank and some exemplary edition projects are welcome approaches, a non-negotiable obligation to publish the relevant data digitally is overdue for publicly funded edition projects. In addition, medieval studies is at a disadvantage compared to corpus research on newer languages in several ways which are partly due to the non-standardized Middle High German writing and partly to the fact that there are fewer or less powerful tools and resources from the field of automatic language processing for Middle High German.

The second part of the article takes up one such desideratum as an example: Sentiment Analysis. Here, too, medieval studies has some catching up to do, which is why the first Middle High German sentiment dictionary »SentiMhd« is presented. Using automatic methods for sentiment analysis, it is possible to examine large corpora or even sections of text with regard to positive or negative sentiments. Problems are discussed that arise with ambiguous, context-dependent or negated lemmas when different annotators characterize the same text with regard to its sentiment content. Such ambiguities push automatic analysis to its limits in terms of hermeneutics. In addition to the evaluation of SentiMhd using a manually annotated corpus, initial analyses of Hartmann's *Iwein* will be presented. Thanks to annotated figure references, sentiment words can also be collected in the context of figure references. The sentiment model indicates that there are rather two low points in *Iwein* (missed appointment and Lunete's incarceration), whereas research usually sees only one crisis between the first and second part of the novel. After an evaluation of male and female main characters, maids and opponents, an exemplary macro-view of four different text types concludes the study.

Keywords Digital Text Analysis · Sentiment Analysis · Medieval Studies · Corpus Analysis on Middle High German Novels · Hermeneutics · Iwein

1 Mediävistische Korpusprobleme

Diese Studie nimmt Spezifika von korpushermeneutischen Herausforderungen bei mittelhochdeutschen Texten in den Blick. Neben den Problemen, die auch in den



neueren Philologien bestehen, weisen in der Mediävistik vier Problemfelder eine besondere Brisanz auf. Erstens wäre zu bedenken, welcher Text, welche Ausgabe oder gar welche Transkription welcher Handschrift in ein Korpus aufgenommen werden sollte, da oft eine erhebliche Varianz zwischen den Handschriften besteht. Zudem liegen mitunter viele Jahre zwischen der Textentstehung und der Überlieferung, in Extremfällen sogar Jahrhunderte: So sind der *Erec*, die *Kudrun* und der *Mauritius von Craun* nur im *Ambraser Heldenbuch* vollständig bzw. nahezu vollständig überliefert. Das *Ambraser Heldenbuch* ist am Anfang des 16. Jahrhunderts entstanden, doch die genannten Texte stammen aus dem ausgehenden 12. bzw. 13. Jahrhundert.

Auf dieses Problemfeld wird hier nicht weiter eingegangen, da digital arbeitende Mediävist*innen sich in der Regel glücklich schätzen dürfen, wenn überhaupt eine Textfassung auch digital verfügbar ist: Die unzureichende Verfügbarkeit digitaler Texte und Editionen – das zweite Problemfeld – stellt leider noch immer ein virulentes Hindernis für eine mediävistische Korpushermeneutik dar. Der dritte Problembereich liegt in der Varianz der nicht-normierten Graphie, der vierte in der eingeschränkten Verfügbarkeit von Tools und Ressourcen zur automatischen Textanalyse. Im zweiten Teil des Aufsatzes wird anhand der Sentimentanalyse exemplarisch vorgeführt, wie eine mediävistische Korpushermeneutik angesichts solcher Problemlagen dennoch Fortschritte erzielen kann.

1.1 Problemfeld: Unzureichende Verfügbarkeit digitaler Texte

Noch vor gut 20 Jahren war es ausgesprochen mühevoll, digitale Texte zu besorgen: Um das schmale Korpus für meine Dissertation (Dimpel 2004) zusammenzustellen, habe ich zahllose Stunden mit dem Scannen und Korrigieren von OCR-Fehlern zugebracht. Zwar gab es damals durchaus bereits viele digitale Texte – bspw. in den Wörterbuchprojekten in Trier und in der Mittelhochdeutschen Begriffsdatenbank in Salzburg. Doch frei zugänglich waren sie nicht, als Student*in oder Doktorand*in hat man sich nicht unbedingt getraut, um die Texte zu bitten. Viele Editionen wurden etwa mit Tustep (<https://www.tustep.uni-tuebingen.de>) digital angefertigt, doch nur in seltensten Ausnahmefällen wurden die Editionsdaten veröffentlicht. In den letzten Jahren wurde bei einigen Editionen eine Weboberfläche geschaffen, doch vielfach kann man darin die Texte nur abschnittsweise ansehen¹ und nicht vollständig herunterladen.

Mit der Zugänglichkeit ist es seitdem etwas besser geworden.² Allmählich spricht sich herum, dass die freundlichen Kolleg*innen von der Mittelhochdeutschen Begriffsdatenbank (MHDBDB, <http://mhdbdb.sbg.ac.at>) seit einigen Jahren die Texte, die in Salzburg vorhanden sind, gern für Forschungszwecke bereitstellen; in Arbeit ist derzeit ein groß angelegter Relaunch mit zahlreichen Download-Optionen (vgl.

¹ Dabei spielen selbst bei älteren Editionstexten offenbar immer noch Sorgen über die Urheberrechtslage eine Rolle. Teilweise muss man bei den Texten, die unter <http://www.mhdwb-online.de/quellenverzeichnis.php?buchstabe=A> verlinkt sind, für jede Zeile einen »Weiter«-Button betätigen.

² Ein Vorreiter bei digitalen Editionen ist das Berner Parzivalprojekt, das inzwischen über eine eigene Textexport-Seite (<https://parzival.unibe.ch/parzdb/index.php#Textexport>) verfügt. Als best-practice-Projekt kann das Projekt Lyrik des deutschen Mittelalters (<https://www.ldm-digital.de/>) gelten.

Zeppezauer-Wachauer 2022). Dieser Politikwechsel ist ein wichtiger Meilenstein für die mediävistische Literaturanalyse: Dank Katharina Zeppezauer-Wachauer, Alan van Beek und Peter Hinkelmans kann man erstmals von einer mediävistischen Korpusanalyse sprechen. Auch wenn allein das neugermanistische Prosa-Korpus im Textgrid-Repository (<https://textgridrep.org>) knapp fünfzimal so groß ist, sind in der MHDDB doch immerhin 666 Texte mit 9,3 Millionen Token vorhanden. Damit ist nun auch für mittelhochdeutsche Texte endlich eine Arbeitsgrundlage vorhanden, auch wenn hier nicht in allen Fällen die optimalen und aktuellsten Ausgaben zu finden sind.

Inzwischen hat die DFG eine digitale Publikation bei Editionen zur Soll-Bestimmung gemacht,³ auch wenn eine Muss-Bestimmung ohne Schlupflöcher angebracht wäre. Wenn Editionsdaten auch tatsächlich digital zugänglich wären, könnten digital arbeitende Literaturwissenschaftler*innen Ressourcen in ganz anderer Weise nutzen.

Dazu kommt ein weiteres Problem: Sonja Glauch hat darauf hingewiesen, dass Editionen nur so lange existieren, »wie sich jemand aktiv um die Softwarebausteine der Funktionsschicht kümmert, plus schätzungsweise fünf bis fünfzehn Jahre.« (Glauch 2022, S. 73) Viele Editionsprojekte verfügen über eigene Web-Oberflächen mit editionsspezifischen Zusatzfunktionen. Allerdings werden diese Oberflächen nach dem Auslaufen der Projekte nicht immer weiter gepflegt. Wenn etwa am Webserver ein Update bei der PHP-Version erfolgt, ist unter Umständen ein Zugriff auf die Edition nicht mehr möglich. Im schlimmsten Fall ist die digitale Edition für die Öffentlichkeit verloren.⁴

Auch wenn man sich lange Zeit allein durch den Einsatz von XML und TEI erhofft hat, dass solche Daten eine sehr lange Zeit nachnutzbar sind, mangelt es doch oft an der Dokumentation der Editionsstrukturen. Selbst für digital erfahrene User kann es Monate dauern, komplexe TEI-Daten nutzbar zu machen:⁵ Die Existenz von TEI-Daten allein gewährleistet nicht, dass eine Edition auch tatsächlich für primär

³ Vgl. die das Merkblatt »Förderkriterien für wissenschaftliche Editionen in der Literaturwissenschaft« (https://www.dfg.de/download/pdf/foerderung/grundlagen_dfg_foerderung/forschungsdaten/foerderkriterien_editionen_literaturwissenschaft.pdf): »Nicht für jedes Editionsprojekt wird eine digitale Veröffentlichungsform erwartet« (S. 2) »Sofern keine Veröffentlichung im Open Access geplant ist, bedarf dies der Begründung.« (S. 3) »Die Veröffentlichung als Druckausgabe kann sowohl in Ergänzung zu einer digitalen Veröffentlichung geschehen (so genannte Hybridedition) als auch die alleinige Publikationsform sein« (S. 4).

⁴ Vgl. Glauch 2022, S. 66: »Häufig müssen im Laufe der Jahre Daten und Ressourcen auf neue Server umgezogen werden; dabei werden oft interne Verzeichnisstrukturen umgestellt, was oft zu solchen Fehlern [404] führt.« Vgl. Glauch 2022, S. 67: »Wir haben oft auf der Serverseite eine Datenbank laufen mit ihrer Software sowie eine Skriptsprache wie PHP [...], dazu weitere Skriptsprachen auf der Clientseite wie JavaScript, XSLT [...]. Jedes dieser Elemente ist eventuell wiederum eingebettet in oder angebunden an ein Framework oder größere Komplettsysteme wie Bootstrap, WordPress, [...]. Dazu kommt der Einsatz fertiger Module für schwierige Aufgaben wie einen Bildviewer, eine Textsynopse usw. Jeder dieser softwaretechnischen Bausteine stellt eine Abhängigkeit dar, jedes dieser Systeme kann theoretisch außer Betrieb gehen oder von seinen Entwicklern aufgegeben werden.«

⁵ Diese Erfahrung durfte ich durchleben, als ich eine Anastasia-Oberfläche für die damaligen Parzival-Projekt-Daten erstellt habe; ein Bericht dazu steht in Dimpel and Stolz 2006.

philologisch orientierte Forscher*innen nutzbar ist, selbst wenn sie als Download verfügbar sein sollten.⁶

1.2 Problemfeld: Nicht-normierte Graphie im Mittelhochdeutschen

Das mittelhochdeutsche Wort *juncvrouwen* findet man in vielen graphischen Varianten vor – bereits mein Normalisierungswörterbuch⁷ enthält derzeit 68 Varianten:

iucvrowen iuncfrauwen iuncfrawen iuncfrouvon iuncfrouwen iuncfrovgen
iuncfrovwen iuncfrowen iunchfrawen iunchfrovwen iunchfrowen iunchu-
rovwen iunchvrovwen iuncurowen iuncurovwen iuncurowen iuncvrouwen
iuncvrouwyn iuncvrovwen iuncvrowen iuncvrowyn iuncwrowen iuncwrow-
wen iunfrowen iung vrowen iungfrawen iungfrovwen iungfrowen iungurowen
iungvrowen iunk frouuon iunkfrouuon iunkfrouvon iunwrowen iuoncfrouwen
iuoncfrouwen iuoncfroen ivncfrauwen ivncfrouwen ivncfrovwen ivncfrov
ivncfrovwen ivncfrowen ivnchfrovwen ivnchfrovwen ivnchfrowen ivnchfrow
ivnchvrovwen ivncvrouwen ivncvrowen ivnkfrouuon ivnuronwin ivnwrowin
jucfrauwen juncfroiwen juncfrouwen juncfrovwen junckfrauen junckfrauwen
juncvrouwen juncvrowin jungfrowen jungkfrawen junkfrawen junkfrouwen
juoncfrauowen jvncfrawen uinc frowen

Diese Varianz könnte dazu führen, dass die Verteilung der häufigsten Wörter im Korpus eine völlig andere ist als im Neuhochdeutschen. Bekanntlich machen recht wenige sehr häufige Types einen Großteil der Token aus. Nach dem Häufigkeitswörterbuch von Kaeding (1897), das auf 11 Millionen Token aus Texten um 1900 beruht, decken bereits 207 Types 54,4 % aller Wortformen im Korpus ab (König 2004, S. 114f.; vgl. auch Kaeding 1897). Angesichts der vogelwildenen Schreibung bei *juncvrouwen* läge die Vermutung nahe, dass im Mittelhochdeutschen die häufigsten 207 Types eine niedrigere Prozentzahl der Token abdecken. Eine rasche Analyse der gesamten MHDBDB-Texte zeigt allerdings, dass die häufigsten 207 Types auch hier 53,9 % der Token ausmachen. Aus dem Textgrid-Prosakorpus habe ich via Zufallsgenerator ein neuhochdeutsches Korpus mit gleichem Umfang erstellt: Dort entsprechen 207 Types 55,1 % der Token. Auch sonst sind die Abweichungen bei den häufigeren Wortformen gering. Bei den Hapax legomena (nur einmal vorkommenden Wortformen) gibt es größere Unterschiede. In der MHDBDB sind unter Hapax legomena nur wenige Lemmata. Ganz überwiegend handelt es sich um graphische Varianten von Wortformen, die bereits weiter oben in der Häufigkeitsliste vorkommen (Tabelle 1).

Für diese Ähnlichkeit könnten drei Aspekte wichtig sein: Erstens liegt ein Teil der Texte in der MHDBDB in normalisierten Ausgaben vor. Zweitens findet man auch im Textgrid-Auszug Abweichungen von einer standardisierten Graphie, da sich dort

⁶ Wichtig wäre daher, dass künftige Editionen, die mit öffentlichen Mitteln finanziert werden, auch digital auf einem Server, dessen Betreuung dauerhaft gesichert ist, publiziert werden. Je nach Art der Edition wären Basistext oder Transkriptionen in mehreren zukunftssicheren Formaten (pdf, xml und html) zugänglich zu machen: Dadurch wäre im Fall einer Funktionsstörung der Projekt-Weboberfläche nicht alles verloren.

⁷ Vgl. Dimpel 2017, S. 100f. In den Handschriften dürfte es noch weit mehr Varianten geben.

Tab. 1 Type-Token-Verteilungen in der MhdBDB und einem gleichlangen Textgrid-Auszug

Aus je 9,3 Millionen Token	MhdBDB	Textgrid-Auszug
Häufigste 10 Types	17,3 %	18,0 %
Häufigste 50 Types	37,6 %	38,5 %
Häufigste 100 Types	46,1 %	47,0 %
Häufigste 1.000 Types	70,1 %	70,3 %
Häufigste 10.000 Types	89,4 %	88,4 %
Hapax legomena (absolut)	133.073	124.687
Types gesamt	253.616	248.764

Texte vom Beginn des Buchdrucks bis ins 20. Jahrhundert finden.⁸ Drittens sind in der Entwicklung zum Neuhochdeutschen hin deutlich weniger mittelhochdeutsche Wörter verschwunden, als neue Wörter dazugekommen sind.

Betrachtet man nur die 10.000 häufigsten Types, findet man deutlich weniger Graphie-Varianten – zu *juncvrouwen* etwa nur 6 Varianten statt 68 Varianten im Normalisierungswörterbuch, bei *vrouwe* bleiben immerhin 11 Varianten statt 63. Dass bei den häufigeren Wortformen die mittelhochdeutsche Vokabular-Verteilung nicht grundlegend andersartig ist, könnte ein Grund dafür sein, dass Analyseverfahren wie etwa Burrows' Delta, mit dem man Texte etwa nach Autorschaft kategorisieren kann, auch im Mittelhochdeutschen funktionieren (Dimpel 2018).

Wenn man zwei Texte mit statistischen Methoden vergleicht, von denen im ersten Text die Wortform *juncvrouwen* 20 Mal vorkommt, im zweiten Text fehlt diese Wortform jedoch, dafür kommt je 10 Mal *iucvrouwen* und *iuncfrauwen* vor, wird ein Vergleichstool, das einen Zeichen-für-Zeichen-Vergleich durchführt, hinsichtlich dieses Parameters eine Abweichung statt einer Identität feststellen, auch wenn diese Unterschiede womöglich nur auf den Präferenzen der Abschreiber beruhen. Das erschwert nicht nur Autorschaftsstudien, sondern auch Studien zur literarischen Analyse, wenn beispielsweise danach gefragt wird, ob im Kontext von *juncvrouwen* besonders häufig bestimmte axiologisch relevante Vokabeln wie *werdekeit*, *kluocheit*, *triuwe* oder überhaupt semantisch interessantes Vokabular beispielsweise im Kontext von Figurenreferenzen analysiert werden soll.

Um solche Probleme abzumildern, habe ich ein Normalisierungswörterbuch zusammengestellt. Es beruht auf manueller Zuordnung von Most-Frequent-Words durch studentische Hilfskräfte, auf Daten des Erlanger Lyrik-Projektes (<https://www.ldm-digital.de>) und auf Daten des Referenzkorpus Mittelhochdeutsch (<https://www.linguistics.rub.de/remf>). Problematisch ist, dass das Referenzkorpus Mittelhochdeutsch anders normalisiert, als es sonst bei im Standardwörterbuch von Matthias Lexer oder in der Standardgrammatik Paul/Wiehl/Grosse üblich ist, so dass sich im Normalisierungswörterbuch zahlreiche normalisierte Formen finden, die man intuitiv nicht als konventionell normalisiertes Mittelhochdeutsch betrachten würde.⁹

⁸ Zudem finden sich hier viele originelle Komposita (»verliebtseyne«, »blondgezöpft«, »zärtlichkeits-schuld«).

⁹ Sehr häufig »b«, »g« oder »d« statt »p«, »c« oder »t« im Auslaut oder bei Komposita; vor Suffixen auch im Inlaut. Oft »æ« statt »e«, »tz« statt »z«, »v« statt »f«, »-heit« statt »-keit«-Suffixe, Vokaländerungen wie »vrome« statt »vrume«, u. v. m.

Es gibt viele ambige Wortform-Varianten, die in einem Kontext zu Lemma A und in anderem Kontext zu Lemma B oder C gehören; im Normalisierungswörterbuch wird jedoch nur eine Normalform zugeordnet; angestrebt wurde, die häufigste Normalform zuzuordnen. Damit ist das Normalisierungswörterbuch vollständig ungeeignet etwa zur Normalisierung von Editionen. Mit seiner Hilfe kann aber dort die Erfolgsquote enorm verbessert werden, wo es darum geht, zu statistischen Zwecken Daten zu vergleichen – wie bei Studien zur Autorschaftsattribuion (vgl. Büttner et al. 2017, S. 10–21). Immerhin können wir inzwischen vorrechnen, dass Konrad von Würzburg mit hoher Wahrscheinlichkeit der Autor der ›Halben Birne‹ ist (Dimpel et al. 2019) und dass die Nürnberger Weingröße mit hoher Wahrscheinlichkeit Rosenplüts Feder entstammen (Dimpel/Wagner 2022).

Trotzdem muss man dabei weiterhin mit vielen ungenauen Zuordnungen leben. In einer Welt, die aus Sicht der digitalen Literaturanalyse eine ideale Welt wäre, würde sich die Altgermanistik auf ein einheitliches Set an Normalisierungsregeln einigen und konsequent bei allen Editionen zumindest zusätzlich zu den Transkriptionen und anderen Textfassungen auch eine digitale Fassung in einer einheitlichen Standardgraphie anbieten. Hier wäre dann auch eine kontextsensitive Disambiguierung von nicht-normalisierten Wortformen vorhanden, die über das Normalisierungswörterbuch nicht realisiert wird.

1.3 Problemfeld: Leistungsschwächere Tools und weniger Ressourcen

Weitere Probleme bringen fehlende oder leistungsschwächere Tools und Ressourcen mit sich. Ohne weitere Voraussetzungen ist es möglich, Studien auf Vokabularbasis vorzunehmen. Dazu gehören neben stylometrischen Zugriffen (vgl. etwa Kestemont 2018; Viehhauser 2015) auch Studien zum Topic Modelling (Viehhauser 2017) oder Studien zur Netzwerkanalyse auf Basis des nicht weiterverarbeiteten Wortmaterials.¹⁰ Neugermanisten*innen oder Computerlinguist*innen allerdings verfügen schon länger über Part-of-Speech-Tagger und über Tools zur Erkennung von Figurenreferenzen,¹¹ so dass Studien zur digitalen Literaturanalyse auch mit solchen Daten operieren können. Mithin ist die Zahl der vorhandenen Tools für Sprachanalyse sehr viel größer; dort, wo es für das Mittelhochdeutsche ähnliche Tools gibt, erreichen die Tools beim Mittelhochdeutschen nicht die gleichen Erfolgsquoten wie bei neueren Sprachen. Eine automatische Erkennung von Referenzen auf Figuren schien lange außer Reichweite zu sein, inzwischen gibt auf diesem Gebiet jedoch Erfolge in Stuttgart.¹²

Die mittelhochdeutschen Wörterbücher in Trier sind seit gut 20 Jahren online. Neu ist, dass man über eine Web-Schnittstelle Abfragen vornehmen kann (Burch

¹⁰ Viehhauser-Mery 2020. Dramen haben bei der literarischen Netzwerkanalyse den Vorteil, dass meist Sprecherangaben vorhanden sind, während Figurenreferenzen in Romanen erst annotiert werden müssen.

¹¹ Krug 2020 berichtet über die Erkennung von Eigennamen mit einem Erfolg von $F1=0,93$; die Koreferenzauflösung inklusive Pronomina gelingt im literarischen Korpus DROC mit $F1=0,55$ (S. iii). Vgl. auch Puppe et al. 2015 sowie Krug et al. 2019.

¹² Ketschik et al. 2020, S. 212–216, nennen für Figurenreferenzen samt Appellativa, jedoch ohne Pronomina einen $F1$ -Wert von 0,32 für Wolframs *Parzival*.



2022, S. 103–106). Eigentlich würde man sich wünschen, dass die vollständigen XML-Daten der Wörterbücher veröffentlicht werden, so dass man diese Daten ohne Umwege und ohne Einschränkungen in eigene Analyse-Pipelines integrieren kann.

Eine automatische Erkennung von Part-of-Speech-Daten schien lange ein Traum zu sein. Nun gibt es erste Mittelhochdeutsch-Modelle für den TreeTagger in Stuttgart (Echelmeyer et al. 2017), für den Spacy-Tagger ein Modell von Peter Hinkelmanns (Hinkelmanns 2021) und ein Modell von Helmut Schmid für den RNN-Tagger, das bei Wortart 96%, bei der Morphologie 89,5% und bei der Lemmatisierung 91,9% erreicht.¹³ Auch wenn bei den Tools und Ressourcen weiterhin ein enormer Vorsprung bei den neueren Sprachen besteht, der auch damit zusammenhängen könnte, dass Informatiker*innen und Computerlinguist*innen oft zu neueren Sprachstufen einen engeren Bezug haben als zu älteren Sprachstufen, stehen inzwischen doch zumindest einige wenige der ganz elementaren Tools zur Verfügung.

2 Das Sentimentwörterbuch SentiMhd

Als exemplarische Anwendung korpushermeneutischer Modellierungen stelle ich im zweiten Teil dieses Beitrags das Sentimentwörterbuch SentiMhd vor – soweit ich weiß, handelt es sich um das erste Sentimentwörterbuch für das Mittelhochdeutsche.

Mit SentiMhd wird es möglich, auch in großen Korpora, bei denen das »too-big-to-read«-Argument¹⁴ relevant wird, Vokabular zu finden und auszuwerten, das als positiv oder negativ eingestuft wird – es geht also um Wertungen und Emotionen. In der Sentiment-Forschung zu neueren Sprachen gibt es zahlreiche Studien; neben wörterbuchbasierten Ansätzen wird dort mit maschinellem Lernen, neuronalen Netzen und mit Word-Embeddings gearbeitet. Einen guten Forschungsüberblick geben Kim und Klinger.¹⁵ Für das Mittelhochdeutsche steht sinnvollerweise ein wörterbuchbasierter Ansatz am Anfang, weil einerseits zunächst erprobt werden muss, ob eine Sentimentanalyse hier überhaupt erfolgsversprechend ist und andererseits haben wir hier den Nachteil, dass für das Trainieren von Sprachmodellen etwa für Word-Embeddings kein vergleichbar großes mittelhochdeutsches Korpus verfügbar ist.

2.1 Annotation, Annotationsregeln und Inter-Annotator-Agreement

Ob ein Wort oder ein Satz positiv oder negativ zu bewerten ist, ist eine Frage der Wahrnehmung, also subjektiv. Eine Person mag das Wort »Stolz« als positiv, eine zweite als neutral und eine dritte als negativ einordnen; vor allem aber wird die Einordnung kontextabhängig unterschiedlich ausfallen. Zunächst soll daher über-

¹³ Schmid 2019. Die Erfolgsquoten hat Helmut Schmid 2022 bei einem Vortrag in Erlangen genannt.

¹⁴ Vgl. hierzu etwa Glawion 2023, S 31 f.

¹⁵ Kim and Klinger 2019. Dennerlein et al. 2022, S. 194, berichten über eine Erfolgsquote bei Polarität von 90% mit einem gbert-Modell; vgl. weiterhin Dennerlein et al. 2023. Für das Neuhochdeutsche steht mit SentText sogar ein Tool mit grafischer Benutzeroberfläche bereit (Schmidt et al. 2021).

prüft werden, inwieweit zwei Personen unabhängig zu einer gleichen Einstufung gelangen.

Dazu habe ich ein Korpus aus wertungsreichen Passagen aus *Nibelungenlied*, *Erec*, *Iwein*, *Tristan* und *Gregorius* mit 3.860 Wortformen und 580 Versen zusammengestellt. Die Annotationsregeln besagen, dass nur bei klarer positiver oder negativer Orientierung annotiert werden soll. Entscheidend ist die Wortbedeutung im aktuellen Kontext. Allerdings sollte auch bedacht werden, ob diese Wortbedeutung nur hier singular positiv oder negativ ist; im Fall einer exceptionell seltenen Bedeutung sollte das Wort eher neutral (ohne Annotation) bleiben.

Probleme entstehen etwa durch Negation oder Konjunktiv-2: Wortformen, die in anderem Kontext häufiger eine gegenteilige Orientierung aufweisen dürften, werden mit einem Polaritätsumkehr-Tag bedacht – ein Beispiel: *niemen was ir gram*+#Negation. Hier bedeutet das »+«, dass es um eine positive Aussage geht, »#Negation« bedeutet jedoch, dass *gram* andernorts negativ ist und hier aufgrund der Negation positiv wird. Bei einem Satz wie »Nun endet unser Leid« würde »Leid« mit »+#Sonstiges« annotiert, da die Polarität durch »endet« verändert wird.

Bestimmte Wortgruppen (*vrouwe*, *ritter*, *küene*, *grôz*, ...) werden nicht annotiert, Kampf-Vokabular ebenfalls nicht – wenn der Held etwa zuschlägt, ist es etwas Anderes, als wenn ein Opponent zuschlägt, zudem gibt es auch böse Damen oder Könige. Bei Polyvalenzverdacht oder Polysemie wird eher nicht annotiert – *wirde* kann oft auch als Flexionsform zu *werden* gebraucht werden und wäre dann neutral, während *wirde* zu neuhochdeutsch »Würde« positiv einzustufen wäre. Im Zweifelsfall sollen falsche Wertattributionen vermieden werden.

Annotiert wurde das Evaluierungskorpus von zwei promovierten Altgermanisten mit über 30 Semestern Lehrererfahrung.¹⁶ Einige eindeutige Verstöße gegen die Annotationsregeln und Flüchtigkeitsfehler wurden berichtigt (etwa vergessene Polaritätsumkehr-Tags). Die Annotatoren haben 380 bzw. 363 positive und 266 bzw. 274 negative Tags vergeben.

Sichtet man die Abweichungen zwischen beiden Annotatoren, zeigen sich Differenzen, die auf unterschiedlichen Interpretationen beruhen. Hier einige Beispiele aus dem *Erec*, bei denen das fragliche Wort von einem Annotator als negativ und vom anderen Annotator als neutral eingestuft wurde: 2933 *wante zuo gemache*; 2946 *diz [der Kirchgang] was sîn meistiu arbeit*; 2984 *ein wandelunge an im geschach*. Dass es zu solchen Abweichungen kommt, ist recht verständlich: Ist *gemache* (nhd. »Bequemlichkeit«) nur im *Erec* negativ, sonst aber eher nicht? *arbeit* heißt Mühe oder Qual, aber ist das hier negativ, gerade dann, wenn man den Vers als ironisch auffasst? Ist eine *wandelunge* (nhd. »Änderung«) auch per se etwas Schlechtes? Bei diesen Beispielen treten aufgrund der mehrdeutigen bzw. interpretationsbedürftigen Textinformationen hermeneutische Differenzen der Annotatoren zu Tage, die zugleich eine Grenze für automatische Annotationen markieren dürfte.

Gemessen wird, inwieweit die Annotatoren übereinstimmende Ergebnisse erzielen: das Inter-Annotator-Agreement. Dafür wird wie üblich Cohens Kappa verwendet, bei dem der Wert 1 eine vollständige Übereinstimmung anzeigt – ein kaum

¹⁶ Eine breite Lektüererfahrung mit mittelalterlichen Texten ist wichtig, um etwa beurteilen zu können, in welcher Bedeutung ein Wort überwiegend gebraucht werden dürfte.

erreichbarer Wert, wenn es um Annotationen zu interpretationsabhängigen Phänomenen geht. Kappa erreicht jedoch den recht guten Wert 0,84.¹⁷

2.2 Wörterbuch-Experimente

Das Vorbild für meine Wörterbuch-Experimente ist das Leipziger SentiWS von Robert Remus, Uwe Quasthoff und Gerhard Heyer, die ein englisches Sentiment-Wörterbuch automatisch ins Neuhochdeutsche übersetzt und grundlegend manuell nachbearbeitet haben.¹⁸ Mit Hilfe der PMI-Methode (dazu unten mehr) erhalten die Wörterbucheinträge dort je nach ihrer Nähe zu positiven oder negativen Leitwörtern wie »gut« oder »böse« einen positiven oder negativen Score.

Um SentiMhd zu erstellen, habe ich vier Experimente durchgeführt: Den Versuch einer automatischen Übersetzung mit Nachbearbeitung, zweitens eine manuelle Wörterbucherstellung und schließlich zwei verschiedenen große Wörterbücher, die auf der PMI-Methode beruhen.

2.2.1 Experiment 1: Automatische Übertragung von SentiWS mittels Trierer BMZ-API

Auch wenn man bei automatischen Übersetzungsversuchen skeptisch sein darf, wollte ich die neue Web-Schnittstelle (API) der Trierer Wörterbücher (vgl. Burch 2022, S. 103–106) testen und versuchsweise das neuhochdeutsche SentiWS ins Mittelhochdeutsche (Mhd.) übersetzen. Über die Schnittstelle zu dem mhd. Wörterbuch von Benecke/Müller/Zarncke (<http://woerterbuchnetz.de/BMZ>) wurden Angaben zu Bedeutungen und Wortarten heruntergeladen und auf Übereinstimmung mit dem neuhochdeutschen SentiWS überprüft. Vorläufig berücksichtigt wurde das mhd. Lemma dann, wenn ein Wort aus SentiWS in den Bedeutungsangaben zu diesem Lemma gefunden wurde.

Die manuelle Nachbearbeitung hat sich als zeitaufwendig erwiesen – auch, weil in dem Wörterbuch aus dem 19. Jhd. die Graphie abweicht, etwa bei »Sanftmuth«. Deshalb habe ich sämtliche Types in allen Bedeutungsangaben mit einem Rechtschreibprogramm geprüft und eine Korrekturliste erstellt. Probleme gibt es u. a. dadurch, dass in den Wörterbuch-Bedeutungsangaben umschreibende Formulierungen oder gar Erläuterungen durch Antonyme stehen, so dass auch falsche Zuordnungen

¹⁷ Wenn man die Polaritätsumkehr-Tags ausblendet, also Fälle, in denen die Annotation etwa durch Negation oder Konjunktiv 2 komplexer wird, erreicht Kappa sogar 0,88.

Da die meisten Wortformen ohne positive oder negative Annotation bleiben, wurden für die Berechnung von Kappa nicht sämtliche Wortformen verwendet, andernfalls würde man unrealistisch hohe Kappa-Daten erhalten. Beispielsweise bei Funktionswörtern ist es nicht relevant, abzuwägen, ob das Wort positiv oder negativ sein könnte. Da jedoch bei anderen Wortformen die Einstufung als »neutral« mitunter nicht nur eine Formsache, sondern tatsächlich auf einer echten Entscheidung beruht, die für das Inter-Annotator-Agreement relevant ist, wurde festgelegt, dass pro Vers 1,5 Entscheidungen hinsichtlich des Sentiment-Gehalts notwendig sind. Bei 580 Versen hat Annotator 1 insgesamt 646 positive oder negative Tags vergeben. Die Setzung »1,5 Entscheidungen pro Vers« bedeutet, dass ungefähr nach zwei Einstufungen einer Wortform als positiv oder als negativ bei einer weiteren Wortform tatsächlich eine aktive Entscheidung zugunsten einer neutralen anstelle einer positiven oder negativen Einstufung anfällt.

¹⁸ Remus et al. 2010. Das Wörterbuch ist unter <https://wortschatz.uni-leipzig.de/de/download> verfügbar.

generiert werden.¹⁹ Die Ergebnislisten werden deshalb mit automatischen Verfahren (etwa: Ist ein Lemma zugleich als positiv und negativ eingestuft?) und manuell geprüft. Insgesamt werden deutlich größere Listen erzeugt als im neuhochdeutschen SentiWS. Manuell wird mit einer zurückhaltenden policy aussortiert; das bedeutet, es werden nur Lemmata entfernt, die eindeutig als nicht positiv oder als nicht negativ eingestuft werden. Übrig bleiben in Experiment 1 immerhin 2.663 positive und 2.924 negative Lemmata.

2.2.2 Experiment 2: ›Manuelles Wörterbuch‹

Für das manuelle Wörterbuch habe ich alle Types aus der gesamten MHDBDB mit einer Häufigkeit von >20 (etwas über 20.000 Types) manuell gesichtet und ggf. mit einer positiven oder negativen Zuordnung annotiert. Types, die sich durch mein Normalisierungswörterbuch oder durch mein Vollformenwörterbuch²⁰ auf eine Normal- bzw. Grundform zurückführen lassen, werden aus ökonomischen Gründen gruppiert.²¹ Hier wird konkordant zu den Annotationsregeln bei der manuellen Annotation eine strenge policy verfolgt: es werden also nur Wortformen gewählt, die ich klar als positiv oder negativ einschätze. Es verbleiben 1.705 Einträge zzgl. Flexionsformen (1.095 positive / 610 negative).

2.2.3 Experiment 3: ›Kleines PMI-Wörterbuch‹

Wie beim neuhochdeutschen SentiWS²² wird die Pointwise-Mutual-Information-Methode (PMI) verwendet, bei der die Semantische Orientierung (SO) einer Wortform im Kontext von Leitwörtern (wie »gut«, »schön«, »schlecht«, »böse«) berechnet wird. Ein Wort erhält dann einen höheren positiven SO-Wert, wenn es häufig kurz vor oder kurz nach einem positiven Leitwort auftritt. Wenn es häufig kurz vor oder kurz nach einem negativen Leitwort auftritt, erhält es einen negativen SO-Wert.²³

¹⁹ Ein Beispiel: Zu *âmaht* steht in der Bedeutungsangabe »mangel an kraft«.

²⁰ Dimpel 2017, S. 100f. Für die Studien zur Autorschaftsattribuion hatten sich bessere Erkennungsquoten ergeben, wenn Vokale mit Zirkumflex durch den reinen Vokal ersetzt und Ligaturen aufgelöst wurden. Im Nachhinein war es ein Fehler, im Normalisierungswörterbuch die ursprünglichen Formen nicht zumindest zusätzlich noch beizubehalten. Auf Grundlage der Formen in der MhdBDB wurde versucht, nachträglich die normalisierte Form im Normalisierungswörterbuch automatisch regelbasiert wieder mit Zirkumflex bzw. Ligatur zu versehen. Unter anderem dadurch, dass zu flektierten nicht-normalisierten Wortformen, die in der MhdBDB häufiger als 20x vorkommen, normalisierte flektierte Formen ermittelt wurden, stehen auch Wortformen im manuellen Wörterbuch, die in der MhdBDB weniger häufig vorkommen. Zudem werden manche Wortformen mehrfach erfasst, wenn etwa eine Form als Variante oder Flexionsform berücksichtigt wird, die Variante oder Flexionsform jedoch zugleich auch in den 20.000 häufigsten MFWs vorkommt.

Teilweise ergeben sich durch den Abgleich mit Vollformenwörterbuch Absurditäten, weil die Zuordnungen nicht häufigkeitsbasiert gewichtet sind – bspw. bei *nehen*>*nicht* oder *diuhen*>*dâhte*. Solche Probleme wurden bei der manuellen Sichtung aussortiert.

²¹ Wenn sich eine normalisierte Form oder eine nicht-flektierte Grundform ermitteln hat lassen, wurden die flektierten oder nicht-normalisierten Formen in der gleichen Zeile der Grundform beibehalten.

²² Remus et al. 2010, S. 1169f. Vgl. auch Flüh 2019, Abs. 30.

²³ Diese Werte werden auf Werte zwischen -1 und +1 skaliert. Berechnet werden SO-Werte für die Wortformen des manuellen Wörterbuchs, wenn es in der Mittelhochdeutschen Begriffsdatenbank mehr als 20

In der Regel liefert die PMI-Methode plausible Ergebnisse, auch wenn es vereinzelt Überraschungen gibt: *freude* wird als leicht negativ eingestuft, andere Schreib- und Flexionsformen wie *vreude* sind aber wie erwartet positiv. Wortformen, die im manuellen Wörterbuch gegenpolig annotiert sind, werden nur bei einer deutlich problematischen Zuordnung aussortiert oder, wie bei *freude*, umgepolt und der SO-Wert wird auf 0,1 (bzw. ggf. auf -0,1) korrigiert. Berechnet werden hier die SO-Werte nur für die Einträge im manuellen Wörterbuch. Das kleine PMI-Wörterbuch ist also eine Teilmenge des manuellen Wörterbuchs. Enthalten sind hier 434 positive und 195 negative Einträge.

2.2.4 Experiment 4: ›Großes PMI-Wörterbuch‹

Hier werden zusätzlich auch SO-Werte für die 20.000 häufigsten Types der Mittelhochdeutschen Begriffsdatenbank einbezogen. Das PMI-Verfahren kann also mit einem kleinen Set, das auf manueller Selektion basiert (Experiment 3), und mit einem großen Set (Experiment 4) in zwei Konstellationen evaluiert werden. Manuell aussortiert habe ich hier wie auch schon bei Experiment 3 mit einer zurückhaltenden policy – ich sortiere nur klar falsche oder klar neutrale Wortformen aus. Es bleiben 1.446 positive und 597 negative Einträge.²⁴

2.3 Evaluation der Wörterbuch-Experimente

Das Evaluationskorpus wird automatisch auf Basis der Wörterbücher annotiert und mit den manuellen Annotationen verglichen. Für alle vier Sentiment-Wörterbücher wird die mittlere Übereinstimmung zu den beiden menschlichen Annotatoren (A1, A2) berechnet. Das übliche Maß hierfür ist der F1-Wert, der Precision und Recall kombiniert. Ein F1-Wert von 1 zeigt eine vollständige Übereinstimmung an, der Wert 0 zeigt eine vollständig fehlende Übereinstimmung an. Die rechte Spalte zeigt den Mittelwert (MW) beider F1-Werte.

Bei der automatischen Annotation erfolgt der Abgleich mit dem Wörterbuch nicht nur über die (ggf. flektierte) Wortform im Text, sondern auch über das Lemma, das vom Mittelhochdeutsch-Modell zum RNN-Tagger von Helmut Schmid (Schmid 2019) ausgegeben wird (Spalte 2), so dass sowohl nicht-normalisierte Formen als auch Flexionsformen zugeordnet werden können. In Tabelle 2 kann man etwa bei

Kookkurrenzen gibt. Als Kookkurrenz-Bereich zwischen Leitwort und der fraglichen Wortform werden fünf Worte vor und nach dem Leitwort gewählt. Wortformen, die nahe am neutralen Wert »0« liegen, werden aussortiert, wenn der Betrag des SO-Wertes $\leq 0,015$ ist. Dieser Parameter hat sich als geeignet erwiesen: Wortformen, deren SO-Wert zu nahe am Nullpunkt liegt, weisen oft keine klare Polarität auf. Setzt man den Parameter deutlich höher, bleiben relativ wenige Einträge übrig und es werden auch zahlreiche Wortformen aussortiert, die nach meiner Einschätzung durchaus eine klare Polarität tragen.

²⁴ Nach diesen Experimenten werden noch Konsistenzprüfungen durchgeführt, bei denen auch für das manuelle Wörterbuch und für das Übersetzungswörterbuch SO-Werte berechnet werden, anhand derer beide Wörterbücher erneut auf Plausibilität überprüft werden. Zudem prüfe ich die Wörterbücher untereinander auf abweichende Polaritäten. Vereinzelt habe ich dabei noch Korrekturen und Ergänzungen im manuellen Wörterbuch vorgenommen. Da sich die manuelle Einstufung in der Evaluation als verlässlicher als der PMI-Score erweist, werden in der automatischen Textannotation in den Folgeabschnitten nur binäre Werte (positiv/negativ) eingetragen und nicht die graduell abgestuften PMI-Scores.

Tab. 2 Evaluation: F1-Werte der wörterbuchbasierten Annotation

	RNN- Abgleich	Generierte Flexion	F1 (A1)	F1 (A2)	F1 (MW)
Übersetzungs-WB	nein	nein	0,31	0,31	0,31
Übersetzungs-WB	nein	ja	0,31	0,31	0,31
Übersetzungs-WB	ja	nein	0,32	0,31	0,32
ÜB nur ISOI>0,33	ja	nein	0,34	0,34	0,34
Manuelles WB	nein	nein	0,72	0,7	0,71
Manuelles WB	ja	nein	0,73	0,71	0,72
Manuelles WB	nein	ja	0,72	0,7	0,71
Manuelles WB	ja	ja	0,73	0,71	0,72
Kleines PMI-WB	nein	nein	0,53	0,53	0,53
Kleines PMI-WB	ja	nein	0,64	0,64	0,64
Kleines PMI-WB	nein	ja	0,53	0,53	0,53
Kleines PMI-WB	ja	ja	0,65	0,64	0,65
Großes PMI-WB	nein	nein	0,52	0,52	0,52
Großes PMI-WB	ja	nein	0,54	0,54	0,54
Großes PMI-WB	ja	ja	0,42	0,41	0,42
SentiMhd	ja	nein	0,73	0,71	0,72
SentiMhd – Negationen und Konjunktiv-2 übergehen	ja	nein	0,75	0,73	0,74

dem fett-kursiv markierten kleinen PMI-Wörterbuch sehen, dass der Einsatz des RNN-Taggers zu einer erheblichen Verbesserung führt.

Weniger erfolgreich war mein Versuch, Flexionsformen mit einem regelbasiert erzeugten Vollformenwörterbuch (Dimpel 2017, S. 100f.) zu erfassen, das nicht manuell korrigiert ist. Die Regeln sind dort für einen maximalen Recall formuliert, sie produzieren dabei auch viele Wortformen, die nicht vorkommen. Beim großen PMI-Wörterbuch werden die F1-Werte schlechter, wenn man zusätzlich zum RNN-Abgleich auch das Vollformenwörterbuch einbezieht; bei SentiMhd und bei den späteren Textanalysen beschränke ich mich daher auf den RNN-Abgleich.

Der Versuch mit der automatischen Übersetzung führt zu desaströsen F1-Werten. Die PMI-Wörterbücher liefern weniger gute Werte als das manuell erstellte Wörterbuch. Die Probleme bei der PMI-Implementierung und bei der Übersetzung beruhen zum Teil auch auf Schwierigkeiten der Normalisierung und der Flexionsformgenerierung. Bei den besseren Werten des manuellen Wörterbuchs ist einschränkend zu bedenken, dass die Annotationsregeln sowohl bei der Wörterbucherstellung als auch bei der manuellen Annotation verwendet wurden, so dass eine gewisse Zirkularität im Spiel ist.²⁵ Allerdings geht in das Wörterbuch bei ambigen Wörtern die mut-

²⁵ Andererseits wäre es widersinnig, für die manuelle Annotation und für den Wörterbuchbau ganz verschiedene oder gar widersprüchliche Annotationsrichtlinien zu verwenden. In einer idealen Welt mit unbegrenzten Ressourcen wäre es spannend, eine sehr große Zahl an gut ausgebildeten Altgermanist*innen Texte zunächst ohne Annotationsregeln annotieren zu lassen. Nachdem ausreichend Erfahrungen damit gesammelt wären und nachdem über die Frage der Konsistenz von Annotationsentscheidungen reflektiert worden wäre, wäre es spannend, in weiteren Annotationsphasen zu ermitteln, inwieweit sich die Annotator*innen an gleichen oder abweichenden Grundsätzen orientieren würden; Annotationsregeln wären also

maßlich überwiegende Wortbedeutung ein; im Evaluierungskorpus wurde bei der manuellen Annotation meist die Bedeutung im jeweiligen Kontext erfasst.

Für das neue Wörterbuch »SentiMhd«, das für Textanalysen weiterverwendet wird, verwende ich nach der Evaluierung der vier Experimente als Basis das manuelle Wörterbuch, das jedoch um 200 Wortformen aus dem großen PMI-Wörterbuch mit sehr hohen Scores erweitert wird. Zwar ergeben sich beim Evaluierungskorpus keine weiteren Verbesserungen durch diese Erweiterung, doch lässt der größere Umfang bei größeren Korpora einen höheren Recall erwarten.²⁶

»SentiMhd« ist im Dariah-Repository zugänglich: <https://doi.org/10.20375/0000-0010-05bb-6>.

Wenn man die Problemfälle mit Konjunktiv-2, Negation oder semantischer Polaritätsumkehr ignoriert (also Wortformen übergeht, die bei der manuellen Annotation mit dem Polaritätsumkehr-Tag versehen wurden), ergibt sich sogar ein F1-Mittelwert von 0,82 – solche Tags sind jedoch beim Einsatz bei nicht annotierten Texten nicht vorhanden. Auf Grundlage der Daten des RNN-Taggers kann man immerhin ermitteln, ob Konjunktiv-2 oder Negation vorliegt. Wenn man in diesen Fällen eine gegenteilige Polarität zuweist, werden die F1-Werte jedoch schlechter. Der mittlere F1-Wert steigt jedoch auf 0,74, wenn Zeilen mit Konjunktiv-2 übersprungen werden und wenn ein Wort, das unmittelbar auf eine Negation folgt, übersprungen wird.²⁷ Dieses Verfahren wird auch für die weitere Textanalyse verwendet.

2.4 Sentimentanalyse und Hermeneutik

Mit Hilfe von SentiMhd wird der *Iwein* mit positiven und negativen Tags automatisch annotiert, dabei werden wieder die Lemmatisierungsinformationen von Schmidts RNN-Tagger-Modell verwendet. Segmentweise wird gezählt, wie viele positive und negative Tags vorhanden sind. Damit sei eine Annäherung (zum Begriff Moretti 2013, S. 2–5) an die Frage modelliert, inwieweit man von einer positiven oder negativen Stimmung im Segment sprechen kann. Tatsächlich gemessen werden natürlich nur Wortkurrenzen aus SentiMhd; selbstverständlich kann keine fehlerfreie und eine objektivierbare Sentiment-Zuordnung erfolgen. Bei einer großen Zahl von An-

in einer idealen Welt auf einer breiteren Basis zu finden, wenn man (wohl sinnvollerweise) das Bedürfnis unterstellt, dass Annotationen möglichst konsistent erfolgen sollen. Vgl. zu Annotationsproblemen etwa Gius und Jacke 2015, die zwar abweichende Annotationsentscheidungen zulassen, zugleich aber fordern, dass die Entscheidungen zufriedenstellend begründbar sein müssen; schlecht begründete Entscheidungen sind zu revidieren.

²⁶ Da der RNN-Tagger auf dem ReM trainiert wurde, gibt er meist als normalisiertes Lemma das ReM-Lemma aus; daher habe ich nach Abschluss dieser Studie dort, wo es bei den Lemmata in SentiMhd Abweichungen zwischen der ReM-Form und der Lexer-Form gibt, zusätzlich noch das ReM-Lemma ergänzt. Im Abgleich mit dem manuell annotierten Evaluierungskorpus haben sich dadurch zwar keine Veränderungen ergeben, bei längeren Texten ist jedoch ein höherer Recall zu erwarten. Die Analysen im letzten Abschnitt dieser Studie verwenden bereits diese Erweiterung. SentiMhd enthält damit nunmehr 1372 positive und 737 negative Wortformen zuzüglich Flexionsformen.

²⁷ Die Polaritätsumkehr-Tags bestehend überwiegend aus Tags zu sonstigen semantischen Problemen; die Negations- und Konjunktiv-2-Tags kommen nicht ganz auf ein Drittel der Tags, so dass der F1-Wert ohne Polaritätsumkehr-Tags von 0,82 deutlich verfehlt wird.

notationen ist immerhin zu hoffen, dass Fehler oder ambige Bedeutungen statistisch nicht dominant werden.

Die Ergebnisdaten können vor dem Hintergrund von konventionellen hermeneutischen Überlegungen studiert werden – etwa in Hinblick auf die Frage, ob die digitalen Daten plausibel sind. Wenn überraschende Befunde vorliegen, wird man nach den Ursachen fragen: Neben einer begrenzten Präzision von digitalen Modellierungen können solche Daten auch Anlass sein, bisherige hermeneutische Befunde nochmals zu durchdenken und auf hermeneutischem Weg zu überlegen, ob auch alternative Erklärungsmöglichkeiten möglich sind. Zudem können überraschende digitale Befunde Phänomene sichtbar machen, die bislang nicht aufgefallen sind bzw. nach denen nicht gesucht wurde (Serendipitätseffekte). Kurzschlüsse von digitalen Befunden auf Thesen, die narratologische oder hermeneutische Überlegungen zum Gegenstand haben, müssen jedoch vermieden werden. Zwar können sich digitale und hermeneutische Erkenntnisse durchaus gegenseitig plausibilisieren. Man wird aber kaum je davon sprechen können, dass ein digitaler Befund die Gültigkeit einer These, die durch eine konventionelle Interpretation zustande gekommen ist, beweisen kann.²⁸

2.4.1 Erste Anwendungen

Ein höherer positiv-negativ-Quotient und somit ein höherer Datenpunkt in Abbildung 1 indiziert – unter den eben genannten Einschränkungen – eine positivere Stimmung. Die schwarze Linie repräsentiert den Quotienten für alle Sentimentwörter. Da ich aus einem anderen Projekt über Annotationen zu Figurenreferenzen²⁹ verfüge, können figurespezifische Quotienten erhoben werden.³⁰ Damit wird modellierbar, inwieweit Figuren eher im Kontext von positiven oder negativen Wortformen vorkommen – gleichsam eine Annäherung an hermeneutische Aussagen zu der Frage, ob eine Figur in einer bestimmten Passage von einer eher positiven oder eher negativen Stimmung umgeben ist. Solche Aussagen sind jedoch freilich durch die hier vorliegende Modellierung limitiert.

Der Quotient unterschreitet den Wert 1 nur selten. Bei dem Wert 1 wären ebenso viele positive wie negative Sentimentwörter vorhanden. Allerdings sind in SentiMhd 1,86 Mal so viele positive wie negative Wortformen enthalten, daher wird die

²⁸ Eibl 2013 hat ähnlich gelagerte Befunde, die mit verschiedenen Methoden erzielt wurden, als »Kontrollpeilung« bezeichnet. Zu empirischen Teilstrecken in der digitalen Literaturanalyse vgl. Dimpel 2015.

²⁹ Dimpel et al. 2022. Durchgehend annotiert sind Eigennamen und Appellativa. Leider sind pronominale Referenzen nur dann annotiert, wenn zehn Verse zuvor keine Referenz auf diese Figur annotiert wurde. Mit Blick auf Sentiment-Studien wäre eigentlich eine vollständige Annotation aller Figurenreferenzen wünschenswert.

³⁰ Ein Sentiment-Wort wird dann als figurespezifisch gezählt, wenn der Abstand zwischen ihm und der folgenden Figurenreferenz <5 Wörter beträgt bzw. wenn der Abstand der Figurenreferenz und dem folgenden Sentiment-Wort <15 Wörter beträgt, da Eigenschaften einer Figur häufiger nach der Referenz als vor der Referenz zugeschrieben werden.

Falls in einem Segment nur eine positive und keine negative Distanz vorhanden ist oder keine positive und nur eine negative Distanz vorhanden ist, wird kein Quotient berechnet und kein Datenpunkt geplottet. Falls keine negative Distanz vorhanden ist, aber mehrere positive Distanzen vorhanden sind, wird der Divisor zur Vermeidung einer Teilung durch Null auf 0,8 gesetzt.

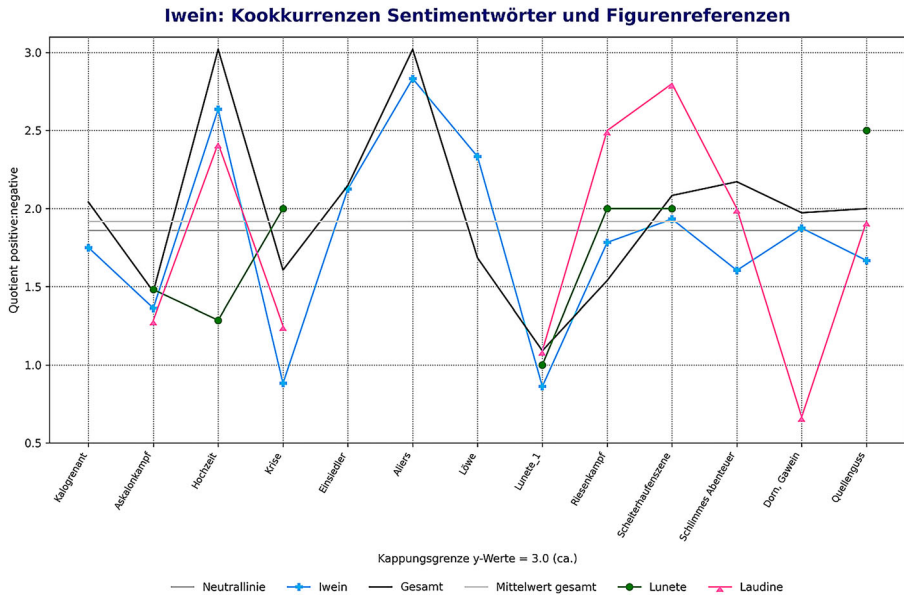


Abb. 1 Sentimentwörter im *Iwein*. In der farbigen Online-Version des Beitrags unter <https://doi.org/10.1007/s41244-024-00339-0> sind die *Figurenlinien* besser zu erkennen

Neutrallinie bei 1,86 als untere graue Linie eingezeichnet. Die obere graue Linie bei 1,92 zeigt den Mittelwert des Quotienten im gesamten *Iwein*.

Wer den *Iwein* kennt, wird das eher positive Vokabular im Segment »Hochzeit« für erwartbar halten. Man könnte aber annehmen, dass die negative Stimmung bei der »Krise« noch deutlicher ausfallen müsste. Hier zeigt sich der Vorteil von figurenspezifischen Auswertungen: Wenn man den Sentiment-Werten folgt, ist die Krise offenbar stark auf Iwein und relativ stark auf Iweins Ehefrau Laudine bezogen, während die Zofe Lunete weniger betroffen ist. Auch der schwarze Durchschnittswert liegt deutlich über Iweins und Laudines Wert.

Im Segment »Aliers« stehen Iweins Höchstwerte – hier gelangen seine Aktionen problemlos. Dagegen unterschreiten die Tiefstwerte im Segment »Lunete_1« noch die niedrigen Werte im Segment »Krise«, es gibt also ausweislich des Sentiment-Modells zwei Tiefpunkte. In diesem Segment wird darüber berichtet, dass Lunete hingerichtet werden soll, weil sie Iwein zur Landesherrschaft verholphen hat. Die Existenz eines zweiten Tiefpunkts ist deshalb interessant, weil die Forschung vielfach einen doppelten Kursus ansetzt, bei dem der erste Kursus vom zweiten Kursus von »der« Krise abgetrennt wird.³¹ Selbst Ludger Lieb, der ein Strukturmodell vorlegt, das vom Mainstream abweicht, spricht im Singular von einer Katastrophe

³¹ Vgl. exemplarisch Kuhn 1948. Das Modell lässt sich vereinfacht etwa wie folgt skizzieren: Im ersten Kursus erwirbt der Ritter durch Bewährungsaventuren Frau und Ehre; in der Krise verliert er nach einem defizitären Verhalten seine Ehre und es kommt zu einer Störung in der Minnebeziehung. Im zweiten Kursus gelingt es dem Ritter, Frau und Ehre zurückzugewinnen; der Aufwand dafür ist größer als im ersten Kursus. Zur Kritik am Modell vgl. Schmid 1999.

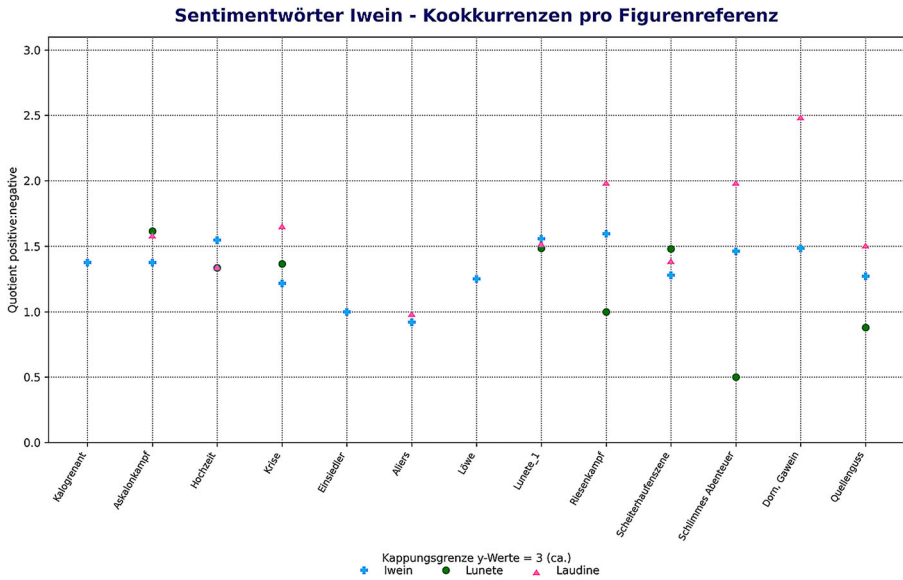


Abb. 2 Kookkurrenzen von Figurenreferenzen und Sentimentwörtern pro Figurenreferenz

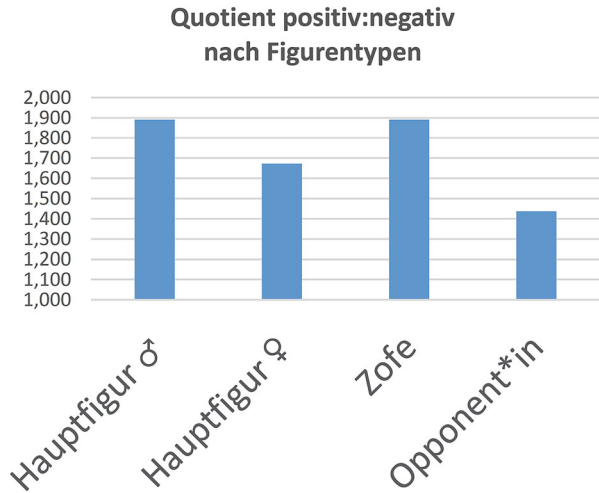
(Lieb 2020, S. 88). Danach führt, so Lieb (2020, S. 86), die Aventurefahrt »den Protagonisten vom Tiefpunkt über mehrere Stationen langsam wieder bis zu einem zweiten Höhepunkt, der den ersten übertrifft«. In einer Skizze zeichnet Lieb eine linear ansteigende Linie vom Tiefpunkt (hier: Segment »Krise«) zum Romanende – allerdings unter der Achsenbeschriftung »Ehre-Schande«. ³² Nach dem Sentiment-Modell erfährt die Krise jedoch eine Wiederholung, als auch Lunete in den Strudel von Iweins Versagen gerät.

Auch wenn der doppelte Kursus nach Hugo Kuhn jenseits von Ludger Liebs Nuancierungen gemäß der *Communis Opinio* als das grundlegende arthurische Strukturmuster (zumindest aber für *Erec* und *Iwein*) gelten kann, sei angemerkt, dass es sich bei diesem Strukturmuster ebenso um eine Abstrahierung und um eine Modellierung handelt, die nicht etwa selbst den Texten als Entität inhärent wäre. Solche Parallelen zwischen digitalen Modellierungen und strukturalistischen oder bzw. und hermeneutischen Modellierungen werden bisweilen übersehen, wenn gegen digital erhobene Daten der Einwand erhoben wird, dass in einer Studie das Zielphänomen selbst gar nicht direkt und unmittelbar gemessen wird.

Am Ende des Sentiment-Diagramms für den *Iwein* steht kein neuer Höhepunkt: Iwein erreicht in acht anderen Segmenten höhere Werte als im Schlussegment – ein Happy End kann also anhand der Sentiment-Daten nicht nachvollzogen werden. Dieser Befund lässt sich jedoch im Sinne einer Kontrollpeilung gut mit Interpretationen in Verbindung bringen, die von einem harten Iwein-Schluss ausgehen: Laudine nimmt Iwein, ihren früheren Mann, nur widerwillig und nur deshalb wieder auf, weil sie auf die listige Eidforderung ihrer Zofe Lunete hereingefallen ist und nicht

³² Eine ähnlich linear ansteigende Linie bietet das Diagramm in Wolf 2007, S. 45.

Abb. 3 Figurentypen



eidbrüchig werden will. Zwei verschiedene *Iwein*-Schlüsse sind überliefert: In der Mehrheit der Handschriften konstatiert Laudine nach Iweins Kniefall nur, dass sie Iwein aufgrund des Eides wieder aufnehmen muss; in einigen wenigen Handschriften stehen Zusatzverse, in denen Laudine ebenfalls vor Iwein auf die Knie fällt und sein Leid bedauert (vgl. Schröder 1997; Hausmann 2001).

Abschließend wird sicherheitshalber, um abschätzen zu können, ob Schief lagen vorhanden sind, die darauf beruhen, dass womöglich nur sehr wenige Sentimentwörter oder nur sehr wenige Figurenreferenzen in einem Segment vorkommen, in Abbildung 2 ausgegeben, wie viele Kookkurrenzen von Figurenreferenzen und Sentimentwörtern pro Figurenreferenz in welchem Segment vorhanden sind:

Da erwartbar ist, dass Hauptfiguren häufiger als Nebenfiguren genannt werden, wird hier anstelle von absoluten Zahlen der Quotient gewählt. Der Mittelwert für alle Figuren in allen Segmenten liegt bei 1,1, die Verteilung von Figurenreferenzen und Sentimentwörtern ist also nicht allzu asymmetrisch.

Da auch zu *Tristan*, *Partonopier* und *Mauritius von Craun* annotierte Figurenreferenzen zu den Hauptfiguren, Opponent*innen und Zofen vorliegen, ist eine Auswertung zu diesen Figurentypen möglich. Abbildung 3 zeigt, dass die männlichen Hauptfiguren deutlich positiver kontextualisiert sind als die weiblichen – in diesen vier Texten geht es auch um Probleme der Hauptfigur mit seiner Partnerin. Die Zofen, die dem Titelhelden beistehen, erreichen ebenso positive Werte wie die Titelhelden, während die Opponent*innen erwartungsgemäß negativere Werte erhalten.

Interessant ist die Binnendifferenzierung bei den Opponent*innen in Abbildung 4: *Tristan* und *Partonopier* neigen insgesamt eher zu positivem Vokabular,³³ hier gibt es selbst bei den Opponent*innen etwas mehr positive als negative Sentimentwörter im Figurenkontext. Zudem haben Partonopiers Mutter und König Marke als nahe Angehörige der Hauptfigur ebenso einen Sonderstatus unter den Opponent*innen

³³ Quotienten gesamt (also nicht nur Sentimentwörter im Figurenkontext): Iwein: 1,93, Mauritius: 1,84, Tristan: 2,60, Partonopier: 2,25.

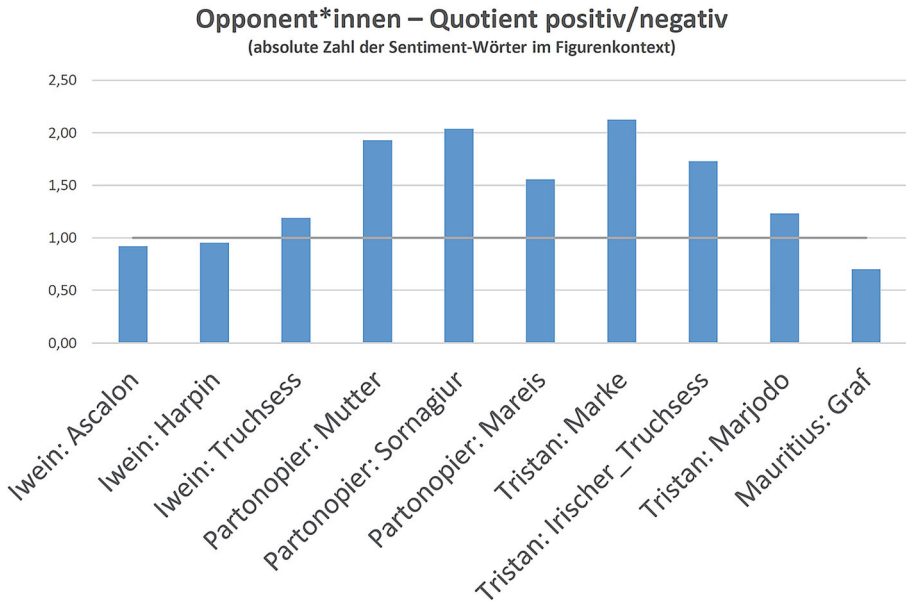


Abb. 4 Opponent*innen

wie Sornagiur, der sich nach dem Verrat von Mareis freiwillig in die Gefangenschaft von Partonopiers Partei gibt. Dagegen finden sich im *Iwein* bei klaren Gegenspielern wie Ascalon deutlich negativere Werte.

3 Vier Textsorten in zehn Segmenten

Zum Abschluss noch ein exemplarischer Makro-Blick auf vier Textsorten in jeweils zehn Segmenten (Abb. 5, 6, 7 und 8).

Bei nur zehn gleichlangen Segmenten bleiben feinere Informationen auf Kosten der Übersichtlichkeit auf der Strecke; die Kappungsgrenze wurde hier aufgrund hoher Werte auf 6 angehoben. Bei den klassischen höfischen Romanen gibt es mit Ausnahme von *Tristan* und *Iwein* zum Romanende wieder einen Anstieg der Werte; im *Parzival* wird im letzten Buch nochmals das Leid von Anfortas thematisiert (vgl. Dimpel 2023). Bei *Erec* und *Iwein* lässt sich die Krise ausmachen (zur doppelten Krise im *Iwein* vgl. oben). Bei *Parzival* und *Tristan* ist noch die Existenz der Elternvorgeschichte zu bedenken.

Bei den nachklassischen höfischen Romanen weisen die Linien nach dem Anfangssegment zunächst nach unten; zum Schluss ist zumindest bei *Lanzelet*, *Gauriel* und *Daniel* ein Anstieg sichtbar, der noch deutlicher ausfällt als bei den klassischen höfischen Romanen – womöglich ein Indikator für ein ausgeprägtes Happy End. Beim *Partonopier* ist allerdings das Ende nicht überliefert. Im *Wigalois* kommt mit dem Namur-Feldzug nach der Befreiung von Korntin eine neue Problemlage ins Spiel.

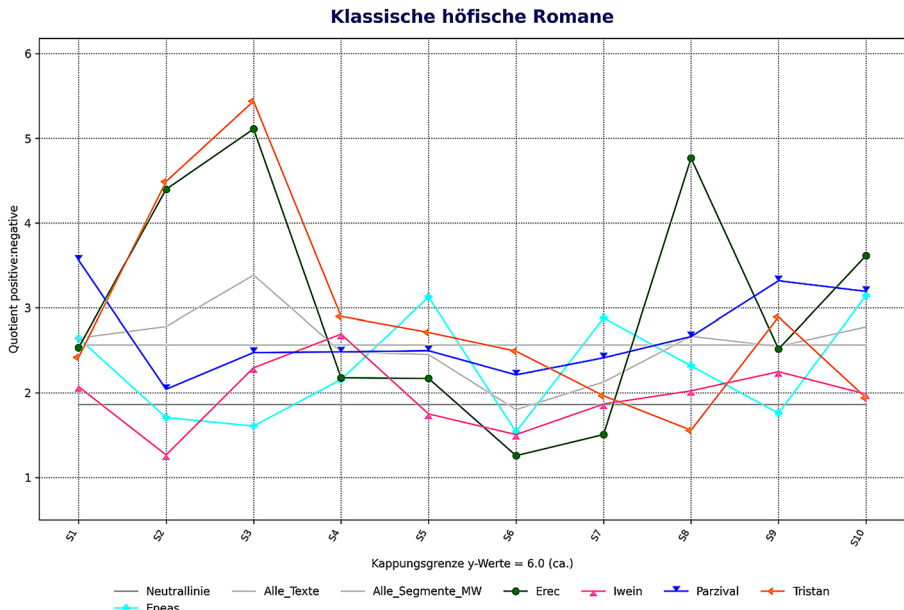


Abb. 5 1. Klassische höfische Romane

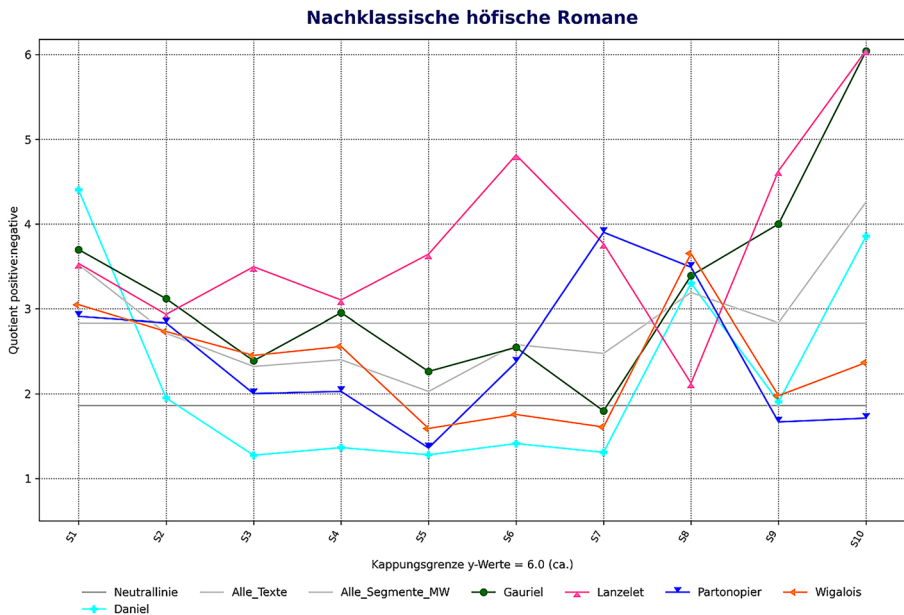


Abb. 6 2. Nachklassische höfische Romane



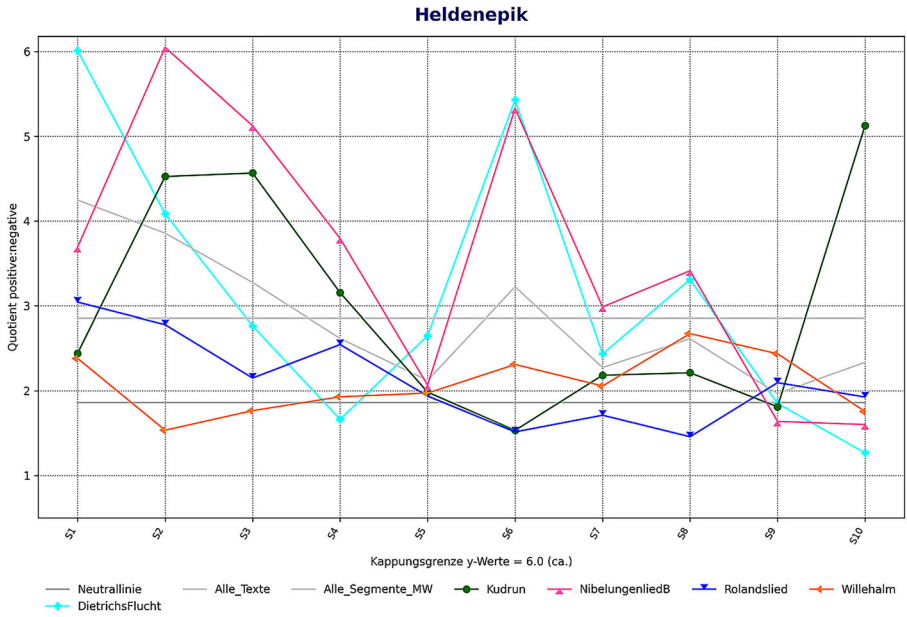


Abb. 7 3. Heldenepik

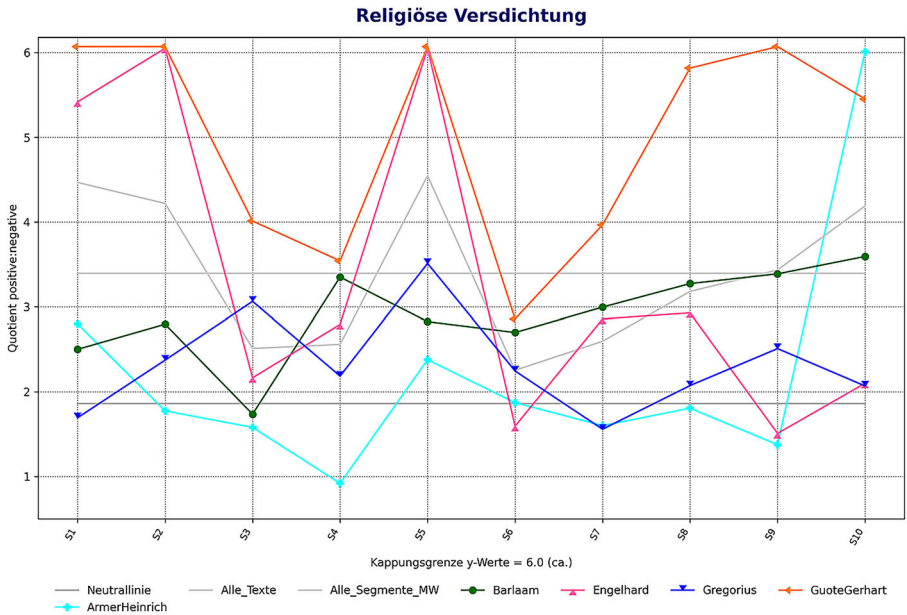


Abb. 8 4. Religiöse Versdichtung

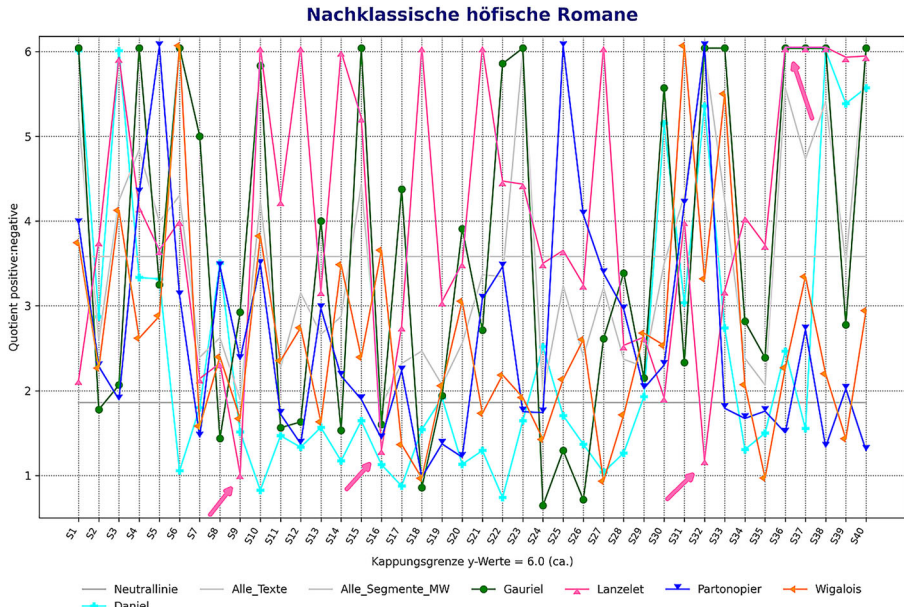


Abb. 9 Nachklassische höfische Romane in 40 Segmenten

In der Heldenepik ist mit Ausnahme der *Kudrun*, die eher ein Operettenende aufweist, am Ende erwartungsgemäß eine Entwicklung zu negativem Vokabular zu sehen. Überraschend ist die in weiten Teilen hohe Linie des *Nibelungenliedes* und ihre partielle Konvergenz mit der Linie zu *Dietrichs Flucht*.

Bei der religiösen Versichtung fällt der *Guote Gerhart* auf: Der Titel ist Programm – zumindest für das Sentiment-Vokabular. Im *Engelhard* fällt das Opfer der Söhne in das letzte Segment; das Wunder am Textende nimmt wenig Raum ein; das Happy End ist beispielsweise in einer Darstellung mit 40 Segmenten besser sichtbar, für die hier kein Raum ist – ähnlich im *Gregorius*.

Eine Gliederung mit 40 Segmenten wird hier noch zum nachklassischen höfischen Roman gezeigt:

Lanzelet jagt, so Kragl (2009, S. 552), von einem Höhepunkt zum nächsten; in Abbildung 9 erreicht er zwölf Mal Werte um 6 und darüber; nach dem *Guoten Gerhart* weist er insgesamt die zweithöchsten Sentiment-Werte auf. Die Tiefpunkte – hier mit Pfeilen markiert – lassen sich ebenfalls gut nachvollziehen: Segment S8 enthält den Kampf gegen Linier, S15 den Umkehrzauber von *schâtel le mort*, in S31 erklären sich Erec und Walwein bereit, sich zu Malduc in Gefangenschaft zu begeben. Im ebenfalls mit Pfeil markierten Segment S36 wird Lanzelet nach der Heimkehr nach Genewis nun auch die Herrschaft in Dodone angetragen, hier erreicht Lanzelet einen Quotienten von 13,2.

Die exemplarische Analyse von vier Textsorten soll künftig durch die Untersuchung von größeren Korpora fortgeführt werden, um etwa Sonderfälle wie den *Guoten Gerhart* (vgl. Abbildung 10) klarer innerhalb der Textsorte verorten zu können. Dabei sollen auch Studien zu weiteren Fragestellungen folgen; weiterhin ist

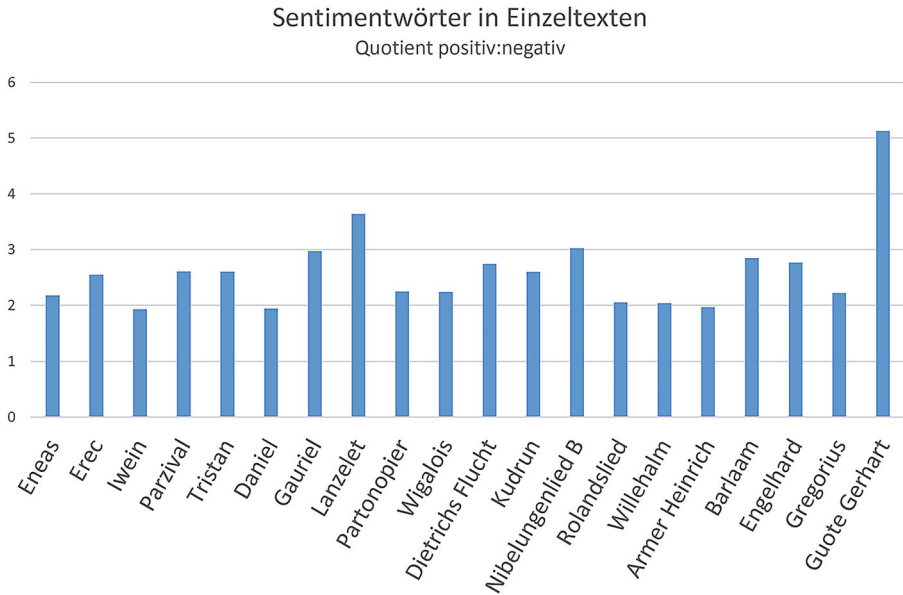


Abb. 10 Gesamtübersicht Sentiment-Werte zu 20 Texten

zu prüfen, ob mit Hilfe von Word-Embedding-Verfahren die Probleme mit Negationen, Konjunktiv-2 und semantischen Polaritätsveränderungen besser gelöst werden können.

Die Ergebnisse der Sentimentanalyse dürfen aus literaturwissenschaftlicher Sicht als plausibel gelten – die Relativierung des Happy Ends im *Iwein* fällt sogar deutlicher aus, als ich angenommen hätte. Traditionelle hermeneutische Befunde können in diesem Fall mit korpusermeneutischen Befunden in Einklang gebracht werden. Insoweit mit der Beobachtung eines zweiten Tiefpunkts im *Iwein* die Doppelweg-Struktur hinterfragt werden kann, bei der konventionell nur eine Krise angenommen wird, ist diese Beobachtung doch auch anschlussfähig an narratologische Überlegungen, die Lunete eine zentralere Rolle zuschreiben als Iweins Ehefrau Laudine (Dimpel 2011, S. 203–228).

Die Experimente zur Sentimentanalyse dürfen zugleich als ein Versuch gelten, den Problemen einer mediävistischen Korpushermeneutik zu begegnen, die im ersten Abschnitt ausgeführt wurden: Die unzureichende Verfügbarkeit digitaler Texte hat sich durch die open-text-policy der Mittelhochdeutschen Begriffsdatenbank in den letzten Jahren erheblich verbessert, auch wenn noch immer viele Texte nicht oder nicht in aktuellen Ausgaben digital frei verfügbar sind. Probleme der ›wilden‹ mittelhochdeutschen Graphie lassen sich durch ein Normalisierungswörterbuch oder durch Schmidts RNN-Tagger zumindest teilweise kompensieren. Das Problem der eingeschränkten Verfügbarkeit von Tools und Ressourcen zur automatischen Textanalyse wird, so ist zu hoffen, mit weiteren Forschungsfortschritten allmählich gemildert; mit SentiMhd steht dem Fach nunmehr auch eine Ressource zur Sentimentanalyse des Mittelhochdeutschen mit einer Creative-Commons-Lizenz zur Nachnutzung und Optimierung bereit.

Funding Open Access funding enabled and organized by Projekt DEAL.

Open Access Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Artikel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.

Weitere Details zur Lizenz entnehmen Sie bitte der Lizenzinformation auf <http://creativecommons.org/licenses/by/4.0/deed.de>.

Literatur

- Burch, Thomas (2022): Infrastrukturprojekte zur digitalen Lexikographie. Vorgestellt am Beispiel des Zentrums für Historische Lexikographie. In: Elisabeth Lienert/Joachim Hamm/Albrecht Hausmann/Gabriel Viehhauser (Hg.): *Digitale Mediävistik. Perspektiven der Digital Humanities für die Altgermanistik (=Beiträge zur mediävistischen Erzählforschung, Themenheft 12)*. Oldenburg, S. 97–108. <https://doi.org/10.25619/BmE20223199>.
- Büttner, Andreas/Dimpel, Friedrich Michael/Evert, Stefan/Jannidis, Fotis/Pielström, Steffen/Proisl, Thomas/Reger, Isabella/Schöch, Christof/Vitt, Thorsten (2017): »Delta« in der stilometrischen Autorschaftsattribuion. In: *Zeitschrift für digitale Geisteswissenschaften*. https://doi.org/10.17175/2017_006.
- Dennerlein, Katrin/Thomas Schmidt/Christian Wolff (2022): Emotion courses in German historical comedies and tragedies. In: *Digital Humanities 2022. Responding to Asian Diversity. 25-29 July 2022, Toshi Center Hotel, Tokyo. Conference Abstracts*, S. 193–197. <https://dh2022.dhii.asia/dh2022bookofabsts.pdf>.
- Dennerlein, Katrin/Schmidt, Thomas/Wolff, Christian (2023): Computational emotion classification for genre corpora of German tragedies and comedies from 17th to early 19th century. In: *Digital Scholarship in the Humanities* 38(4), S. 1466–1481. <https://doi.org/10.1093/lc/fqad046>.
- Dimpel, Friedrich Michael (2004): *Computergestützte textstatistische Untersuchungen an mittelhochdeutschen Texten*. Tübingen: Francke.
- Dimpel, Friedrich Michael (2011): *Die Zofe im Fokus. Perspektivierung und Sympathiesteuerung durch Nebenfiguren vom Typus der Confidente in der höfischen Epik des hohen Mittelalters*. Berlin: ESV.
- Dimpel, Friedrich Michael (2015): Der Computerphilologe als Interpret – ein Teilzeit-Empiriker? In: Jan Borkowski/Stefan Descher/Felicitas Ferder/Philipp Heine (Hg.): *Literatur interpretieren: Interdisziplinäre Beiträge zur Theorie und Praxis*. Münster: Mentis, S. 339–359.
- Dimpel, Friedrich Michael (2017): Autorschaftsattribuion bei nicht-normalisiertem Mittelhochdeutsch. Bessere Erkennungsquoten durch ein Normalisierungswörterbuch. In: Michael Stolz (Hg.): *Konferenzabstracts DHD 2017 Bern. Digitale Nachhaltigkeit*. Bern, S. 100–103. <https://doi.org/10.5281/zenodo.3684825>.
- Dimpel, Friedrich Michael (2018): Stabile Autorschaft trotz handschriftlicher Varianz? Die Erfolgsquote von Burrows' Delta bei nicht-normalisierten mittelhochdeutschen Texten optimieren. In: *Zeitschrift für deutsches Altertum und deutsche Literatur* 147, S. 341–363.
- Dimpel, Friedrich Michael (2023). Erste Versuche einer Sentiment-Analyse zu Wolframs »Parzival« mit SentiMhd. In: Sabine Bartsch/Philipp Hegel/Luise Borek (Hg.): *Computer im Musenhain: Die Träume der alten Bücher und die Aura des Digitalen*. Darmstadt, im Druck.
- Dimpel, Friedrich Michael/Michael Stolz (2006): Computergestütztes Kollationieren und dynamische Textpräsentation. Ein Werkstattbericht aus dem Parzival-Projekt. In: *Computergestützte Kollationierung und ihre Integration in den editorischen Arbeitsfluss (Tagung an der BBAW am 26./27. Mai 2006)*. Berlin. http://edoc.bbaw.de/frontdoor.php?source_opus=516&la=de.

- Dimpel, Friedrich Michael/Wagner, Silvan (2022): Rosenplüts Weingrüße. Stilistische und digitale Studien zur Autorschaft einer Nürnberger Kleinstgattung. In: Sylvia Jurchen/Silvan Wagner (Hg.): *Wetter und Grenze, Zeitschrift Brevitas* 2, BmE Sonderheft. Oldenburg. <https://doi.org/10.25619/BmE20221165>.
- Dimpel, Friedrich Michael/Schlager, Daniel/Zeppezauer-Wachauer, Katharina (2019): Der Streit um die Birne. Autorschafts-Attributionstest mit Burrows' Delta und dessen Optimierung für Kurztexte am Beispiel der ›Halben Birne‹ des Konrad von Würzburg. In: Roman Bleier/Franz Fischer/Thorsten Hiltmann/Gabriel Viehhauser/Georg Vogeler (Hg.): *Digitale Mediävistik, Das Mittelalter. Perspektiven mediävistischer Forschung, Zeitschrift des Mediävistenverbandes* 24, S. 71–90.
- Dimpel, Friedrich Michael/Blessing, André/Zeppezauer-Wachauer, Katharina /Hinkelmanns, Peter/Ketschik, Nora (2022): Figures and ihr Handeln. Eine computergestützte Untersuchung von Figurenaktivitäten im Kontext von Figurenreferenzen mit Hilfe des Begriffssystems der MHDDBD. In: Elisabeth Lienert/Joachim Hamm/Albrecht Hausmann/Gabriel Viehhauser (Hg.): *Digitale Mediävistik. Perspektiven der Digital Humanities für die Altgermanistik* (=Beiträge zur mediävistischen Erzählforschung, Themenheft 12). Oldenburg, S. 283–328.
- Echelmeyer, Nora/Nils Reiter/Sarah Schulz (2017): Ein PoS-Tagger für »das« Mittelhochdeutsche. In: Michael Stolz (Hg.): *Konferenzabstracts DHd 2017 Bern. Digitale Nachhaltigkeit*, S. 141–147. (<http://www.dhd2017.ch/programm>).
- Eibl, Karl (2013): Ist Literaturwissenschaft als Erfahrungswissenschaft möglich? Mit einigen Anmerkungen zur Wissenschaftsphilosophie des Wiener Kreises. In: Philip Ajouri/Katja Mellmann/Christoph Rauen (Hg.): *Empirie in der Literaturwissenschaft*. Münster: C.H. Beck (=Poetogenesis. Studien zur empirischen Anthropologie der Literatur 8), S. 19–45.
- Flüh, Marie (2019): Sentimentanalyse. In: *forTEXT. Literatur digital erforschen*. <https://fortext.net/routinen/methoden/sentimentanalyse>.
- Gius, Evelyn/Jacke, Janina (2015): Informatik und Hermeneutik. Zum Mehrwert interdisziplinärer Textanalyse. In: Constanze Baum/Thomas Stäcker (Hg.): *Grenzen und Möglichkeiten der Digital Humanities, Sonderband der Zeitschrift für digitale Geisteswissenschaften 1*. https://doi.org/10.17175/sb001_006.
- Glauch, Sonja (2022): Welche Lebenserwartung haben digitale Editionen? In: Elisabeth Lienert/Joachim Hamm/Albrecht Hausmann/Gabriel Viehhauser (Hg.): *Digitale Mediävistik. Perspektiven der Digital Humanities für die Altgermanistik* (=Beiträge zur mediävistischen Erzählforschung, Themenheft 12). Oldenburg, S. 65–75.
- Glawion, Anastasia (2023): *Remembering World War II: a mixed-methods exploration of memory practices on an online forum*. Berlin: J.B. Metzler (=Digitale Literaturwissenschaft). <https://doi.org/10.1007/978-3-662-66708-8>.
- Hausmann, Albrecht (2001): Mittelalterliche Überlieferung als Interpretationsaufgabe. »Laudines Kniefall« und das Problem des »ganzen Textes«. In: Ursula Peters (Hg.): *Text und Kultur. Mittelalterliche Literatur 1150–1450*. Stuttgart/Weimar, S. 72–95.
- Hinkelmanns, Peter (2021): Middle High German pos tagger model for spacy. <https://github.com/Middle-High-German-Conceptual-Database/Spacy-Model-for-Middle-High-German>.
- Kaeding, Friedrich Wilhelm (1897): *Häufigkeitwörterbuch der deutschen Sprache. Festgestellt durch einen Arbeitsausschuß der deutschen Stenographie-Systeme*. Faksimile-Druck der Originalausgabe, Selbstverlag des Herausgebers, Steglitz bei Berlin 1897. Quickborn bei Hamburg.
- Kestemont, Mike (2018): Stylometric Authorship Attribution for the Middle Dutch Mystical Tradition from Groenendaal. In: *Dutch crossing* 42, S. 203–237. <https://doi.org/10.1080/03096564.2016.1252077>.
- Ketschik, Nora/André Blessing/Sandra Murr/Maximilian Overbeck/Axel Pichler (2020): Interdisziplinäre Annotation von Entitätenreferenzen. In: Nils Reiter/Axel Pichler/Jonas Kuhn (Hg.): *Reflektierte algorithmische Textanalyse. Interdisziplinäre(s) Arbeiten in der CRETA-Werkstatt*. Berlin/Boston: De Gruyter, S. 203–263. <https://doi.org/10.1515/9783110693973>.
- Kim, Evgeny/Klinger, Roman (2019): A Survey on Sentiment and Emotion Analysis for Computational Literary Studies. In: *Zeitschrift für digitale Geisteswissenschaften*. https://doi.org/10.17175/2019_008_v2.
- König, Werner (2004): *Dtv-Atlas Deutsche Sprache*. München: Deutscher Taschenbuch Verlag.
- Kragl, Florian (2009): *Veni, vidi, basiavi*. Pathologisches Erzählen im ›Lanzelet‹. In: Florian Kragl (Hg.): *Ulrich von Zatzikhoven: Lanzelet. Text – Übersetzung – Kommentar*. Studienausgabe. Berlin/New York: De Gruyter, S. 533–554.
- Krug, Markus (2020): *Techniken zur automatischen Extraktion von Figurennetzwerken aus deutschen Romanen. Techniques for the Automatic Extraction of Character Networks in German Historic Novels*. Würzburg. <https://doi.org/10.25972/OPUS-20918>.

- Krug, Markus/Sebastian Kempf/David Schmidt/Frank Weimer/Frank Puppe (2019): Detecting Character References in Literary Novels using a Two Stage Contextual Deep Learning approach. In: Patrick Sahle (Hg.): *DHd 2019. Digital Humanities: multimedial & multimodal. Konferenzabstracts*. De Gruyter Open, S. 109–111. <https://doi.org/10.5281/zenodo.2600812>.
- Kuhn, Hugo (1948): Erec. In: *Festschrift für P. Kluckhohn und H. Schneider*, S. 122–147. Wieder in: Hugo Kuhn/Christoph Cormeau (Hg.): *Hartmann von Aue*. Darmstadt 1973 (=Wege der Forschung 359), S. 17–48.
- Lieb, Ludger (2020): *Hartmann von Aue: Erec-Iwein-Gregorius-Armer Heinrich*. Berlin: Erich Schmidt Verlag.
- Moretti, Franco (2013): »Operationalizing«: or, the function of measurement in modern literary theory. Pamphlets of the Stanford Literary Lab 6. Stanford. <https://litlab.stanford.edu/LiteraryLabPamphlet6.pdf>.
- Puppe, Frank/Krug, Markus/Jannidis, Fotis/Macharowsky, Luisa/Reger, Isabella/ Weimer, Lukas (2015): Rule based Coreference Resolution in German Historic Novels. In: *Proceedings of the Fourth Workshop on Computational Linguistics for Literature*. Denver. <http://www.aclweb.org/anthology/W15-0711>.
- Remus, Robert/Uwe Quasthoff/Gerhard Heyer (2010): SentiWS – a Publicly Available German-language Resource for Sentiment Analysis. In: *Proceedings of the 7th International Language Resources and Evaluation (LREC'10)*, S. 1168–1171.
- Schmid, Elisabeth (1999): Weg mit dem Doppelweg. Wider eine Selbstverständlichkeit der germanistischen Artusforschung. In: Friedrich Wolfzettel (Hg.): *Erzählstrukturen der Artusliteratur*. Tübingen: De Gruyter, S. 69–85.
- Schmid, Helmut (2019): RNNTagger – a Neural Part-of-Speech Tagger. <https://www.cis.uni-muenchen.de/~schmid/tools/RNNTagger/>.
- Schmidt, Thomas/Johanna Dangel/Christian Wolff (2021): SentText: A Tool for Lexicon-based Sentiment Analysis in Digital Humanities. In: Thomas Schmidt/Christian Wolff (Hg.): *Information between Data and Knowledge. Information Science and its Neighbors from Data Science to Digital Humanities. Proceedings of the 16th International Symposium of Information Science (ISI 2021) 74*, Schriften zur Informationswissenschaft. Glückstadt: Werner Hülsbusch, S. 156–172. <https://doi.org/10.5283/epub.44943>.
- Schröder, Werner (1997): *Laudines Kniefall und der Schluß von Hartmanns ›Iwein‹*. Stuttgart: Akad. der Wiss. und der Literatur Steiner.
- Viehhauser, Gabriel (2015): Historische Stilometrie? Methodische Vorschläge für eine Annäherung textanalytischer Zugänge an die mediävistische Textualitätsdebatte. In: Constanze Baum/Thomas Stäcker (Hg.): *Grenzen und Möglichkeiten der Digital Humanities, Sonderband der Zeitschrift für digitale Geisteswissenschaften 1*. https://doi.org/10.17175/sb001_009.
- Viehhauser, Gabriel (2017): Digitale Gattungsgeschichten. Minnesang zwischen generischer Konstanz und Wende. In: *Zeitschrift für digitale Geisteswissenschaften*. https://doi.org/10.17175/2017_003.
- Viehhauser-Mery, Gabriel (2020): Fastnachtspiel und digitale Dramenanalyse. In: *Comparatio* 12, S. 21–37.
- Wolf, Jürgen (2007): *Einführung in das Werk Hartmanns von Aue*. Darmstadt: WBG.
- Zeppzauer-Wachauer, Katharina (2022). 50 Jahre Mittelhochdeutsche Begriffsdatenbank (MHDBDB). Eine Jubiläums-Zeitreise zwischen Lochkarten, Pixel-Drachen, relationaler Datenbank und Graphdaten. In: Elisabeth Lienert/Joachim Hamm/Albrecht Hausmann/Gabriel Viehhauser (Hg.): *Digitale Mediävistik. Perspektiven der Digital Humanities für die Altgermanistik* (=Beiträge zur mediävistischen Erzählforschung, Themenheft 12). Oldenburg, S. 283–328. <https://doi.org/10.25619/BmE20223203>.

Hinweis des Verlags Der Verlag bleibt in Hinblick auf geografische Zuordnungen und Gebietsbezeichnungen in veröffentlichten Karten und Institutsadressen neutral.