



Separatum aus:

THEMENHEFT 12

*Elisabeth Lienert / Joachim Hamm
Albrecht Hausmann / Gabriel Viehhauser (Hrsg.)*

Digitale Mediävistik

Perspektiven der Digital Humanities für die Altgermanistik

Publiziert im Juli 2022.

Die BmE Themenhefte erscheinen online im BIS-Verlag der Carl von Ossietzky Universität Oldenburg unter der Creative Commons Lizenz [CC BY-NC-ND 4.0](https://creativecommons.org/licenses/by-nc-nd/4.0/). Die ›Beiträge zur mediävistischen Erzählforschung‹ (BmE) werden herausgegeben von PD Dr. Anja Becker (München) und Prof. Dr. Albrecht Hausmann (Oldenburg). Die inhaltliche und editorische Verantwortung für das einzelne Themenheft liegt bei den jeweiligen Heftherausgebern.

<http://www.erzaehlforschung.de> – Kontakt: herausgeber@erzaehlforschung.de
ISSN 2568-9967

Zitiervorschlag für diesen Beitrag:

Brandes, Phillip/Marshall, Sophie/Schneider, Felix: Stilfiguren aus der Distanz gelesen. Zur automatischen Detektion von Wortstellungsfiguren und deren Nutzen für die qualitative Analyse, in: Lienert, Elisabeth/Hamm, Joachim/Hausmann, Albrecht/Viehhauser, Gabriel (Hrsg.): Digitale Mediävistik. Perspektiven der Digital Humanities für die Altgermanistik, Oldenburg 2022 (BmE Themenheft 12), S. 247–282 (online).

Phillip Brandes / Sophie Marshall / Felix Schneider

Stilfiguren aus der Distanz gelesen

Zur automatischen Detektion von Wortstellungsfiguren und deren Nutzen für die qualitative Analyse

Abstract. Im Projekt ›Anomaly-based large-scale analysis of style and genre reflected in the use of stylistic devices in medieval literature‹ befragen wir mittelhochdeutsche Texte der *Trois Matières* danach, welche Aussagen zu Textähnlichkeiten sich aufgrund des Gebrauchs von Stilmitteln der Wortstellung, wie etwa Parallelismus und Chiasmus, sowie der Tropen, wie etwa der Metapher, treffen lassen. Der hier vorliegende Artikel stellt die aktuellen Ergebnisse unserer bisherigen Parallelismus- und Chiasmusdetektion in 30 mittelhochdeutschen Texten vor. Es wird zu diskutieren sein, ob unsere Ergebnisse eine Korrelation des Stils mit der gemeinhin nach Stoffkreis definierten Gattung bezeugen, ob andere Faktoren wie Verfasserschaft oder Abfassungszeit eine übergeordnete Rolle spielen und ob darüberhinausgehende Aussagen zu den Ähnlichkeitsverhältnissen der Texte getroffen werden können.

1. Einleitung

»Was braucht das Fach?« war die Leitfrage der Tagung »Digitale Mediävistik«, in deren Rahmen der hier vorliegende Beitrag erstmals diskutiert wurde. Die Frage bezieht sich auf die Herausforderungen, denen sich die Germanistische Mediävistik aufgrund der wachsenden Bedeutung und Re-

levanz der Digital Humanities und angrenzender Nachbardisziplinen entgeggestellt sieht. Hier sollen aus der Praxis eines Projektes, das gleichermaßen aus Informatik und Mediävistik heraus betrieben wird, Bedürfnisse mitformuliert werden, die sich aus der computationalen Textanalyse ergeben.

Ziel des Projektes ›*Anomaly-based large-scale analysis of style and genre reflected in the use of stylistic devices in medieval literature*‹¹ ist es, computationale Methoden zur Detektion von rhetorischen Stilmitteln zu entwickeln und diese Methoden auf ein Corpus, das aus mittelhochdeutschen Texten der *Trois Matières* besteht, anzuwenden und die daraus resultierenden Funde zu analysieren.

Das Projekt bewegt sich damit auch im Bereich der Stilometrie. Stilometrie ist, so kann man es kurzfassen, die computationale Untersuchung von Stil.² Bisher ist die Attribution von Autorschaft das vorherrschende Anwendungsgebiet der Stilometrie (Büttner [u. a.] 2017).³ Während Autorstil durch diverse *most-frequent-words*-Ansätze bereits hinreichend untersucht werden konnte, ist noch weitestgehend unklar, wie ein Gattungsstil untersuchbar wäre (oder überhaupt zu bestimmen ist). Gattung scheint nicht etwas zu sein, das man allein aufgrund von Worthäufigkeiten erkennen kann (Allison [u. a.] 2017, insb. S. 31–33). Dies liegt wohl daran, dass ›Gattung‹ ein Begriff ist, der seine jeweilige Geltung aus den Konventionen zieht, die ihn hervorgebracht haben (Underwood 2019; Remele 2021) und Gattungseinteilungen nicht immer auf der systematisch gleichen Ebene vorgenommen werden (Viehhauser 2017).

Für Texte der *Trois Matières* erfolgt die Zuordnung nach Jean Bodel (vor 1200) zu einer ›Materie‹ aufgrund des Inhalts (einleitend dazu Herweg 2013). Was Artus und seine Tafelrunde zum Gegenstand hat, wird so der *Matière de Bretagne* zugeordnet. Formal sind zumindest die mittelhochdeutschen Vertreter dieser drei Stoffkreise über weite Teile identisch – alle Texte sind in kurzen Reimpaarversen verfasst (von kleinen Abweichungen wie dreiversigen Abschnittschlüssen, wie sie bspw. der ›Wigalois‹ kennt,

abgesehen). Die formale Gleichförmigkeit wird mit Wolframs ›Titurel‹ zunächst nur punktuell und erst später auch von Prosa-Texten herausgefordert. Möchte man sagen, dass es sich bei den *Trois Matières* also um jeweils eigenständige Gattungen handelt, hat man bisher nur inhaltliche Argumente. Wir fragen nun danach, ob die Quantität der Verwendung bestimmter Stilmittel – nämlich der Metapher sowie der Wortstellungsfiguren Chiasmus und Parallelismus – mit dem jeweils adaptierten Stoffkreis distinkt korreliert oder ob bezüglich der Stilmittel andere Parameter wie Verfasser, Entstehungszeit oder Szenentyp bestimmend sind. Den quantitativen Analysen folgen qualitative Anschlussuntersuchungen zu Form und Funktion der Stilmittel. Im vorliegenden Artikel wird nur auf bisherige Ergebnisse zu Chiasmen und Parallelismen eingegangen.

Wenn sich die jeweiligen Stoffkreise der *Trois Matières* – Antikenroman, Karlsepic und Artusroman (samt der Tristan-Texte) – hinsichtlich des Gebrauches von Wortstellungsfiguren in Form, Funktion und Häufigkeit unterscheiden und Einflüsse anderer Parameter wie Verfasserschaft und Entstehungszeit weniger zu beobachten sein sollten, könnte dies als ein Gattungssignal gewertet werden. Zu erwägen wäre dann also, ob für bestimmte Stoffkreise, möglicherweise aufgrund des Usus der altfranzösischen Vorlagen, etwa Parallelismen dem Chiasmus vorgezogen wurden. Das in mittelalterlicher Bildung zentrale Konzept des rhetorischen *aptum* legt auch nahe, dass bestimmte Themen innerhalb der Texte wie etwa Kämpfe rhetorisch elaborierter sind als andere (Ehrismann 1919). Dann würde nicht der Stoffkreis die Verwendung von Wortstellungsfiguren mitbestimmen, sondern einzelne Szenen oder Textabschnitte wie möglicherweise Kampf- oder Festszenen.⁴ Dass die Ebene der Stilmittel nur einen Bruchteil von möglichen szenentypischen und/oder gattungsrelevanten Merkmalen bildet (neben beispielsweise narratologischen Kriterien von Zeit, Modus und vielen weiteren), ist selbstverständlich.

Das Corpus unseres Projektes umfasst folgende dreißig Texte⁵:

Matière de Bretagne	Matières de France	Matière de Rome
Crône (CRO)	Karl (KAR)	Ulrich v. Etzenbach: Alexander (AXU)
Daniel (DA)	Rennewart (REN)	Rudolf von Ems: Alexander (AXR)
Erec (ER)	Rolandslied (ROL)	Straßburger Alexander (AXS)
Garel (GAR)	Willehalm (WH)	Vorauer Alexander (AXV)
Gauriel (GL)	Alischanz (WUT)	Eneas (ENE)
Iwein (IW)		Trojanerkrieg (TRO)
Lanzelet (LZ)		Lied von Troja (TRY)
Mantel (MAN)		Göttweiger Trojanerkrieg (GWTK)
Meleranz (MEL)		
Parzival (PZ)		
Tandareis und Flordibel (TAN)		
Wigalois (WGL)		
Wigamur (WGM)		
Eilharts Tristan (EIL)		
Tristan (TR)		
Heinrich v. Freiberg Tristan (TRH)		
Ulrich von Türheim Tristan (TRU)		

Tabelle 1: Corpus

Im Folgenden sollen die quantitativen Befunde analysiert werden. Dabei werden wir zunächst auf die Herausforderungen, die die Modellierung literaturwissenschaftlicher Begriffe wie in unserem Fall Chiasmus und Parallelismus mit sich bringen, und auf unsere technische Umsetzung eingehen. Anschließend werten wir die gefundene Menge an Stilmitteln in den oben genannten dreißig Texten aus. Wir werden zeigen, dass Fragen, die auf Basis computationeller Methoden gestellt werden können, nicht zwangsläufig von traditionelleren Textuntersuchungsmethoden differieren. Dafür kann aber die Hinzuziehung computationeller Methoden den Blick für solche Fragen neu öffnen und scheinbar selbstverständliche Annahmen mit überraschenden Befunden zur Diskussion stellen.

2. Wortstellungsfiguren der Ähnlichkeit

Müssen computationelle Methoden für eine quantitative Untersuchung von Stilmitteln erst entwickelt werden – wie es in unserem Projekt der Fall ist –, liegt nahe, dass eine Eingrenzung auf bestimmte Stilmittel erfolgen

muss.⁶ Dies hat Einschränkungen auf die Aussagekraft der Funde zur Folge. Denn es ist ja durchaus möglich, dass sich Texte bei der Verwendung etwa von Chiasmen ähneln, hinsichtlich der Verwendung von Apokoinu oder auch Paralipsen hingegen stark unterscheiden. Was Textähnlichkeit also ist, ab wann sich Texte auf Ebene der Stilmittelverwendung – darüber hinaus gäbe es ja, wie gesagt, noch die Ebene des Inhalts, des Metrums, der Zeit und viele weitere – ähneln (ab drei Stilmitteln? wenn die Menge an Stilmitteln ähnlich ist, oder doch eher die ›Art‹ der Verwendung?), ist diskutabel.

Was folgt, ist die ressourcenbedingte Einschränkung auf einige wenige Stilmittel: Neben Parallelismen und Chiasmen untersuchen wir derzeit Metaphern, wobei Letztere kein Gegenstand des hier vorliegenden Beitrags sind. Entsprechend ist die Aussagekraft der Befunde als womöglich bedingt einzustufen. Das soll nicht heißen, es könnten keine validen Beobachtungen angestellt werden. Es heißt nur, dass die hier getroffenen Aussagen vorläufig sein können (die Weiterentwicklung unserer Methoden zur Detektion von Stilmitteln bleibt ein Ziel). Denn grundsätzlich gilt, je mehr Variablen zur Untersuchung eines literarischen Phänomens – hier eben mögliche Abgrenzungskriterien der *Trois Matières* – herangezogen werden, desto wahrscheinlicher die Signifikanz und Plausibilität der Ergebnisse (Underwood 2019). Allerdings legen andere Studien nahe,⁷ dass durchaus als wahrscheinlich gelten kann, dass unsere Ergebnisse ein Indiz für einen grundsätzlichen Trend in Bezug auf die Stilmittelhäufigkeit im Allgemeinen in den von uns untersuchten Textgruppen sein können – aber eben nicht zwingenderweise sein müssen. Unsere Untersuchung liefert also keineswegs keine Ergebnisse, nimmt aber eben – vorerst, das muss hier betont werden, – ›nur‹ Parallelismen und Chiasmen in den Blick.

Die Stilmittel, auf die wir uns beschränkt haben, erfüllen allerdings den zweifachen Anspruch, es wert zu sein, eine computationelle Methode zu entwickeln (Alliterationen etwa wären aus technischer Perspektive nicht lohnend gewesen), und gleichzeitig besonders ›wirksam‹ zu sein.

Beide, Parallelismus und Chiasmus, sind jedenfalls prägende Merkmale mittelhochdeutscher Dichtung und sind auch heute noch Bestandteil von (literarischen) Texten. Andere Stilmittel wie etwa Hyperbaton lassen nur schwerlich textübergreifende Untersuchungen zu. Die Bedeutung, die Parallelismen und Chiasmen haben, gleicht die nur geringe Menge an untersuchten Stilmittelarten partiell aus; es kann wohl mit einigem Recht davon ausgegangen werden, dass eine Untersuchung der drei Stilmittel, die unser Projekt (die Metapher einbezogen) anstrebt, aussagekräftiger ist als eine Untersuchung von mehr Stilmitteln mit geringerer literargeschichtlicher Persistenz.

2.1 Das Verhältnis von Chiasmus und Parallelismus in mittelhochdeutscher Literatur

Chiasmus und Parallelismus sind zwar neuzeitliche Begriffe, dennoch ist das, was sie bezeichnen, bereits in antiken und mittelalterlichen Texten beobachtbar. Es handelt sich um Phänomene der syntaktischen, aber auch semantischen Parallelität und Gegensätzlichkeit – es fand keine begriffliche Trennung zwischen beiden Phänomenen statt, am ehesten treffen die Begriffe ›Isokolon‹ und ›Parison‹ zu. Als solche sind sie nach den im Mittelalter normsetzenden Rhetoriken Mittel der *compositio* (Scaglione/Marvin 1994, Sp. 303).

Sowohl Parallelismus als auch Chiasmus haben eine a) syntaktische und b) semantische Ebene. Im Falle des Parallelismus ist mit syntaktischer Ebene die »syntaktische Äquivalenz zweier oder mehrerer aufeinanderfolgender Sätze oder Satzteile« (Ostrowicz 2003, Sp. 546) gemeint. ›Semantisch parallel‹ meint die Wiederholung »eines Gedankens in zwei oder mehr Satzteilen oder ganzen Sätzen« (Ostrowicz 2003, Sp. 546). Sonderformen des semantischen Parallelismus sind der synonyme und der synthetische Parallelismus. Bei Ersterem wird ein Gedanke mit variierendem Vokabular wiederholt, während sich ein synthetischer Parallelismus dadurch

auszeichnet, dass die im ersten Glied des Stilmittels getroffene Aussage nicht wiederholt, sondern fortgeführt und damit genauer bestimmt wird. Ein weiterer Sonderfall ist der antithetische Parallelismus, der syntaktisch parallele Strukturen aufweist, semantisch jedoch oppositionell aufgestellt ist.⁸ Zuletzt kann bei weiter voneinander entfernten Textstellen und Strukturanalogien auch von einem Parallelismus im weiten Sinne gesprochen werden (Ackermann 2007).

Der Chiasmus wird als komplementär zum Parallelismus gesehen (Fauer 1994, Sp. 171). Die abhängigen Satzteile oder Sätze sind im Gegensatz zum Parallelismus syntaktisch überkreuzt (a–b – b–a), gleiches gilt für den semantischen Chiasmus. Ein Sonderfall des Chiasmus ist die Antimetabole: Die syntaktischen Einheiten sind in diesem Fall nicht nur gespiegelt, sondern erfordern eine exakte Wiederholung der Lexeme.

Die neuzeitliche Distinktion von Wortstellungsfiguren der Spiegelung und (kontrastierenden) Wiederholung in Chiasmus einerseits und Parallelismus andererseits galt für die Rhetoriken, vor deren Hintergrund die mittelalterlichen Texte verfasst wurden, nicht. Dass eine scharfe Trennung auch nur bedingt möglich ist, mag folgendes Beispiel verdeutlichen:

In Gottfrieds ›Tristan‹ heißt es im Prolog programmatisch: *ein man ein wîp, ein wîp ein man, / Tristan Îsolt, Îsolt Tristan* (V. 129f.). Diese zwei Verse decken bereits ein breites Spektrum der oben vorgestellten Stilmittel ab. Einerseits ist jeder der Verse für sich genommen eine Antimetabole – die Lexeme werden gespiegelt. Andererseits liegt auch – liest man die Nomina *man* und *wîp* als zu den Namen *Tristan* und *Îsolt* zugehörig – ein Parallelismus vor: *man* steht zu *Tristan* wie *wîp* zu *Îsolt*. Möchte man die Geschlechter Mann und Frau als ›gegensätzlich‹ lesen, liegt ein antithetischer Parallelismus – der mit einigem Recht auch semantischer Chiasmus heißen könnte – vor; liest man die Spiegelungen als Zeichen der Zusammengehörigkeit, eine Lesart, die angesichts der Handlung wohl plausibel ist, könnte auch von einem synthetischen Parallelismus gesprochen werden. Allein diese zwei Verse zeigen also, dass die Trennung in Chiasmus und Pa-

rallelismus für mittelalterliche Texte (und Dichtung allgemein) nicht immer und nur bedingt Geltung beanspruchen kann. Wir betrachten die mit computationellen Methoden erarbeiteten Ergebnisse daher nicht nur im Einzelnen für das jeweilige Stilmittel, sondern werten sie auch für Aussagen über (natürlich noch weiter zu ergänzende) Wortstellungsfiguren im Allgemeinen aus.

2.2 Modellierung: Was ›ist‹ ein Chiasmus?

Für computationelle Untersuchungen ist die Trennung in Chiasmus und Parallelismus durchaus sinnvoll. Denn um literarische Phänomene computationell untersuchbar zu machen, ist es nötig, sie und die zur Untersuchung herangezogenen Variablen mittels eines Modells genauestens zu profilieren. Das untersuchte ›Phänomen‹ ist hier im weiten Sinne Textähnlichkeit, konkret meint es für uns die Möglichkeit einer Gattungstrias; die Wortstellungsfiguren sind die Variable, anhand derer Aussagen über das Phänomen getroffen werden sollen; im Kontext der Modellierung sind die Wortstellungsfiguren bereits ›Phänomen‹. Verschiedene Modelle des gleichen Phänomens haben je nach intendiertem Zweck unterschiedliche Eigenschaften (Jannidis 2017, S. 100). Die Modellierung literarischer Sachverhalte bewegt sich dabei auch oftmals in einem Spannungsverhältnis aus technischer Machbarkeit und Aussagekraft für literaturwissenschaftliche Fragestellungen.

Exemplarisch sei hier auf eine Textstelle im ›Straßburger Alexander‹ verwiesen. Im Konflikt zwischen Alexander und Darius schicken sich beide Herrscher Gewürze in Form von Mohn bzw. Pfeffer – die Reaktionen auf die ›Geschenke‹ könnten unterschiedlicher nicht ausfallen. Während Alexander Darius' Mohn fast schon genussvoll zu sich nimmt, krampft Darius nach Verzehr weniger Pfefferkörner (vgl. ›StA‹, V. 2075, 2097). Die Passagen umspannen über 50 Verse (vgl. ›StA‹, V. 2046–2117). Dazu ist der Ablauf zwar nach dem Muster Gewürzerhalt-Freude – Gewürzerhalt-Wut

gespiegelt, so dass gewissermaßen von einem antithetischen Parallelismus im weiteren Sinne geredet werden könnte. Hinzu kommt, dass dies nicht der erste Gabenaustausch der beiden Kontrahenten war. Auch im ersten Gabenaustausch geht Alexander als ›Sieger‹ hervor – hier liegt also wiederum eine parallelistisch anmutende Wiederholung vor. Mehrere hundert Verse Umfang und das Erkennen-Können-Müssen, dass bei dieser Distanz Äquivalenzbeziehungen wie hier ›mehrmaliger Gabenaustausch‹ vorliegen, in das Stilmittelmodell zu integrieren, wäre aufgrund technischer Hürden nicht zielführend gewesen. Nicht zuletzt aufgrund eines fehlenden Syntax-Parsers für das Mittelhochdeutsche sind komplexere Konfigurationen der Wortstellungsfiguren, die über Wortarten hinausgehen, wohl (noch) nicht als Programm umsetzbar. Beide Textstellen sind jedoch auch auf der Mikroebene stark durch Wortstellungsfiguren überformt (Brandes 2022); diese automatisiert zu erkennen genügt, um die grundsätzliche Tendenz – dass, wie weiter unten genauer ausgeführt, der ›Straßburger Alexander‹ als Vertreter der Antikenromane wesentlich parallelistisch organisiert ist – zu erfassen, auch wenn nicht alle Strukturen, in denen der Text organisiert ist, erfasst werden können.

Da unsere Modellierungsarbeiten bei den Vorarbeiten der Stilmittel-detektion ansetzten – die sich primär mit der Antimetabole, wie oben beschrieben einem Spezialfall des Chiasmus, auseinandersetzen (Dubremetz/Nivre 2017) –, erläutern wir hier zunächst unser Chiasmus-Modell. In unserem Fall sollte also das Modell von ›Chiasmus‹ einerseits so ausgestaltet sein, dass eine spätere automatisierte Detektion tatsächlich machbar ist, und musste andererseits der literarischen Komplexität der Texte gerecht werden können. Relevant ist hier, dass unser Chiasmus-Modell nicht zwangsläufig auch für andere Arbeiten, die sich mit Chiasmen befassen, richtig sein muss. Ein Modell enthält immer nur die Eigenschaften des modellierten Gegenstandes, die für die Untersuchung maßgeblich sind (Jannidis 2017, S. 100).

Schließlich galten für unser Chiasmus-Modell folgende Merkmale als wesentlich: 1) ein Chiasmus ist eine Textstelle mit bis zu 30 Token Umfang,⁹ 2) innerhalb dieser 30 Token gibt es vier ›stilmitteltragende‹ Wörter; stilmitteltragend meint, dass diese vier Wörter notwendig sind, um als betreffendes Stilmittel erkannt zu werden, 3) zwischen dem ersten und vierten sowie zweiten und dritten Wort herrscht eine syntaktische Ähnlichkeitsbeziehung und/oder eine semantische Ähnlichkeits- bzw. Kontrastbeziehung.

2.3 Technische Umsetzung

Die technische Umsetzung wird detailliert in Schneider ([u. a.] 2021) erläutert. Der bis dahin aktuelle *State-of-the-Art* der Chiasmusdetektion galt nur für den Spezialfall der Antimetabole. Für die Detektion von Chiasmen im Allgemeinen kann eine Wiederholung von Lexemen nützlich sein, reicht allein jedoch nicht aus. Unser Stilmittelmodell (s. o.) umfasst daher noch weitere syntaktische und semantische Informationen.

Auf syntaktischer Ebene war es möglich, mittels eines Part-of-Speech-Taggers für das Mittelhochdeutsche (Echelmeyer [u. a.] 2017) nach invertierten Wortarten gemäß dem Muster $a-b - b-a$ zu suchen. Um die semantische Ebene berücksichtigen zu können, trainierten wir *Word Embeddings* für die mittelhochdeutsche Sprache.¹⁰ Um die Informationen, die die *Word Embeddings* bereitstellen, zu nutzen, haben wir Beispiele der gesuchten Stilmittel manuell in mittelhochdeutschen Texten annotiert. Zur Annotation gehörte auch die Markierung der ›stilmitteltragenden‹ Wörter, also jener Wörter, die konstitutiv für das jeweilige Stilmittel sind. Von diesen Wörtern haben wir den Cosinus-Abstand der Wortvektoren (s. Abb. 1) als vom Klassifikator, der die von der Maschine ausgegebenen Kandidaten mit einem Score versieht, zu berücksichtigendes Merkmal angegeben.

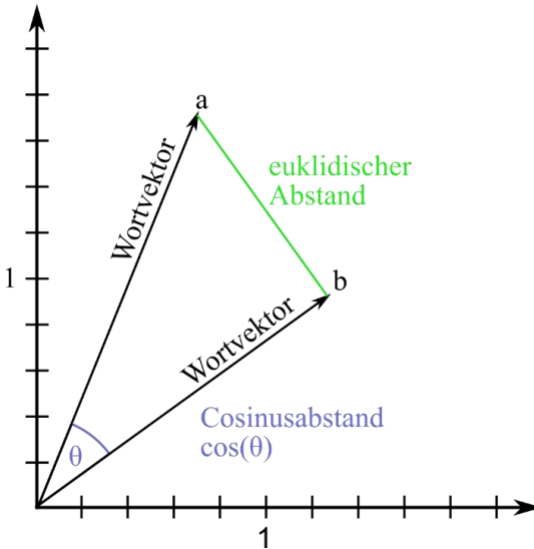


Abb. 1: Cosinusabstand und Wortvektoren

Die Idee war es, bei einer automatisierten Detektion von stilmitteltragenden Wörtern jenen Kandidaten ein höheres Ranking einzuräumen, bei denen der Cosinus-Abstand der Wortvektoren der stilmitteltragenden Wörter möglichst nah an dem unserer manuell annotierten Stilmittel war. Die These, die für dieses Vorgehen sprach, lautet also: Die Cosinus-Abstände der Wortvektoren von stilmitteltragenden Wörtern in Chiasmen sind bei allen Chiasmen tendenziell ähnlicher als die Cosinus-Abstände zufällig ausgewählter Wörter mit invertierter Wortart.

Es handelt sich bei der Nutzung unserer *Word Embeddings* also um ein *Machine-Learning*-Verfahren. Diese Verfahren haben sich in den *computational literary studies* (und auch in benachbarten Disziplinen wie der Computerlinguistik) als besonders leistungsfähig erwiesen (Kuhn 2020, S. 13f.). *Machine-Learning*-Verfahren setzen ›Trainingsmaterial‹ voraus, ein für die jeweiligen Zwecke geeignetes Set an Datensätzen, das Muster enthält, anhand derer die Maschine bessere Ergebnisse erzielen kann. In unserem Fall handelt es sich dabei um die manuell annotierten Daten.

Grundsätzlich gilt: je größer und korrekter der Datensatz, desto besser sind spätere Resultate. *Machine-Learning*-Verfahren sind also wesentlich mit von den Ressourcen, die zur Verfügung stehen, und den Datensätzen, die eingegeben werden, abhängig. Wären mehrere Millionen Texte vorab manuell annotiert worden – was eine *low-resource-language* wie das Mittelhochdeutsche nicht zulässt, selbst unannotierte Daten für unüberwachtes Lernen gibt es nicht in der dafür benötigten Größenordnung –, wären spätere Resultate wohl schneller und/oder besser erreicht worden. Auch hieraus folgt also die ressourcenbedingte Vorläufigkeit der Ergebnisse. Glücklicherweise gilt dies jedoch stärker für die Details der Ergebnisse als für die allgemeinen Trends, die aufgezeigt werden.

2.4 Optimierung des Verfahrens

Den *State-of-the-Art* der Antimetabole-Detektion konnten wir verbessern (Schneider [u. a.] 2021). Mit einer Quote von 35% innerhalb der besten einhundert Funde war die Aussagekraft über literarische Texte dennoch gering. Ziel war es in der Folge nicht mehr, den zugrundeliegenden technischen Hintergrund zu verändern, sondern die Erfolgsrate durch Entfernen von dem, was einen negativen Fund wahrscheinlicher macht, zu steigern.

Eine der ausgegebenen Textstellen ist folgende aus dem ›Tristan‹: *der gottinne Minne. / zer fossiure [oben] inne / [dâ] wâren cleiniu vensterlîn / durch daz licht gehouwen in, / diu lûhten [dâ] unde hie. / [dâ] man ûz und in gie* (V. 16723–16728) – sie steht exemplarisch für eine Vielzahl von Funden, die keinem der gesuchten Stilmittel entsprechen. Die Wörter in eckigen Klammern wurden maschinell mit der Wortart ›Adverb‹ als stilmitteltragend markiert. Es fiel aufgrund zahlreicher ähnlicher Beispiele auf, dass sich die gesuchten Stilmittel überproportional häufig aus Wortarten wie Nomen, Adjektiv und Verb zusammensetzten und dass hingegen Wortarten wie Pronomina, Hilfsverben und eben auch Adverbien

nahezu immer in ausgegebenen Textstellen vorkamen, die den gesuchten Stilmitteln nicht entsprachen. Ein Problem mit Funktionswörtern wie etwa Hilfsverben war die häufige Ausgabe von Textstellen nach folgendem Muster: *[Dâ] [wâren] gesazt under / starkir radere viere. / Starker elfentiere sehs unde drîzich / (daz [was] [vi] hêrlîch)* (›StA‹, V. 6108–6112) – die Spiegelung aus Adverb–Hilfsverb–Hilfsverb–Adverb veranlasste, einen relativ hohen Score anzusetzen, obwohl keines der gesuchten Stilmittel vorlag.

Um diesem Problem zu begegnen, haben wir in einem iterativen Verfahren eine Liste von POS-Tags und auch einzelner Lemmata zusammengetragen, die überproportional häufig in falschen Ergebnissen mündeten (entweder weil sie häufig falsch getaggt wurden oder schlicht keine Beispiele bekannt wurden, in denen sie den gesuchten Stilmitteln entsprachen, dazu gehört beispielsweise das Lemma ›ër‹) und die wir daher nicht mehr berücksichtigt haben. Dieses Vorgehen birgt freilich die Gefahr, Texten, die bevorzugt Stilmittel mit der Zusammensetzung ›Adverb–Pronomen – Pronomen–Adverb‹ enthalten, nicht gerecht zu werden. Denn Stilmittel mit dieser Zusammenstellung würden selten bis gar nicht in den Top 100-Ergebnissen angezeigt – so hätten wir, ohne es zu wissen, unseren Ansatz systematisch an Texten ausgerichtet, die ohnehin bereits weiter oben im Ranking platziert waren. Um derartige Fehler zu vermeiden, haben wir die Ergebnisse von Einzeltexten, die niedrige Scores auswiesen, analysiert, ohne bestimmte POS-Tags und Lemmata auszuschließen, und konnten feststellen, dass keiner der untersuchten Texte durch unsere gefilterte POS-Tag-Liste benachteiligt wurde.

Parallelismen und Chiasmen, das kann als – wenn auch wenig überraschendes – Zwischenergebnis festgehalten werden, scheinen sich in überwiegender Mehrheit durch bedeutungstragende Wortarten zu konstituieren, während Wortarten wie Pronomina tendenziell kein konstitutiver Bestandteil der Stilmittel sind. Dabei bleibt natürlich zu betonen, dass es durchaus Stilmittel mit diesen Wortarten geben kann, sie bei einer Trend-

analyse auf Basis einer großen Textmenge allerdings keinen oder einen zu vernachlässigenden Einfluss auf die zu beobachtenden Trends haben.

Die Parallelismusdetektion wurde erst im Anschluss an die funktionierende Chiasmusdetektion begonnen. Einziger Unterschied ist hier, dass nicht auf eine invertierte Reihenfolge der Wortarten, sondern eine sich wiederholende Reihenfolge nach dem Muster A-B-A'-B' geachtet wurde. Für die Top-100-Ergebnisse erzielten wir so eine Erfolgsrate von über 90%.

3. Quantitative Analyse

Die folgenden Tabellen und Abbildungen stellen aktuelle Ergebnisse der Stilmitteldetektion dar. Ausgegeben wurden Textstellen, die gewichtet nach den oben genannten Features – syntaktische und semantische Relation, Wortumfang, Lemma-Wiederholung – mittels eines Klassifikators mit einem bestimmten Score gerankt wurden. Je höher der Score, desto eher handelt es sich bei der ausgegebenen Textstelle um eines der gesuchten Stilmittel. Bis zu einem Score von 11 können wir davon ausgehen, dass es sich bei den ausgegebenen Stilmitteln mit einer Wahrscheinlichkeit von mindestens 50% um ein korrekt erkanntes Stilmittel handelt. Um die Ergebnisse aussagekräftig zu halten, haben wir also nur solche berücksichtigt, die eben einen Score von ≥ 11 aufweisen.

Rang	Text	Vers/Stilmittel
1	MAN	497
2	TRH	861
3	TR	1222
4	TRU	1250
5	ER	1274
6	TRY	1318
7	KAR	1356
8	AXR	1546

Score Relativ	Text	Rang
25,64	WGM	1
22,55	GAR	2
18,4	MAN	3
8,1	TR	4
7,17	TRU	5
7,06	TRY	6
6,78	TRH	7
5,27	DA	8

Brandes [u. a.]: Stilfiguren aus der Distanz gelesen

9	TRO	1608	4,66	ER	9
10	DA	1696	4,63	KAR	10
11	GAR	1938	3,84	ROL	11
12	GL	2083	3,01	AXR	12
13	IW	2722	2,97	TRO	13
14	PZ	2757	2,94	REN	14
15	WGL	2928	2,79	AXS	15
16	AXU	3500	2,74	IW	16
17	WUT	3535	2,58	ENE	17
18	TAN	3610	2,44	GL	18
19	MEL	4281	1,62	WUT	19
20	WGM	6115	1,56	MEL	20
21	ENE	6784	1,51	CRO	21
22	WH	7001	1,38	TAN	22
23	AXS	7098	1,34	PZ	23
24	REN	7300	1,28	AXU	24
25	ROL	9094	1,27	WGL	25
26	CRO	30077	1,06	WH	26
27	GWTK	–	0	GWTK	27
28	LZ	–	0	LZ	28
29	AXV	–	0	AXV	29
30	EIL	–	0	EIL	30

Tabelle 2: Chiasmusdetektion

Die linke Hälfte der Tabelle ist nach Versanzahl je Stilmittel sortiert. Erkennbar ist, dass weder Entstehungszeit noch Autorschaft (auch wenn es mit Hartmann, Wolfram und dem Pleier nur drei Autoren gibt, die mehr als einen der hier aufgeführten Texte verfasst haben) oder Stoffkreis einen Einfluss darauf zu haben scheinen, wie viele Chiasmen im Text vorkommen. Allenfalls sind leichte Tendenzen, was den Stoffkreis anbelangt, erkennbar. So sind in drei der vier Tristan-Texte vergleichsweise viele Stil-

mittel gefunden worden, lediglich der Text Eilharts weicht hier stark ab. Auch dass die karlsepischen Texte (mit Ausnahme des ›Karl‹) weiter unten in der Tabelle platziert sind, ist erkennbar.

Die rechte Hälfte der Tabelle ist nach dem relativen Score sortiert. Dieser sagt aus, wie viele aller gefundenen Stilmittel mit einem Score ≥ 11 bewertet wurden. Die Texte mit einem hohen relativen Score weisen also im Verhältnis viele mit erhöhter Wahrscheinlichkeit richtige Funde auf. Auch hier ist erkennbar, dass Texte des Bretagne-Stoffkreises das obere Drittel dominieren. Solche Resultate sind gemeint, wenn in Abschnitt 2 von der Notwendigkeit möglichst vieler Variablen, die zur Beantwortung einer literaturwissenschaftlichen Fragestellung herangezogen werden, die Rede war. Obwohl die Reihenfolge der Texte in beiden Tabellenhälften im Detail unterschiedlich ist, ist der grundsätzliche Trend – Texte aus dem Bretagne-Stoffkreis weisen mehr Chiasmen auf – in beiden Tabellen erkennbar.

Auffällig ist auch, dass in einigen Texten keine der ausgegebenen Textstellen einen Score ≥ 11 aufweist. Zu diesen gehört auch der ›Vorauer Alexander‹. Bedenkt man, dass dieser in weiten Teilen den ersten rund 1500 Versen des ›Straßburger Alexander‹ entspricht, ließe sich für Letzteren folgern, dass die ›Chiastizität‹ des Textes im zweiten Textabschnitt begründet liegt. Tatsächlich zeigt eine unserer qualitativen Arbeiten (Brandes 2022), dass der Beginn des ›Straßburger Alexander‹ eher von (auch antithetischen) Parallelismen geprägt ist.

Rang	Text	Verse/ Stilmittel
1	TRH	492
2	DA	652
3	AXV	767
4	AXR	1082
5	AXS	1183
6	KAR	1220
7	TRY	1230

Score Relativ	Text	Rang
26,9	AXV	1
15,5	AXS	2
12,3	DA	3
11,2	TRH	4
10,9	ROL	5
8,3	ENE	6
7,6	GAR	7

8	ER	1274	7,3	TRY	8
9	IW	1361	6,3	TR	9
10	TR	1504	5,9	REN	10
11	TRO	1511	5	IW	11
12	ENE	1938	4,8	KAR	12
13	PZ	2481	4,4	ER	13
14	WGL	2928	4	AXR	14
15	TAN	3008	3	TRO	15
16	ROL	3031	2,3	TRU	16
17	AXU	3111	2	WH	17
18	REN	3318	1,5	TAN	18
19	WH	3501	1,4	PZ	19
20	TRU	3751	1,4	CRO	20
21	GAR	5330	1,3	AXU	21
22	MEL	6421	1,2	WGL	22
23	WUT	10604	1	MEL	23
24	CRO	30077	0,5	WUT	24
25	GL	–	0	GL	25
26	MAN	–	0	MAN	26
27	GWTK	–	0	GWTK	27
28	WGM	–	0	WGM	28
29	LZ	–	0	LZ	29
30	EIL	–	0	EIL	30

Tabelle 3: Parallelismusdetektion

Betrachtet man die Parallelismen-Ergebnisse, ergibt sich ein etwas anderes Bild. Antikenromane sind nun mehrheitlich in der oberen Hälfte zu finden – drei Alexanderromane sogar innerhalb der ersten fünf Plätze. Die Platzierungen der Artusromane fallen nun – konträr zu den Antikenromanen – insgesamt niedriger aus. In Bezug auf karlsepische Texte bleibt das Bild ähnlich. Der ›Karl‹ des Strickers weist erneut die meisten Funde auf; in den

übrigen Texten dieses Stoffkreises ließen sich vergleichsweise wenige der gesuchten Stilmittel finden.

Sortiert nach relativem Score, zeichnet sich der gleiche Trend – bei Unterschieden in den Details – ab: Texte der *Matière de Rome* rangieren deutlich weiter oben als die übrigen Texte. Die quantitativen Funde bestätigen damit unsere qualitative Analyse (Brandes 2022). Das ist insofern gut, als es zeigt, dass unser doch ›reduziertes‹ Modell der Stilmittel – so entfallen etwa die Möglichkeiten eines Syntax-Parsers – uns nicht daran hindert, die grundsätzliche Tendenz eines Textes bezüglich seiner Stilmittelhäufigkeit erkennen zu können.

Die hier dargelegten Beobachtungen sind noch leichter erkennbar, wenn man die Funde nach Stoffkreisen sortiert:

Stoffkreis	rel. Score Parallelismen	rel. Score Chiasmen	Stoffkreis
Antikenroman	8,3	6,5	Bretagne
Karlsepik	4,8	2,8	Karlsepik
Bretagne	3,3	2,5	Antikenroman

Tabelle 4: Relativer Score der Stoffkreise

Hier ist deutlich erkennbar, dass die Antikenromane zum Parallelismus zu tendieren scheinen, während die Artustexte vorsichtig als ›Stoffkreis des Chiasmus‹ bezeichnet werden könnten. Die Karlsepischen Texte nehmen – wohl dank Strickers ›Karl‹ – stets einen Mittelplatz ein. Generell wurden mehr Parallelismen als Chiasmen detektiert. Das liegt – so unsere Vermutung – in einem Detail unserer Detektion begründet: Geachtet wird, wie oben erläutert, darauf, ob das erste und dritte sowie das zweite und vierte (oder im Fall des Chiasmus das erste und vierte bzw. zweite und dritte, aber das ist hier nicht relevant) stilmitteltragende Wort von gleicher Wortart sind. Es gibt jedoch keinen Mechanismus, der ausschließt, dass alle Wörter

die gleiche Wortart haben können – einfach, weil Beispiele wie *ein man ein wîp, ein wîp ein man* (>TR<, V. 129) dann ausgeschlossen würden. Dies hat zur Folge, dass bei der Parallelismus-Detektion einige synonyme und synthetische Parallelismen detektiert wurden, obwohl diese Sonderform des Parallelismus nicht zwangsläufig im Fokus unserer Bemühungen lag. Solche >Reihungen< (von der Aufzählung unterscheiden sie sich nur durch das Merkmal der syntaktischen Abhängigkeit) sind natürlich >einfach< anzusetzen und kommen in den Texten vielfach vor.

Wie in Abschnitt 2.1 erwähnt, wird man den mittelhochdeutschen Texten jedoch nur bedingt gerecht, wenn man die untersuchten Wortstellungsfiguren streng in Parallelismus und Chiasmus einteilt. Dies zeigen auch zahlreiche Funde, die – je nach Auslegung der Textstelle – sowohl als Chiasmus als auch als Parallelismus bezeichnet werden könnten. Zusätzlich zum oben genannten Beispiel aus dem >Tristan< sei folgende Passage genannt, die maschinell detektiert wurde: *der [arme] und der [reiche] der [junge] und der [alte]*. Diese für das Mittelhochdeutsche fast schon topische Formel, die schlicht >alle< meint, stammt in diesem Fall aus dem >Rennewart< (V. 25806f.). Natürlich liegt hier syntaktisch gesehen ein Parallelismus vor. Semantisch ist der Fall nicht ganz so eindeutig. Für sich genommen stellen *arme* und *reiche* sowie *junge* und *alte* jeweils eine Antithese dar. Den gesamten Textauszug semantisch chiastisch zu deuten – nämlich so, dass >arm und alt< also >reich und jung< umklammern, wäre plausibel, wenn man die Setzung, dass >arm< und >alt< hier pejorativ konnotiert sind, wohingegen >reich< und >jung< meliorativ verwendet werden, mitträgt.

Dass die Trennung in Chiasmus und Parallelismus nur bedingt ihren Sinn hat, macht auch der Sonderfall des antithetischen Parallelismus deutlich. Während die Syntax parallelistisch organisiert ist, stellt sich die Semantik – in einigen Fällen auch chiastisch – entgegen. Wir haben uns daher – wegen der im Mittelalter nicht vorhandenen Trennung in Parallelismus und Chiasmus und den diese Dichotomie ohnehin unterlaufenden Sonder-

fällen wie dem antithetischen Parallelismus – dafür entschieden, unsere Ergebnisse auch zu Wortstellungsfiguren insgesamt darzustellen:

Rang	Text	Vers/Stilmittel
1	TRH	405
2	MAN	497
3	DA	565
4	AXR	722
5	AXV	767
6	TRY	769
7	TR	782
8	ER	849
9	KAR	872
10	TRU	938
11	TRO	1061
12	IW	1167
13	AXS	1183
14	PZ	1460
15	GAR	1640
16	WGL	1673
17	ENE	1696
18	AXU	2000
19	TAN	2005
20	GL	2083
21	WH	2334

Score Relativ	Text	Rang
18,8	AXV	1
17,4	GAR	2
16,9	WGM	3
12,2	MAN	4
10,6	AXS	5
9,6	TRH	6
9,5	DA	7
8,7	TRY	8
8,6	TR	9
7,9	ROL	10
6,4	ENE	11
5,8	TRU	12
5,6	REN	13
4,6	KAR	14
4,4	ER	15
4,3	AXR	16
3,9	IW	17
3,1	TRO	18
2,2	WH	19
1,9	CRO	20
1,7	PZ	21

22	REN	2433	1,6	GL	22
23	WUT	2651	1,5	WGL	23
24	ROL	3031	1,5	TAN	24
25	MEL	3211	1,4	AXU	25
26	WGM	6115	1,4	WUT	26
27	CRO	15039	1,3	MEL	27
28	GWTK	–	0	GWTK	28
29	LZ	–	0	LZ	29
30	EIL	–	0	EIL	30

Tabelle 5: Wortstellungsfiguren

Doppelte Funde, also solche, die sowohl bei der Suche nach Chiasmen als auch nach Parallelismen detektiert wurden, wurden nur einfach – mit dem je höchsten Score – gezählt. Das bisherige Bild bestätigt sich. Es gibt kein eindeutiges Gattungssignal, allerdings ansatzweise eine Tendenz. Im oberen Drittel gibt es mit dem ›Karl‹ nur einen karlsepischen Text, die übrigen neun Plätze werden von Antikenromanen und Artusromanen besetzt. Auch wenn karlsepische Texte nicht weit oben platziert sind, so gibt es doch zumindest für jeden der karlsepischen Texte Resultate – keine Funde kommen bei karlsepischen Texten nicht vor. Mit Ausnahme des ›Karl‹ folgen diese Texte in der linken Tabellenhälfte sogar direkt aufeinander (Rang 21–24), sind sich also ähnlich; die drei Texte, die zum ›Willehalm‹-Umkreis gehören (Rang 21–23), weisen besonders geringe Abstände zueinander auf. – Hier kann sehr vorsichtig vermutet werden, dass die gemeinsame *materia* der Texte – denn Autorschaft und Entstehungszeit können es nicht sein – Auswirkungen auf die Stilmittelhäufigkeit hat. Dieses Bild wird allerdings von den weiter auseinander platzierten Alexandertexten unterlaufen. Doch nur weil es kein oder nur ein schwaches Gattungs- oder Stoffsignal in den Antiken-

romanen gibt, heißt das nicht, dass dies auch für karlsepische Texte gelten muss. Zu bedenken ist in diesem Zusammenhang auch die formal von den anderen Gattungen abweichende Gruppe der altfranzösischen Karlsepik (zunächst in Laissen gedichtete *chansons de geste*). Möglicherweise hat die intensive Auseinandersetzung der deutschen Verfasser mit diesen Texten auf die eigene Wahl von Stilmitteln eingewirkt.¹¹

In Bezug auf den relativen Score fällt auf, dass zwei der vier oben platzierten Texte, der ›Vorauer Alexander‹ und das ›Mantel‹-Fragment, zu den kürzesten Texten des Corpus gehören. Nur sehr wenige Funde mit einem Score ≥ 11 genügen also, um einen hohen relativen Score aufzuweisen. Hier kann (zumindest teilweise) Verzerrung aufgrund der Textlänge vermutet werden. Auch hier gilt: Die Reihenfolge der Texte mag im Detail unterschiedlich sein, die grundsätzliche Verteilung der Stoffkreise bleibt ähnlich.

Insgesamt, so kann vorerst konstatiert werden, ist allein die Häufigkeit von Stilmitteln der Wortstellungsfiguren kein eindeutiger Indikator für die Zugehörigkeit zu einem der Stoffkreise. So haben Texte der Karlsepik etwa vergleichsweise wenig Funde, aber nicht alle Texte mit verhältnismäßig wenigen Funden gehören der Karlsepik an. Hinzu kommt der geringe Anteil karlsepischer Texte am Corpus sowie der mit ›Karl‹ selbst betitelte Text, der innerhalb seines Stoffkreises immer die Ausnahme bildet. Und doch, so kann das Ergebnis zumindest mit Blick auf die Texte Wolframs und des Pleiers auch vorsichtig formuliert werden, gruppieren sich die einander ähnlichen Texte eher nach dem Stoffkreis als nach der Autorschaft.

Wenn auch die bloße Quantität der untersuchten Stilmittel wohl kein Gattungssignal erkennen lässt, so lassen die Funde möglicherweise doch noch weitere Aussagen über die Texte zu. Hierzu folgende Abbildung:

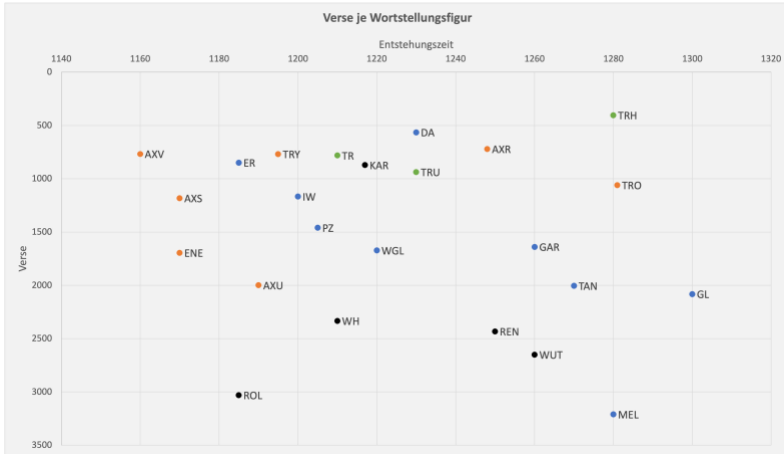


Abb. 2: Verse je Wortstellungsfigur

In dieser Abbildung fehlen die Texte ohne Funde (›Göttweiger Trojanerkrieg‹, ›Lanzelet‹ und Eilharts ›Tristrant‹) sowie ›Wigamur‹ und ›Crône‹, da beide Texte die übrigen in einer linearen Darstellung zu nah aneinanderrücken würden. Die vollständige Abbildung kann im [hier](#) eingesehen werden.

Abgebildet ist, mit welcher Anzahl von Versen Abstand (y-Achse) ein Stilmittel detektiert wurde und wann¹² (x-Achse) der Text entstanden ist. Erkennbar ist, dass früheste Texte ein gewisses Mindestmaß an Stilmitteln zu enthalten scheinen. Erst im ›Rolandslied‹ und in nur wenigen Texten nach ihm wurde \geq alle 2000 Verse ein Stilmittel entdeckt. In der überwiegenden Mehrheit der Texte konnte alle 500 bis 2000 Verse ein Stilmittel detektiert werden. Davon weichen nur neun Texte ab. In acht dieser Texte ist die Zahl ›benötigter‹ Verse größer, davon sind vier Texte der Karlsepiik zuzuordnen, die übrigen der Artusepiik, wobei hier keiner der ›Klassiker‹ und mit ›Meleranz‹ und ›Gauriel‹ die zwei spätesten Artusromane betroffen sind.

Die Dimension ›Entstehungszeit‹ prädeterminiert zwar nicht, ob ein Text viele oder wenige Stilmittel aufweist, dennoch ist ein grundsätzlicher

Trend erkennbar: Später geschriebene Texte enthalten weniger Stilmittel,¹³ wobei auch hier die Ausnahme in Form von Heinrichs ›Tristan‹ die Regel bestätigt. Es ist jedoch nicht so, dass die späten Texte mit wenigen Stilmitteln, etwa der ›Meleranz‹, erheblich weniger Stilmittel hätten als frühere Texte mit wenigen Funden, etwa das ›Rolandslied‹; beide liegen auf der y-Achse nah beieinander – es fehlt in der zweiten Hälfte des 13. Jh. also schlicht eine gewisse Anzahl an Texten, in denen alle 1000 oder 1500 Verse ein Stilmittel gefunden wurde. Separiert man die Abbildung nach Stoffkreis, wird klar, dass der ›abnehmende‹ Trend nicht für alle Stoffkreise gilt:

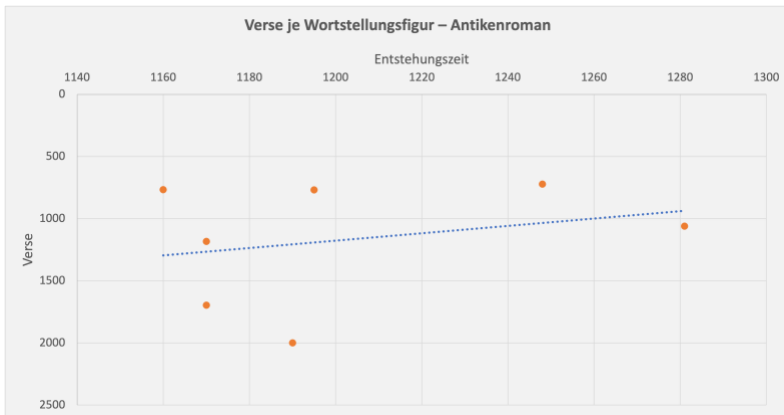


Abb. 3: Verse je Wortstellungsfigur – Antikenroman (lineare Trendlinie)

Im Gegenteil ist in dieser Abbildung, die nur noch die Datenpunkte der Antikenromane enthält, zu sehen, dass sogar ein leichter Aufwärtstrend vorhanden ist. Aufgrund der geringen Textmenge von nur sieben sollte diesem leichten Trend allerdings nicht allzu viel Gewicht beigemessen werden. Noch weniger Datenpunkte weist die Abbildung zur Karlsepike auf:

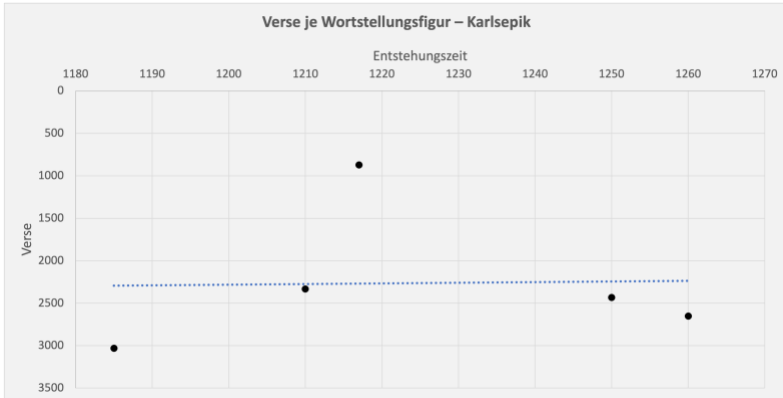


Abb. 4: Verse je Wortstellungsfigur – Karlsepi (lineare Trendlinie)

Hier ist nun kein Trend, allenfalls ein minimaler, so man ihn so nennen möchte, erkennbar. Dass in vier der fünf Texte vergleichsweise wenig Stilmittel gefunden werden konnten, zeigten auch die bisherigen Ausführungen. Die größte Textmenge weist die *Matière de Bretagne* auf, der die folgende Abbildung – hier mit den Ausreißern ›Crône‹ und ›Wigamur‹ – gilt.

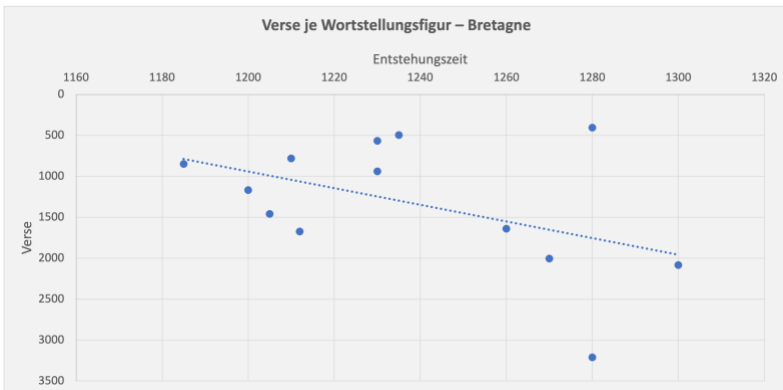


Abb. 5: Verse je Wortstellungsfigur – Bretagne (lineare Trendlinie)

Der in Abbildung 1 sichtbare Abwärtstrend über die Zeit hinweg basiert also im Wesentlichen auf den Texten der Artusromane. Hier ist auch der ›stärkste‹ Trend erkennbar.

Zu fragen wäre nun, wie mit den hier erzielten Ergebnissen weiter verfahren werden soll. Einerseits muss auf ihnen aufgebaut werden, indem die Genauigkeit des Detektors weiter verbessert wird – dazu ist bereits ein iteratives Verfahren geplant, bei dem die Ergebnisse von Durchgang zu Durchgang besser werden, was auch zur Folge hätte, dass alle POS-Tags berücksichtigt werden können und der Fokus auf die Zusammensetzung der Stilmittel erweitert wird. Denn wenn die Häufigkeit nicht oder nur geringfügig an die Stoffkreiszugehörigkeit geknüpft ist, ist dies möglicherweise bei einer quantitativen Auswertung der Wortarten der Fall, aus denen die jeweiligen Stilmittel bestehen. Auch eine Ergänzung um unsere Metaphern-Detektion steht aus.¹⁴

Andererseits bieten die jetzigen Ergebnisse bereits einen explorativen Zugang zu den Texten. Intuitiv würde man Texten wie dem ›Parzival‹, dem ›Willehalm‹, aber auch dem ›Eneas‹ und dem ›Tristan‹ wohl einen rhetorisch ›artifiziellen‹ Stil attestieren – gerade Wolfram und Gottfried gelten als Dichter, deren Sprachvirtuosität heraussticht. Auch sind späte oder ›postklassische‹ Texte wie der ›Garel‹, der ›Gauriel‹ oder auch der ›Lanzelet‹ (auf den das Attribut ›spät‹ gar nicht zutrifft) in Bezug auf ihre literarische Qualität – wozu eben auch Stil gehört – als ›schlicht‹ abgewertet worden (exemplarisch zum ›Gauriel‹ die Zusammenfassung von Achnitz 1997). Die hier gezeigten Tabellen decken diese Positionen der Forschung jedoch nur bedingt ab – wieso?

Zunächst einmal – das ist eingangs bereits gesagt worden – besteht Stil aus mehr als der bloßen Menge der verwendeten Stilmittel. Darüber hinaus beziehen sich die hier vorliegenden quantitativen Ergebnisse auf zwei Stilmittel, nicht auf alle. Nicht zuletzt umfasst unser Modell der Stilmittel eben nicht alle möglichen Ausprägungen der Stilmittel. So könnte etwa für Wolframs Texte vermutet werden, dass die in ihnen verwendeten Wort-

stellungfiguren deutlich komplexer sind, als es unsere Stilmitteldetektion bisher erfassen kann.

Autor, Entstehungszeit, Stoffkreis und ›Szenen‹ stehen, wie oben erläutert, im Verdacht, den Stilmittelgebrauch zu beeinflussen. Für die Texte von ein und demselben Autor – Hartmann, Wolfram und Pleier – war erkennbar, dass durchaus Unterschiede in der Menge der Stilmittelverwendung vorhanden sind. Ein klares Autorsignal ist in dieser Stichprobe nicht bemerkbar. Dies gilt auch für die Entstehungszeit. Auch wenn es – von Stoffkreis zu Stoffkreis unterschiedlich stark ausgeprägt – leichte Tendenzen zu geben scheint, gibt es zu jeder Zeit Texte, in denen besonders viele, aber auch besonders wenige Stilmittelfunde vorliegen. Auch die Stoffkreise selbst lassen keine eindeutigen Cluster erkennen. Karlsepische Texte und zum Tristan-Stoffkreis gehörende Texte scheinen allerdings eine Tendenz zu wenigen bzw. zu vielen Stilmitteln zu haben – eine Tendenz, die dann von den jeweiligen Ausnahmen der Textgruppen (›Karl‹ respektive ›Tristrant‹) unterlaufen wird. Szenen, etwa Figurenrede oder Kämpfe, könnten den Stilmittelgebrauch wohl beeinflussen. Zu einzelnen Szenentypen und deren Verteilung in den Einzeltexten können hier (noch) keine quantitativen Aussagen getroffen werden. Allgemeine Beobachtungen könnten aber erste Vermutungen zulassen. So beinhalten Texte der *chansons de geste* vergleichsweise viele Schlachten. Möglich ist etwa, dass Schlachten weniger Wortstellungsfiguren, aber umso mehr Metaphern enthalten. Dies zu überprüfen, gehört zu den nächsten geplanten Schritten unseres Projektes.

Dieses Resümee sollte neben allen Potenzialen der neuen Methoden die besondere Herausforderung deutlich gemacht haben: Auszuloten, welche Bedeutung den computationell gewonnenen Befunden beigemessen werden kann, ist aus literaturwissenschaftlicher Sicht die wichtigste und schwierigste Aufgabe. Und auch wenn in der Forschung teils stark gemacht wird, dass es nicht nur auf die Menge von Stichproben und Variablen ankommt, da schon einzelne bereits sehr plausibel auf große Trends hin-

weisen könnten, ist klar, dass bei einem literaturwissenschaftlichen Thema wie ›Stil und Gattung‹ mit der quantitativen Untersuchung von Wortstellungsfiguren allenfalls ein wichtiger Anfang gemacht sein kann. Literarische Texte sind zu komplex, als dass eventuell genrebedingte Unterschiede zwischen ihnen an derart wenigen Merkmalen festzumachen wären. Differenzen zwischen den Stoffkreis-Bearbeitungen wird quantitativ wohl nur über Merkmalsbündel beizukommen sein – diese in kleinen Schritten zu erarbeiten, verstehen wir aber als lohnenden Weg.

4. Literaturwissenschaft als Experiment?

Literaturwissenschaft erhält – davon konnte hier nur ein flüchtiger Eindruck vermittelt werden – in Verbindung mit den Möglichkeiten der Statistik und der Informatik (samt benachbarter Disziplinen wie der Computerlinguistik) Experimentcharakter. Bis in die 2010er-Jahre hinein schien das Paradigma der Exploration vorherrschend zu sein. Es galt umzusetzen, was technisch umsetzbar war, um erst dann zu überprüfen, ob und was mit den Ergebnissen über Literatur ausgesagt werden kann (als technische Machbarkeitsstudie kann etwa auch Jockers 2013 gelesen werden).¹⁵

Spätestens mit Ted Underwoods ›Distant Horizons‹ (2019) galten rein explorative Vorgehensweisen, bei denen vor allem auch technische Potenz im Vordergrund stand, als obsolet. Ausgangspunkt für ein *Distant Reading* oder generell für computergestützte Literaturwissenschaft sollte stets literaturwissenschaftliches Hintergrundwissen sein (Underwood 2019, S. 17). Aus diesem bilden sich Interessen und Fragestellungen, die dann experimentell untersucht werden. Es gilt danach zu fragen, welche Variable Aussagekraft für welches Phänomen hat (also hier die Variable ›Stilmittelhäufigkeit‹ für das Phänomen ›Stoffkreiszugehörigkeit‹). Je mehr verschiedene Variablen in Bezug auf ein bestimmtes Phänomen untersucht werden, umso aussagekräftiger sind die zu beobachtenden Trends.

Zum Abschluss soll kurz der Versuch unternommen werden, eine weitere Variable zur Untersuchung von Textähnlichkeiten heranzuziehen. Braun und Ketschik (2019) unternehmen es, anhand einer Netzwerkanalyse, die Figurenbeziehungen untersucht, Aussagen über die Märchenhaftigkeit von Artusromanen treffen zu können. Dabei wird gezeigt, dass der ›Iwein‹ weniger komplex in der Interaktion der Figuren ist als der ›Erec‹ (Braun/Ketschik 2019, S. 59); dies fällt vor allem auf, wenn man beide Figurennetzwerke betrachtet.¹⁶ Der ›Parzival‹ ist wiederum deutlich komplexer als die vorigen Texte. Die Unterschiede fallen so deutlich aus, dass »zunächst einmal die beiden Artusromane Hartmanns dem ›Parzival‹ Wolframs gegenüber [gestellt werden]« (Braun/Ketschik 2019, S. 70). Diese Gegenüberstellung trifft partiell auch auf die von uns gemachten Stilmittelfunde zu, wenn auch auf den ersten Blick in umgekehrter Weise. Denn während ›Erec‹ und ›Iwein‹ einen relativen Score von ~4 aufweisen, kommt der ›Parzival‹ nur auf einen relativen Score von 1,7. Es könnte also sein, dass ›Komplexität‹ gemessen an der Figureninteraktion mit einer spezifischen Verwendung der Stilmittel korreliert, so dass diese bei höherem ›Schwierigkeitsgrad‹ entweder seltener vorhanden sind oder – was bei Wolfram wahrscheinlicher scheint – mit unseren bisherigen Ansätzen schwieriger detektiert werden können, weil sie ›komplexer‹ sind.

Wird unsere Variable also um weitere ergänzt, wie es eben nur angedeutet wurde und wie wir es mit Metaphernhäufigkeiten, Zusammensetzung der Wortstellungsfiguren, inhaltlichem Kontext der Metaphern und Wortstellungsfiguren, Szenentypen-Analysen sowie dem Vergleich mit Texten außerhalb der drei Stoffkreise planen,¹⁷ kann eine für das Mittelhochdeutsche umfassende Studie zu Textgruppenkonstitutionen erwartet werden. Eine solche Studie wäre kein Experimentieren um des Experimentierens willen. Sie wäre den Anforderungen an eine Literaturwissenschaft, die nun immer mehr von computationellen und statistischen Methoden Gebrauch machen kann und sollte, angemessen – denn komplexe Phänomene wie ›Gattung‹ und ›Stil‹ werden nur dann hinreichend erforscht

werden können, wenn von den ebenso komplexen Untersuchungsmethoden, die zur Verfügung stehen (und zukünftig zur Verfügung stehen werden), Gebrauch gemacht wird.

Anmerkungen

- 1 Das Projekt unter der Leitung von Prof. Dr.-Ing. Joachim Denzler und Jun.-Prof. Dr. Sophie Marshall ist Teil des von der DFG finanzierten Schwerpunktprogramms 2207 ›Computational Literary Studies‹. Informationen zu allen Projekten und unseren publizierten Ergebnissen sind hier einsehbar: <https://dfg-spp-cls.github.io/>.
- 2 Was Stil genau ist, variiert in entsprechenden Untersuchungen je nach zugrundeliegender Theorie. Stil wird hier als den Ausdrucksformen funktionszuweisendes Konzept (Hübner 2015) verstanden. Stilfiguren sind demnach Teil dieser Ausdrucksformen. Zur weiteren Diskussion von Stil und Stilkonzepten s. Andersen [u. a.] (2015) und Gumbrecht (1986). Für die nachfolgende Untersuchung soll betont sein, dass Stil nicht allein mit der Verwendung bestimmter Stilmittel gleichgesetzt wird.
- 3 Weitere Schwerpunkte der Stilometrie sind u. a. – wenn auch bei weitem nicht so etabliert wie Autorschaftsattribution – *topic modelling* (Viehhauser 2018) und das noch neue Feld der automatisierten Szeneneinteilung (Zehe [u. a.] 2021). Für die Mediävistik und unser Projekt von Interesse ist vor allem die Beobachtung, dass es nicht den einen Indikator gibt, anhand dessen alle Gattungen untersucht werden können (Viehhauser 2015).
- 4 Eine ›klassisch-analoge‹ Untersuchung aus dem Projekt beschäftigt sich mit Szenen entsprechender Texte, die Gabellogisches verhandeln (Brandes 2022).
- 5 Bereitgestellt werden die maschinenlesbaren Dateien der Texte von der Mittelhochdeutschen Begriffsdatenbank (MHDBDB), der dafür an dieser Stelle herzlich gedankt wird.
- 6 Dass das Detektieren der drei Stilmittel Chiasmus, Parallelismus und Metapher bereits durchaus ambitioniert ist, zeigen allein bisherige Arbeiten in der Chiasmusdetektion. Diese befassten sich nahezu ausschließlich mit dem Sonderfall der Antimetabole (Dubremetz/Nivre 2017) und waren aus computerlinguistischer Sicht zweifellos ein wichtiger Entwicklungsschritt; für literaturwissenschaftliche Untersuchungen sind sie aufgrund der geringen Genauigkeit aber irrelevant.

- 7 Underwood macht klar, dass Korrelation noch längst keine Kausalität darstellt. Ebenso auffällig scheint aber zu sein, dass beobachtbare Trends auf sehr spezieller Ebene (hier die Ebene zweier Wortstellungsfiguren) sich in der Regel auf größerer Ebene wiederfinden. In Underwoods Untersuchung gilt dies etwa für die spezielle Ebene der Farbadjektive, deren häufigeres Vorkommen auf grundsätzlich häufigeres Vorkommen von beschreibenden Adjektiven hinweist (Underwood 2019, insb. S. 11–14). Für unser Experiment würde das bedeuten, der Spezialfall weniger Stilmittel der Wortstellung lässt durchaus Rückschlüsse auf den allgemeinen Gebrauch von Stilmitteln im jeweiligen Werk zu.
- 8 Als Antithese kann hingegen bereits eine bloße Zusammenführung gegensätzlicher Inhalte gelten (Villwock 1992, Sp. 722–750). Hier liegt also die Struktur A–B vor, die Antithese ist in diesem Sinne damit keine der hier untersuchten Wortstellungsfiguren.
- 9 30 Token sind in der Computerlinguistik ein etablierter Umfang, wenn es um die Untersuchung mikrostruktureller Textphänomene geht, und solch ein Umfang kommt auch bei der bisherigen Antimetabole-Detektion zum Einsatz (Dubremetz/Nivre 2017).
- 10 *Word Embeddings* ermöglichen es, Wörter im Hinblick auf ihre Semantik repräsentieren zu können. Die semantische Relation, die bei *Word Embeddings* beachtet wird, ist das Umfeld von Wörtern. *Word Embeddings* liegt also die Annahme zugrunde, die Bedeutung eines Wortes ergäbe sich aus dessen Kontext. Um maschinenlesbar zu sein, erfolgt die Abbildung von Wörtern in einem mehrdimensionalen Vektorraum (Pilehvar/Camacho-Collados 2020, S. 5f.). Jedes Wort erhält aufgrund seiner Relation zu anderen Wörtern seinen eigenen Wert; in einem 4-dimensionalen Vektorraum beispielsweise (2,7,5,9), üblich sind deutlich höherdimensionale Abbildungen über 750 Dimensionen. Zur Entstehung eines Vektors s. Pilehvar/Camacho-Collados 2020). Unsere *Word Embeddings* wurden mit *fastText* erzeugt.
- 11 Dies möchte unser geplantes Anschlussprojekt näher in den Blick nehmen.
- 12 Die Entstehungszeiten der Texte sind Herweg (2013) entnommen. Ist eine Zeitspanne angegeben, wurde der frühere Zeitpunkt ausgewählt. Ausnahmen bilden das ›Rolandslied‹, für das angenommen werden kann, dass es nach dem ›Straßburger Alexander‹ verfasst wurde, sowie die Texte des Pleiers, die alle um 1260/80 datiert werden (Herweg 2013, S. XXII). Es kann bei der Entstehung der Texte von der Reihenfolge ›Garel‹, ›Tandareis und Flordibel‹, ›Meleranz‹ ausgegangen werden, weshalb sich diese in der Abbildung widerspiegelt. Wichtig ist

dann nicht die exakte Jahreszahl, sondern die Beobachtung, dass in späteren Texten weniger Stilmittel gefunden worden sind.

- 13 Dies ist eine Tendenz, die wohl auch die ›analoge‹ Forschung stützen würde. Lange Zeit – und für einige Texte gilt dies bis heute – galten Texte, die nicht zu den ›Klassikern‹ gehören, als weniger ›kunstvoll‹, als weniger ›gut gemacht‹. Häufigkeit von Stilmitteln und Qualität sind nun freilich verschiedene Dinge, doch auf beiden Ebenen kann von ›wenig‹ Stil (›Stil‹ gewichtet dann das bloße Vorhandensein von Stilmitteln stark) gesprochen werden. Die Ausnahme dieser Tendenz bildet dann der ›Tristan‹ Heinrichs von Freiberg, der als einer der letzten Texte die meisten Funde aufweist.
- 14 Erste Metaphern-Ergebnisse werden auf der [DH 2022](#) vorgestellt. Thema ist vor allem die Metaphernverwendung in mittelhochdeutschen Texten mit thematischem Bezug zu ›Asien‹.
- 15 Ein Paradebeispiel für eine explorative Tool-Sammlung ist [Voyant](#) – wie sehr etwa Wordclouds die Literaturwissenschaft bereichern, ist dort nicht die Fragestellung; lediglich ein ›anderes‹ Kennenlernen der Texte steht im Vordergrund. Statistisch reichhaltiger ist etwa das Tool ›[Correlations](#)‹.
- 16 Die Figurennetzwerke können unter folgendem Link eingesehen werden (Braun/ Ketschik 2019, S. 61 Anm. 23): [10.17879/55189456328](https://doi.org/10.17879/55189456328).
- 17 Denn denkbar wäre ja auch, dass die deutschsprachigen Varianten der französischen Heldenepik, die *chansons de geste*, in Bezug auf die Stilmittelverwendung näher an der ›deutschen‹ Heldenepik sind als an den Stoffkreisen *de Rome* und *de Bretagne*. Ebenso gut möglich ist aber auch, dass die *Trois Matières* ein gemeinsames Cluster gegenüber den anderen Gattungen bilden.

Literaturverzeichnis

Primärliteratur

- Gottfried von Straßburg: *Tristan und Isold*, hrsg. von Walter Haug und Manfred Günter Scholz, Berlin 2012.
- Pfaffe Lambrecht: *Alexanderroman*. Mittelhochdeutsch/Neuhochdeutsch, hrsg., übers. und komm. von Elisabeth Lienert, Stuttgart 2007.
- Ulrich von Türheim: *Rennewart*. Aus der Berliner und Heidelberger Handschrift hrsg. von Alfred Hübner, Berlin/Zürich 1968.

Sekundärliteratur

- Achnitz, Wolfgang: Interpretationsansätze, in: Ders. (Hrsg.): Der Ritter mit dem Bock. Konrads von Stoffeln ›Gauriel von Muntabel‹. Neu hrsg., eingeleitet und komm. von dems., Tübingen 1997 (Texte und Textgeschichte 46), S. 195–232.
- Ackermann, Irmgard: Art: Parallelismus, in: Metzler Lexikon Literatur. Begriffe und Definitionen, (2007), S. 570 Sp. 2.
- Allison, Sarah/Heuser, Ryan/Jockers, Matthew/Moretti, Franco/Witmore, Michael: Quantitativer Formalismus. Ein Experiment, in: Algee-Hewitt, Mark/Allison, Sarah/Gemma, Marissa/Heuser, Ryan/Jockers, Matthew/Pestre, Dominique/Steiner, Erik/Tevel, Amir/Walser, Hannah/Witmore, Michael/Yamboliev, Irena (Hrsg.), unter der Leitung von Franco Moretti: Literatur im Labor, Paderborn 2017, S. 17–48.
- Andersen, Elizabeth/Bauschke-Hartung, Ricarda/McLelland, Nicola/Reuvekamp, Silvia (Hrsg.): Literarischer Stil. Mittelalterliche Dichtung zwischen Konvention und Innovation. XXII. Anglo-German Colloquium Düsseldorf, Berlin/Boston 2015.
- Brandes, Phillip: Von Gaben mit Stil erzählen. Zur Korrelation von reziproken Logiken in Inhalt und Form im ›Straßburger Alexander‹, ›Lanzelet‹ und ›Willehalm‹, in: PBB 144 (2022), S. 368–395.
- Braun, Manuel/Ketschik, Nora: Soziale Netzwerkanalysen zum mittelhochdeutschen Artusroman oder: Vorgreiflicher Versuch, Märchenhaftigkeit des Erzählens zu messen, in: Das Mittelalter 24 Nr. 1 (2019), S. 54–70 ([online](#)).
- Büttner, Andreas/Dimpel, Friedrich Michael/Evert, Stefan/Jannidis, Fotis/Pielström, Steffen/Proisl, Thomas/Reger, Isabella/Schöche, Christof/Vitt, Thorssten: ›Delta‹ in der stilometrischen Autorschaftsattribuion, in: Zeitschrift für digitale Geisteswissenschaften (2017) ([online](#)).
- Dubremetz, Marie/Nivre, Joakim: Machine Learning for Rhetorical Figure Detection: More Chiasmus with Less Annotation, in: Proceedings of the 21st Nordic Conference on Computational Linguistics (2017), S. 37–45 ([online](#)).
- Ehrismann, Gustav: Studien über Rudolf von Ems. Beiträge zur Geschichte der Rhetorik und Ethik im Mittelalter, Heidelberg 1919 (Sitzungsberichte der Heidelberger Akademie der Wissenschaften 8).
- Fauser, Markus: Art. Chiasmus, in: Historisches Wörterbuch der Rhetorik, Bd. 2 (1994), Sp. 171–173.
- Gumbrecht, Hans Ulrich/Pfeiffer, K. Ludwig (Hrsg.): Stil. Geschichten und Funktionen eines kulturwissenschaftlichen Diskurselements, Frankfurt a. M. 1986 (Suhrkamp Taschenbuch Wissenschaft 633).
- Herweg, Mathias: Volkssprachige Großepik im deutschen Mittelalter. Stoffe, poetologische Konzepte, diskursive Profile im Überblick, in: Achnitz, Wolfgang

- (Hrsg.): Deutsches Literatur-Lexikon. Das Mittelalter. Bd. 5: Epik (Vers – Strophe – Prosa) und Kleinformen, Berlin/Boston 2013, S. VII–XXVI.
- Hübner, Gert: Historische Stildiskurse und historische Poetologie, in: Andersen [u. a.] 2015, S. 17–38.
- Jannidis, Fotis: Grundlagen der Datenmodellierung, in: Jannidis, Fotis/Kohle, Hubertus/Rehbein, Malte (Hrsg.): Digital Humanities. Eine Einführung. Stuttgart 2017, S. 99–108.
- Jockers, Matthew L.: Macroanalysis: Digital Methods and Literary History, Urbana [u. a.] 2013 (Topics in the Digital Humanities).
- Kuhn, Jonas: Einleitung, in: Reiter, Nils/Pichler, Axel/Kuhn, Jonas (Hrsg.): Reflektierte Algorithmische Textanalyse. Interdisziplinäre(s) Arbeiten in der CRETA-Werkstatt, Berlin/Boston 2020, S. 9–40.
- Ostrowicz, Philipp: Art. Parallelismus, in: Historisches Wörterbuch der Rhetorik, Bd. 6 (2003), Sp. 546–552.
- Pilehvar, Mohammad Taher/Camacho-Collados, Jose: Embeddings in Natural Language Processing. Theory and Advances in Vector Representations of Meaning, Toronto 2020 (Synthesis Lectures on Human Language Technologies 47).
- Remele, Florian: Theorie und Methode der Gattungsgeschichtsschreibung. Mediävistische Perspektiven, in: Journal of Literary Theory 15 (2021), S. 53–80 ([online](#)).
- Scaglione, Aldo/Marvin, William P: Art. Compositio, in: Historisches Wörterbuch der Rhetorik, Bd. 2 (1994), Sp. 300–305.
- Schneider, Felix/Barz, Björn/Brandes, Phillip/Marshall, Sophie/Denzler, Joachim: Data-Driven Detection of General Chiasmi Using Lexical and Semantic Features, in: SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature (2021), S. 96–100 ([online](#)).
- Underwood, Ted: Distant Horizons. Digital Evidence and Literary Change, London 2019.
- Viehhauser, Gabriel: Historische Stilometrie? Methodische Vorschläge für eine Annäherung textanalytischer Zugänge an die mediävistische Textualitätsdebatte, in: Baum, Constanze/Stäcker, Thomas (Hrsg.): Grenzen und Möglichkeiten der Digital Humanities, Sonderband der Zeitschrift für digitale Geisteswissenschaften 1 (2015) ([online](#)).
- Viehhauser, Gabriel: Digitale Gattungsgeschichten. Minnesang zwischen generischer Konstanz und Wende, in: Zeitschrift für digitale Geisteswissenschaften (2017) ([online](#)).
- Viehhauser, Gabriel: Digital Humanities ohne Computer? Alte und neue quantifizierende Zugänge zum mittelhochdeutschen Tagelied, in: Bernhart, Toni/Willand, Marcus/Richter, Sandra/Albrecht, Andrea (Hrsg.): Quantitative An-

- sätze in den Literatur- und Geisteswissenschaften. Systematische und historische Perspektiven, Berlin/Boston 2018, S. 173–203.
- Villwock, Jörg: Art. Antithese, in: Historisches Wörterbuch der Rhetorik, Bd. 1 (1992), Sp. 722–750.
- Zehe, Albin/Konle, Leonard/Dümpelmann, Lea Katharina/Gius, Evelyn/Hotho, Andreas/Jannidis, Fotis/Kaufmann, Lucas/Krug, Markus/Puppe, Frank/Reiter, Nils/Schreiber, Annekea/Wiedmer, Nathalie: Detecting Scenes in Fiction: A new Segmentation Task, in: Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics Nr. 1 (2021), S. 3167–3177 ([online](#)).

Online-Ressourcen

- CLS (Computational Literary Studies): <https://dfg-spp-cls.github.io/>.
- DH 2022 (Digital Humanities 2022): <https://dh2022.adho.org/>.
- Digitale Mediävistik. Perspektiven der Digital Humanities für die Altgermanistik. Internationale Tagung Bremen, vom 9. bis 11. Februar 2022: <https://www.uni-bremen.de/fb-10/forschung/perspektiven-der-digital-humanities-fuer-die-altgermanistik>.
- fastText. Library for efficient text classification and representation learning: <https://fasttext.cc/>.
- MHDBDB (Mittelhochdeutsche Begriffsdatenbank): <http://www.mhdbdb.sbg.ac.at>.
- Voyant Tools: <https://voyant-tools.org/>; <https://voyant-tools.org/docs/#!/guide/correlations>.

Anschrift der Autorin und Autoren:

Phillip Brandes M.A.
Friedrich-Schiller-Universität Jena
Institut für Germanistische Literaturwissenschaft
Leutragraben 1, Raum 21N08
07743 Jena
E-Mail: phillip.brandes@uni-jena.de

Jun.-Prof. Dr. Sophie Marshall
Friedrich-Schiller-Universität Jena
Institut für Germanistische Literaturwissenschaft
Fürstengraben 18
07743 Jena
E-Mail: sophie.marshall@uni-jena.de

Felix Schneider M.A.
Friedrich-Schiller-Universität Jena
Fakultät für Mathematik und Informatik
Ernst-Abbe-Platz 2
07743 Jena
E-Mail: felix.schneider@uni-jena.de