

Biblisch erneuerte Theologie.
Jahrbuch für Theologische Studien

SCM

Stiftung Christliche Medien

SCM R.Brockhaus ist ein Imprint der SCM Verlagsgruppe,
die zur Stiftung Christliche Medien gehört, einer gemeinnützigen
Stiftung, die sich für die Förderung und Verbreitung christlicher
Bücher, Zeitschriften, Filme und Musik einsetzt.

© 2022 SCM R.Brockhaus in der SCM Verlagsgruppe GmbH
Max-Eyth-Straße 41 · 71088 Holzgerlingen
Internet: www.scm-brockhaus.de; E-Mail: info@scm-brockhaus.de

Satz: Daniel Keil, Jena
Druck und Bindung: Print Group Sp. z o.o.
Gedruckt in Polen
ISSN 2570-1975
ISBN 978-3-41-724176-1
Bestell-Nr. 224176000

Biblisch erneuerte Theologie. Jahrbuch für Theologische Studien (BeTh)

Band 6 (2022)

Schwerpunktthema: Ist Gott wandelbar?

Herausgegeben für den Arbeitskreis für evangelikale Theologie
und die Arbeitsgemeinschaft für biblisch erneuerte Theologie

*von Christoph Raedel und Jürg Buchegger-Müller
Henrik Homrighausen (Redaktion)*

Wissenschaftlicher Beirat (Advisory Board)

Andreas Beck (Leuven); Roland Deines (Bad Liebenzell); Roland
Gebauer (Reutlingen); Rolf Hille (Gießen); Lydia Jaeger
(Nogent-surMarne); Karsten Lehmkuhler (Strasbourg); Eckhard
Schnabel (South Hamilton); Stefan Schweyer (Basel); Julius
Steinberg (Ewersbach); Christian Stettler (Zürich/Basel); Ulrike
Treich (Gießen); Beat Weber (Basel); Peter Zimmerling (Leipzig).

Computergestützte Verfahren für die Narrative Exegese

Jens Dörpinghaus

1. Einleitung

Diese Arbeit untersucht mögliche Schnittmengen zwischen der computergestützten Analyse von natürlicher Sprache und der narrativen Exegese biblischer Texte, versucht einen Überblick über bereits etablierte Verfahren und mögliche Synergien zu geben und diese kritisch zu würdigen.

Die Narrative Exegese ist nicht nur in ihren Methoden, sondern auch in ihrer Benennung äußerst vielseitig¹. Allen gemein ist allerdings der Fokus auf den biblischen Text als literarisches Werk und sie bedienen sich primär für die Auslegung von Erzähltexten Methoden der Literaturwissenschaften. Dabei lassen sich verschiedene Strömungen identifizieren²:

An den Texten orientierte Methoden	An den Lesern orientierte Ansätze
New (Literary) Criticism, Formalism, Narratology,	Affective, Phenomenological and Psychoanalytical Criticism,
Semiotik, Strukturalismus,	Poststrukturalismus und Dekonstruktion
Rhetorical Criticism,	Rezeptionsästhetik
Linguistik	Reader Response Criticism

¹ Oeming spricht etwa von „New (Literary) Criticism, Literary Approach/Criticism, Holistic Approach, Rhetorical Criticism, Close Reading, Synchronical Approach, Narratology und Narrative Criticism“, Oeming, Manfred, *Biblische Hermeneutik: Eine Einführung*. Darmstadt 2007, 70.

² In Anlehnung an Cornils, Anja, *Vom Geist Gottes erzählen. Analysen zur Apostelgeschichte*. Tübingen 2006.

Es ist zu beachten, dass es – sozusagen vertikal dazu – auch noch verschiedene Schulen gibt, die nicht nur unterschiedlichen literaturwissenschaftlichen Theorien verpflichtet, sondern auch mit verschiedenen geografischen Bereichen verbunden sind³. Dies führt zu einem enorm komplexen Feld, das hier nicht einmal ansatzweise vollständig dargestellt werden kann.

Es sind zwei Beobachtungen zu machen: Zum einen werden computergestützte Verfahren in der biblischen Exegese immer beliebter⁴, zum anderen werden aber die Fortschritte der digitalen Sprachverarbeitung in der Theologie kaum rezipiert. Hier sind in den letzten Jahren gerade auch für die Verarbeitung von literarischen Texten, auch an der Schnittmenge zu allgemeinen religiösen Daten und Texten, Fortschritte zu verzeichnen⁵ und auch in den Literaturwissenschaften bedient man sich immer öfter digitaler Methoden⁶.

Ziel dieser Arbeit ist nun weder eine abschließende Darstellung der narrativen Exegese noch der quantitativen Ansätze oder Methoden aus den *Digital Humanities* für die Literaturwissenschaft, sondern einen Überblick über eine Auswahl möglicher Schnittmengen: Welche Teilaspekte dieser computerge-

³ Oeming, *Biblische Hermeneutik: Eine Einführung*, 70–71; Dabei ist eine genaue Zuordnung oft gar nicht möglich. Oeming spricht von Amerika, „Angelsachsen“, den Niederlanden und Israel; Cornils betont auch die Unterschiede im frankophonen und deutschsprachigen Raum, vgl. Cornils, *Vom Geist Gottes erzählen. Analysen zur Apostelgeschichte*, 30ff.

⁴ So z. B. für Wortstudien, Worthäufigkeiten oder Übersetzungen. In der Praxis findet man häufig lediglich das Enumerieren von einzelnen Wörtern (etwa in griechischer, hebräischer oder aramäischer Sprache) und das Erarbeiten von Schaubildern.

⁵ Man vergleiche z. B. die aktuellen Bände *Quantitative Ansätze in den Literatur- und Geisteswissenschaften: Systematische und historische Perspektiven*, Toni Bernhart/Marcus Willand/Sandra Richter/u. a. (Hg.) 2018; *Reflektierte algorithmische Textanalyse: Interdisziplinäre(s) Arbeiten in der CRETA-Werkstatt*, Nils Reiter/Axel Pichler/Jonas Kuhn (Hg.) 2020; *Digital Humanities and Christianity: An Introduction*, Tim Hutchings/Claire Clivaz (Hg.) 2021.

⁶ Etwa im Bereich der digitalen Geisteswissenschaften, den *Digital Humanities*. Kuhn beschreibt die Situation wie folgt: „Die Computerlinguistik und die Sprachtechnologieforschung entwickeln ihre Modelle und Methoden überwiegend für Gebrauchstexte wie Zeitungsartikel, Produktbesprechungen auf Internetseiten, Forenbeiträge in den Sozialen Medien etc. Dennoch üben literarische Texte mit ihren vielfältigen Herausforderungen an die Textanalyse eine große Anziehungskraft auf Computerlinguistinnen und -linguisten aus und in den wichtigsten Publikationsorganen, den Tagungsbänden der großen Computerlinguistikkonferenzen, erscheinen seit vielen Jahren vereinzelt, aber immer wieder Beiträge zur Erweiterung von computerlinguistischen Analysemodellen, die auf Charakteristika literarischer Texte abzielen.“ Kuhn, Jonas, *Computerlinguistische Textanalyse in der Literaturwissenschaft? Oder: »The Importance of Being Earnest« bei quantitativen Untersuchungen*, in: *Quantitative Ansätze in den Literatur- und Geisteswissenschaften*, 2018 11–44, 11.

stützten Verfahren sind für die biblische Exegese relevant, welche Methoden werden hier wie dort schon verwendet und welche zusätzlichen Synergien könnten sich ergeben? Dabei sollen aber auch methodische Vorbehalte und mögliche Probleme diskutiert und gewürdigt werden.

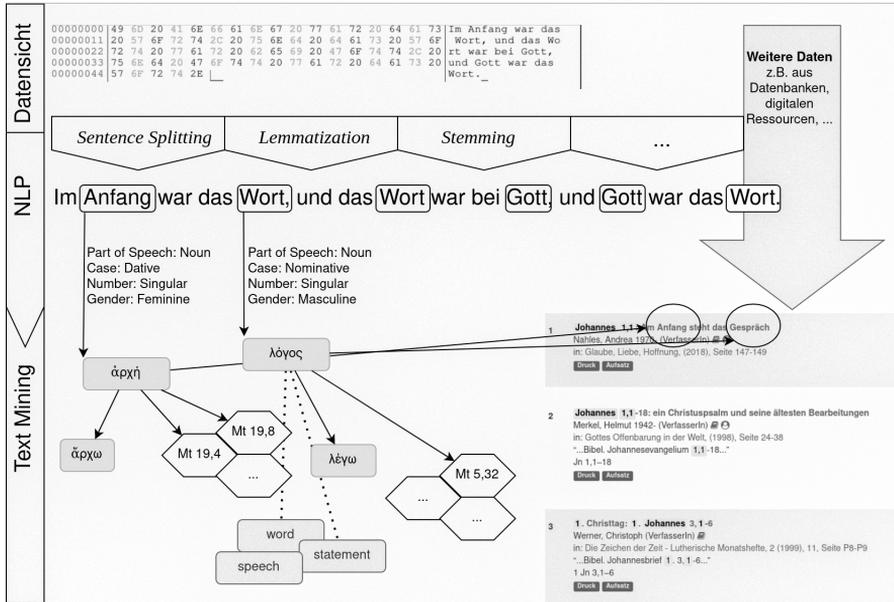


Abbildung 1: Schematische Abbildung, wie mittels NLP und Text Mining aus unstrukturierter Daten (d.h. Text, in der Datensicht Binärdaten) mit verschiedenen Methoden wie der Satz- und Worterkennung Sprachbausteine erkannt werden und strukturiert mit weiteren Informationen verknüpft werden können. Weitere Daten, neben einer Konkordanzfunktion zum Beispiel Daten aus Literaturdatenbanken, Wörterbüchern oder anderen digitalen Ressourcen können damit verknüpft werden und bilden ein Wissensnetzwerk.

Die Themengebiete der computergestützten Analyse von natürlicher Sprache (NLP, *Natural Language Processing*), des automatischen Übersetzens und der Extraktion von Informationen aus Texten (*Text Mining*) sind keine jungen Disziplinen innerhalb der Informatik⁷. Menschliche Sprache und Texte bestehen in digitalisierter Form zunächst nur aus einzelnen Buchstaben,

⁷ Vgl. weiter Berry, M.W./Kogan, J., *Text Mining: Applications and Theory*. 2010; Miner, G./Elder, J./Fast, A./u. a., *Practical Text Mining and Statistical Analysis for Non-structured Text Data Applications*. 2012; Biemann, C./Mehler, A., *Text Mining: From Ontology Learning to Automated Text Processing Applications*. 2014.

bzw. ihrer binären Referenz und sind somit unstrukturierte Daten. Um sie verarbeiten zu können, müssen diese in strukturierte Daten überführt werden. Dies übernimmt das NLP. Die dadurch erzeugte Struktur mit weiteren Informationen (z.B. Wörterbüchern) zu verknüpfen oder weitere Informationen zu extrahieren ist Aufgabe des Text Minings, vergleiche Abbildung 1. Beide Themen haben in den letzten Jahrzehnten durch wachsende Speicher- und Rechenressourcen einen großen Aufschwung erlebt. Durch die Digitalisierung und Vernetzung der verschiedenen Datenbanken und Archive ist es nun allerdings auch möglich auf eine ständig wachsende Anzahl von Daten zuzugreifen und diese elektronisch zu verarbeiten. Auch die Theologie kann auf immer mehr digitale Ressourcen zurückgreifen. Die direkte Verknüpfung der einzelnen Daten (engl. *linked data*) führt nicht nur zu bequemen Abfragemöglichkeiten („Welche theologische Fachliteratur beschäftigt sich mit Bibelstelle X unter Berücksichtigung der griechischen Form Y?“), sondern ermöglicht durch offene Datenformate auch ihre Visualisierung. Ein solches Wissensnetzwerk (engl. *Knowledge Graph*) stößt – nicht nur in der Theologie – wegen mancher Probleme, die im Folgenden näher dargestellt werden sollen, an Grenzen: Datenverfügbarkeit, Datenkomplexität und Datengeschwindigkeit.

Grundsätzlich besteht die Analyse und Verarbeitung von natürlicher Sprache immer aus mehreren Schritten. Am Anfang steht das Erkennen von Sätzen, Wörtern und Lemmata (*Sentence Splitting, Lemmatization, Stemming*). Hierbei können auch Wortarten (engl. *Part of speech*) und ihre grammatische Form untersucht werden. Aus $\Theta\epsilon\acute{o}\nu$ würde neben der Grundform $\theta\epsilon\acute{o}\varsigma$ dann auch die Form des Substantivs, Akkusativ Singular, extrahiert. Weiter ist das Erkennen von Entitäten oder Namen (NER, *Named-entity recognition*) eine wichtige Aufgabe, vor allem da sie verschiedenen Begriffen genormte Einträge in Wörterbüchern oder strukturierten Listen zuordnet, siehe auch Abbildung 1. Der Form $\Theta\epsilon\acute{o}\nu$ würde dann zum Beispiel die Strong-Nummer 2316 zugeordnet.

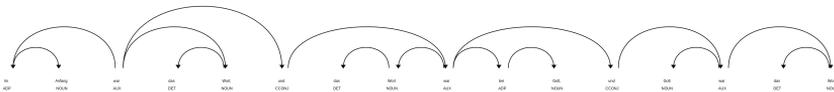


Abbildung 2: Ein Plot des *Dependency Parsing* (DP) von Joh 1,1, der die Abhängigkeiten zwischen den Bausteinen eines Satzes um dessen grammatische Struktur darstellt.

Diese Entitäten in strukturierte Zusammenhänge zu setzen (also Subjekt-Verb-Objekt-Relationen zu erkennen) ist noch schwieriger (*Relation Extraction*). Dies kann als ein sogenanntes *Dependency Parsing* auch visualisiert werden, siehe Abbildung 2. Allein schon bei diesen (aus Sicht der Informatik) schwierigen Themen stehen nicht nur eine Fülle von Programmbibliotheken, sondern auch von (KI und nicht-KI) Modellen für verschiedenen Sprachen zur Verfügung⁸. Denn allgemeingültige Heuristiken und Modelle gibt es kaum, nicht nur die verschiedene Sprachen, sondern auch verschiedene Literaturarten haben ganz unterschiedliche Anforderungen⁹.

Neben der Textübersetzung gibt es auch vielfältige weitere Anwendungsfälle, die grob unter die Kategorisierung von Texten fallen: die Verschlagwortung von Dokumenten, automatisches Sortieren und Filtern (z. B. von Mails) oder das Erkennen von Emotionen (z. B. guter, schlechter oder aggressiver Onlinebewertungen).

Ist allerdings schon in der Informatik eine Vielzahl von Methoden und Ansätzen für die algorithmisch äußerst komplexen Aufgaben zu finden, so kommt für Fragestellungen aus anderen Domänen noch hinzu, dass nun interdisziplinär gearbeitet werden muss. Dabei lässt sich beobachten, dass das Zusammenspiel von Forschungsfragen und Methoden „in hohem Grade fachspezifisch beeinflusst [ist]: Wonach und auf welche Art und Weise gefragt wird, hängt von einer Vielzahl disziplinärer Faktoren ab, die im Regelfall nicht vollständig freigelegt werden.“¹⁰ So bleibt also neben der Hürde, computerbasierte Verfahren einzusetzen, für die Theologie und Narrative Exegese auch die Hürde, die schon existierenden Verfahren aufgrund ihrer interdisziplinären Hindernisse einzusetzen – gleichwohl sie den Grundfragen ähnlich nahe stehen wie die Literaturwissenschaft.

Nun sollen ausgewählte Themen aus dem Bereich *Digital Humanities* und der Computerlinguistik mit Bezug zur Narrativen Exegese vorgestellt werden und auch ihr möglicher Wert für die Exegese und in der Theologie adressiert werden. In einem weiteren Abschnitt werden kritische Beobachtungen zu

⁸ Allein für Python stehen etliche Bibliotheken zur Verfügung. Die populärsten sind derzeit NLTK <https://www.nltk.org/> und spaCy <https://spacy.io/>. Weiterhin wird vor allem Java oder R verwendet.

⁹ Vgl. dazu bspw. Blessing, Andre/Bockwinkel, Peggy/Reiter, Nils/u. a., Dramenwerkbank – Automatische Sprachverarbeitung zur Analyse von Figurenrede, in: DHd 2016.

¹⁰ Pichler, Axel/Reiter, Nils, Reflektierte Textanalyse, in: Reflektierte algorithmische Textanalyse. Interdisziplinäre(s) Arbeiten in der CRETA-Werkstatt, 2020 43–60, 46.

digitalen Methoden in der Narrativen Theologie und Exegese angesprochen. Eine Zusammenfassung und ein Fazit bilden den Abschluss dieser Arbeit.

2. Ausgewählte Themen

Die computerbasierte literaturwissenschaftliche Textanalyse kann – ebenso wie die Narrative Exegese – auf verschiedene Arten strukturiert werden. Ivanovic geht hier zunächst von einer Dreiteilung aus: Es gehe um eine strukturelle Erfassung, ein hermeneutisches Verstehen und um eine kontextuelle Verortung eines Textes¹¹. Es ist dabei offensichtlich, dass das hermeneutische Verstehen eines Textes derzeit lediglich ein Randthema¹², die strukturelle Erfassung und das kontextuelle Verorten allerdings Hauptthemen der Forschung sind. Es bietet sich allerdings an, das zweistufige Methodenmodell von Ivanovic durch weitere Dimensionen aus der Arbeit von Gius zu einem fünfschrittigen zu erweitern:

(1) Fragestellung

(2) Datengrundlage

Ein Text	Mehrere Texte	Kontextualisierung?
----------	---------------	---------------------

(3) NLP Methoden

Isolierung, Beschreibung und Klassifizierung von Phänomenen oder Eigenschaften

(4) Zuordnung zu „bedeutungstragenden Systemen“

(5) Interpretation der Ergebnisse

¹¹ Vgl. Ivanovic, Christine, Die Vernetzung des Textes: Im Möglichkeitsraum digitaler Literaturanalyse, in: Zeitschrift für digitale Geisteswissenschaften (2017); Einen anderen Fokus setzen Jockers, Matthew L., Macroanalysis: Digital Methods and Literary History. 2013; Gius, Evelyn, „Computationelle Textanalysen als fünfdimensionales Problem: Ein Modell zur Beschreibung von Komplexität“, in: LitLab Pamphlet 8 (2019).

¹² Vgl. Krüger, Reinhard/Nickel, Beatrice, Schäfer, Computer, Internet: Digital Humanities und frühneuzeitliche Pastoralliteratur. 2016, 22; Erb, Maurice, Alles oder gar nichts lesen? Foucault, Moretti und die Verheißungen des Algorithmus, in: Le foucauldien 2/1 (2016) 6.

Dabei ergeben sich die Schritte (1) und (5) jeweils in der originären wissenschaftlichen Disziplin, also entweder in den Literaturwissenschaften oder der Theologie¹³. Aus der Fragestellung ergibt sich (2) die Datengrundlage. Bezieht sich diese auf einen einzelnen Text, mehrere Texte bzw. einen Text mit unterschiedlichen Teilen oder bezieht die Fragestellung auch weitere Quellen zur Kontextualisierung ein, benötigt also externes Wissen bzw. Datenquellen¹⁴. Die Teilaspekte (3) und (4) sind jeweils als Teil der Computerlinguistik zu verstehen. Zu diesem vertikalen Schema liegen horizontal die verschiedenen Methodenstränge (Handlung und Figur, Redewiedergabe, Gattungsanalysen, Stilometrie, ...), wobei eine Zuordnung zu verschiedenen Fragestellungen nicht eindeutig ist¹⁵.

Daraus ergibt sich die Schwierigkeit der folgenden Auswahl: Sortiert man sie nach Fragestellungen, ergibt sich eine Vielfalt von Methoden; sortiert man nach Methoden, können diese für vielfältige Fragestellungen genutzt werden. Deshalb wurde ein leicht anderer Ansatz gewählt: In einem ersten Abschnitt soll es um die kontextuelle und strukturelle Einordnung von Texten und ihre Textanalyse gehen: Hier soll von den bisher in der Theologie verwendeten Methoden ausgehend ein Blick auf die Computerlinguistik geworfen werden. Dabei sollen, um einige grundsätzliche Problematiken zu verdeutlichen, auch einige tiefergehende Aspekte angerissen werden. Da es hierbei vor allem um abzählbare Elemente geht, soll sich ein nächster Abschnitt der quantitativen Dramenanalyse widmen und weiter der für die Narrative Exegese nutzbaren Teilaspekte. Die Auswahl der Methoden und Fragestellungen muss dabei zwangsläufig willkürlich und unvollständig bleiben, sie wird allein von einer möglichen Anwendung in der Theologie geleitet. Dabei soll, wann immer möglich, weiterführende Literatur angegeben werden.

¹³ Nicht unerwähnt bleiben darf ein Verweis auf die *Operationalisierung*, die bspw. Fragestellungen gliedert und so unterteilt, dass eine Überprüfung und Evaluation stattfinden kann: „Operationalisierung bezeichnet [...] dementsprechend sowohl die (Teil-)Schritte bzw. Aktivitäten, die notwendig sind, um einen Begriff empirisch überprüfbar bzw. messbar zu machen, als auch das Resultat dieser Aktivitäten.“ Pichler, Axel/Reiter, Nils, *Reflektierte Textanalyse*, in: *Reflektierte Algorithmische Textanalyse*, 2020 43–60, 46–47.

¹⁴ Vgl. Gius, *Computationelle Textanalysen als fünfdimensionales Problem*, 16.

¹⁵ So verwendet man das sogenannte Clustering als Methode zur Gruppierung ähnlicher Elemente nicht nur für das Finden von ähnlichen Dokumenten, sondern auch zur Textanalyse (Clustering von Wortvektoren) etwa zum *Topic Modeling*.

2.1 Texte und Textanalysen: Das kontextuelle und strukturelle Einordnen

Bei der Analyse von Texten und Textausschnitten steht oft die Frage nach dem Inhalt im Vordergrund, beispielsweise: Worum geht es? Was ist das Thema? Hierbei werden textimmanente Aspekte behandelt, es geht also um die strukturelle Erfassung von Texten. Beispielsweise kann man auch fragen: Was charakterisiert den Text¹⁶?

Methodisch werden diese Fragestellungen häufig mittels *Topic-* oder *Corpus-Detection* und im weitesten Sinne mit *Clustering* oder *Categorization* gelöst. So ergibt sich, zum Beispiel als ein Anwendungsfall, durch die wachsende Anzahl digital zugänglicher Fachveröffentlichungen die Frage, wie ein Forscher möglichst zügig einen Überblick über die für ihn relevanten Informationen bekommen kann¹⁷. Dabei gilt allerdings für die Analyse von Literatur noch ein weiterer methodischer Schritt, der die Besonderheiten dieser Texte berücksichtigt¹⁸. Doch ist die Bilanz grundsätzlich eher ernüchternd: „the uptake of CL [*computational linguistics*] in literary studies has

¹⁶ Aber auch: Wie kann der Stil bzw. Charakter eines Textes verglichen werden? Hierbei geht es natürlich auch um die Frage nach Gattungen, Autoren, aber auch um historische Verläufe und „sprachlich und kulturell bedingte Ordnungen“, vgl. Ivanovic, Die Vernetzung des Textes. Diese Methodik wird auch als *Stilometrie* bezeichnet.

¹⁷ Als Beispiel für die beständige Forschung in diesem Bereich kann TREC– Text REtrieval Conference – genannt werden. Diese wird seit 1992 organisiert und bündelt Kompetenzen und Forschung im Bereich Information Retrieval, vgl. Roberts, Kirk/Simpson, Matthew S./Voorhees, Ellen M./u. a., Overview of the TREC 2015 Clinical Decision Support Track, in: TREC 2015. Aber die Menge an Arbeiten zum Thema Corpus oder Topic detection ist schlichtweg erschlagend, vgl. u.a. Miner/Elder/Fast/u. a., Practical Text Mining and Statistical Analysis for Non-structured Text Data Applications; Dörpinghaus, Jens/Schaaf, Sebastian/Jacobs, Marc, Soft document clustering using a novel graph covering approach, in: BioData Mining 11/1 (2018), 11; Dörpinghaus, Jens/Jacobs, Marc, Knowledge Detection and Discovery using Semantic Graph Embeddings on Large Knowledge Graphs generated on Text Mining Results, in: 2020 15th Conference on Computer Science and Information Systems (FedCSIS) 2020, 169–178.

¹⁸ „Die Digitalisierung von Texten, die Erarbeitung komplexer digitaler Editionen und die Herstellung maschinell analysierbarer Korpora sind Hauptaufgaben gegenwärtiger digitaler Philologie. Diese baut vielfach auf Methoden und Erkenntnissen auf, wie sie in der Computerlinguistik seit Jahrzehnten erprobt werden. Sie müssen jedoch für die Anforderungen der literaturwissenschaftlichen Analyse beispielsweise in ihrer Granularität angepasst werden.“ Ivanovic, Die Vernetzung des Textes.

been slow so far, and very few of the immense number of literary publications show use of CL.“¹⁹

Dennoch ist gerade bei dieser Fragestellung eine starke Verwendung von enumerativen Verfahren in der biblischen Exegese zu beobachten, grundsätzlich werden Worthäufungen bzw. die Nennung einzelner Begriffe zur Analyse verwendet²⁰. Zunächst liefern Worthäufigkeiten natürlich eine erste Übersicht, gerade auf kleinen Textausschnitten. Allerdings ist gerade bei der Anwendung von (methodisch beliebig ausgewählten) Normierungen und der Verwendung im größeren Stil die Frage nach der Verortung in der Computerlinguistik zu stellen²¹.

Für diese Aufgaben, z. B. um Ähnlichkeiten zwischen Dokumenten und Texten oder ihre strukturelle oder kontextuelle Einordnung zu messen, müssen bestimmte Eigenschaften, sogenannte *features*, der Textelemente bzw. Dokumente in einem Dokumentenraum \mathbb{D} mit einer Funktion in eine reelle Zahl abgebildet werden. Daraus ergibt sich ein sogenanntes *Vector Space Model*. Dieses Modell ist so grundlegend, dass es zur Einordnung vielfältiger Methoden genutzt werden kann und an dieser Stelle zumindest in groben Zügen dargestellt werden muss. Es ergeben sich daraus nämlich zwei wichtige Beobachtungen: Zum einen findet das Modell innerhalb der Theologie in Form von Worthäufigkeiten auch manuell Verwendung, zum anderen beschreibt es nachdrücklich die Komplexität der hier vorgestellten Methoden, auch wenn sie teilweise durch weiterentwickelte Ansätze ersetzt werden. Statt der Worthäufigkeiten kann man den gewichteten Vektor der Worte

¹⁹ Frank, Andrew/Ivanovic, Christine, Building literary corpora for computational literary analysis—a prototype to bridge the gap between CL and DH, in: Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018) 2018, 798.

²⁰ So zum Beispiel die „Pneuma-Nennungen in der Figuren-Rede“ in der Apg, siehe Cornils, Vom Geist Gottes erzählen. Analysen zur Apostelgeschichte, 92–93; Dabei ist oft eine gewisse methodische Beliebigkeit zu finden, so untersucht z. B. Martella oft nur die Anzahl von Versen, in denen das Wort auftritt, oder normiert auf die Anzahl der Verse im Kapitel, vgl. Martella, Immanuel Matthäus, Geschwisterrivalitäten im Buch Genesis: Eine literaturwissenschaftliche Analyse ausgewählter Geschwisterrivalitäten. 2015, 108.

²¹ Vgl. dazu auch Blessing, André/Kliche, Fritz/Heid, Ulrich/u. a., Computerlinguistische Werkzeuge zur Erschließung und Exploration großer Textsammlungen aus der Perspektive fachspezifischer Theorien, in: Grenzen und Möglichkeiten der Digital Humanities (Sonderband der ZfdG 1) (2015); Aber auch die präzise Aussage „Nevertheless, there is a consensus that the sense enumeration found in a standard dictionary is not the kind of lexicon required for linguistic or psycholinguistic theory.“ Miller, George A., On knowing a word, in: Annual review of psychology 50/1 (1999), 1–19.

eines Textes – oder von anderen Entitäten – als eine Koordinate in einem n -dimensionalen Vektorraum darstellen²². Dadurch würde man das oben beschriebene Szenario der reinen Worthäufigkeiten modellieren.

Allerdings findet sich in der Theologie vor allem die Analyse bestimmter Wörter. In diesem Zusammenhang kann die Methode genannt werden, dass seltene Wörter höher gewichtet werden sollten und Standardwörter (sogenannte „*stop words*“) normalerweise ignoriert werden. Dabei entspricht die *term frequency document frequency* der Anzahl der Verwendungen über einem oder allen Dokumenten. Das Verfahren, das Wörter mit einer kleinen *document frequency* belohnt, wird *inverse document frequency* genannt²³. Die Berechnung kann für N Dokumente oder Texte in \mathbb{D} folgendermaßen erfolgen: Es sei f_{ij} die Anzahl des Wortes i in Dokument j . Dann ist die *term frequency* TF_{ij} wie folgt mit einer Normalisierung über die maximale Anzahl eines Wortes als Abbildung nach $[0, 1]$ definiert²⁴:

$$TF_{ij} = \frac{f_{ij}}{\max_k f_{kj}} \quad (1)$$

Kombiniert ergibt sich das TF.IDF-Maß (*term frequency–inverse document frequency*). Daneben werden allerdings auch andere Maße – mit jeweils eigenen methodischen Vor- und Nachteilen – verwendet. Dabei ist zu beachten, dass in der Regel eine Reduzierung der Dimensionen erfolgt, da die Berechnung mit Wortvektoren ansonsten zu aufwendig wird.

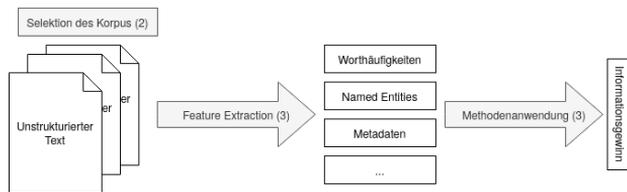


Abbildung 3: Schematische Darstellung der Schritte (2) und (3): Durch die Selektion verschiedener *features* können diese z. B. in einem Vektorraummodell abgelegt werden und durch eine bestimmte Methode kommt es zu einem Informationsgewinn.

²² Vgl. Jannidis, Fotis/Kohle, Hubertus/Rehbein, Malte/u. a., *Digital Humanities: Eine Einführung*. 2017, 275.

²³ Vgl. Zhai, ChengXiang/Massung, Sean, *Text data management and analysis: a practical introduction to information retrieval and text mining*. 2016, 99.

²⁴ Vgl. Psallidas, Fotis/Becker, Hila/Naaman, Mor/u. a., *Effective Event Identification in Social Media*, in: *IEEE Data Eng. Bull.* 36/3 (2013), 42–50, 8f.

Dies führt die Komplexität dieser Aufgabe vor Augen: Nicht nur ist ein (mathematisches) Maß nach der Überführung einer Repräsentation des Textes in einen Vektorraum zu finden, sondern auch eine (mathematische) Ähnlichkeitsfunktion. Dieses Verfahren ist an sich schon hochkomplex und führt häufig zu verzerrten (allerdings mathematisch korrekten) Ergebnissen²⁵, vgl. die Darstellung in Abb. 3. Wenn also beispielsweise Martella Worthäufigkeiten in einem bestimmten Teil des Buches Genesis betrachtet, muss er nicht nur seine Textauswahl begründen, sondern auch, warum er normiert, womit und warum. Offensichtlich ergibt sich hier eine Verzerrung, die der Autor bereinigen möchte²⁶.

Dabei ist zu beobachten, dass stilometrische Studien in der Bibelwissenschaft durchaus zu finden sind; ihren Höhepunkt hatten diese Ansätze in den 1980er Jahren, aber auch einige aktuelle Studien lassen sich finden. Dabei geht es vor allem um die Frage der Autorenschaft biblischer Texte, aber auch um Klassifizierung von bestimmten Quellen (z. B. Q)²⁷. Gerade zum Alten Testament sind aber nur wenige Ansätze publiziert worden. Hier sehen wir eine Besonderheit, die sich durch die gesamte Computerlinguistik zieht: Die meisten Modelle und Verfahren finden sich für die gegenwartsnahen Sprachen mit vielen Sprechern. Es gibt derzeit keine oder nur sehr wenige

²⁵ Vgl. dazu auch Dörpinghaus, Jens/Darms, Johannes/Jacobs, Marc, What was the Question? A systematization of Information Retrieval and NLP Problems, in: 2018 Federated Conference on Computer Science and Information Systems (FedCSIS) 2018 471–478.

²⁶ Dies ist eine methodische Kritik, die sich aber breit verallgemeinern lässt. Die inhaltlichen, praktischen und technischen Schwierigkeiten der Worthäufigkeiten wurden in der Literatur häufig genannt, vgl. bspw. Viehhauser, Gabriel, Historische Stilometrie? Methodische Vorschläge für eine Annäherung textanalytischer Zugänge an die mediävistische Textualitätsdebatte, in: ZfdG (2015). Hierzu ist auch die lesenswerte Studie von Schruhl zu nennen, die folgert: „In dieser Hinsicht lenkt die Digitalisierung die Aufmerksamkeit nicht nur auf den technischen Fortschritt. Vielmehr gerät die beeindruckende Komplexität literaturwissenschaftlicher Praxiszusammenhänge in den Fokus. Was durch die Quantifizierungen offensichtlich wird, ist, also der Sachverhalt und nicht die ›Sache selbst‹“, Schruhl, Friederike, Quantifizieren in der Interpretationspraxis der Digital Humanities, in: Quantitative Ansätze in den Literatur- und Geisteswissenschaften, 2018 235–268, 261.

²⁷ Vgl. bspw. Kenny, Anthony, A stylometric study of the New Testament. Oxford 1986; Erwin, Harry/Oakes, Michael, Correspondence analysis of the New Testament, in: Workshop Organizers 2012, 30; Mealand, David L., Hellenistic Greek and the New Testament: A Stylometric Perspective, in: Journal for the Study of the New Testament 34/4 (2012), 323–345; Erwin/Oakes, Correspondence analysis of the New Testament.

Modelle für antike Sprachen²⁸. Dabei zeigen diese Studien auch einige methodische Fehler in den Bereichen der Fragestellung und Auswertung²⁹. Auch ist zu bemerken, dass der Ansatz der Wortvektoren in der Computerlinguistik immer seltener verwendet wird, und stattdessen eher die Extraktion bzw. Annotation sogenannter *Named Entities* (d. h. von Wörtern und Synonymen aus kontrolliertem Vokabular oder Ontologien) als zielführend betrachtet wird³⁰.

Es ist also die Frage nach der Relevanz dieser Verfahren zu stellen, zumal die automatisierten Verfahren oft im Nachhinein nicht mehr nachvollziehbar arbeiten. So fassen z. B. Methoden zur Dimensionsreduktion Wörter zusammen; ein Verfahren, das in der Theologie ebenfalls, allerdings primär manuell, Anwendung findet. Auch Verfahren des Maschinellen Lernens lassen nach ihrem Training keinerlei Rückschlüsse auf Regeln zu. Allerdings hilft ein sorgfältiges und methodisch valides Vorgehen, entsprechende Parameter zu

²⁸ Es zeigt sich ein Mangel an Modellen für das Altgriechische und das Fehlen von Modellen für das Bibelhebräisch und Aramäisch. Dabei ist das Problem des Arbeitens mit Antiken Schriften und der „scriptio continua“ noch gar nicht eingenommen. Erste Ansätze gibt es allerdings zum Beispiel durch das ETCBC in Amsterdam unter <http://etcbc.nl/data/>. Vgl. dazu Perrone, Valerio/Palma, Marco/Hengchen, Simon/u. a., GASC: Genre-aware semantic change for Ancient Greek, in: arXiv preprint arXiv:1903.05587 (2019); Dörpinghaus, Jens/Düing, Carsten, Automated creation of parallel Bible corpora with cross-lingual semantic concordance, in: 2021 16th Conference on Computer Science and Intelligence Systems (FedCSIS) 2021, 111–114; Bamman, David, Natural Language Processing for the Long Tail., in: Book of Abstracts of Digital Humanities 2017 2017; Vijayarani, S/Janani, R/others, Text mining: open source tokenization tools-an analysis, in: Advanced Computational Intelligence: An International Journal (ACIJ) 3/1 (2016), 37–47; Roorda, Dirk, Coding the Hebrew Bible: Linguistics and Literature, in: Research Data Journal for the Humanities and Social Sciences 3/1 (2018), 27–41.

²⁹ So schreibt beispielsweise Eder: „It can be safely assumed, then, that the problems concerning the authorship of subsequent books of the New Testament have been thoroughly examined from linguistic, historical, theological, and rhetorical points of view. For this reason, the Scripture seems to be an ideal material for stylometric benchmarks, because the traditional scholarship can serve as a straightforward validation of the results obtained by using the computational approach“, Eder, Maciej, Computational stylistics and Biblical translation: How reliable can a dendrogram be, in: The translator and the computer (2013), 155–170, 6. Allerdings ist dabei zum einen festzuhalten, dass es oft nicht eine These zu Autorenschaften gibt. Zum anderen macht es aus Sicht der Aussagelogik keinen Sinn, diese Thesen zur Validierung der Ergebnisse zu nutzen.

³⁰ Vgl. bpsw. Gius, Evelyn/Reiter, Nils/Strötgen, Jannik/u. a., SANTA: Systematische Analyse Narrativer Texte durch Annotation., in: DHd 2018; Dörpinghaus, Jens/Stefan, Andreas, Knowledge extraction and applications utilizing context data in knowledge graphs, in: 2019 Federated Conference on Computer Science and Information Systems (FedCSIS) 2019, 265–272.

extrahieren und einer weiteren Diskussion zur Verfügung zu stellen³¹. Für die Narrative Exegese ergeben sich in diesem Zusammenhang vielfältige Möglichkeiten, die über das bloße Enumerieren von Wörtern hinausgehen: Neben den Stilometrischen Studien könnten auch Inhalte, Themen und Emotionen extrahiert und analysiert, oder sogar Gattungsstudien durchgeführt werden³². Diese Methoden betreffen im oben beschriebenen Schema die Schritte 3 und 4. Sie müssen aber selbstverständlich im Sinne von 1 und 5 methodisch validiert werden³³. Doch auch wenn hier methodisch noch einige Pioniertaten zu vollbringen sind, gibt es doch auch ein weiteres für die Narrative Exegese wichtiges Feld: die quantitative Dramenanalyse.

2.2 Ausgewählte Themen

Ein Drama mag aufgrund seiner literarischen Gattung zunächst irrelevant für die Analyse biblischer Texte erscheinen: Dramen sind in der Regel durch wörtliche Rede und eine starke Textstrukturierung geprägt. Das macht sie für die automatische Sprachverarbeitung gleichzeitig attraktiv wie schwierig³⁴. Dabei sind zwei Aspekte aus Sicht der Theologie besonders wichtig: Zum einen tragen einige biblische Bücher definitiv Züge eines modernen Dramas³⁵. Zum anderen bietet das Drama als höchst strukturierte Textform Anknüpfungspunkte für die Narrative Exegese: Nicht nur der Sprachstil, Zeit

³¹ Vgl. bspw. die lesenswerte Studie von Barth, Florian, *Zwischen Elisabeth Hauptmann und Bertolt Brecht: Stilometrische Studien einer Zusammenarbeit*, in: *Quantitative Ansätze in den Literatur- und Geisteswissenschaften*, 2018, 95–120.

³² Vgl. bspw. Kim, Evgeny/Klinger, Roman, *A survey on sentiment and emotion analysis for computational literary studies*, in: *arXiv preprint arXiv:1808.03137* (2018); Viehhauser, Gabriel, *Digitale Gattungsgeschichten. Minnesang zwischen generischer Konstanz und Wende*, in: *ZfdG* (2017).

³³ So gibt z. B. Greenwood zu bedenken, „that the frequencies of common words, sentence lengths, and other simple parameters offer a prescription for determining patterns of authorship within texts is not comfortably assimilated within the spirit of classical scholarship“, Greenwood, HH, *St Paul revisited—word clusters in multidimensional space*, in: *Literary and linguistic computing* 8/4 (1993), 211–219, 217.

³⁴ Vgl. Blessing, Andre/Bockwinkel, Peggy/Reiter, Nils/u. a., *Dramenwerkbank – Automatische Sprachverarbeitung zur Analyse von Figurenrede.*, in: *DHd* 2016, 281.

³⁵ So finden sich etwa in den Evangelien besondere Strukturelemente und auch Dialoge – wenn auch nicht in der Form klassischer Dramen. Vgl. Hitchcock, Francis RM, *Is the Fourth Gospel a Drama?*, in: *The Gospel of John as Literature*, 1993, 15–24; Johnson, Luke Timothy/Penner, Todd C., *The Writings of the New Testament: An Interpretation*. 1999, 183, 533; Stacey, David, *Prophetic Drama in the Old Testament*. 2018.

und Raum sind von Bedeutung, sondern auch Personen, ihre Charakterisierungen, der Erzähler, aber eben auch der Handlungsverlauf bzw. Plot³⁶. Eine Auswahl soll im Folgenden diskutiert werden.

2.3 Figurenanalyse

Die Figurenanalyse lässt sich in eine quantitative und eine qualitative Fragestellung trennen: Wann ist eine Figur präsent (in Dramen spricht man von der „Bühnenpräsenz“) und mit wem interagiert sie? Qualitativ kann man auch von Inhalten (der „Figurenrede“) oder nach Charakterisierungen fragen. Das erste wird durch die „Figurenkonstellation“ und ihre „Konfigurationsstruktur“ beantwortet: In der ersten werden die Person und ihre Interaktionen erschlossen, in der zweiten werden diese gegenübergestellt. Während die Extraktion von Figuren als Wortentität nicht sehr schwierig ist, ist die genaue Analyse der Interaktionen eine Herausforderung. In aktuellen Studien wird deshalb vor allem die „Kopräsenz“ untersucht, die angibt „welche Figuren zu welchem Zeitpunkt gemeinsam auftreten“³⁷. Nur selten werden dabei Modelle untersucht, um diese Interaktionen präzise zu beschreiben³⁸.

³⁶ Dabei ist die Analyse sowohl der Grundelemente in der Narrativen Exegese als auch in der Literaturwissenschaft von großem Interesse. So widmet Pfister ein ganzes Kapitel der „Informationsvergabe“ und weist bspw. mit der „Personenkonfiguration“ auf Strukturelemente hin, die eine große methodische Überschneidung bietet, vgl. dazu bspw. Pfister, Manfred, *Das Drama*. München 1988, 67ff; Cornils, *Vom Geist Gottes erzählen. Analysen zur Apostelgeschichte*, 75. Berlin beschreibt beispielsweise nur Charakteren und Erzählperspektiven, Uspenskij als Vertreter der Semiotik präsentiert eine wesentlich feinere Unterteilung und auch die Entlehnung von Ansätzen aus der Poetik ist möglich, vgl. Berlin, A., *Poetics and Interpretation of Biblical Narrative*. Winona Lake 1994, 23ff; Uspenskij, B.A., *Poetik der Komposition*. Frankfurt am Main 1975; Seybold, Klaus, *Poetik der prophetischen Literatur im Alten Testament*. Stuttgart 2010, 69ff; Alter, Robert, *The Art of Biblical Narrative*. New York 2011. Es ist also zu untersuchen, wo genau sich Überschneidungen und die Möglichkeit zur Methodenapplication ergeben.

³⁷ Krautter, Benjamin/Pagel, Janis/Reiter, Nils/u. a., »[E]in Vater, dächte ich, ist doch immer ein Vater«. Figurentypen und ihre Operationalisierung, in: *Zeitschrift für digitale Geisteswissenschaften* (2020).

³⁸ Vgl. bspw. Elson, David/Dames, Nicholas/McKeown, Kathleen, *Extracting social networks from literary fiction*, in: *Proceedings of the 48th annual meeting of the Association for Computational Linguistics 2010*, 138–147; Wiedmer, Nathalie/Pagel, Janis/Reiter, Nils, *Romeo, Freund des Mercutio: Semi-Automatische Extraktion von Beziehungen zwischen dramatischen Figuren.*, in: *DH 2020*; Elsner, Micha, *Character-based kernels for novelistic plot structure*, in: *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics 2012*, 634–644; Frank/Ivanovic, *Building literary corpora*

In neutestamentlichen Studien werden Figurenkonstellationen bislang händisch erzeugt³⁹.

Spannend ist nun der Hinweis Pfisters, darin „erschließt sich die Möglichkeit, Methoden der Soziometrie, wie sie für die Untersuchung sozialer Gruppenstrukturen entwickelt wurden, auf die Analyse der Struktur des Dramenpersonals zu applizieren.“⁴⁰ Und tatsächlich können Methoden der sozialen Netzwerkanalyse (deren Vorläufer als Soziometrie bezeichnet wurden) darauf angewendet werden, was wieder neue computerbasierte Analyseverfahren ermöglicht⁴¹. Die spannende und bis jetzt nicht final beantwortete Fragestellung lautet: Kann man ein soziales Netzwerk aus literarischen Texten automatisiert extrahieren oder lediglich eine Figurenkonstellation?

Für die qualitative Figurenanalyse wird die Analyse der Handlungsverben, des Sentiment (d. h. der Gefühle) und weiterer Figureninformationen vorgeschlagen⁴². Auch eine mögliche Klassifizierung in Protagonisten und Nebenfiguren wurde erfolgreich untersucht⁴³. Allerdings ist hier noch weitere Forschungsarbeit zu leisten:

Narratologische Basiskategorien wie Ereignis, Plot oder indirekte Charakterisierung sind bislang kaum automatisch extrahierbar. Das gilt umso mehr für höherstufige Operationen, z. B. die Identifikation und Deutung von komplexen Formen uneigentlicher Rede wie Symbolen oder die Verbindung von Erzähltexten mit Kontexten wie der Ideen-, Diskurs-, Sozial- oder Kulturgeschichte.⁴⁴

for computational literary analysis—a prototype to bridge the gap between CL and DH; Elson, David K., *Modeling narrative discourse*. 2012.

³⁹ Vgl. Cornils, *Vom Geist Gottes erzählen. Analysen zur Apostelgeschichte*, 75.

⁴⁰ Pfister, *Das Drama*, 233.

⁴¹ Dies ist eine absolute Randerscheinung in der Forschung. Vgl. methodisch dazu Dörpinghaus, Jens, *Die soziale Netzwerkanalyse: Neue Perspektiven für die Auslegung biblischer Texte?*, in: *BeTh 1* (2021); Trilcke, Peer, *Social network analysis (SNA) als Methode einer textempirischen Literaturwissenschaft*, in: *Empirie in der Literaturwissenschaft*, 2013, 201–247.

⁴² Vgl. Krautter/Pagel/Reiter/u. a., »[E]in Vater, dächte ich, ist doch immer ein Vater«. Figurentypen und ihre Operationalisierung.

⁴³ Vgl. Krautter, Benjamin/Pagel, Janis/Reiter, Nils/u. a., „Titelhelden und Protagonisten – Interpretierbare Figurenklassifikation in deutschsprachigen Dramen“, in: *Litlab Pamphlet 7* (2018).

⁴⁴ Jannidis, Fotis, *Computergestützte Erzähltextforschung*, in: *Grundthemen der Literaturwissenschaft: Erzählen*, 2017, 597–612, 606.

Eine Untersuchung biblischer Texte steht noch aus, wobei hier die später noch zu diskutierende Frage der Textgrundlage zu stellen ist: Erfolgt die Analyse auf den Grundtexten oder auf deutschsprachigen Übersetzungen mit allen daraus resultierenden Problemen? In jedem Fall ist hierbei gerade die Nennung von Plot und Handlung interessant, die nun als weiteres Element näher betrachtet werden soll.

2.4 Szenen und Perspektiven als Strukturelemente

Szenen und Perspektiven sind elementare Strukturelemente von literarischen Texten; in Dramen sind diese besonders prominent. Allerdings ist selbst die automatische Detektion dieser Elemente nicht trivial, auch wenn es bereits sehr vielversprechende Ergebnisse gibt⁴⁵. Grundsätzlich können Verfahren entweder mit sogenannten „Diskurshinweisen“ arbeiten – wobei die Perspektive auf die Handlung verloren geht – oder versuchen Inhaltsverzeichnisse zu generieren – wobei die lokalen Handlungsstrukturen aus dem Fokus geraten⁴⁶. Grundsätzlich bieten die meisten deutschen Bibelübersetzungen – genauso wie Dramen – bereits Überschriften und Einteilungen, allerdings ergeben sich dadurch zwei neue Fragestellungen: Wie sind diese aus literarischer Sicht zu interpretieren und sind diese überhaupt digital für eine Verarbeitung verfügbar? Unmittelbar damit verbunden ist die Verfügbarkeit computerlinguistischer Methoden und Modelle für biblische Sprachen.

Auch wenn es vielfältige methodische Ansätze in diesem Bereich gibt, so sind diese doch hochspezifisch und kaum generalisiert anwendbar⁴⁷. Hier ergibt sich das Problem der äußerst komplexen Modellbildung „in der Differenz zwischen dem, was im Text, verstanden als eine Folge von Buchstaben,

⁴⁵ „While the CRF-based model sets a high baseline for the tasks of act, scene and speaker recognition, the BERT-based models outperform the baseline in all measures.“ Pagel, Janis/Sihag, Nidhi/Reiter, Nils, Predicting Structural Elements in German Drama, in: Proceedings <http://ceur-ws.org> ISSN 1613-0073 (2021), 225.

⁴⁶ Vgl. Hearst, Marti A., Text Tiling: Segmenting text into multi-paragraph subtopic passages, in: Computational linguistics 23/1 (1997), 33–64; Doucet, Antoine/Kazai, Gabriella/Colutto, Sebastian/u. a., Overview of the ICDAR 2013 Competition on Book Structure Extraction, in: Twelfth International Conference on Document Analysis and Recognition (ICDAR'2013), 2013, 1438–1443; Pagel/Sihag/Reiter, Predicting Structural Elements in German Drama, 218.

⁴⁷ Vgl. bspw. auch Hanser, Eva/Mc Kevitt, Paul/Lunney, Tom/u. a., Text-to-Animation: Affective, Intelligent and Multimodal Visualisation of Natural Language Scripts, in: School of Computing and Intelligent Systems, Univ. Ulster (2009).

gegeben ist, und der vollständigen Repräsentation und Interpretation einer narrativen Welt“⁴⁸. Wir sehen hier also die Grenzen des derzeit Machbaren.

2.5 Handlungsverlauf und -analyse

Die Grenzen des Machbaren erreicht man auch bei dem Versuch, Handlungen automatisiert zu analysieren. Dies ist „sehr viel komplexer, da das narratologische Konzept eines Plots als Folge von Ereignissen aufgrund des problematischen Ereignisbegriffs nur schwer operationalisierbar ist.“⁴⁹ Doch gilt auch hier, dass bereits unterschiedliche Methoden genutzt wurden und diese unterschiedliche Ergebnisse liefern⁵⁰. Aus Sicht der Narrativen Exegese ist der Handlungsverlauf auch nicht eindeutig definierbar⁵¹. Wichtige Bemerkungen sind die Schwierigkeit, besonders in alttestamentlichen Erzähltexten Handlungsverläufe festzustellen,⁵² und die Hypothese, dass Handlung eng mit Charakterisierungen zusammenhänge⁵³. Auch die Zeit „est un trait caractéristique non seulement du récit cinématographique, mais aussi du récit oral, à tous ses niveaux d’élaboration esthétique“⁵⁴. Diese Erzählzeit verhält sich aber weder linear zum Text noch zur Handlung und ist oft schon für den menschlichen Leser schwer zu erkennen. Gleiches gilt auch für Orte, die mögliche Anhaltspunkte für die Analyse von Erzähleinheiten liefern könnten⁵⁵. Es ergibt sich daraus zunächst die Frage an die Narrative Exegese, ob der Handlungsverlauf überhaupt operationalisierbar ist oder ob sich ein Hand-

⁴⁸ Jannidis, *Computergestützte Erzähltextforschung*, 607.

⁴⁹ Ebd., 599; Vgl. aber auch Lahn, Silke/Meister, Jan Christoph, *Einführung in die Erzähltextanalyse*. Stuttgart 2016, 215.

⁵⁰ Vgl. die Analyse von Werken Mark Twains Kuhn, *Computerlinguistische Textanalyse in der Literaturwissenschaft?*, 22ff; Krautter, Benjamin, *Über die Attribution hinaus*, in: *Quantitative Ansätze in den Literatur- und Geisteswissenschaften*, 2018 289–314, 307.

⁵¹ Vgl. die Definitionen von „Plot“, „Handlungsverlauf“ etc. bei Bar-Efrat, Shimon, *Wie die Bibel erzählt*. Gütersloh 2006, 107; Golka, Friedemann W., Jakob, *Biblische Gestalt und literarische Figur: Thomas Manns Beitrag zur Biblexegese*. Stuttgart 1999, 19; Berlin, *Poetics and Interpretation of Biblical Narrative*, 39.

⁵² Das ist allerdings ein grundsätzliches Problem antiker Literatur: „Dans ce cas, les marques de composition sont toutes intérieures au texte, c’est-à-dire linguistiques.“ Meynet, Roland/Pouzet, Louis/Farouki, Naïla/u. a., *Rhétorique sémitique*. Paris 1998, 82. Dieser Aspekt soll im nächsten Abschnitt noch ausführlich diskutiert werden.

⁵³ Vgl. Berlin, *Poetics and Interpretation of Biblical Narrative*, 39.

⁵⁴ Genette, G., *Figures III*. Paris 2014, 75.

⁵⁵ Vgl. Bar-Efrat, *Wie die Bibel erzählt*, 200; 202; aber z. B. auch Hausendorf, Heiko/Schmitt, Reinhold, *Sprachliche Interaktion im Raum*, in: Arnulf Deppermann/Silke

lungsverlauf nicht primär aus der Fragestellung (1) ergibt und im Sinne der Auswertung (5) gedeutet werden muss. Solange hier keine Klarheit herrscht, kann auch die Computerlinguistik nur bedingt Resultate liefern.

Doch ergibt sich hier auch eine andere Perspektive: Gerade durch die vielfältigen Methoden der Computerlinguistik ergeben sich neue Ansätze für Handlungsverläufe. Wird zum Beispiel die Figurenkonstellation als Hilfe verwendet⁵⁶, um Handlungen zu analysieren, so muss dies in die Fragestellung eingefügt werden. Das Gleiche ergibt sich, wenn man Ketten von Verben („chains of verbs“) verwendet⁵⁷. Insofern muss eine neue Forschungsfrage gestellt werden: Ergibt sich eine sinnvolle Deutung des Textes durch die Verwendung mehrerer Handlungsverläufe, die zu einem vollständigen zusammengefasst werden können?

3. Narrative Theologie und digitale Methoden: kritische Beobachtungen

In dieser kurzen Auswahl von Methoden mussten naturgemäß einige Aspekte ausgelassen werden. Doch die Eingangsaspekte (1) und (2) betrifft vor allem die möglichen Forschungsfragen und auch die mögliche Datengrundlage, mit denen diese Fragen beantwortet werden können. Dabei zeigt sich: Digitale Inhalte sind auch für die theologische Forschung, Ausbildung und das Studium wichtiger geworden. Doch beschreibt Clivaz noch 2017 mit wenig enthusiastischen Worten, dass „[e]inige Forscher[] bereits Überblicksartikel zur Begegnung von Bibelwissenschaft und digitaler Wende vorgelegt“⁵⁸ hätten. Das Ergebnis fällt dann auch im Jahr 2020 ernüchternd aus: „Theologians have shown scant interest to this point in the tools for linking data, mapping, network analysis, text mining, and visualizing information that are fueling

Reineke (Hg.) Sprache im kommunikativen, interaktiven und kulturellen Kontext, 2018, 87–118.

⁵⁶ Vgl. Lehnert, Wendy G., Plot units: A narrative summarization strategy, in: *Strategies for natural language processing* (1982), 375–414.

⁵⁷ Vgl. Chambers, Nathanael/Jurafsky, Dan, Unsupervised learning of narrative schemas and their participants, in: *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP 2009*, 602–610.

⁵⁸ Clivaz, Claire, Die Bibel im digitalen Zeitalter: Multimodale Schriften in Gemeinschaften, in: *Zeitschrift für Neues Testament* 20/39/40 (2017), 35–57, 36.

digital scholarship in other disciplines.“⁵⁹ Lediglich im Bereich der digitalisierten Handschriften und im *Digital Academic Research and Publishing* sind Fortschritte sichtbar⁶⁰.

Allein für die Analyse biblischer Texte stellt sich ja schon die Frage: Welches Textes? Die Analyse des griechischen, hebräischen oder aramäischen Grundtextes ist – wie oben diskutiert – äußerst problematisch⁶¹. Die Analyse deutscher Bibelübersetzungen wiederum hat ihre eigenen Vorbedingungen, die sich in der Forschungsfrage niederschlagen müsste. Da aber auch in den verschiedenen theologischen Teildisziplinen sowohl mit digital erfassten empirischen Daten als auch mit digitalen Quellen (z. B. historischen Texten, hebräischen und griechischen Grundtexten oder anderer inzwischen frei verfügbarer Literatur) gearbeitet wird, ist auch die Frage nach ihrer Verfügbarkeit zu stellen. Dabei ist zu beobachten, dass diese höchstens in äußerst teuren kommerziellen Softwareprodukten angeboten werden und es bei deutschen Bibelübersetzungen meist eine restriktive Lizenzpolitik gibt⁶². Allerdings bieten diese Softwareprodukte keine Schnittstelle zu den herkömmlichen Methoden der quantitativen Textanalyse, des NLP und des Text Minings⁶³. Es gibt lediglich eine *Open Source* Bibliothek, die dafür genutzt werden kann; hier fehlen allerdings fast alle lizenzpflichtigen Grund- und Bibeltexte⁶⁴. Insofern ergibt sich bereits vor der Methodendiskussion schon eine ernsthafte Hürde, die Texte diesen überhaupt zugänglich zu machen.

⁵⁹ Anderson, Clifford, *Digital Humanities and the Future of Theology*, in: *Cursor_ Zeitschrift für explorative Theologie* (2018), <https://cursor.pubpub.org/pub/anderson-digitalhumanities-2018/release/4>. Zuletzt geprüft am: 27.8.2020, 2.

⁶⁰ Clivaz, Claire, *Digital Humanities in Biblical, Early Jewish and Early Christian Studies*, in: Andrew Gregory/David Hamidović/Claire Clivaz (Hg.) *Digital Humanities in Biblical, Early Jewish and Early Christian Studies*, 2013 Internet: <https://brill.com/view/title/24956>. Zuletzt geprüft am: 28.8.2020, 1–8.

⁶¹ Wobei die Ansätze der sogenannten Korpuslinguistik hier eventuell sinnvoll anwendbar wären, vgl. Mair, Christian, *Erfolgsgeschichte Korpuslinguistik?*, in: Marc Kupietz/Thomas Schmidt (Hg.) *Korpuslinguistik*, 2018, 5–26.

⁶² Man vergleiche die Informationen der DBG <https://www.die-bibel.de/ueber-uns/verlag/lizenzen/> und die offenen Prinzipien des Projekts „Offene Bibel“ unter https://offene-bibel.de/wiki/Unsere_Ziele#Freie_Verf.C3.Bcgbarkeit.

⁶³ Vgl. die ausführliche Darstellung bei Dörpinghaus/Düing, *Automated creation of parallel Bible corpora with cross-lingual semantic concordance*. Die bekannten Softwareprodukte lassen keinen automatisierten externen Zugriff auf die Daten zu.

⁶⁴ Es handelt sich um das SWORD Projekt, das eine vollständige API zur Verfügung stellt und unter GNU Lizenz steht. Siehe <http://crosswire.org/sword/>.

Ein weiterer Aspekt ist zu beachten: Liegt die Fragestellung in Schritt (1) in der Theologie, so muss Schritt (2) schon den Transfer in die Computerlinguistik vollziehen. Gleiches gilt aber für die Rückrichtung: In Schritt (5) sind die Ergebnisse der Computerlinguistik in der Theologie zu interpretieren. Übernimmt man Modelle der *Digital Humanities*, z. B. für andere Gattungen oder Sprachen, so ist letztlich das interdisziplinäre Gespräch zwischen drei Disziplinen zu suchen. Dies gilt – wie oben diskutiert – für jede Verwendung computergestützter Verfahren über die bloße Enumeration hinaus. Denn die von Jannidis genannten Herausforderungen in der „Modellbildung, der Repräsentativität, der Domänenadaption und der Evaluation“⁶⁵ gelten ebenso für die Theologie. Trotz allem können natürlich etablierte Verfahren der Computerlinguistik auch in der Narrativen Exegese unter Beachtung der methodischen Grundlagen genutzt werden.

Gleichzeitig ergeben sich aber auch weitere Möglichkeiten, die im interdisziplinären Gespräch zwischen *Digital Humanities*, der Computerlinguistik und der Theologie ausgelotet werden können. Erstens verfügt die Theologie über mehrsprachige, gut erforschte Texte und Fragestellungen. Wenn hier ein offenes wissenschaftliches Miteinander gefunden werden könnte, würden sicherlich alle Seiten davon profitieren⁶⁶. Zweitens würde die systematische digitale Verfügbarkeit theologischer Quellen die kommunikativen wissenschaftlichen Möglichkeiten der Theologie erweitern und im letzten Schritt auch die Anwendung von Technologien auf Wissensgraphen ermöglichen⁶⁷. Drittens ergibt sich eine weitere spannende Komponente mit Blick auf die Praktische Theologie und Gemeindepraxis, da Methoden der Computerlinguistik das digitale Leben bereits beherrschen und es hier nicht nur kommu-

⁶⁵ Jannidis, Computergestützte Erzähltextforschung, 607.

⁶⁶ Vergleiche die Arbeiten an parallelen Bibelkorpora, Wälchli, Bernhard, Similarity semantics and building probabilistic semantic maps from parallel texts, in: *Linguistic Discovery* 8/1 (2010), 331–371. Auch als Grundlage für automatisierte Übersetzungen werden biblische Texte genutzt, vgl. Christodouloupoulos, Christos/Steedman, Mark, A massively parallel corpus: the bible in 100 languages, in: *Language resources and evaluation* 49/2 (2015), 375–395.

⁶⁷ Da die Methodendiskussion der Informatik meist eine untergeordnete Rolle in der Theologie spielt und der Begriff Open Data bzw. FAIR Data noch relativ unbekannt ist, können die Ergebnisse nur selten reproduziert werden bzw. stehen für Data Fusion, d. h. der Verbindung mit anderen Datenquellen, zur Verfügung, vgl. Wilkinson, Mark D./Dumontier, Michel/Aalbersberg, IJsbrand Jan/u. a., The FAIR Guiding Principles for scientific data management and stewardship, in: *Scientific Data* 3 (2016), 160018.

nikative Anknüpfungspunkte gibt, sondern auch für weitere wissenschaftliche Fragestellungen jenseits der Narrativen Exegese bis hin zur Narrativen Theologie⁶⁸.

4. Schlussfolgerungen

Als Fazit dieser Arbeit muss zunächst festgehalten werden, dass die computergestützte Analyse von natürlicher Sprache ein emergentes Thema in der Computerlinguistik und den Literaturwissenschaften ist. Es ist darum nicht verwunderlich, dass einfache Methoden dieser Disziplinen bereits in der Theologie genutzt werden. Allerdings wird dieser Schritt methodisch bislang zu wenig in seinem interdisziplinären Spannungsfeld wahrgenommen. Gleichzeitig ergeben sich – neben den oben genannten weiteren Möglichkeiten – bereits mit den jetzt schon etablierten Methoden spannende Aspekte für die Narrative Exegese: Wie lassen sich Raum und Zeit mit Methoden der Computerlinguistik untersuchen? Ergäbe eine Sentiment Analysis neue Perspektiven auf Erzählungen? Wie verhält sich die soziale Netzwerkanalyse methodisch zu der Narrativen Exegese? Welche Möglichkeiten der verschiedenen Teilspannungsbögen, die mittels computerbasierter Verfahren bestimmt werden, ergeben sich für einen Blick auf den Spannungsbogen von Erzähltexten? Es ist wichtig, zu beobachten, dass nicht nur die Ergebnisse der Computerlinguistik neue Perspektiven ermöglichen, sondern auch die Methodendiskussion an sich schon Impulse für die Bibelwissenschaften liefern kann. Fasst man das Feld noch weiter, so ist auch das interdisziplinäre Feld der *Digital Humanities* bzw. *Digital Theology* als möglicher Impulsgeber zu nennen.

Dr. Jens Dörpinghaus, u21829927@tuks.co.za

⁶⁸ Vgl. z. B. die Arbeit des ODEC Research Centre at Durham University, Phillips, Peter/Schiefelbein-Guerrero, Kyle/Kurlberg, Jonas, Defining digital theology: Digital humanities, digital religion and the particular work of the Codec Research Centre and Network, in: *Open theology* 5/1 (2019), 29–43.

Abstract

This paper discusses possible intersections between computer-based natural language processing, analysis of narrative and literary texts, and narrative exegesis of biblical texts. It provides an overview of already established methods and discusses a critical evaluation of possible synergies with linked data and knowledge graph approaches. Despite limitations regarding data management and accessibility several methods could be suitable for interdisciplinary research: The detection of time and space, sentiment analysis and the integration of social network analysis into narrative exegesis, and the discussion with computer linguistics may give new perspectives for narrative exegesis.