Institut für Visualisierung und Interaktive Systeme

Universität Stuttgart
Universitätsstraße 38
D–70569 Stuttgart

Bachelorarbeit

Interaktive, vergleichende Visualisierungen von Textveränderungen

Wolfgang Knopki

Studiengang:	Informatik
Prüfer/in:	Prof. Dr. Thomas Ertl
Betreuer/in:	Dipl. Inf. Hermann Pflüger Dipl. Math. Martin Baumann, M.A. Markus John, M.Sc.
Beginn am:	2018-03-27
Beendet am:	2018-09-27

Kurzfassung

Die Analyse der Textentstehung anhand mehrerer überlieferter Versionen ist mit einer der ältesten Zweige der literaturgeschichtlichen Forschung. Insbesondere bei großen Texten ist jedoch schon allein des Umfangs wegen die Nutzung informationstechnischer Methoden unumgänglich, insbesondere die Visualisierung der Texte und der Unterschiede zwischen ihnen ist von großer Bedeutung.

Im Rahmen dieser Bachelorarbeit wurde ein Werkzeug zum visuellen Vergleich zweier Texte als Prototyp entwickelt und wird hier am Beispiel eines Legendariums, also einer Sammlung von Heiligenlegenden, aus dem 15. Jahrhundert vorgestellt. Das Werkzeug wurde in enger Kooperation des Institutes für Visualisierung und Interaktive Systeme (VIS) und der Abteilung für germanistische Mediävistik des Institutes für Literaturwissenschaft der Universität Stuttgart entwickelt. Es werden sowohl *Close*- als auch *Distant-Reading*-Ansätze implementiert und verschiedene Interaktions- und Filtermöglichkeiten bereitgestellt.

Im Abschluss wird die Evaluation durch entsprechende Experten aus der Literaturwissenschaft vorgestellt.

Abstract

Analysis of textual formation using multiple witnesses is amidst the oldest branches of literary research. Especially as large texts are concerned using IT methods is inevitable. Visualization of texts and the differences between them is vitally important.

This bachelor thesis aims at developing a tool to visual-analytically compare two texts using the example of a legendarium, a collection of legends of saints dating back to the 15th century. The institute for visualisation and interactive systems (VIS) and the department of germanic medieval studies of the institute of literary studies cooperated in developing the framework. Close-reading and distant-reading-approaches were implemented as well as different possibilities for interaction and filtering.

The prototype being developed was later evaluated by experts on the field of literary studies.

Inhaltsverzeichnis

Abbildungsverzeichnis

1. Einführung

1.1. Motivation

Anfang des 15. Jahrhunderts, vermutlich um 1406, entstand in Nürnberg eine Legendensammlung, die mit ca. 400 Legenden das wohl umfangreichste deutschsprachige Legendar des Mittelalters darstellt. Sie integriert sämtliche Texte des in Europa erfolgreichsten Legendars seiner Zeit, dem Legendar „Der Heiligen Leben“, das 251 Texte umfasst, und fügt diesem noch ca. 150 weitere Legenden hinzu, sodass für jeden Tag des Jahres mindestens eine solche Legende zur Verfügung steht. Dieses umfangreiche, dreibändige Werk, das in der Forschung als „Der Heiligen Leben, Redaktion“, im folgenden HL-Red I, geführt wird, wurde kurz nach seiner Entstehung vollständig neu bearbeitet (noch vor 1447): Jede einzelne Legende wurde wiedererzählt, also stilistisch umgestaltet, zum Teil durch weitere Details ergänzt, mitunter auch gekürzt oder neu arrangiert. Eine systematische Untersuchung der Bearbeitungstendenzen dieses Legendars (HL-Red II) im Vergleich mit HL-Red I ist in der Germanistischen Mediävistik dringend gewünscht.

Schon allein der Umfang des Legendars stellt jedoch ein grundlegendes Problem bei seiner Analyse dar, weshalb eine systematische Untersuchung unter philologischen, insbesondere literaturhistorischen, Fragestellungen bislang überaus aufwändig war. Um eine vollständige Analyse nun erstmals zu ermöglichen, wurde in der vorliegenden Bachelorarbeit anhand ausgewählter Legenden ein Visualisierungswerkzeug als Prototyp entwickelt, das sich mit den Textveränderungen der HL-Red II gegenüber der HL-Red I beschäftigt.

1.2. Zielsetzung

Das Ziel dieser Arbeit war, ein Vergleichswerkzeug zur visuellen, interaktiven Exploration von Textveränderungen prototypisch zu entwickeln. Dabei wurde besonderes Augenmerk auf den visuellen Vergleich beider Texte gelegt. Außerdem werden unterschiedliche Ansichten und Interaktionsmöglichkeiten zur Verfügung gestellt, um die Experten bei der Analyse der Textveränderungen zu unterstützen. Das Werkzeug implementiert *Close- and Distant-Reading*-Ansätze, Experten haben durch Textverlinkung die Möglichkeit, zwischen der Gesamtansicht der Texte und der jeweiligen Textpassage zu springen.

Die Entwicklung des Ansatzes und die prototypische Implementierung erfolgte in Kooperation des Instituts für Visualisierung und Interaktive Systeme mit der Abteilung Germanistische Mediävistik des Instituts für Literaturwissenschaft. Die enge

Zusammenarbeit des Instituts für Visualisierung und Interaktive Systeme mit der Abteilung Germanistische Mediävistik des Instituts für Literaturwissenschaft sollte zu einem Werkzeug führen, das sowohl für den beschriebenen Einsatzbereich als auch für die entsprechende Zielgruppe geeignet ist. Die Evaluation sollte gemeinsam erfolgen.

2. Grundlagen

Im folgenden Kapitel werden die Grundlagen der in der Arbeit behandelten Themengebiete angerissen. Es wird dabei zwischen Grundlagen der Informationsvisualisierung auf Seiten der Informatik und der Digital Humanities auf Seiten der Literaturwissenschaft unterschieden.

2.1. Grundlagen der Informationsvisualisierung

Visualisierung bezeichnet einen Teilbereich der Informatik, welcher sich mit der bildhaften Darstellung von Daten beschäftigt. Diese fallen heutzutage meist in Mengen an, die eine Analyse im klassischen Sinne ohne HIlfsmittel unmöglich machen. In der Visualisierung wird im Allgemeinen zwischen der Informationsvisualisierung, der wissenschaftlichen Visualisierung und der *Visual Analytics* unterschieden.

In der Informationsvisualisierung werden Informationen, die sonst nur abstrakt vorliegen, durch Visualisierung aufbereitet. Eng damit verwandt ist die Wissenschaftliche Visualisierung, in der die, beispielsweise durch Experimente gewonnenen, Daten aufbereitet und dann dargestellt werden. Charakteristisch für diese Daten ist dabei meist, dass sie eine räumliche Komponente enthalten, wie beispielsweise bei der Volumen- oder Flussvisualisierung. Die *Visual Analytics* erweitern nun diesen klassischen Visualisierungsbegriff um weitere Bereiche, wie Datenanalyse oder Mensch-Computer-Interaktion.

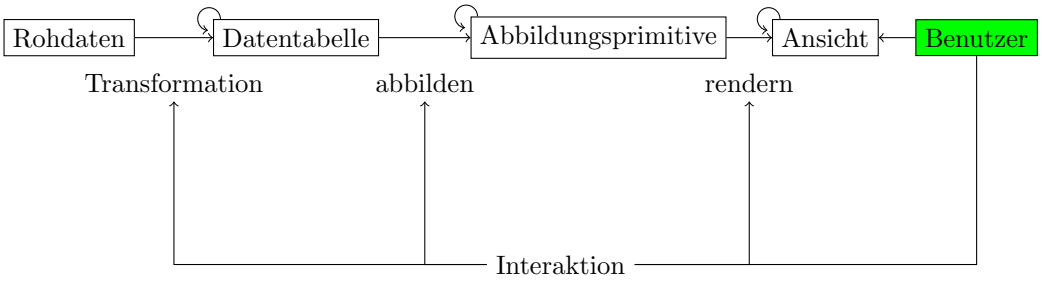


Abbildung 2.1.: Visualisierungspipeline nach Card [1]

Jeder Darstellung von abstrakten Daten gemein ist die in Abbildung 2.1 zu sehende Visualisierungspipeline: Die Daten liegen zumeist als Rohdaten vor. Die Form ist dabei beliebig, es kann sich beispielsweise um Ergebnisse aus Experimenten oder um Text handeln. Durch Transformation wird nun aus diesen Daten eine Datentabelle

abgeleitet. Eine Datentabelle bezeichnet hierbei eine Funktion auf Objekten in ihre Eigenschaften, also eine Funktion der Form:

$$f : \Omega^m \to E^{m \times n}, \quad m, n \in \mathbb{N}$$

$$\texttt{Objekt}_i \mapsto (\texttt{Eigenschaft}_{i,0}, \ldots, \texttt{Eigenschaft}_{i,n})$$

wobei Ω^m den Raum aller Objekte, $E^{m \times n}$ den Raum ihrer Eigenschaften bezeichnet. Diese Funktion wird nun auf geeignete Abbildungsprimitive bzw. Visualisierungswerkzeuge abgebildet. Diese werden letztlich durch Rendering als Visualisierung dem Benutzer angezeigt. Der Benutzer wiederum kann in jedem einzelnen Schritt der Pipeline durch Interaktion auf die Visualisierung Einfluss nehmen.

Folgendes Beispiel aus der vorliegenden Arbeit soll diese abstrakte Pipeline nun verdeutlichen: Zu Beginn liegen beide Texte als einzelne Token (Wörter) vor. In der Datentabelle wird nun für jedes Token festgehalten, welchem Text das Token zugehörig ist, ob es eine Verbindung zu Token im anderen Text gibt, welcher Kategorie diese Verbindung mit welcher Sicherheit angehört, sowie weitere Eigenschaften. Abhängig von diesen Eigenschaften wird das Token nun im Text markiert, also entsprechend hervorgehoben. Diese Markierung wird dann entsprechend gerendert und im Browser dargestellt. Die Interaktionen des Benutzers, wie Auswahl gewisser Kategorien oder Anwahl des entsprechenden Tokens haben dabei direkten Einfluss zunächst auf die Datentabelle („wurde das Token angewählt“ ist beispielsweise eine dort abgespeicherte Eigenschaft) als auch auf das Rendering und damit die endgültige Ansicht.

Wichtige grundlegende Techniken der Informationsvisualisierung wurden in den 80er Jahren des vergangenen Jahrhunderts unter anderem von Cleveland und McGill entwickelt. In obiger Visualisierungspipeline entsprechen diese Techniken den Abbildungsprimitiven, die entsprechend der Natur der Daten und ihnen zugeordneten Datentabellen eingesetzt werden. Cleveland und McGill beschäftigten sich mit Studien zur visuellen Wahrnehmung [2] und den sich daraus ergebenden Schlussfolgerungen für die Visualisierung wissenschaftlicher Daten [3]. Sie forschten ferner auf dem Gebiet der Graphvisualisierung und entwickelten folgende mittlerweile zum Standard eines jeden Visualisierers gehörenden Werkzeuge:

Säulendiagramme

Säulendiagramme erfordern zum Auslesen ein Abschätzen der Länge und zugehörigen Fläche der Säulen. Sie sind besonders für die Visualisierung diskreter Daten geeignet. Zur Visualisierung von mehrdimensionalen Daten können auch geteilte oder zusammengefügte (siehe Abbildung 2.2) Säulendiagramme Anwendung finden:

Tortendiagramme

Die Wahrnehmung von Tortendiagrammen basiert auf der Schätzung der zugehörigen Winkel im Diagramm. Sie werden besonders für die Visualisierung von Verhältnissen verwendet.

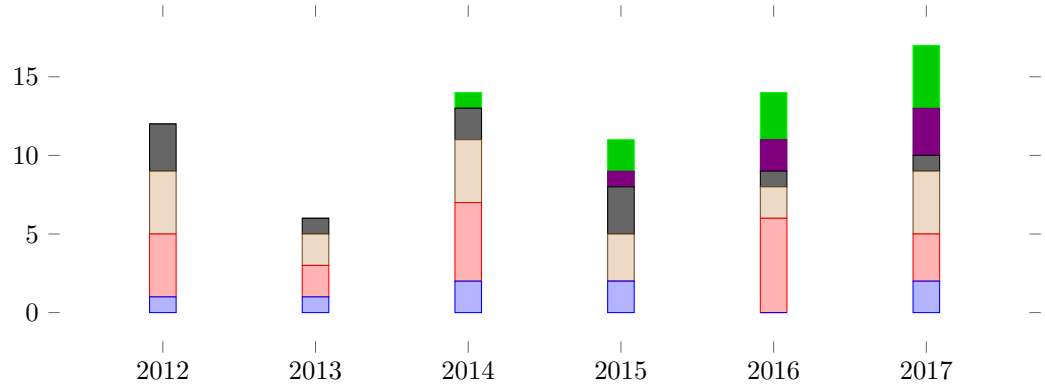


Abbildung 2.2.: zusammengefügte Säulendiagramme

Einfärbung

Die Verwendung von Farben oder Schraffur stellt einen der offensichtlichsten Wege der Informationskodierung dar. In Bildern oder Karten wird sie meist zur Kodierung von Informationen in Abhängigkeit von geospatialen Koordinaten genutzt. Ein Beispiel ist in Abbildung 2.3 zu sehen.

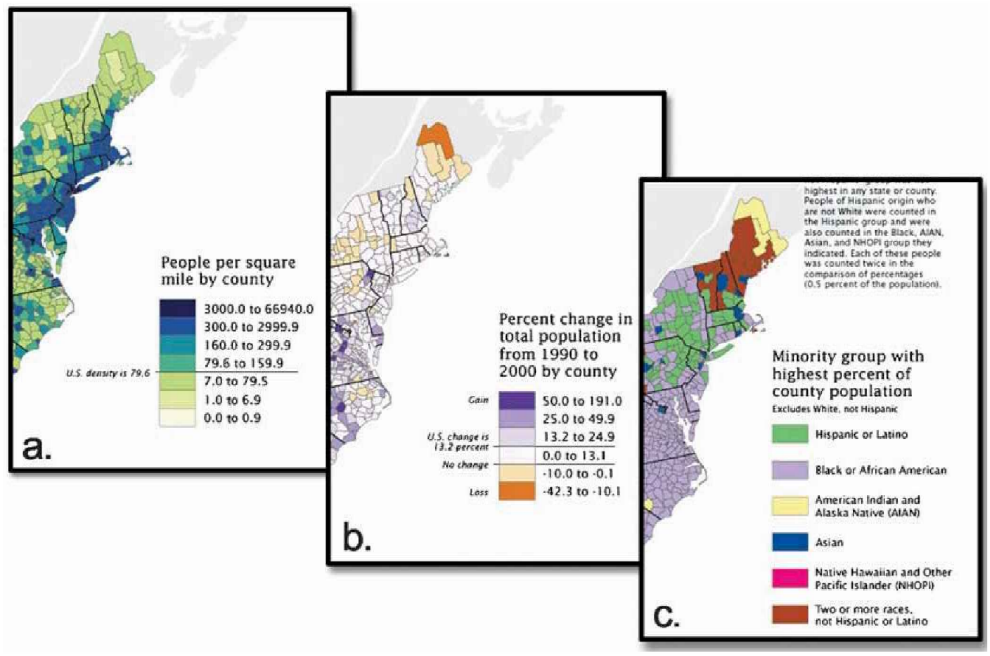


Abbildung 2.3.: Karten mit Einfärbung [4]

Die Geschichte der Visualisierung von Texten reicht zurück in die 90er Jahre, als sich Eick et al. [5] bei der Entwicklung von Seesoft mit der Visualisierung der Codebasis großer Softwareprojekte auseinandersetzten. Dabei wurde die Visualisierung von Texten mit der Repräsentation der Textzeilen als Linien (siehe Abbildung 2.4) maßgeblich entwickelt.

In der Textvisualisierung wurden in den letzten Jahren noch weitere Visualisierungstechniken eingeführt: So könnnen in der Schriftgröße Informationen kodiert werden, oder durch den Einsatz von Glyphen beispielsweise Auslassungen angezeigt werden. Auch die Verwendung von Verbindungslinien ist mittlerweile gängige Praxis. Einen Überblick über den aktuellen Stand der Textvisualisierung geben Jänicke et al [6].

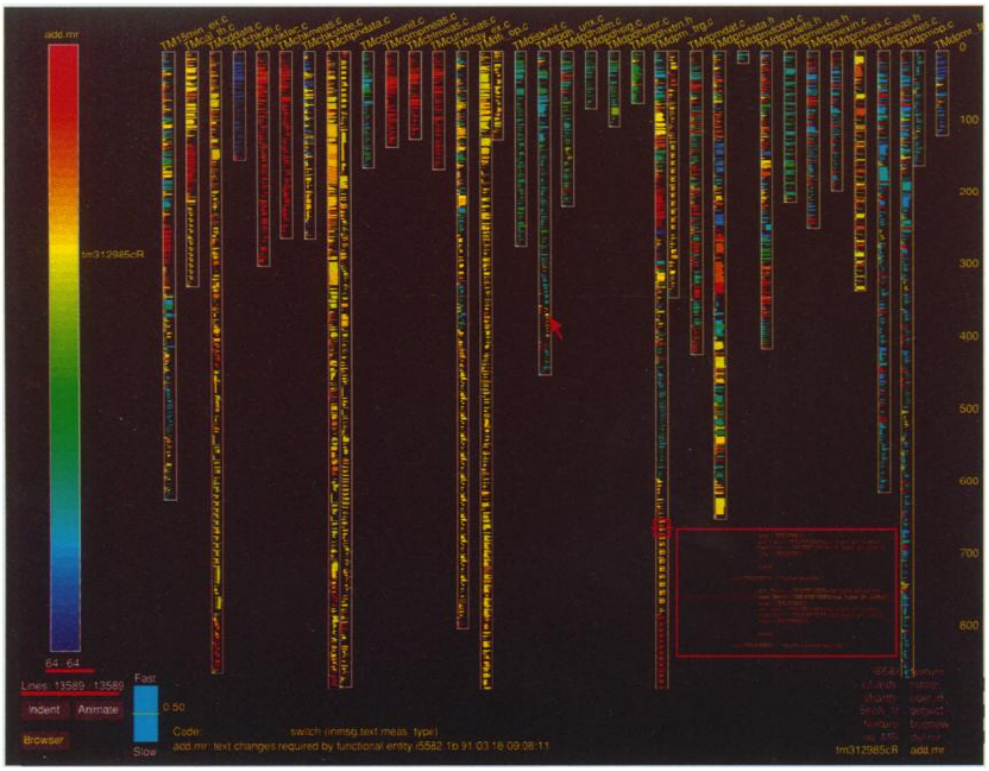


Abbildung 2.4.: Beispiel in SeeSoft: Codezeilen werden nach Alter farbig markiert: Alte Zeilen in Blau, neueste Zeilen in Rot, dazwischen ein Spektrum [7]

2.2. Grundlagen der DH

Abbildung 2.5.: Beispiel für *Distant-Reading*: Wordcloud mit Worten aus fünf Stücken Shakespeares [6]

Die Digital Humanities sind ein verhältnismäßig junges Forschungsgebiet, welches zwischen der Informatik und den klassischen Geisteswissenschaften angesiedelt ist. Es beschäftigt sich dabei vor allem damit, inwieweit informationstechnische Methoden und Verfahren auf Forschungsfragen aus der Geisteswissenschaft angewandt werden können. Bezüglich der Anwendung solcher Verfahren auf die Literaturwissenschaft gibt es seit Jahren bereits Anstrengungen, digitale Archive aufzubauen, um die literaturwissenschaftlichen Urtexte und auch Forschungsergebnisse effizient und nachhaltig zugänglich aufbewahren zu können. Im Rahmen der Archivierung und Digitalisierung solcher Texte wurde auch eine gewisse Standardisierung angestrebt. Die Text Encoding Initiative (TEI) [8] stellt hierbei einen de-facto Standard für die Übertragung von

Texten in XML-basierte Datenformate dar.

Auch in der direkten Anwendungsentwicklung wurden nach und nach Fortschritte erzielt. So wurden beispielsweise im Rahmen des Projektes ePoetics [9] von 2013 bis 2016, welches sich mit der Analyse deutschsprachiger Poetiken des 18. bis 20. Jahrhunderts auseinandersetzte, unter anderem Werkzeuge zur Textanalyse und Textvisualisierung entwickelt. Eines der dabei entwickelten Werkzeuge ist der Varifocal Reader (siehe 3.2) In ihm vereinen sich *Close-* und *Distant-Reading*-Ansätze.

Close-Reading bezeichnet den klassischen Analyseansatz der Literaturwissenschaft: Texte werden Zeile für Zeile und Wort für Wort analysiert, die dabei gewonnenen Erkenntnisse belaufen sich meistens auf Vokabular und grammatische Figuren, sowie inhaltliche Punkte, wie Figurenentwicklungen oder entwickelte Ideen. Dem gegenüber steht das 2005 von Franco Moretti [10] eingeführte *Distant-Reading*: Texte werden nicht mehr Zeile für Zeile analysiert, sondern abstrahiert in ihrer Gesamtheit unter Einsatz quantitativer Verfahren betrachtet. Wichtige Hilfsmittel sind dabei die Verwendung von Visualisierungen, wie beispielsweise Karten zur Visualisierung von Geodaten, Wordclouds (siehe Abbildung 2.5) oder Verbindungsgraphen.

3. Verwandte Arbeiten

Zur Visualisierung von Textveränderungen, auch von mittlelalterlichen Texten, existiert bereits eine Vielzahl von Ansätzen. Einige davon werden im Folgenden vorgestellt und in ihrer Relevanz für diese Arbeit diskutiert:

3.1. TRAViz: A Visualization for Variant Graphs

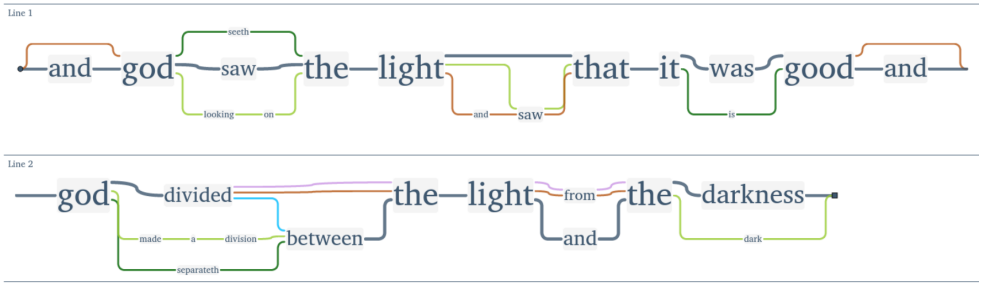


Abbildung 3.1.: Vergleich von Genesis 1:4 unter Verwendung von TRAViz [11]

Jänicke et al [11] entwickelten mit TRAViz eine Bibliothek zur Visualisierung von Veränderungsgraphen und damit zur Visualisierung von Textunterschieden. Dabei werden fünf grundlegende Designentscheidungen implementiert:

1. Kodierung der Häufigkeit durch Größe des Textes. Ein Wort, welches also in einer Vielzahl der untersuchten Texte vorkommt, wird entsprechend größer erscheinen.
2. Keine rückwärtigen Kanten. Es wird genau eine Leserichtung des Graphen angenommen, was entsprechend der Natur der Graphen als Textrepräsentation geschuldet ist.
3. Vermeidung der Kantenbeschriftung. Diese wird als zu störend empfunden. Stattdessen werden unterschiedliche Urtextversionen durch unterschiedliche Kantenfarbe kodiert.
4. Benutzung von Kantenbündelung, um die visuelle Belastung möglichst gering zu halten. Gebündelte Kanten werden entsprechend durch graue Farbe dargestellt
5. Implementierung von Zeilenumbrüchen, um die bequeme Analyse auch großer (sprich langer) Graphen zu ermöglichen.

Es werden auch Interaktionsmöglichkeiten mit der Visualisierung angeboten, so kann beispielsweise der Graph angepasst werden und neue Kanten hinzugefügt bzw. bestehende Kanten zusammengefasst werden.

TRAViz ist allerdings ein Tool, welches beinahe ausschließlich im Bereich des *Close-Readings* auf Satz- bzw. Wortebene eingesetzt werden kann. Für *Distant-Reading*-Ansätze muss auf andere Technologien zurückgegriffen werden (siehe 3.3).

3.2. Varifocal Reader

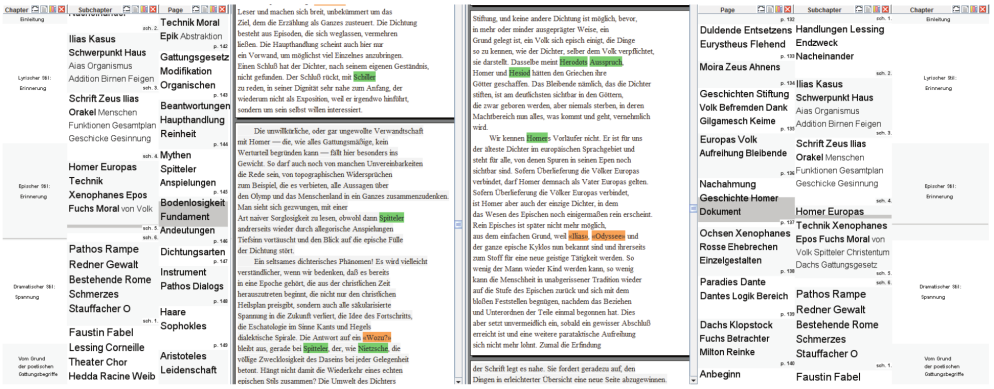


Abbildung 3.2.: Vergleich zweier Dokumente im Varifocal Reader [12]

Koch et al [12] stellten mit ihrem Varifocal Reader ein Werkzeug für die DH zusammen, welches zum einen *Close-* und *Distant-Reading*-Ansätze vereint, und zudem auch die Möglichkeit der automatischen Annotation mittels *Natural Language Processing* (NLP) bietet. Der Varifocal Reader wurde im Rahmen des Projektes ePoetics entwickelt und ist für die Analyse eines Korpus aus Poetiken zugeschnitten. Somit bietet er eine Vielzahl an Analyseebenen: Wordclouds auf Kapitelebene unterstützen die Analyse des Inhalts, es ist ebenso möglich, die originalen Handschriften (bzw. im Fall der Poetiken Scans) neben den Texten darzustellen. Auch ist die händische Annotation des Korpus möglich.

Für den in dieser Bachelorarbeit angestrebten Vergleich zweier Texte ist der Varifocal Reader allerdings nach Aussage der Entwickler nur bedingt geeignet („Comparing multiple text sources, however, is not well supported by our approach.“[12]). Um zwei Texte vergleichen zu können (siehe Abbildung 3.2), müssen zwei Fenster des Varifocal Readers nebeneinander gestellt werden, eine direkte Verlinkung ist damit also nicht möglich.

Es wurden jedoch einige Konzepte übernommen, insbesonders die Verwendung von *Close-* und *Distant-Reading*-Ansätzen.

3.3. Interactive Visual Alignment of Medieval Text Versions

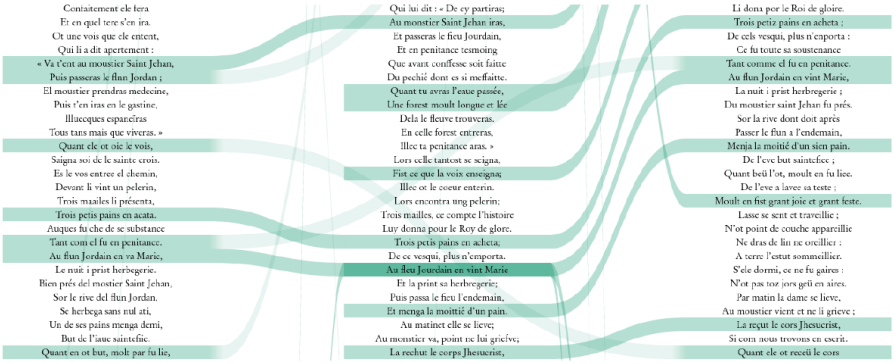


Abbildung 3.3.: Meso-Reading Ansatz aus Jänicke und Wrisley: Interactive Visual Alignment of Medieval Text Versions [13]

Jänicke und Wrisley [13] verfolgen einen Ansatz, der sich ebenso auf die Implementierung von *Close-* und *Distant-Reading* fokussiert. Sie erweitern diese jedoch noch um den sogenannten *Meso-Reading*-Ansatz, welcher nochmals eine Betrachtungsebene zwischen dem *Distant-Reading*, also der Betrachtung des gesamten Textes in Abstraktion, und dem *Close-Reading*, also der Betrachtung der einzelnen Textstelle, hinzufügt.

Somit ist es möglich, die in ihrer Arbeit betrachteten mittelalterlichen französischen Gedichte auch auf Versebene zu analysieren.

In ihrer Arbeit wurden ferner Algorithmen zur semiautomatischen Alignierung der Texte entwickelt. Diese Texte und ihre entsprechende Alignierung wurden dann Seite an Seite dargestellt. In der *Close-Reading*-Ansicht wurde dabei auf die Bibliothek TRAViz (siehe 3.1) zurückgegriffen. Für die Alignierung wird dabei die zeilenweise Ähnlichkeit der zu vergleichenden Gedichte computergestützt berechnet. Dabei wird der Text auf ähnliche N-Gramme (auch unter Berücksichtigung unterschiedlicher Orthographie) hin untersucht, um darüber dann die Alignierung mit einer gewissen Sicherheit zu berechnen. Das dabei entwickelte Verfahren zur computergestützten Textalignierung konnte allerdings nicht auf die vorliegende Arbeit angewandt werden, da die untersuchten Texte aus dem Legendar sich schon strukturell sehr von mittelalterlicher französischer Poesie unterscheiden.

3.4. Juxta Commmons

Juxta Commons ist ein webbasiertes Werkzeug zur Annotation und Visualisierung mehrerer Texte.

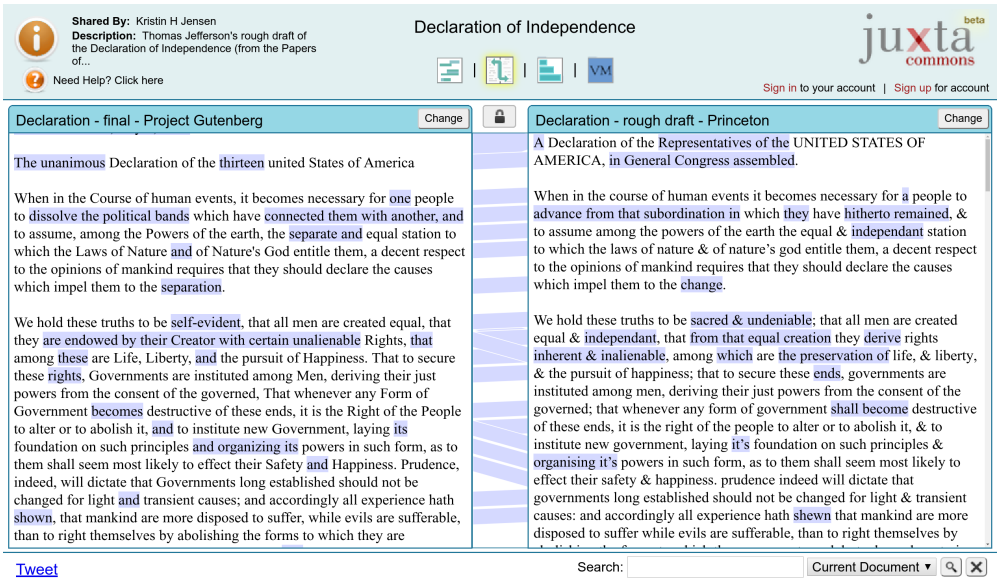


Abbildung 3.4.: Textvergleichsansicht in Juxta Commons [14]

Es bietet dabei weitreichende Visualisierungsmöglichkeiten, wie die Verwendung sogenannter *Heatmaps*, also einer Markierung der unterschiedlichen Textstellen, für eine allgemeine Übersicht der Textunterschiede. Mit einem Histogramm ist es möglich, die Abweichung der Texte je Zeile zu visualisieren. Ferner existiert auch eine direkte Vergleichsansicht, in der die Texte nebeneinander dargestellt werden können.

In der aktuellsten Version ist auch eine Versioning Machine [15], also ein Browser für TEI-konformes XML, eingebaut.

Für die vorliegende Arbeit ist Juxta Commons allerdings nur bedingt geeignet, da die geforderte Differenzierung nach Kategorien nicht möglich ist.

4. Konzept

Um Aufgabenbereich, Art und Umfang des Datenmaterials und der Fragestellungen zu ermitteln, wurden initale Interviews mit den Experten des Instituts für Literaturwissenschaft geführt. Dabei kristallisierten sich die im Folgenden beschriebenen Anforderungen an das zu entwickelnde Werkzeug heraus. Der Entwurf des Werkzeuges erfolgte dann in Zusammenarbeit mit den Experten in mehreren Schritten.

4.1. Anforderungen

Von Seiten der Literaturwissenschaft wurden folgende Anforderungen formuliert. Das Tool sollte

- beide Texte darstellen.
- Annotationen in unterschiedlichen Kategorien unterstützen.
- gegebenenfalls auch die originalen Handschriften neben den transkribierten Texten darstellen.
- XML-basierte Datenformate (De-Facto Standard der DH) verwenden.
- später: die Texte auch nur teilweise (z.B. nur Phänomene einzelner Kategorien) darstellen.

In Tabelle 4.1 sind die zwölf Kategorien zu sehen, die zur Annotation vorgegeben wurden.

Von Seiten der Entwickler führte dies zu folgenden Anforderungen an das zu entwickelnde Werkzeug:

- Implementierung von *Close-* und *Distant-Reading*
- zeitgleicher direkter Vergleich beider Texte
- Verlinkung aller Vorkommen in allen Ansichten untereinander
- Visualisierung sich überlappender Textannotationen verschiedener Kategorien
- Bereitstellung von Unsicherheitsinformationen (Wie sicher ist die Annotation in Kategorie A und nicht in Kategorie B?)
- Filter nach Annotationskategorien

	Kategorie	Beispiel	
0	zusätzliche Adjektive	Er hatte eine edle Frau.	Er hatte eine seelige edle Frau.
1	Zusätze innerhalb des Satzes	Von Sankt Klara	Von der heiligen Jungfrau Sankt Klara
2	zusätzliche Episode	Sie bat für die Sünder.	Sie bat für die Sünder. Sodann bat sie auch für die Alten und Kranken.
3	Synonyme	Es kam der böse Geist über sie.	Es kam der Teufel über sie.
4	ähnliche Semantik	Und es kam, so wie ihr gesagt wurde.	Und es kam, so wie es verkündet wurde.
5	veränderte Semantik	Da sprach eine himmlische Stimme zu ihr.	Da kam eine himmlische Stimme zu ihr.
6	andere Namensform	Und da kam der Kaiser Friedrich zu ihr.	Und da kam der Kaiser Friederich zu ihr.
7	zusätzliche Personen	Und sie bat für ihre Schwester.	Und sie bat für ihre Schwester St. Agnes.
8	zusätzliche Ortsnamen	Sie ging ins Kloster.	Sie ging ins Kloster St. Barbara.
9	syntaktische Umstellung	Und sie war so rein und keusch.	Und sie war so keusch und rein.
10	genera Verbi (Passiv vs. Aktiv)	Er sprach von der Liebe.	Es wurde von der Liebe gesprochen.
11	Oratio (direkte vs. indirekte Rede)	Er sprach: „Es werde Licht“	Er sprach, dass es Licht werde.

Tabelle 4.1.: Annotationskategorien

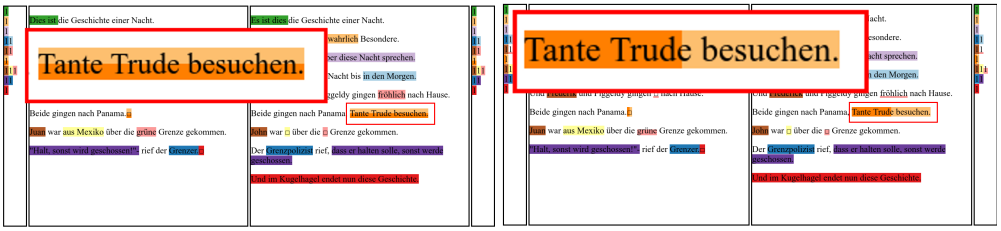
4.2. Entwurf

Basierend auf diesen Anforderungen wurde nun ein entsprechendes Konzept entwickelt und schrittweise optimiert. Um den direkten Vergleich beider Texte schon auf den ersten Blick zu erreichen, sollten diese nebeneinander dargestellt werden. Auf diese Weise konnte die Vereinigung von *Distant-* und *Close-Reading*-Ansätzen verwirklicht werden.

In ersten Mockups wurde die Übersicht über die Gesamttexte, also die *Distant-Reading-View*, außen zu beiden Seiten der Texte verortet. Dies war jedoch im Hinblick auf die Verknüpfung beider Texte ungünstig, sodass sie in späteren Versionen mittig in einer eigenen Spalte platziert wurde. Hier konnte nun die Verknüpfung beider Texte durch Verbindungslinien zwischen den beiden Vorkommen im jeweiligen Text realisiert werden (siehe dazu Abbildung 4.3 mit dem ersten Prototypen).

Es sollte ferner möglich sein, zu einer Darstellung der originalen Handschriften zu wechseln. Diese wurde zunächst im Konzept und dem parallel entwickelten Datenformat (siehe 5.1) zwar noch berücksichtigt, in der letztlichen Implementierung des Prototypen jedoch aus Gründen des Umfangs fallengelassen. Die Konzeptualisierung dieser Ansicht und deren Verknüpfung ist somit zukünftiger Arbeit (siehe Kapitel 8.2) vorbehalten.

Die Anwahl der einzelnen Kategorien sollte ebenfalls in der Hauptansicht möglich sein. Die Kodierung der einzelnen Annotationskategorien erfolgte mittels Farben. Die hierfür gewählte Farbpalette musste Farben beinhalten, welche zum einen gut voneinander zu unterscheiden waren und auf der anderen Seite dennoch eine gewisse Zugehörigkeit zu Kategoriengruppen vermitteln. Hier konnte auf die Arbeit von Harrower und Brewer [4] zurückgegriffen werden, welche mit dem „ColorBrewer“ ein Werkzeug für die Auswahl von Farbpaletten nach gewissen photometrischen Eigenschaften bereitstellten. So können hier Farbverläufe, welche sich für die Visualisierung kontinuierlicher Daten eignen, oder aber Farbpaletten zur Visualisierung diskreter Daten angewählt werden, welche dann auch noch im Bezug auf ihre Erkennbarkeit in schwarzweißer Photokopie, auf LCD-Bildschirmen, oder gar bei rot-grün-Blindheit gefiltert werden können. Für die vorliegende Arbeit wurde zunächst auf eine diskrete Farbpalette zurückgegriffen, welche jedoch im späteren Verlauf ihre Schwächen offenbarte. Zum einen waren die gewählten Farben zu kräftig, als dass sie für eine Markierung des schwarzen Textes eingesetzt hätten werden können, und auch die Diskretisierung der Kategorien durch die zugehörigen Farben stellte ein Problem dar. Letztlich wurde die Farbpalette entsprechend angepasst: Ein roter Farbverlauf wurde für sämtliche Zusätze gewählt (zusätzliche Adjektive, zusätzliche Episoden, Zusätze innerhalb des Satzes, zusätzliche Personen, zusätzliche Ortsnamen). Semantische Kategorien (Synonyme, veränderte Semantik, gleiche Semantik) wurden durch einen blauen Verlauf kodiert, die verbliebenen Kategorien erhielten dann noch jeweils eigene Farben, welche gut von den bisher vergebenen Verläufen unterscheidbar waren (andere Namensform – grün, syntaktische Umstellung – braun, Genera Verbi – türkis, Oratio – violet).

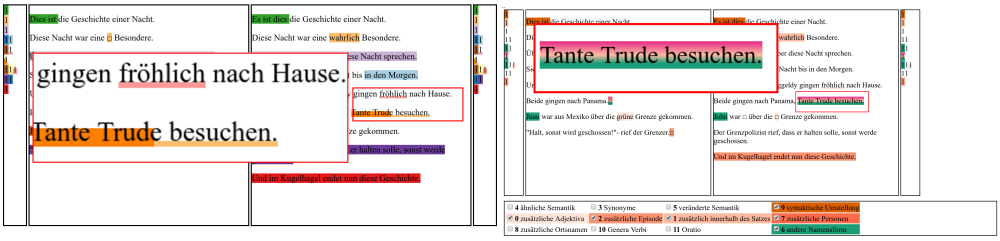


(a) vertikale Doppelannotation (b) horizontale Doppelannotation

Abbildung 4.1.: Doppelannotationen

Überlappende Kategorien wurden zunächst durch eine vertikale (siehe Abbildung4.1a) oder horizontale (siehe Abbildung 4.1b) Überlappung der zugehörigen Farben dargestellt. Der vertikale Verlauf hat dabei den Vorteil, dass es zu keiner Verwechslung zwi-

schen Verlaufsgrenzen und Annotationsgrenzen kommen kann. Im horizontalen Verlauf kann es vorkommen, dass die Grenze zwischen den beiden Farben auf eine Wortgrenze fällt und somit der Eindruck entsteht, dass eine neue Annotation einer anderen Kategorie an dieser Stelle beginnt. Allerdings lassen sich im horizontalen Verlauf gegebenenfalls Unsicherheiten besser kodieren (siehe Abbildung 4.2a), sodass diese Variante letztlich favorisiert wurde. Ab drei unterschiedlichen Kategorien wird dieses Konzept jedoch zu unübersichtlich (siehe Abbildung 4.2b), sodass in der Implementierung hier andere Lösungen gefunden werden mussten. Solche Stellen werden dort dann durch eine rote Umrandung markiert, um somit den Benutzer zu einer eingehenderen Untersuchung der Stellen aufzufordern.



(a) Annotation von Unsicherheit (b) Drei Kategorien

Abbildung 4.2.: Annotation mehrer Kategorien

Die Filterung der Vorkommen nach Kategorien war eine ebenfalls sehr wichtige Anforderung. Zunächst sollte die Legende und Kategorienauswahl in einem Menü unterhalb beider Texte erfolgen (siehe Abbildung 4.2b). Sie wurde dann jedoch über beide Texte verlagert. Dies eröffnete nun die Möglichkeit, den nun unterhalb der Texte frei gewordenen Raum für weitere Informationen, wie eine Liste aller Annotationen, zu nutzen.

Basierend auf den bislang entwickelten Ideen und Mockups wurde dann ein Prototyp implementiert.

In Abbildung 4.3 ist die erste funktionierende Version dargestellt. Hier sind bereits alle wesentlichen Elemente enthalten: Zwei Textspalten zur direkten Gegenüberstellung, sowie die Mittelspalte zur Ansicht der vollständigen Texte.

In der Mittelspalte wurden die Texte jeweils zeilenweise aufgeschlüsselt und gegenübergestellt. Die entsprechenden Phänomene werden dabei durch Verbindungen der einzelnen Zeilen kodiert. Die Länge der der jeweiligen Zeile zugehörigen Linie ist dabei logarithmisch von der Tokenanzahl abhängig. Dabei werden die Token, die zu einem Phänomen gehören, dessen Kategorie angewählt ist, in der entsprechenden Farbe dargestellt.

In dieser Version sind noch die ursprünglich für zukünftige Erweiterungen mit den Originalen Handschriften und quantitativer Auswertung der Texte vorgesehenen Schaltflächen zu sehen. Diese wurden in späteren Versionen entfernt, da diese Schaltflächen zum einen am oberen Bildschirmrand zu prominent waren und zum anderen insbesondere die Implementierung des Vergleichs mit den Originalen Handschriften den

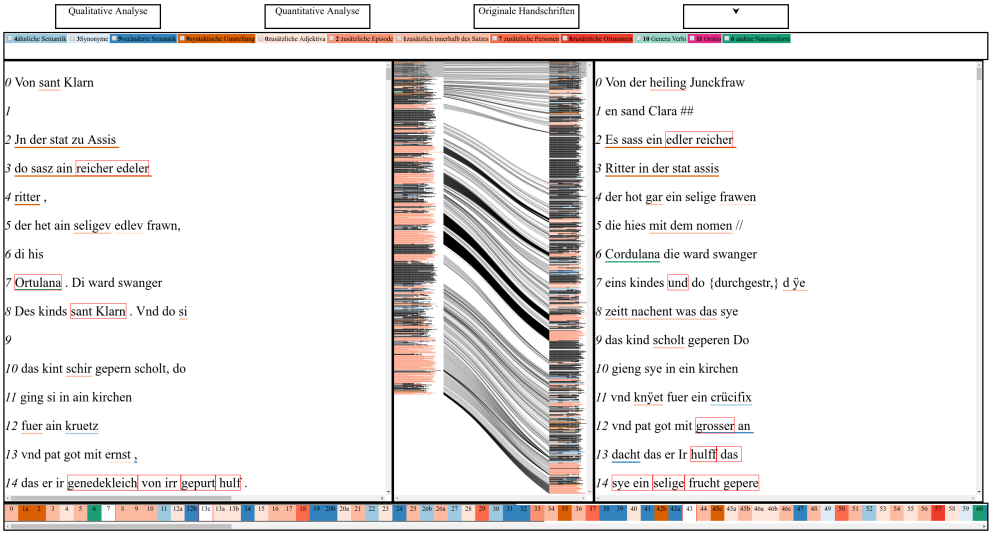


Abbildung 4.3.: Version 0.1

zeitlichen Rahmen der Arbeit gesprengt hätte.

In Abbildung 4.4 ist nun Version 0.2 zu sehen. Die Schaltflächen am oberen Bildschirmrand wurden entfernt und in einen Menübutton am rechten oberen Rand integriert. Ebenso wurde die Legende der Vorkommensarten massiv reduziert, da die lange Beschreibung diesen Bereich unnötig vergrößert und damit unnötig in den Vordergrund gerückt hätte. Die ausführliche Beschreibung wurde in tooltips und das Menü oben rechts verschoben. Der Wegfall der Legende ermöglichte es nun, die Anzahl der Vorkommen je Kategorie in der Länge des entsprechenden Balkens zu kodieren. Die Kodierung erfolgt hierbei linear in Abhängigkeit von der absoluten Anzahl. Die Kategorieanwahl fungiert somit auf den ersten Blick als zusammengefügtes Säulendiagramm zur Gesamtübersicht über die Kategorieverteilung, was bereits als ein Ansatz des *Distant-Readings* betrachtet werden kann. In das Menü wurde ferner auch die Möglichkeit zur Zuschaltung einer quantitativen Analyse integriert. Zur Visualisierung der jeweiligen Anzahl der Phänomene je Kategorie und der Anzahl an Phänomenen mit Kategorieüberlappungen wurde auf Säulendiagramme zurückgegriffen, da diese die gegebenen diskreten Daten am besten wiedergeben können und die Verhältnisse der Kategorieanzahlen untereinander nicht von großer Bedeutung für die Untersuchung des Legendars sind.

Der untere Balken mit der Liste aller Phänomene lenkte jedoch zu sehr vom eigentlichen Text ab, sodass er optional ausgeblendet werden konnnte. Auch die Mittelspalte wurde in den darauffolgenden Versionen verschlankt: Es werden nur noch diejenigen Phänomene angezeigt, deren Kategorie angewählt wurde. Zugleich wurde die Intensität der Mittelspalte nochmals verringert, um sie nicht mehr derart prominent in den Vordergrund zu rücken. Somit ist nun eine Fokussierung auf die beiden Texte deutlich

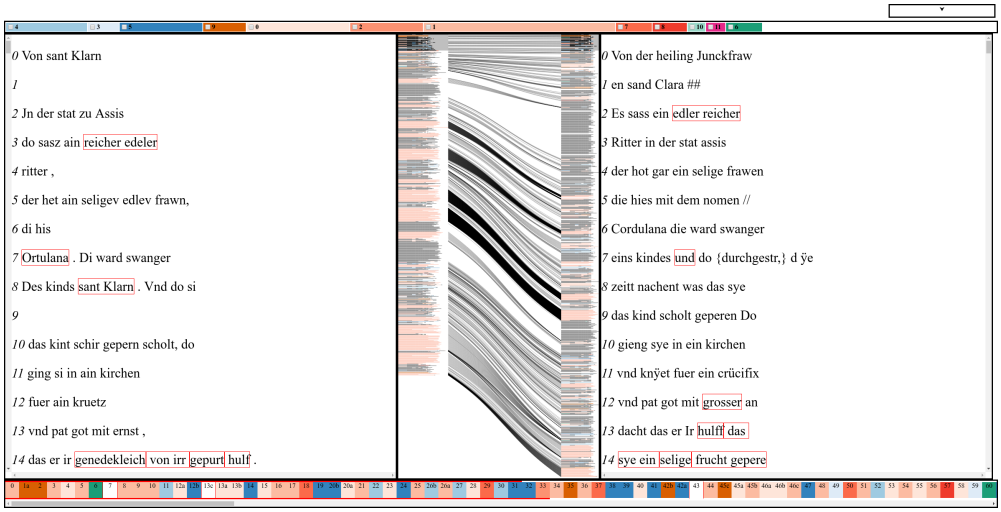


Abbildung 4.4.: Version 0.2

einfacher.

In den Texten wiederum sollten die einzelnen Phänomene markiert werden. Dabei wurde rasch auf eine Unterstreichung in der zur jeweiligen Kategorie gehörigen Farbe gesetzt; eine Hervorhebung durch farbigen Hintergrund war schon aufgrund der gewählten Farben nicht machbar, da dadurch die Lesbarkeit des Textes zu sehr eingeschränkt worden wäre. Auch wurde es so möglich, die Stellen des jeweils gerade betrachteten Phänomens (sei es durch Klicken markiert oder durch Schweben mit dem Mauszeiger gerade angewählt) durch einen grauen Hintergrund hervorzuheben. Falls ein Phänomen zu mehreren Kategorien gehören sollte, so wird dies durch eine entsprechend zweifarbige Unterstreichung angezeigt. Die Unsicherheitskodierung erfolgt hier über die Dicke der Unterstreichung. Sichere Kategorien werden entsprechend dicker unterstrichen als unsichere. Falls ein Token zu mehreren Phänomenen gehört, sodass eine Zuordnung der Darstellung nicht eindeutig erfolgen kann, wird diese Stelle durch einen roten Rahmen markiert, um so dem Benutzer anzuzeigen, dass an dieser Stelle noch detaillierte Untersuchungen notwendig sind.

Für eine detailliertere Untersuchung der einzelnen Kategorien wurde im Verlauf der Implementierung auf Wunsch der beteiligten Literaturwissenschaftler die Kategorieansicht entwickelt, in der nur die Textzeilen mit Phänomenen der jeweils angewählten Kategorie zu sehen sind. Zusammengehörende Zeilen, die zu einem Phänomen gehören, werden dabei durch graue Balken am Rand markiert. Die Auswahl des entsprechenden Phänomens wird durch Rotfärbung des zugehörigen Balkens und des zugehörigen Verbindungsbogens in der Mitte, sowie durch graue Hervorhebung der betrachteten Stelle angezeigt. Ein Beispiel für diese Ansicht findet sich in Abbildung 5.2.

5. Implementierung

Die Implementierung des Werkzeuges erfolgte webbasiert in HTML 5 und JavaScript. Für die Darstellung der beiden Texte in der Mittelspalte wurde auf d3.js [16] zur Erstellung von SVGs zurückgegriffen.

Um die Text- und Annotationsdaten verarbeiten zu können, wurde ein eigenes XML-basiertes Datenformat entwickelt. Das in der DH-Community weitestgehend als Standard akzeptierte TEI-Format wurde dabei bewusst nicht verwendet, da eine Einarbeitung in das mittlerweile recht komplexe System für den gegebenen Zweck als zu aufwändig angesehen wurde, insbesondere, da in TEI-konformen Texten auch zahlreiche Merkmale annotiert worden wären, welche für die vorliegenden Analysen nicht von Bedeutung sind. Eine Umwandlung des Arbeitsformates in TEI ist allerdings sehr wohl möglich und für den Produktivbetrieb durchaus anzustreben.

Im Folgenden erfolgt nun zunächst ein Überblick über das Arbeitsformat:

5.1. Datenformat

Zunächst wird jeder Text in Tokens bzw. Wörter aufgeteilt. Diese Tokens erhalten als Attribut eine ID, welche innerhalb des jeweiligen Textes einzigartig ist. Zur späteren Darstellung im Originalbild wird der Text zudem noch mit Seiten, Spalten und Zeilen ausgezeichnet. Hierbei sind die Seiten-IDs innerhalb des Textes einzigartig, ebenso die Zeilen-IDs. Die Spalten-IDs wiederum sind nur innerhalb der jeweiligen Seite einzigartig.

Zur Auszeichnung werden zunächst Kategorien mittels des `<category>` Tags angelegt. Auch diese erhalten als Attribut eine ID, welche innerhalb des Korpus einzigartig ist. Ferner enthalten sie eine Beschreibung der Kategorie.

Die einzelnen Relationen zwischen den Texten (Phänomene) werden mittels `<relation>` ausgezeichnet. Die hier vergebene ID ist innerhalb des Korpus einzigartig. Es wird eine Liste der Kategorie-IDs angegeben, zu denen dieses Phänomen gehört, sowie eine Liste an Gewichtungen pro Kategorie angegeben. Ferner werden ebenfalls zwei Listen mit den zugehörigen Token-IDs aus beiden Texten angegeben. Eine Übersicht über alle verwendeten Tags findet sich in Tabelle 5.1.

5.1.1. Phänomenannotation

Für die Annotation der Phänomene werden immer die betroffenen Tokens in beiden Texten gelistet. Im Falle hinzugefügter Teile wird im Urtext immer auch das Token vor der ausgelassenen Stelle aufgelistet, dieses Token muss im hinzugefügten Text ebenfalls

Tags	Bedeutung	Scope der ID
`<body>`	Hauptelement, enthält das Korpus	–
`<text id=n>`	Texte	Korpusweit
`<page id=n>`	Seite	Textweit
`<column id=n>`	Spalte	Seitenweit
`<line id=n>`	Zeile	Textweit
`<t id=n>`	Token, Wort	Textweit
`<category id=n>`	Auszeichnung der Kategorien	Korpusweit
`<description>`	Beschreibung der Kategorie	–
`<relation id=n>`	Unterschied zwischen beiden Texten	Korpusweit
`<type>`	Liste zugehöriger Kategorien	–
`<weight>`	Liste zugehöriger Gewichtungen $\in \{1, 2, 3\}$	–
`<tokens_text0>`	Liste der Token-IDs	–
`<tokens_text1>`	Liste der Token-IDs	–
`<comments_text0>`	Kommentare	–
`<comments_text1>`	Kommentare	–

Tabelle 5.1.: mögliche Tags

gelistet sein. Unter `<weight>` kann ein Wert zur Gewichtung der Kategorie im jeweiligen Phänomen (zwischen 1 und 3) angegeben werden. Damit ist beispielsweise eine Annotation der Unsicherheit, also eines Maßes dafür, mit welcher Wahrscheinlichkeit das Phänomen in der entsprechenden Kategorie liegt, möglich. Es ist ebenfalls möglich, Kommentare zum Phänomen in den Tags `<comments_text0>` und `<comments_text1>` zu hinterlegen.

Beispiele für die Verwendung dieses Standards finden sich im Anhang an diese Arbeit. Textbeispiele zu den gegebenen Kategorien wurden bereits in Kapitel 4 gegeben (siehe 4.1), für die entsprechende Standarddefinition siehe A.1. Für eine Annotation des Textbeispiels im entsprechenden XML siehe A.2.

In der Praxis zeigte sich rasch, dass den mit der Annotation der Texte betrauten Literaturwissenschaftlern der Umgang mit reinem XML-Code nicht zuzumuten war. Somit wurde als Zwischenstufe zunächst noch ein Parser entwickelt, welcher die bereitgestellten docx-Dateien, in denen die Phänomene durch Kommentare hinterlegt waren, in entsprechendes XML umwandelte. Dieser ist in Python 3 geschrieben.

5.2. Prototyp

Im Folgenden erfolgt nun eine detaillierte Beschreibung des entwickelten Prototypen:

In Abbildung 5.1 ist der Prototyp nach Laden der Texte zu sehen. Es sind ferner sowohl das Menü am oberen rechten Rand als auch die Phänomenleiste am unteren Bildschirmrand ausgeklappt.

Hier nun die Elemente der Ansicht:

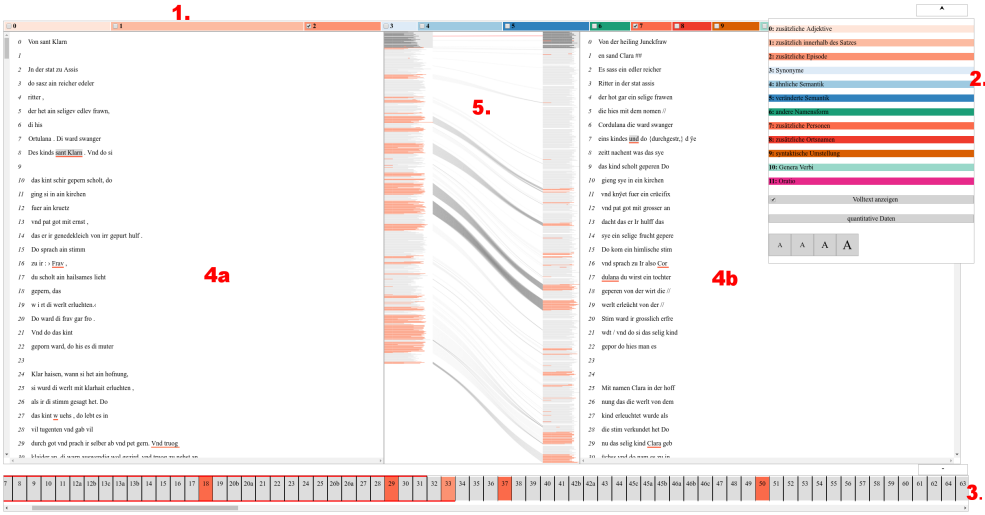


Abbildung 5.1.: Prototyp

5.2.1. Kategorieanwahl (1.)

Am oberen Bildschirmrand findet sich die Anwahl der auszuwählenden Kategorien. Der vollständige Titel der Kategorie ist mittels tooltips zuschaltbar. Die Länge des jeweiligen Balkens ist dabei proportional zur Häufigkeit der Kategorie. Eine Auswahl der entsprechenden Kategorie zeigt alle Phänomene der Kategorie in den diversen Ansichten an.

5.2.2. Menü (2.)

Im Menü, welches über eine Schaltfläche am rechten oberen Bildschirmrand zuschaltbar ist, findet sich zum einen eine Legende aller Kategorien, und sodann diverse Schaltflächen: Mit der Schaltfläche „Volltext anzeigen“ kann zwischen der Anzeige der Texte als Volltext oder nur der Anzeige einzig der Textstellen der ausgewählten Kategorien umgeschaltet werden. Die Schaltfläche „quantitative Daten“ schaltet die Ansicht der quantitativen Daten des Textes (Anzahl der Phänomene nach Kategorie, Anzahl der Phänomene nach zugeordneter Kategorieanzahl) zu. Eine Ansicht des quantitativen Fensters findet sich in Abbildung 5.4.

Über weitere Schaltflächen lässt sich die Schriftgröße der Textanzeige (4.) anpassen.

5.2.3. Phänomenleiste (3.)

Am unteren Bildschirmrand lässt sich eine Phänomenleiste mit einer Übersicht aller Phänomene zuschalten. Dies ist somit einer der Orte, an denen der *Distant-Reading*-Ansatz verwirklicht wird. Die in der Kategorieanwahl (1.) angewählten Kategorien

werden jeweils durch die zugehörige Farbe hervorgehoben. Gehört das Phänomen zu zwei angewählten Kategorien, so wird ein horizontaler Farbverlauf angezeigt. Durch zwei rote Balken am oberen und unteren Rand der Leiste werden zudem die Phänomene angezeigt, welche sich gerade in der Textansicht befinden. Ein Klick auf das Phänomen fokussiert die Textansicht automatisch auf das Phänomen und markiert dieses ebenfalls in der Mittelspalte durch rote Hervorhebung der entsprechenden Verbindungslinie. Das ausgewählte Phänomen wird grau hinterlegt. Ein Rechtsklick öffnet die Detailansicht, die in Abbildung 5.3 zu sehen ist.

5.2.4. Textansicht (4.)

Volltextansicht

In der Textansicht werden beide Texte Zeile für Zeile gegenüber gestellt. Somit ist in dieser Ansicht der *Close-Reading*-Ansatz implementiert.

Die Phänomene der angewählten Kategorie werden durch Unterstreichung der entsprechenden Stellen markiert. Die Dicke der Unterstreichung gibt dabei die Sicherheit der Kategoriezuordnung an. Falls dem Phänomen zwei Kategorien zugeordnet werden, so wird dies durch einen horizontalen Verlauf in der Unterstreichung dargestellt. Bei unterschiedlicher Gewichtung der Kategorien wird der Verlauf entsprechend zur höher gewichteten Farbe verschoben. Dies stellt eine Abweichung zur Konzeptualisierung (siehe Kapitel 4) dar, ist jedoch technischen Gründen in der Implementierung geschuldet.

Textstellen, an denen sich mehrere Phänomene überschneiden, werden durch rote Umrandung markiert, um den Benutzer zu einer eingehenden Untersuchung aufzufordern. Ein Klick öffnet ein Dropdown-Menü, in dem das zu untersuchende Phänomen angewählt werden kann. Rechtsklick öffnet wiederum die Detailansicht.

Wird ein Phänomen ausgewählt, so wird es grau hinterlegt und in den anderen Ansichten entsprechend markiert. Ein Rechtsklick auf das entsprechende Phänomen öffnet die Detailansicht.

Kategorieansicht

Im Menü kann die Volltextanzeige abgeschalten werden. Wird diese abgeschaltet, so werden nur noch Zeilen angezeigt, die Phänomenen zugehörig sind, die angewählt wurden. In Abbildung 5.2 ist diese Ansicht zu sehen: Die Texte sind zu beiden Seiten reduziert, es werden auch in der Mittelspalte nur noch die Linien hervorgehoben, die auch tatsächlich sichtbar sind. Zusammengehörige Zeilen werden dabei durch graue Markierungen am Rand angegeben. Ein Zeigen auf ein entsprechendes Phänomen hebt dieses zum Einen durch die entsprechende rote Linie in der Mittelspalte und zum Anderen durch Rotfärbung der entsprechenden Markierung hervor. Gleiches geschieht entsprechend bei Auswahl des Phänomens durch normalen Linksklick.

Abbildung 5.2.: Hauptfenster in Kategorieansicht

5.2.5. Mittelspalte (5.)

Die Mittelspalte ist der primäre Ort, an dem der *Distant-Reading*-Ansatz verwirklicht wird. Beide Texte werden dabei zeilenweise in ihrer Gesamtheit dargestellt. Die Länge der zur jeweiligen Zeile gehörenden Linie ist dabei logarithmisch von der Tokenanzahl je Zeile abhängig. Wird eine Kategorie angewählt, so werden zum einen alle Tokens, die einem Phänomen einer entsprechenden Kategorie angehören, farblich hervorgehoben, und zudem die der entsprechenden Zeile zugehörigen Verbindungsbögen eingeblendet. Somit wird sofort ersichtlich, welche Textstellen jeweils einander zugeordnet werden. Wird ein Phänomen in der Textansicht oder der Phänomenleiste angewählt, so wird der entsprechende Bogen durch Rotfärbung deutlich hervorgehoben. In der Mittelspalte sind zudem zwei Scrollbalken implementiert, die anzeigen, welcher Teil des Textes gerade in der Textansicht (4.) zu sehen ist. Durch Klick auf die jeweilige Textrepräsentation wird die Ansicht auf die entsprechende Stelle fokussiert.

5.2.6. Detailansicht

Durch Rechtsklick auf ein Phänomen oder ein Phänomencluster öffnet sich die in Abbildung 5.3 zu sehende Detailansicht. Sie besteht aus folgenden Elementen: Mit der Phänomenanwahl (1.) ist es möglich, falls an der ausgewählten Textstelle mehrere Phänomene parallel auftreten, zwischen den verschiedenen Phänomenen auszuwählen. In der Textstellenansicht (2.) werden die jeweiligen Textzeilen direkt gegenübergestellt. Die zum Phänomen gehörenden Token werden grau hinterlegt. Außerdem werden die zugehörigen Kommentare angezeigt. Sollte kein Kommentar hinterlegt sein, wird der Text „none“ angezeigt. In der Kategorietabelle (3.) werden die Kategorien, zu denen das angewählte Phänomen gehört, gelistet. Dabei wird auch die vollständige Beschreibung und die Sicherheit, also ein Maß, wie sicher das Phänomen in der entsprechenden Kategorie ist, angezeigt.

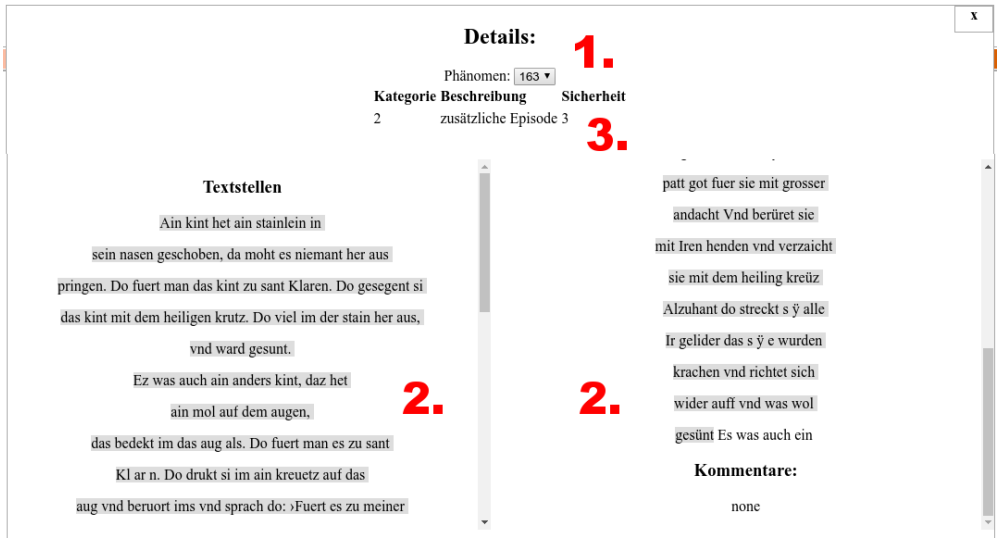


Abbildung 5.3.: Detailansicht

5.2.7. Quantitative Analyse

Bei einem Klick im Menü auf „Quantitative Analyse“ öffnet sich ein Fenster, welches Daten über die Anzahl der Phänomene im Text visualisiert. Dabei wird zum Einen ein Säulendiagramm mit der Anzahl der Phänomene je Kategorie angezeigt, und zum Anderen ein Diagramm mit der Anzahl der Kategorien je Phänomen, also wie vielen Phänomenen mehr als eine Kategorie zugeordnet wurde. Eine Ansicht dieses Fensters findet sich in Abbildung 5.4.

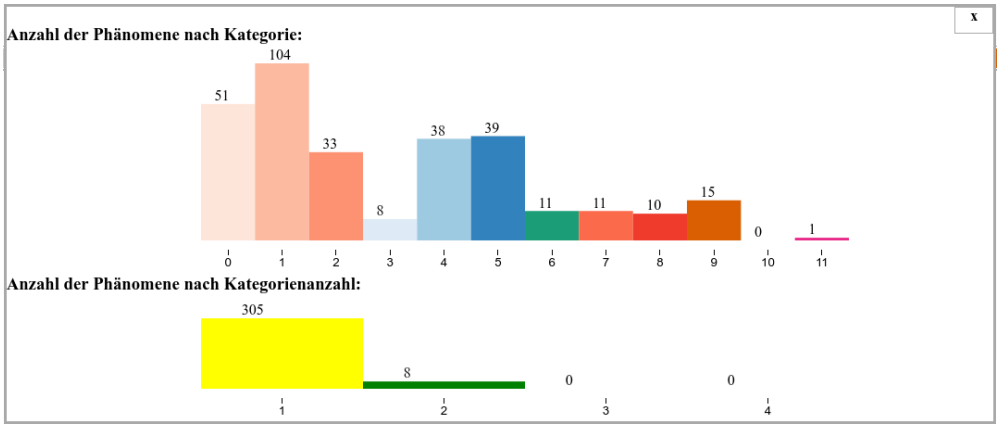


Abbildung 5.4.: Quantitative Ansicht

6. Anwendungsbeispiele

Im Folgenden wird der Nutzen des Tools anhand einiger Anwendungsbeispiele demonstriert.

6.1. Anwendungsbeispiel 1

Ein erstes solches Beispiel dient zunächst dazu, die Einbindung der einzelnen Views in einen klassischen Workflow zu beschreiben. Es beschäftigt sich mit der Fragestellung:

> Welche Synonyme werden in beiden Texten verwendet? Lassen sich dabei Tendenzen erkennen?

Um diese Fragestellung mit den klassischen Methode beantworten zu können, musste ein Forscher bislang beide Texte händisch durchsuchen, mögliche Synonyme auflisten und diese Listen sodann miteinander vergleichen.

Um diese Fragestellung unter Einbeziehung des entwickelten Werkzeuges zu beantworten, wird zunächst die entsprechende, zuvor generierte XML-Datei in das Werkzeug eingelesen. Dies dauert etwa fünf bis zehn Sekunden. Danach wird nach der entsprechenden Kategorie gesucht. Ein Blick auf die Legende im Menü offenbart, dass es sich um Kategorie 3 handelt. Von dieser existieren nur sehr wenige Vorkommen. Ein Blick auf die quantitative Analyse ergibt, dass es genau acht Phänomene dieser Kategorie gibt (siehe dazu auch Abbildung 5.4). Diese gilt es nun, in den Texten zu finden. An die entsprechenden Stellen zu scrollen, ist zwar möglich, ist allerdings bei der geringen Phänomendichte (8 von ca. 300) verhältnismäßig mühsam.

Also schaltet der Anwender zunächst die Phänomenleiste am unteren Bildschirmrand zu, um diese Phänomene besser finden zu können. Sie sind in der Kategoriefarbe markiert und unterscheiden sich somit deutlich vom grauen Hintergrund der nicht markierten Phänomene. Nun kann entweder durch einfachen Klick zum Phänomen gesprungen werden, oder mittels Rechtsklick die Detailansicht zugeschaltet werden und somit das entsprechende Synonym betrachtet werden. Dies ermöglicht aber noch immer keine Übersicht über alle Vorkommen der Kategorie, welche für eine Untersuchung der Tendenzen zwingend erforderlich wäre. Man müsste hier ebenfalls eine händische Liste führen. Daher schaltet der Anwender nun im Menü die Volltextansicht aus und kann somit alle acht Vorkommen gesammelt betrachten. Dies ist in Abbildung 6.1 zu sehen.

Hier können nun auf einen Blick alle Stellen betrachtet werden und somit die Aussage getroffen werden, dass eindeutig keine Tendenzen in der Verwendung von Synonymen zu erkennen sind. Die einzige Wortkombination, die mehrfach auftaucht, ist die Ersetzung des „bösen Geistes“ durch „Feind“ beziehungsweise „Teufel“, allerdings ist

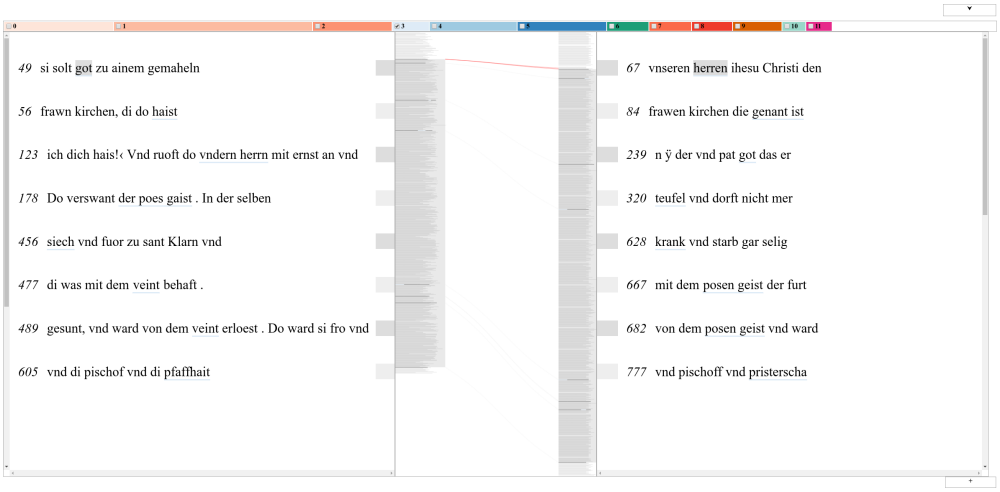


Abbildung 6.1.: Ergebnisse zu Anwendungsbeispiel 1

hier die Ersetzung nicht konsistent. Es kann keine Aussage getroffen werden, dass der „böse Geist“ immer ersetzt wird, da der Begriff sowohl in der HL-Red I als auch in der HL-Red II auftaucht.

6.2. Anwendungsbeispiel 2

Ein zweites Beispiel wird sich nun mehr auf die Verknüpfung der einzelnen Vorkommen und damit auf die Kernstärke des entwickelten Werkzeuges konzentrieren:

Im Zentrum dieses Anwendungsszenarios steht die Untersuchung der Legende im Hinblick auf syntaktische Umstellung des Textes:

> Welche Stellen wurden umgestellt und welchen Umfang haben die Umstellungen?

Um diese Fragestellung händisch zu beantworten, müsste ein Forscher sämtliche Stellen in beiden Handschriften suchen, diese von Hand herausschreiben und könnte erst dann mit der Analyse beginnen. Eine solche Analyse würde sich allerdings notwendigerweise zunächst auf eine einzelne Kategorie beschränken, die Hinzunahme weiterer Kategorien wäre überaus aufwändig, da hier der gesamte Text erneut durchgegangen werden müsste, erneut betroffene Stellen herausgeschrieben werden und auf eventuelle Überlappungen hin untersucht werden müssten. Die Verwendung des entwickelten Werkzeuges erleichtert dies deutlich:

Zunächst betrachtet der Anwender dazu alle Vorkommen der Kategorie 9. Eine Auswahl in der Legende hebt alle entsprechenden Phänomene der Kategorie hervor. Ein Blick auf die Distant-Reading-View in der Mittelspalte offenbart, dass die Vorkommen dieser Kategorie über den gesamten Text gestreut sind, mit leichten Häufungen am

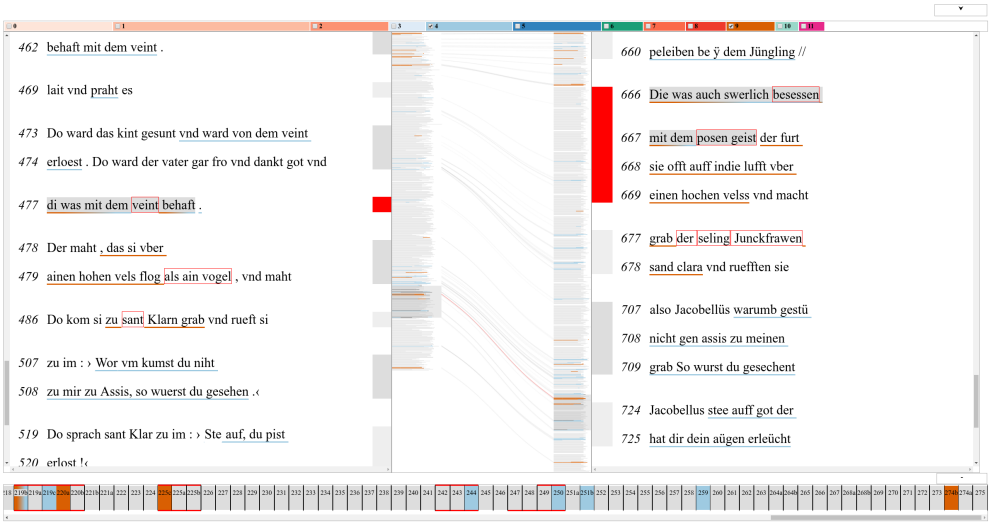


Abbildung 6.2.: Ergebnisse zu Anwendungsbeispiel 2

Anfang und gegen Ende. Die quantitative Analyse offenbart, dass es sich um 15 Stück handelt, diese können also einfach der Reihe nach durchgegangen werden. Um sich die Phänomene der Reihe nach anzuschauen, bietet es sich an, die Phänomenleiste am unteren Rand aufzuklappen, um auf dieser nun mittels Rechtsklick auf die markierten Phänomene umgehend die Detailansicht zu öffnen. Geht man so die Phänomene der Reihe nach durch, so stößt man bereits beim dritten Phänomen auf eine Kategoriendoppelung: Das hier annotierte Phänomen gehört sowohl zur ausgewählten Kategorie 9 als auch zur Kategorie 5, der veränderten Semantik.

Ausgehend von diesem Fund stellt sich nun die Frage nach weiteren Kategoriedopplungen mit dieser Kategorie. Die Quantitative Analyse offenbart acht Kategoriedopplungen im gesamten Dokument. Geht man nun die 15 Vorkommen der Kategorie 9 durch, so stößt man tatsächlich auf zwei weitere Dopplungen: Zwei weitere Phänomene sind ebenfalls der Kategorie 4 (ähnliche Semantik) zugeordnet. Es lohnt sich also nun unter Umständen, auch die Phänomene dieser Kategorie zuzuschalten, um gegebenenfalls so auf Textstellen zu stoßen, deren nähere Untersuchung von Interesse ist.

Die Zuschaltung dieser Kategorie sorgt zunächst dafür, dass noch mehr Phänomene hervorgehoben werden, was die Übersichtlichkeit verschlechtert. Daher konzentriert sich der Anwender nun auf die Phänomenleiste und entdeckt dort im Bereich des zweiten Phänomens mit sich überschneidenden Kategorien einen Phänomencluster aus mehreren Umstellungen und semantisch ähnlichen Stellen. Um dieses Cluster näher zu betrachten, schaltet er zunächst den Volltextmodus aus und betrachtet dann die markierten Stellen. Dabei fällt auf, dass in der Tat an dieser Stelle der Text massiv umgestellt wurde (siehe Abbildung 6.2). Um den Umfang dieser Umstellung besser bewerten zu können, kann der Anwender nun die Kategorie 4 wieder abwählen, um nur

die Stellen der Kategorie 9 angezeigt zu bekommen, und diese in der Kategorieansicht untereinander zu vergleichen.

7. Ergebnisse: Expertenfeedback

Im folgenden Kapitel werden nun die Ergebnisse der Arbeit anhand der von Experten aus der Literaturwissenschaft gegebenen Rückmeldung vorgestellt:

Durch die Kooperation des Instituts für Visualisierung und Interaktive Systeme (VIS) mit der Abteilung Germanistische Mediävistik des Instituts für Literaturwissenschaft der Universität Stuttgart bestand die Möglichkeit, die in dieser Arbeit entwickelten Ansätze direkt durch Experten auf dem Gebiet der Literaturwissenschaft evaluieren zu lassen. Dazu wurden die Experten aus der Literaturwisschenschaft bereits zu Beginn in die Formulierung der Anforderungen mit eingebunden, und sodann nach Abschluss der Implementierung um Evaluation des entwickelten Prototypen gebeten. Im Folgenden werden die Ergebnisse dieser Rückmeldung dargestellt:

Die Rückmeldung bei der Vorstellung des Prototypen fiel durchweg positiv aus. Das Werkzeug wurde als wertvolle Unterstützung bei der Analyse der annnotierten Legenden bewertet, insbesondere seine Usability wurde hervorgehoben. Die Möglichkeit, Vorkommen nach Kategorien gefiltert anzeigen zu lassen, wurde ebenfalls als besonders nützlich gesehen. Die Verlinkung der einzelnen Ansichten untereinander, insbesondere die Markierungen der Textpositionen sowohl in der Mittelspalte als auch in der Phänomenansicht, wurde als sehr informativ und hilfreich bewertet..

Das Werkzeug werde sich ferner auch im Forschungsalltag bewähren, der Kooperationspartner plant, das Werkzeug in den Grundlagenvorlesungen des Masterstudiengangs „Digital Humanities“ einzusetzen, und auf einem Workshop der Abteilung „Digital Humanities“ vorzustellen. Das Werkzeug wurde auch als Grundlage für weitere Entwicklungen angesehen, insbesondere für die Einbindung eines Editors zur Annotation der Texte und für einen weitergehenden Export der gewonnenen Ergebnisse. Weitere Möglichkeiten zukünftiger Arbeit finden sich im zugehörigen Kapitel 8.2.

8. Zusammenfassung und Ausblick

Im folgenden Kapitel wird eine kurze Zusammenfassung der Arbeit gegeben, und im anschließenden Ausblick Erweiterungsmöglichkeiten des Prototypen und der Ansätze diskutiert..

8.1. Zusammenfassung

Im Rahmen dieser Arbeit wurde ein visueller interaktiver Analyseansatz für die Exploration der Veränderung zweier Texte entwickelt. Dabei wurden unterschiedliche Visualisierungen und Interaktionsmöglichkeiten zur Verfügung gestellt, um Experten bei der Analyse der Textveränderungen zu unterstützen. Besonderes Augenmerk wurde dabei auf die Implementierung von *Close-* und *Distant-Reading*-Ansätzen gelegt.

Zunächst wurden Grundlagen sowohl der Informationsvisualisierung im Allgemeinen als auch der Textvisualisierung in den Digital Humanities im Besonderen dargelegt. Ferner wurde einige verwandte Arbeiten detailliert betrachtet und auf ihre Verwendbarkeit für den vorliegenden Fall analysiert. Auf Basis dieser Grundlagen und Arbeiten wurde sodann ein Konzept zur Entwicklung des Ansatzes erstellt und beschrieben. Dieser Ansatz enthält verschiedene Ansichten und Filtermöglichkeiten. Er vereint *Close-Reading* und *Distant-Reading* Möglichkeiten schon in der Hauptansicht und bietet die Möglichkeit der Selektion nach einzelnen Kategorien. Die Verknüpfung der beiden Texte sowohl auf der Ebene des *Distant-Readings*, als auch auf der Textebene des *Close-Readings* ermöglicht dabei, neue Erkenntnisse aus dem vorliegenden Textkorpus zu gewinnen. Es ist außerdem möglich, nur die der jeweiligen Kategorie zugeordneten Phänomene anzuzeigen. Ferner wird auch eine ausschließlich auf die quantitative Analyse des Textes im Hinblick auf die zu untersuchenden Kategorien ausgerichtete Ansicht bereitgestellt. Dieses Konzept wurde anhand einer prototypischen Implementierung webbasiert umgesetzt und ausführlich erläutert. Dabei wurden detailliert die verschiedenen Ansichten, Visualisierungen und Filtermöglichkeiten beschrieben und begründet. Anhand einiger Anwendungsbeispiele wurde der Nutzen des Ansatzes exemplarisch mit Hilfe typischer Problemstellungen aus dem Bereich der Literaturwissenschaft überprüft. Die abschließende Evaluation durch Kooperationspartner vom Institut für Literaturwissenschaft bestätigte die Eignung für den gegebenen Einsatzbereich, eine hervorragende Usability insbesondere im Hinblick auf die Analyse des Legendars und die Eignung zur Einbindung in weiterführende Analysesysteme.

8.2. Ausblick

Die in dieser Arbeit gewonnenen Erkenntnisse und der entwickelte Prototyp bieten Grundlage für weitere Arbeit. Bereits im Konzept Kapitel 4 wurden einige Ansätze angesprochen, die aus Zeitgründen in dieser Arbeit nicht implementiert werden konnten: Ein erster Punkt ist die zusätzliche Verknüpfung des transkribierten Textes mit Scans der originalen Handschriften. Hier wurde in der Entwicklung des Datenformates bereits Vorarbeit geleistet: Der Text wird bereits in Spalten und Zeilen aufgeschlüsselt, was eine spätere Abbildung auf originale Handschriften deutlich erleichtert. Auf diese Art ist es mit verhältnismäßig wenig zusätzlichem Aufwand möglich, auch die Originale miteinander zu vergleichen.

Eine weitere Möglichkeit zur Erweiterung wäre, das Werkzeug auf die Benutzung mit mehreren Texten auszudehnen. Bislang sind hier nur zwei Texte möglich. Die Erweiterung auf beispielsweise drei Texte würde insbesondere für die Analyse des Korpus „Der Heiligen Leben“, welcher ja an sich drei Texte umfasst (HL-Red I, HL-Red II und „Der Heiligen Leben“) weitere Möglichkeiten eröffnen.

Im Rahmen der Expertenrückmeldung wurde die Exportmöglichkeit der gewonnenen Erkenntnisse und Ergebnisse des Tools gewünscht. Ferner wurde der Wunsch nach Unterstützung bei der Annotation der Texte geäußert. Es können hier mehrere Ansätze verfolgt werden: Entweder wird der bestehende Ansatz um die Möglichkeit eines Editors erweitert oder aber ein bestehender Editor in den Workflow eingebaut, welcher direkt in das Arbeitsformat importiert. Das Institut für Visualisierung und Interaktive Systeme forscht hierbei bereits an der Entwicklung solcher Editoren.

Eine weitere Erweiterung wäre die Vorschaltung eines Parsers, der TEI-konformes XML in das Arbeitsformat umwandelt. Alternativ kann auch das in dieser Arbeit entwickelte Arbeitsformat dahingehend modifiziert werden, dass es den Richtlinien der TEI entspricht.

Literatur

[1] M. Card, *Readings in information visualization: using vision to think*. Morgan Kaufmann, 1999.

[2] W. S. Cleveland und R. McGill, "Graphical perception: Theory, experimentation, and application to the development of graphical methods", *Journal of the American statistical association*, Jg. 79, Nr. 387, S. 531–554, 1984.

[3] ——, "Graphical perception and graphical methods for analyzing scientific data", *Science*, Jg. 229, Nr. 4716, S. 828–833, 1985.

[4] M. Harrower und C. A. Brewer, "ColorBrewer. org: an online tool for selecting colour schemes for maps", *The Cartographic Journal*, Jg. 40, Nr. 1, S. 27–37, 2003.

[5] S. Eick, J. L. Steffen und E. E. Sumner, "Seesoft-a tool for visualizing line oriented software statistics", *IEEE Transactions on Software Engineering*, Jg. 18, Nr. 11, S. 957–968, 1992.

[6] S. Jänicke, G. Franzini, M. F. Cheema und G. Scheuermann, "Visual text analysis in digital humanities", in *Computer Graphics Forum*, Wiley Online Library, Bd. 36, 2017, S. 226–250.

[7] S. G. Eick, "Graphically displaying text", *Journal of Computational and Graphical Statistics*, Jg. 3, Nr. 2, S. 127–142, 1994.

[8] (). TEI: Text Encoding Initiative, Adresse: `http://www.tei-c.org/` (besucht am 22. 08. 2018).

[9] (). ePoetics — Korpuserschließung und Visualisierung deutschsprachiger Poetiken (1770- 1960) für den „Algorithmic criticism", Adresse: `index5534.html?page_id=6` (besucht am 22. 08. 2018).

[10] F. Moretti, *Graphs, maps, trees: abstract models for a literary history*. Verso, 2005.

[11] S. Jänicke, A. Geßner, G. Franzini, M. Terras, S. Mahony und G. Scheuermann, "TRAViz: A visualization for variant graphs", *Digital Scholarship in the Humanities*, Jg. 30, Nr. suppl_1, S. i83–i99, 2015.

[12] S. Koch, M. John, M. Wörner, A. Müller und T. Ertl, "VarifocalReader—in-depth visual analysis of large text documents", *IEEE transactions on visualization and computer graphics*, Jg. 20, Nr. 12, S. 1723–1732, 2014.

[13] S. Jänicke und D. J. Wrisley, "Interactive Visual Alignment of Medieval Text Versions",

[14] D. Wheeles und K. Jensen, "Juxta commons", *Proceedings of the Digital Humanities*, Jg. 5, Nr. 6, S. 12, 2013.

[15] S. Schreibman, A. Kumar und J. McDonald, "The versioning machine", *Literary and Linguistic Computing*, Jg. 18, Nr. 1, S. 101–107, 2003.

[16] M. Bostock. (). D3.js - Data-Driven Documents, Adresse: `https://d3js.org/` (besucht am 16. 08. 2018).

[17] A. Don, E. Zheleva, M. Gregory, S. Tarkan, L. Auvil, T. Clement, B. Shneiderman und C. Plaisant, "Discovering interesting usage patterns in text collections: integrating text mining with visualization", in *Proceedings of the sixteenth ACM conference on Conference on information and knowledge management*, ACM, 2007, S. 213–222.

[18] S. Jänicke und D. Joseph Wrisley, "Visualizing Mouvance: Toward a visual analysis of variant medieval text traditions", *Digital Scholarship in the Humanities*, Jg. 32, Nr. suppl_2, S. ii106–ii123, 2017.

[19] M. Wattenberg und F. B. Viégas, "The word tree, an interactive visual concordance", *IEEE Transactions on Visualization & Computer Graphics*, Nr. 6, S. 1221–1228, 2008.

[20] J. Mackinlay, "Automating the design of graphical presentations of relational information", *Acm Transactions On Graphics (Tog)*, Jg. 5, Nr. 2, S. 110–141, 1986.

[21] K. Kucher und A. Kerren, "Text visualization techniques: Taxonomy, visual survey, and community insights", in *Visualization Symposium (PacificVis), 2015 IEEE Pacific*, IEEE, 2015, S. 117–121.

[22] D. A. Keim, F. Mansmann, J. Schneidewind, J. Thomas und H. Ziegler, "Visual analytics: Scope and challenges", in *Visual data mining*, Springer, 2008, S. 76–90.

[23] J. A. Wise, J. J. Thomas, K. Pennock, D. Lantrip, M. Pottier, A. Schur und V. Crow, "Visualizing the non-visual: Spatial analysis and interaction with information from text documents", in *Information Visualization, 1995. Proceedings.*, IEEE, 1995, S. 51–58.

A. Anhang

A.1. xsd-Datei

Listing A.1: Entwickeltes XML-Schema

```
1  <?xml version="1.0"?>
2  <xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema"
3            targetNamespace="ba" xmlns:ba="ba">
4
5  <xs:complexType name="token">
6      <xs:simpleContent>
7          <xs:extension base="xs:string">
8              <xs:attribute name="id" type="xs:integer"/>
9          </xs:extension>
10     </xs:simpleContent>
11 </xs:complexType>
12
13 <xs:complexType name="line">
14     <xs:sequence>
15         <xs:element name="t" type="ba:token" maxOccurs="unbounded"/>
16     </xs:sequence>
17     <xs:attribute name="id" type="xs:integer"/>
18 </xs:complexType>
19
20 <xs:complexType name="column">
21     <xs:sequence>
22         <xs:element name="line" type="ba:line" maxOccurs="unbounded"/>
23     </xs:sequence>
24     <xs:attribute name="id" type="xs:integer"/>
25 </xs:complexType>
26
27 <xs:complexType name="page">
28     <xs:sequence>
29         <xs:element name="column" type="ba:column" minOccurs="1"
               maxOccurs="2"/>
30     </xs:sequence>
31     <xs:attribute name="id" type="xs:integer"/>
32 </xs:complexType>
```

```
33
34 <xs:simpleType name="catList">
35     <xs:list itemType="xs:integer"/>
36 </xs:simpleType>
37
38 <xs:simpleType name="tokenList">
39     <xs:list itemType="xs:integer"/>
40 </xs:simpleType>
41
42 <xs:simpleType name="weightValue">
43     <xs:restriction base="xs:integer">
44       <xs:minInclusive value="1"/>
45       <xs:maxInclusive value="3"/>
46     </xs:restriction>
47 </xs:simpleType>
48
49 <xs:simpleType name="weightList">
50     <xs:list itemType="ba:weightValue"/>
51 </xs:simpleType>
52
53 <xs:complexType name="relation">
54     <xs:sequence>
55         <xs:element name="type" type="ba:catList"/>
56         <xs:element name="weight" type="ba:weightList"/>
57         <xs:element name="tokens_text0" type="ba:tokenList"/>
58         <xs:element name="comments_text0" type="xs:string"/>
59         <xs:element name="tokens_text1" type="ba:tokenList"/>
60         <xs:element name="comments_text1" type="xs:string"/>
61     </xs:sequence>
62     <xs:attribute name="id" type="xs:integer"/>
63 </xs:complexType>
64
65 <xs:complexType name="category">
66     <xs:sequence>
67         <xs:element name="description" type="xs:string"/>
68     </xs:sequence>
69     <xs:attribute name="id" type="xs:integer"/>
70 </xs:complexType>
71
72 <xs:complexType name="text">
73     <xs:sequence>
74         <xs:element name="page" type="ba:page" maxOccurs="unbounded">
75             <xs:key name="columnKey">
76                 <xs:selector xpath="column"/>
77                 <xs:field xpath="@id"/>
```

```
78            </xs:key>
79        </xs:element>
80    </xs:sequence>
81    <xs:attribute name="id" type="xs:integer"/>
82 </xs:complexType>
83
84 <xs:element name="body">
85   <xs:complexType>
86     <xs:sequence>
87         <xs:element name="category" type="ba:category" maxOccurs="
               unbounded"/>
88         <xs:element name="relation" type="ba:relation" maxOccurs="
               unbounded"/>
89         <xs:element name="text" type="ba:text" maxOccurs="2">
90             <xs:key name="pageKey">
91                 <xs:selector xpath="page"/>
92                 <xs:field xpath="@id"/>
93             </xs:key>
94             <xs:key name="lineKey">
95                 <xs:selector xpath="page/column/line"/>
96                 <xs:field xpath="@id"/>
97             </xs:key>
98             <xs:key name="tokenKey">
99                 <xs:selector xpath="page/column/line/t"/>
100                <xs:field xpath="@id"/>
101            </xs:key>
102        </xs:element>
103    </xs:sequence>
104  </xs:complexType>
105  <xs:key name="textKey">
106       <xs:selector xpath="text"/>
107       <xs:field xpath="@id"/>
108  </xs:key>
109  <xs:key name="catKey">
110       <xs:selector xpath="category"/>
111       <xs:field xpath="@id"/>
112  </xs:key>
113  <xs:key name="relKey">
114       <xs:selector xpath="relation"/>
115       <xs:field xpath="@id"/>
116  </xs:key>
117 </xs:element>
118 </xs:schema>
```

A.2. Beispiel-Datei

Listing A.2: Beispieldatei des entwickelten XML-Schemas

```
1  <?xml version='1.0' encoding='UTF-8'?>
2  <x:body xmlns:x="ba" xmlns:xsi="http://www.w3.org/2001/XMLSchema-
       instance" xsi:schemaLocation="ba schema0.xsd">
3    <category id="0">
4      <description>zusätzliche Adjektiva</description>
5    </category>
6    <category id="1">
7      <description>Zusätze innnerhalb des Satzes</description>
8    </category>
9    <category id="2">
10     <description>zusätzliche Episode</description>
11   </category>
12   <category id="3">
13     <description>Synonyme</description>
14   </category>
15   <category id="4">
16     <description>ähnliche Semantik</description>
17   </category>
18   <category id="5">
19     <description>veränderte Semantik</description>
20   </category>
21   <category id="6">
22     <description>andere Namensform</description>
23   </category>
24   <category id="7">
25     <description>zusätzlicher Personenname</description>
26   </category>
27   <category id="8">
28     <description>zusätzlicher Ortsname</description>
29   </category>
30   <category id="9">
31     <description>syntaktische Umstellung</description>
32   </category>
33   <category id="10">
34     <description>Genera verbi: Aktiv vs. Passiv</description>
35   </category>
36   <category id="11">
37     <description>Oratio: direkte Rede vs. indirekte Rede</description>
38   </category>
39   <relation id="rel0">
40       <type>0</type>
```

```
41        <weight>3</weight>
42        <tokens_text0>t0-5</tokens_text0>
43        <tokens_text1>t1-5 t1-6</tokens_text1>
44        <comments_text0> </comments_text0>
45        <comments_text1> </comments_text1>
46        </relation>
47    <relation id="rel1">
48        <type>1</type>
49        <weight>3</weight>
50        <tokens_text0>t0-2</tokens_text0>
51        <tokens_text1>t1-2 t1-11 t1-13 t1-15</tokens_text1>
52        <comments_text0> </comments_text0>
53        <comments_text1> </comments_text1>
54        </relation>
55    <relation id="rel2"><type>2</type><weight>3</weight><tokens_text0>t0
          -21</tokens_text0><tokens_text1>t1-25 t1-26 t1-27 t1-28 t1-29 t1
          -30 t1-31 t1-32 t1-33 t1-34</tokens_text1> <comments_text0> </
          comments_text0> <comments_text1> </comments_text1></relation>
56    <relation id="rel3"><type>3</type><weight>3</weight><tokens_text0>t0
          -25 t0-26</tokens_text0><tokens_text1>t1-37</tokens_text1> <
          comments_text0> </comments_text0> <comments_text1> </
          comments_text1></relation>
57    <relation id="rel4"><type>4</type><weight>3</weight><tokens_text0>t0
          -33 t0-34</tokens_text0><tokens_text1>t1-44 t1-45</tokens_text1>
          <comments_text0> </comments_text0> <comments_text1> </
          comments_text1></relation>
58    <relation id="rel5"><type>5</type><weight>3</weight><tokens_text0>t0
          -36</tokens_text0><tokens_text1>t1-47</tokens_text1> <
          comments_text0> </comments_text0> <comments_text1> </
          comments_text1></relation>
59    <relation id="rel6"><type>6</type><weight>3</weight><tokens_text0>t0
          -46</tokens_text0><tokens_text1>t1-57</tokens_text1> <
          comments_text0> </comments_text0> <comments_text1> </
          comments_text1></relation>
60    <relation id="rel7"><type>7</type><weight>3</weight><tokens_text0>t0
          -53</tokens_text0><tokens_text1>t1-64 t1-65 t1-66</tokens_text1>
          <comments_text0> </comments_text0> <comments_text1> </
          comments_text1></relation>
61    <relation id="rel8"><type>8</type><weight>3</weight><tokens_text0>t0
          -56</tokens_text0><tokens_text1>t1-69 t1-70 t1-71</tokens_text1>
          <comments_text0> </comments_text0> <comments_text1> </
          comments_text1></relation>
62    <relation id="rel9"><type>9</type><weight>3</weight><tokens_text0>t0
          -60 t0-61 t0-62</tokens_text0><tokens_text1>t1-75 t1-76 t1-77</
          tokens_text1> <comments_text0> </comments_text0> <comments_text1>
```

```
        </comments_text1></relation>
63    <relation id="rel10"><type>10</type><weight>3</weight><tokens_text0>
          t0-20 t0-63 t0-64 t0-65 t0-66</tokens_text0><tokens_text1>t1-20
          t1-78 t1-79 t1-80 t1-81 t1-82</tokens_text1> <comments_text0> </
          comments_text0> <comments_text1> </comments_text1></relation>
64    <relation id="rel11"><type>11</type><weight>3</weight><tokens_text0>
          t0-68 t0-69 t0-70</tokens_text0><tokens_text1>t1-84 t1-85 t1-86
          t1-87</tokens_text1> <comments_text0> </comments_text0> <
          comments_text1> </comments_text1></relation>
65    <text id="text0">
66      <page id="p0-0">
67        <column id="c0-0">
68          <line id="l0-0"><t id="t0-0">Er</t><t id="t0-3">hatte</t><t id=
                "t0-5">eine</t><t id="t0-7">edle</t><t id="t0-9">Frau.</t></
                line>
69          <line id="l0-1"><t id="t0-2">Von</t><t id="t0-11">Sankt</t><t
                id="t0-13">Klara</t></line>
70          <line id="l0-2"><t id="t0-4">Sie</t><t id="t0-15">bat</t><t id=
                "t0-17">für</t><t id="t0-19">die</t><t id="t0-21">Sünder</t>
                </line>
71          <line id="l0-3"><t id="t0-6">Es</t><t id="t0-23">kam</t><t id="
                t0-24">der</t><t id="t0-25">böse</t><t id="t0-26">Geist</t><
                t id="t0-27">über</t><t id="t0-28">sie.</t></line>
72          <line id="l0-4"><t id="t0-8">Und</t><t id="t0-29">es</t><t id="
                t0-30">kam,</t><t id="t0-31">so</t><t id="t0-32">wie</t><t
                id="t0-33">ihr</t><t id="t0-34">gesagt</t><t id="t0-35">
                wurde.</t></line>
73          <line id="l0-5"><t id="t0-10">Da</t><t id="t0-36">sprach</t><t
                id="t0-37">eine</t><t id="t0-38">himmlische</t><t id="t0-39"
                >Stimme</t><t id="t0-40">zu</t><t id="t0-41">ihr,</t></line>
74          <line id="l0-6"><t id="t0-12">Und</t><t id="t0-42">da</t><t id=
                "t0-43">kam</t><t id="t0-44">der</t><t id="t0-45">Kaiser</t>
                <t id="t0-46">Friedrich</t><t id="t0-47">zu</t><t id="t0-48"
                >ihr.</t></line>
75          <line id="l0-7"><t id="t0-14">Und</t><t id="t0-49">sie</t><t id
                ="t0-50">bat</t><t id="t0-51">für</t><t id="t0-52">ihre</t><
                t id="t0-53">Schwester.</t></line>
76          <line id="l0-8"><t id="t0-16">Sie</t><t id="t0-54">ging</t><t
                id="t0-55">ins</t><t id="t0-56">Kloster.</t></line>
77          <line id="l0-9"><t id="t0-18">Und</t><t id="t0-57">sie</t><t id
                ="t0-58">war</t><t id="t0-59">so</t><t id="t0-60">rein</t><t
                 id="t0-61">und</t><t id="t0-62">keusch.</t></line>
78          <line id="l0-10"><t id="t0-20">Er</t><t id="t0-63">sprach</t><t
                 id="t0-64">von</t><t id="t0-65">der</t><t id="t0-66">Liebe.
                </t></line>
```

```
79          <line id="l0-11"><t id="t0-22">Er</t><t id="t0-67">sprach:</t><
                t id="t0-68">"Es</t><t id="t0-69">werde</t><t id="t0-70">
                Licht"</t></line>
80        </column>
81      </page>
82    </text>
83    <text id="text1">
84      <page id="p1-0">
85        <column id="c1-0">
86          <line id="l1-0"><t id="t1-0">Er</t><t id="t1-1">hatte</t><t id=
                "t1-3">eine</t><t id="t1-5">seelige</t><t id="t1-7">edle</t>
                <t id="t1-9">Frau.</t></line>
87          <line id="l1-1"><t id="t1-2">Von</t><t id="t1-11">der</t><t id=
                "t1-13">heiligen</t><t id="t1-15">Jungfrau</t><t id="t1-17">
                Sankt</t><t id="t1-19">Klara</t></line>
88          <line id="l1-2"><t id="t1-4">Sie</t><t id="t1-21">bat</t><t id=
                "t1-23">für</t><t id="t1-24">die</t><t id="t1-25">Sünder.</t
                ><t id="t1-26">Sodann</t><t id="t1-27">bat</t><t id="t1-28">
                sie</t><t id="t1-29">auch</t><t id="t1-30">für</t><t id="t1
                -31">die</t><t id="t1-32">Alten</t><t id="t1-33">und</t><t
                id="t1-34">Kranken.</t></line>
89          <line id="l1-3"><t id="t1-6">Es</t><t id="t1-35">kam</t><t id="
                t1-36">der</t><t id="t1-37">Teufel</t><t id="t1-38">über</t>
                <t id="t1-39">sie.</t></line>
90          <line id="l1-4"><t id="t1-8">Und</t><t id="t1-40">es</t><t id="
                t1-41">kam,</t><t id="t1-42">so</t><t id="t1-43">wie</t><t
                id="t1-44">es</t><t id="t1-45">verkündet</t><t id="t1-46">
                wurde.</t></line>
91          <line id="l1-5"><t id="t1-10">Da</t><t id="t1-47">kam</t><t id=
                "t1-48">eine</t><t id="t1-49">himmlische</t><t id="t1-50">
                Stimme</t><t id="t1-51">zu</t><t id="t1-52">ihr.</t></line>
92          <line id="l1-6"><t id="t1-12">Und</t><t id="t1-53">da</t><t id=
                "t1-54">kam</t><t id="t1-55">der</t><t id="t1-56">Kaiser</t>
                <t id="t1-57">Friederich</t><t id="t1-58">zu</t><t id="t1-59
                ">ihr.</t></line>
93          <line id="l1-7"><t id="t1-14">Und</t><t id="t1-60">sie</t><t id
                ="t1-61">bat</t><t id="t1-62">für</t><t id="t1-63">ihre</t><
                t id="t1-64">Schwester</t><t id="t1-65">St.</t><t id="t1-66"
                >Agnes.</t></line>
94          <line id="l1-8"><t id="t1-16">Sie</t><t id="t1-67">ging</t><t
                id="t1-68">ins</t><t id="t1-69">Kloster</t><t id="t1-70">St.
                </t><t id="t1-71">Barbara.</t></line>
95          <line id="l1-9"><t id="t1-18">Und</t><t id="t1-72">sie</t><t id
                ="t1-73">war</t><t id="t1-74">so</t><t id="t1-75">keusch</t>
                <t id="t1-76">und</t><t id="t1-77">rein.</t></line>
```

```
96          <line id="l1-10"><t id="t1-20">Es</t><t id="t1-78">wurde</t><t
                id="t1-79">von</t><t id="t1-80">der</t><t id="t1-81">Liebe</
                t><t id="t1-82">gesprochen.</t></line>
97          <line id="l1-11"><t id="t1-22">Er</t><t id="t1-83">sprach,</t><
                t id="t1-84">dass</t><t id="t1-85">es</t><t id="t1-86">Licht
                </t><t id="t1-87">werde.</t></line>
98        </column>
99      </page>
100   </text>
101 </x:body>
```

Erklärung

Ich versichere, diese Arbeit selbstständig verfasst zu haben. Ich habe keine anderen als die angegebenen Quellen benutzt und alle wörtlich oder sinngemäß aus anderen Werken übernommene Aussagen als solche gekennzeichnet. Weder diese Arbeit noch wesentliche Teile daraus waren bisher Gegenstand eines anderen Prüfungsverfahrens. Ich habe diese Arbeit bisher weder teilweise noch vollständig veröffentlicht. Das elektronische Exemplar stimmt mit allen eingereichten Exemplaren überein.

Ort, Datum, Unterschrift