

Digitale Methoden für die Historische Semantik

Auf den Spuren von Begriffen in digitalen Korpora

von Silke Schwandt*

Abstract: One of the biggest challenges associated with semantic analysis is to obtain a distance from existing semantic knowledge. Methods from the field of Digital Humanities allow us to find new ways to cope with that challenge. Text mining or topic modeling tools offer a new perspective on words and their combination within a text. The following article highlights the relevance of digital methods for historical semantics, using the Latin term *virtus* and its medieval use as an example. It raises the question of genre and diachronic semantic change and demonstrates how digital tools have the potential to not only challenge our knowledge about texts but also help to reorganize what we already know.

Wenn man sich um die Geschichte eines Begriffes und die Erzählung semantischen Wandels bemüht, geht es in erster Linie um die Rekonstruktion von Verwendungskontexten einer sprachlichen Repräsentation dieses Begriffes. Solche Repräsentationen können einzelne Wörter oder längere Phrasen sein, die in ihrer Form als Zeichenfolgen (in ihrer schriftlichen Gestalt) als einigermaßen stabil gesetzt werden. Eine Untersuchung der semantischen Karriere eines Wortes wie *virtus* beginnt daher als Untersuchung der Verwendungskontexte dieses Wortes und wird begleitet von dem Versuch, das Konzept, das wir hinter dem Wort vermuten, nicht als stabil vorauszusetzen.

Eine nicht sehr neue Herausforderung der Begriffsgeschichte besteht darin, sich als Forscherin vom eigenen Gegenstand, in diesem Fall von einem Begriff oder Konzept, und dem Wissen über diesen Gegenstand zu trennen. In gewisser Weise lässt sich diese Herausforderung als eine Herausforderung an die Objektivität beschreiben. An dieser Stelle soll nicht diskutiert werden, welchen Stellenwert das Paradigma der wissenschaftlichen Objektivität in der Geschichtswissenschaft oder in der Historischen Semantik hat. Aber wenn wir verstehen wollen, wie sich Verwendungskontexte und mit ihnen die Bedeutung und Bewertung von Wörtern und Begriffen verändern, dann bleibt der Umgang mit dem Vorwissen der Forscherin oder des Forschers als Heraus-

* Ich bedanke mich bei meinen Mitarbeiterinnen im Projekt „Dateninfrastruktur und Digital Humanities“ des SFB 1288 „Praktiken des Vergleichens“ an der Universität Bielefeld, Anna Neubert und Helene Schlicht, für wertvolle Anregungen und Korrekturen des Manuskripts. Gleiches gilt für Franz-Josef Arlinghaus, Lena Gumpert und Simon Siemianowski für kritische Anmerkungen zu den Beispielen und ihrer Interpretation sowie zur Verständlichkeit und Argumentation.

forderung bestehen. Es dient als Ausgangspunkt und als Horizont für die Interpretation. Es ist die Grundlage jeder Forschungsfrage.

Wie kann es dennoch gelingen, dass wir zu neuen Interpretationen kommen? Dieser Aufsatz folgt der These, dass der Einsatz digitaler Methoden hierbei besonders produktiv sein kann. Die Berechnung von Worthäufigkeiten, von Wortabständen und Kookurrenten liefert einen Überblick über den Text, der frei ist von Interpretationen – der Computer ist semantisch blind und hilft uns, unsere Vorannahmen infrage zu stellen.¹

Gegenwärtig sind die Digital Humanities, die sich der Entwicklung und Anwendung von digitalen Methoden für geisteswissenschaftliche Fragestellungen widmen, im Aufwind. Sie reagieren auf die steigende Verfügbarkeit von digitalisierten Korpora für die Forschung und fragen danach, wie neben der beschleunigten Durchsuchbarkeit immer größer werdender Textmengen auch die Methoden und Methodologien ihrer Erschließung weiterentwickelt werden können.²

Diese Entwicklung beginnt schon mit Roberto Busa in den 1940er Jahren. Der Vater des heutigen „Corpus Thomisticum“,³ einer digitalen Gesamtausgabe der Werke von Thomas von Aquin († 1274), hat für seine Arbeit am Wortschatz des mittelalterlichen Autors bereits mit digitalen Methoden gearbeitet und mit der Unterstützung von IBM Konkordanzen für das Gesamtwerk des Aquinaten erstellt. Die heutige Oberfläche der Internetpräsenz des „Corpus Thomisticum“ erlaubt neben der Suche in den Werken des Autors auch statistische Abfragen über das Korpus.

Schon dieses Beispiel zeigt deutlich, wie sich die Arbeit mit digitalen Methoden in den Geisteswissenschaften entwickelt hat.⁴ Vor allem in der angloamerikanischen Forschung entstehen Impulse, die für viele Fachwissenschaftlerinnen und Fachwissenschaftler prägend sind und sich keineswegs allein auf die Arbeit mit digitalen Textkorpora beschränken. Unter dem Label der Digital Cultural Heritage wird verstärkt auch über die Möglichkeiten der Digitalisierung im Bereich der kulturellen Institutionen, wie zum Beispiel

1 Vgl. Silke Schwandt, *Digitale Objektivität in der Geschichtswissenschaft? Oder: Kann man finden, was man nicht sucht?*, in: *Rechtsgeschichte* 24. 2016, S. 337 f. Vgl. zum Paradigma der computergestützten Hermeneutik Geoffrey Rockwell u. Stéfán Sinclair, *Hermeneutica. Computer-Assisted Interpretation in the Humanities*, Cambridge, MA 2016. Kookurrenten sind Wörter, die in der Umgebung eines Suchwortes verwendet werden und als Hinweise auf die Verwendungsweisen und Bedeutungsaspekte von Wörtern und Begriffen betrachtet werden können.

2 Vgl. Fotis Jannidis u. a., *Digital Humanities. Eine Einführung*, Stuttgart 2017.

3 Vgl. Fundación Tomás de Aquina, *Corpus Thomisticum*, <http://www.corpusthomicum.org>.

4 Überblickswerke, die diese Entwicklung für die letzten 15 Jahre dokumentieren, sind etwa Susan Schreibman u. a., *A Companion to Digital Humanities*, Malden, MA 2004 oder Claire Warwick u. a., *Digital Humanities in Practice*, London 2012.

Museen oder Stiftungen, diskutiert.⁵ In diesen Bereich gehören dann auch vermehrt Fragestellungen der Visualisierung und Darstellung sowie der digitalen Edition. Für die Arbeit mit textlichen Quellen lässt sich festhalten: Auch wenn sich der Trend beobachten lässt, dass immer mehr historische Korpora digitalisiert werden, ist die Aufbereitung und Auszeichnung vor allem am Beispiel von Korpora der modernen Sprachen erprobt.⁶ Je heterogener die Sprachformen und Korpora desto größer wird die Herausforderung.

Um über das beschleunigte Auffinden von Belegstellen für die Begriffsgeschichte hinaus die Leistungsfähigkeit des Computers zu erproben, braucht es in erster Linie annotierte Korpora, die die Informationen, die man zur Interpretation heranziehen will, maschinenlesbar hinterlegen. Seit den 1980er Jahren werden zu diesem Zweck im Rahmen der Text Encoding Initiative Modelle zur Auszeichnung gemäß XML-Standard entwickelt.⁷ Diese Initiative zielt in erster Linie darauf, Formate so zu entwerfen, dass sie für möglichst viele Projekte und Korpora anwendbar sind. Sie arbeitet also sukzessive an der Entwicklung plattformübergreifender Richtlinien. Auf diese Weise wird unter anderem auch die Nachnutzung von in verschiedenen Projekten und Initiativen annotierten Korpora und entwickelten Analysewerkzeugen erleichtert.⁸

Der Aufsatz möchte der Leistungsfähigkeit digitaler Methoden für die Historische Semantik in drei Schritten nachspüren. Als Beispiel dient dazu die Begriffsgeschichte, oder besser: Wortverwendungsgeschichte von *virtus*, einer lateinischen Vokabel, die im Deutschen meist als Tugend wiedergegeben wird, auf der Basis mittellateinischer Korpora. Dabei wird zunächst der Blick auf Worthäufigkeiten und Kookurrenten von *virtus* innerhalb mittelalterlicher Fürstenspiegel gelenkt, um Verwendungsmuster zu identifizieren und

5 Vgl. als eines der ersten Werke einer produktiven Diskussion Fiona Cameron u. Sarah Kenderdine (Hg.), *Theorizing Digital Cultural Heritage. A Critical Discourse*, Cambridge, MA 2007.

6 Ein gutes Beispiel für die Anwendung digitaler Methoden in der neueren Begriffsgeschichte ist zuletzt Alexander Friedrich u. Chris Biemann, *Digitale Begriffsgeschichte? Methodologische Überlegungen und exemplarische Versuche am Beispiel moderner Netzsemantik*, in: *Forum Interdisziplinäre Begriffsgeschichte* 5. 2016, http://www.zfl-berlin.org/tl_files/zfl/downloads/publikationen/forum_begriffsgeschichte/ZfL_FIB_5_2016_2_FriedrichBiemann.pdf. Vgl. ebenso zur Anwendung in der Zeitgeschichte Kathrin Kollmeier, *Begriffsgeschichte und Historische Semantik. Version: 2.0*, in: *Docupedia-Zeitgeschichte*, https://docupedia.de/zg/Begriffsgeschichte_und_Historische_Semantik_Version_2.0_Kathrin_Kollmeier.

7 XML (Extensible Markup Language) ist ein Standard, der den Austausch von Daten erleichtern soll, weil er maschinenlesbar und international anerkannt ist. Für die Auszeichnung von Dokumentstrukturen hat sich der TEI-Standard etabliert. Vgl. *Text Encoding Initiative, P5 Guidelines for Electronic Text Encoding and Interchange*, <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/index.html>.

8 Eine Übersicht über für die Geschichtswissenschaft nützliche und verfügbare Korpora und Tools liefern die Beiträge in Laura Busse u. a. (Hg.), *Clio-Guide. Ein Handbuch zu digitalen Ressourcen für die Geschichtswissenschaften*, Berlin 2016.

den Bedeutungsgehalt, der mit diesen Mustern verbunden wird, darzustellen (I). In einem zweiten Schritt sollen diese Muster mit der Verwendung des Wortes in mittelalterlichen Urkunden konfrontiert werden, um die Grenzen der Übertragbarkeit der Ergebnisse auszuloten (II). Für die Historische Semantik sind neben einer semasiologisch orientierten Begriffsgeschichte auch onomasiologische Vorgehensweisen zur Erforschung von Bedeutungszusammenhängen relevant. Daher soll in einem dritten Schritt nach der Repräsentativität der Ergebnisse mit Blick auf den Diskurs gefragt werden (III). Wie zentral ist das für *virtus* erarbeitete Bedeutungsfeld in den untersuchten Texten?

I. Semantische Momentaufnahmen – *virtus* in den Fürstenspiegeln

Virtus ist ein Wort, das nicht nur in den lateinischen Texten des Mittelalters nahezu omnipräsent ist. Es ist vor allem mit Blick auf seine Interpretation und Übersetzung als Tugend auch im Forschungsdiskurs zur mittelalterlichen Geschichte schon vielfach thematisiert worden. Diese Omnipräsens lässt sich nicht nur in der Forschung, sondern auch in der zeitgenössischen lateinischen Schriftproduktion nachvollziehen. Nimmt man die „Patrologia Latina“ als eine Sammlung von lateinischen Texten des dritten bis dreizehnten Jahrhunderts zur Hand, findet man dort nahezu 150.000 Okkurrenzen von *virtus*.⁹ Dies ist selbstverständlich nur ein Auszug aus der mittelalterlichen Schriftkultur, aber einer, der viele verschiedene Gattungen und Formen umfasst. In diesem Sinne ist die „Patrologia Latina“ eher ein Repositorium als ein kontrolliertes Korpus, und vermittelt daher den Eindruck, dass *virtus* überall verwendet würde. Um aber Aussagen über Verwendungsweisen und Kontexte treffen zu können, ist es notwendig, dass man von einer bloßen Textsammlung zu einem strukturierten Korpus wechselt. Korpora sind Textmengen, die anhand bestimmter Kriterien zusammengestellt worden sind. Mögliche Kriterien zur Erstellung eines Korpus sind Autorschaft, Datierung, Textsorte, Region und dergleichen.

1. Worthäufigkeiten und Kookkurrenten

Bei der begriffsgeschichtlichen Arbeit ist ein semasiologisches Vorgehen oft der erste Schritt. Man begibt sich auf die Suche nach einem Wort, das für die Forscherin oder den Forscher einen Begriff repräsentiert, und wählt dafür ein Korpus, in dem man die Verwendung dieses Begriffes analysieren will.

⁹ Die „Patrologia Latina“ besteht aus insgesamt 106.515.458 Einzelwörtern. Vgl. Alexander Mehler u. a., *Inducing Linguistic Networks from Historical Corpora. Towards a New Method in Historical Semantics*, in: Paul Bennett u. a. (Hg.), *New Methods in Historical Corpora*, Tübingen 2013, S. 257–274, hier S. 260.

Im Folgenden soll daher die Verwendung von *virtus* in einem Korpus von mittelalterlichen Fürstenspiegeln untersucht werden.¹⁰ Methodisch geht es um eine Kookkurrenzanalyse, die als Kookkurrenten diejenigen Wörter in den Blick nimmt, die mit *virtus* in einem Satz verwendet werden. Im Unterschied zur Kollokation ist eine Kookkurrenz dabei definiert als das bloße Erscheinen von zwei Wörtern im gleichen Kontext. Syntaktische Zusammenhänge und Abhängigkeiten werden in diesem Schritt noch nicht in Betracht gezogen. Das Korpus umfasst vier Texte: „De civitate Dei“ (Die Gottesstadt) von Augustinus, vollendet 427 n. Chr., die „Regula Pastoralis“ Gregors des Großen, geschrieben etwa 590, die „Via Regia“ von Smaragd von St. Mihiel, geschrieben vermutlich 810, sowie der „Policraticus“ von Johannes von Salisbury, veröffentlicht 1159. Grundlage der Textauswahl war der Versuch, ein homogenes Textsortenkörpus zu erstellen,¹¹ das es erlaubt, die Verwendungsweisen eines Wortes diachron zu verfolgen und eine Wortgebrauchsgeschichte zu schreiben. Vor diesem Hintergrund soll *virtus* als ein Begriff des politischen Diskurses untersucht werden.¹² Bei der Analyse der Worthäufigkeiten und der Kookkurrenten konnten dann Muster identifiziert werden, die bestimmten Bedeutungsaspekten entsprechen.

2. Verwendungsmuster und Bedeutungsaspekte

Mithilfe der Kookkurrenzanalyse lassen sich Kombinationen von Wörtern beobachten und Zusammenhänge ableiten. Wiederholt verwendete Wortkombinationen bilden dann häufig Verwendungsmuster, die sich verschiedenen Bedeutungsaspekten eines Wortes zuordnen lassen.

Für *virtus* in den genannten Spiegeln konnten drei Bedeutungsaspekte identifiziert werden: moralische Norm, persönliche Kraft und Wirkmacht. Die beiden ersten Aspekte sind personenbezogen in dem Sinne, dass *virtus* in dieser Funktion immer einer Person, dazu gehören auch die Personen der Gottheit (Gott, Christus, Heiliger Geist), zugeschrieben wird. Die Norm richtet sich an den Menschen als Handlungsordnung, als persönliche Kraft bezeichnet

10 Die meisten in diesem Abschnitt präsentierten Ergebnisse stammen aus der Forschung der Autorin zu ihrer Dissertation. Vgl. Silke Schwandt, *Virtus. Zur Semantik eines politischen Konzepts im Mittelalter*, Frankfurt 2014; und Silke Schwandt, *Virtus as a Political Concept in the Middle Ages*, in: *Contributions to the History of Concepts* 10. 2015, H. 2, S. 71–90. Einige der hier abgedruckten Grafiken stammen aus dem gleichen Zusammenhang.

11 Die Homogenität des Korpus basiert auf der Zurechenbarkeit der Texte zu einer Textsorte und ist damit inhaltlich begründet.

12 Die folgenden Ergebnisse wurden mithilfe des eHumanities Desktop (<https://hudesktop.hucompute.org/desktop.jsp>) erarbeitet, der vom Frankfurter Projekt Computational Historical Semantics (<http://comphistsem.org/home.html>) betreut wird. Vgl. auch Alexander Mehler u. a., eHumanities-Desktop. Eine webbasierte Arbeitsumgebung für die geisteswissenschaftliche Fachinformatik, in: *Proceedings of the Symposium „Sprachtechnologie und eHumanities“*, 26.–27.2.2009, Universität Duisburg-Essen 2009.

virtus Handlungen und Eigenschaften. Die Zuordnung von Handlungen und Eigenschaften geschieht über Personal- und Possessivpronomina (*mea, tua, sua, noster, vester*) oder beigeordnete Attribute (*divina*); die Handlungsnorm steht für sich. Der Aspekt der Wirkmacht ist Ausdruck einer abstrakten Kraft, die Personen, Dingen und anderen Abstrakta zugeschrieben werden kann. *Virtus* als moralische Norm ist die Bedeutung, die in den meisten Fällen die Auswahl des deutschen Entsprechungswortes bestimmt: Tugend. Das Wort ist im normativen Moraldiskurs omnipräsent. Tugend ist der Maßstab, der an jedwedes Handeln angelegt wird und lobenswerte Eigenschaften kennzeichnet. Es werden ganze Kataloge solcher Eigenschaften und Handlungsmuster entworfen und diskutiert.¹³ Die moralische Norm übernimmt im politischen Diskurs Funktionen der sozialen Inklusion, der diskursiven Integration und der politischen Legitimation. Die sprachlichen Merkmale dieses Bedeutungsaspekts sind die Verwendung von Pluralformen und Mengenadjektiven sowie die Beiordnung von Tugendnamen. Außerdem finden sich in diesem Zusammenhang der Einsatz von Verben des Strebens sowie Kombinationen mit Wörtern, die einen Weg repräsentieren (*iter, trames*).¹⁴ Ausschlaggebend für die Bedeutungen eines Wortes ist also der Gebrauchskontext, sind die Kookkurrenten und ihre Relationen zum Bezugswort.

Gerade die Musterhaftigkeit im Sprachgebrauch lässt sich nicht allein lesend ermitteln. Die Beobachtung von Mustern oder Modellen fällt beim Lesen schwer – vor allem, weil es in den Geisteswissenschaften nicht üblich ist, den Leseprozess so zu dokumentieren, dass man diese Muster beobachten könnte. Ähnliches gilt für die Verteilung dieser Muster im Text. Die Identifikation der Muster ist ein Akt der Interpretation; diese im Textverlauf zu verfolgen, ist eine Aufgabe, bei der die Arbeit computergestützt erfolgen kann. Auf diese Weise lassen sich Argumentationsmuster im Text verorten, auf deren Basis dann Textstrukturen neu bewertet werden können.

In „De civitate Dei“ wird *virtus* als moralische Norm einerseits im Sinne der moralphilosophischen *areté* und andererseits im Sinne einer christlichen Werkemoral verwendet. Ein Beispiel: Im fünften Buch, Kapitel zwanzig schreibt Augustinus von der *vera virtus* der Bürger der ewigen Stadt und der einfachen *virtus* derer, die über die irdische Stadt nicht hinauskommen werden.¹⁵ *Virtus* gilt hier in Verbindung mit *habere* als Eigenschaft. Ihr Besitz

13 Vgl. hierzu bspw. Martin Seel, 111 Tugenden, 111 Laster. Eine philosophische Revue, Frankfurt 2011.

14 Vgl. Schwandt, Semantik, S. 101 mit Anm. 308.

15 Bernhard Dombart u. Alphons Kalb, Sancti Aurelii Augustini De civitate Dei, tom. I: libri I–X u. tom. II: libri XI–XXII (Corpus Christianorum Series Latina [CCSL] 47/8, Aurelii Augustini Opera pars XIV, 1/2), Turnhout 1955, hier: Buch V, Kap. 20, S. 155–156: „dum illud constet inter omnes veraciter pios, neminem sine vera pietate, id est veri Dei vero cultu, veram posse habere virtutem; nec eam veram esse, quando gloriae servit humanae. Eos tamen qui cives non sint civitatis aeternae, quae in sacris Litteris nostris dicitur civitas Dei [Ps. 45, 5; 47, 3 und 9, etc.], utiliores esse terrenae civitati, quando

macht sie dem Menschen verfügbar. Sie kann durch andere sprachliche Muster und Argumentationszusammenhänge außerdem normativ aufgeladen werden.¹⁶ Ausschlaggebend ist die nähere Bestimmung von *virtus* durch das Adjektiv *vera* (wahr, wahrhaftig). Diese Hervorhebung grenzt *vera virtus* von anderen Formen ab. Es gibt also zwei Formen von *virtus*, die von Augustinus unterschieden werden. *Verus*, das auch als Attribut zu *pietas*, *Deus* und *cultus* gebraucht wird, verbindet diese Wörter wie eine Klammer mit der einen *virtus*, der christlichen *virtus*. Auf der anderen Seite steht die *virtus* der irdischen Stadt, die dem menschlichen Ruhm dient (*gloriae servit humanae*). Hinter diesen Vorstellungen steht das Ideal des ciceronischen Staatsmannes und der ciceronischen Philosophie, in der *virtus* eines der zentralen Wörter ist.¹⁷ Diese Ideale werden nicht negiert, sondern einem spezifischen Bereich der Gemeinschaft, dem nicht-christlichen, nicht-erwählten Teil der Gemeinschaft, zugeschrieben. In dieser Zuordnung liegt deutlich Kritik. Diese wird schon durch die Wahl des kennzeichnenden Adjektivs für die christliche *virtus* zum Ausdruck gebracht. Die christliche Tugend hat den philosophischen Zusammenhang nicht vollständig abgelöst. Augustinus vermittelt beide, gibt dem christlichen Horizont aber größeres Gewicht. Die Anbindung dieser Umdeutung an die Unterscheidung der *civitates* macht *virtus* zu einem Kriterium sozialer Inklusion und Exklusion.

Das Nachverfolgen der Muster sowie deren Quantifizierung erlaubt es also, auch implizite Argumentationsstrukturen im Text des Augustinus aufzuspüren, die durch nichts weiter markiert sind, als durch spezifische Verwendungsmuster von *virtus*. Befunde dieser Art sagen mehr über den Text aus, als die Diskussion berühmter Belegstellen, die von der Forschung mit definitivischer Macht ausgestattet werden. Für den Gebrauch von *virtus* bei Augustinus ist dies eine Stelle aus der Gottesstadt, in der *virtus* vom Kirchenvater selbst als *ordo amoris* definiert wird.¹⁸ Von der Forschung wird diese Stelle zur Grundlage der Interpretation von Sitte und Moral bei Augustinus und auf alle weiteren Instanzen der Verwendung des Wortes übertragen.¹⁹ Auch die

habent virtutem vel ipsam, quam si nec ipsam. Illi autem qui vera pietate praediti bene vivunt, si habent scientiae regendi populos, nihil est felicius rebus humanis, quam si Deo miserante habeant potestatem.“

16 Ebd.

17 Cicero wählt *virtus* als Übersetzungswort für das griechische Ideal der *areté* und macht *virtus* damit zu einer philosophischen Vokabel. Vgl. die gängigen Lexikonartikel zur Tugend, beispielsweise Jean Porter, Art. Tugend, in: Gerhard Müller u.a. (Hg.), *Theologische Realenzyklopädie*, Bd. 34, Berlin 2002, S. 184–197.

18 Dombart u. Kalb, *Sancti Aurelii Augustini De civitate Dei*, hier: Buch XV, Kap. 22, S. 488: „Nam et amor ipse ordinate amandus est, quo bene amatur quod amandum est, ut sit in nobis virtus qua vivitur bene. Unde mihi videtur, quod definitio brevis et vera virtutis ordo est amoris.“

19 Vgl. für diese Forschungsrichtung insb. Norbert Hartmann, *Ordo amoris*. Zur augustininischen Wesensbestimmung des Sittlichen, in: *Wissenschaft und Weisheit* 18, 1955, S. 1–23 u. S. 108–121; Josef Brechtken, *Augustinus Doctor Caritatis*. Sein Liebesbegriff im Widerspruch von Eigennutz und selbstloser Güte im Rahmen der

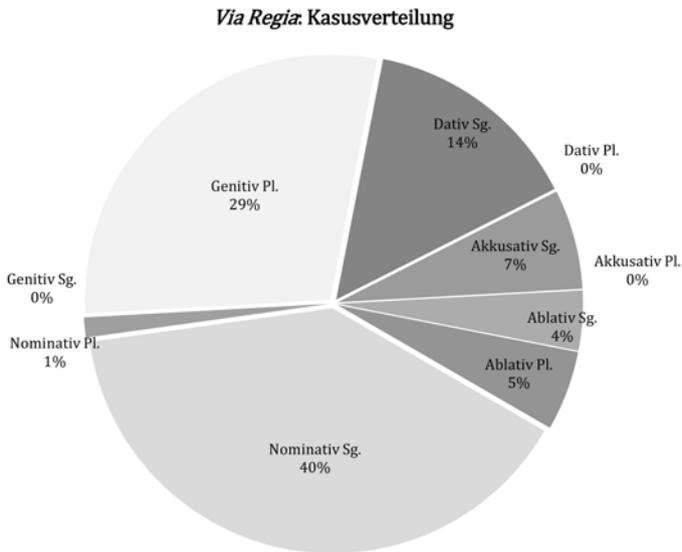


Abbildung 1: Verteilung der Kasus in der „Via Regia“.

übrigen Gebrauchssituationen in den Blick zu nehmen, zeigt schnell, dass Augustinus diese Definition nur ein einziges Mal verwendet und sie auch in den Werken späterer Autoren in dieser Form kaum verwendet wird. Im Rahmen der Betrachtung von Verwendungsmustern wird diese vormals prägende Stelle zu einer von vielen, zu einem Beleg für die Verteilung von Bedeutungsaspekten und Argumentationsstrukturen.

Auch bei der Verwendung der Kasus lassen sich Muster beobachten. Dabei ist es wichtig zu berücksichtigen, dass der Kasusgebrauch im Lateinischen wie in anderen Sprachen durchaus auch Bedeutungsträger sein kann. Nominative beispielsweise belegen die Funktion eines Wortes als Subjekt, das als handelnder Akteur interpretiert werden kann. Dative und Akkusative bezeugen die Objektivierbarkeit eines Wortes und des damit Bezeichneten. Die Verwendung eines Wortes im Genitiv abhängig von einem anderen Wort macht ersteres zu einem Attribut, das seinerseits Bedeutungszuschreibung übernimmt. Auf ähnliche Weise hat auch der Numerus Bedeutung. Abbildung 1 zeigt die Verteilung der Kasus in der „Via Regia“ von Smaragd von St. Mihiel.

antiken Glückseligkeits-Ethik, Meisenheim am Glan 1975; Remo Bodei, *Ordo amoris. Augustinus, irdische Konflikte und himmlische Glückseligkeit*, Wien 1993; oder auch Hannah Arendt, *Love and St. Augustine*, Chicago 1996 sowie Jörg Ulrich, *Virtus est ordo amoris*, in: Klaus Tanner (Hg.), „Liebe“ im Wandel der Zeiten. Kulturwissenschaftliche Perspektiven, Leipzig 2005, S. 53 – 61.

Auf den ersten Blick erkennt man, dass Smaragd von St. Mihiel *virtus* erstaunlich häufig als Agens eines Satzes verwendet. In vierzig Prozent aller Gebrauchssituationen steht *virtus* im Nominativ Singular. Tatsächlich zeigt der Blick in den Text, dass die meisten dieser Sätze demselben Muster folgen: Smaragd beschreibt verschiedene Eigenschaften als *virtus regia* und empfiehlt sie dem Herrscher zur Nachahmung.²⁰ *Virtus* ist also nicht als Akteur imaginiert.

Für die Verwendung des Genitivs Plural zeigt sich ein etwas anderer Befund. Bei der Verwendung des Plurals ist der Aspekt der Zählbarkeit von Normen und Handlungen entscheidend. Die *virtutes* sind Teil einer Handlungsordnung, die nicht selten als Weg oder als Aufstieg entlang der Tugenden imaginiert wird. Vor allem die Kombination mit Mengenangaben spielt darauf an. Die anderen (*aliae*) oder die übrigen (*ceterae*) *virtutes* ergeben als sprachlicher Ausdruck nur dann Sinn, wenn sie Teil einer größeren Menge und – im normativen Sinne – Teil einer Ordnung sind. Insbesondere die Gemeinschaftsspiegel von Gregor dem Großen und Smaragd von St. Mihiel weisen hohe Anteile dieser normativen Sprache auf. Das liegt sicherlich auch daran, dass ihre Werke sich an spezifische Adressatengruppen richten, die sie zu einem bestimmten Handlungsmuster ermahnen wollen. In beiden Schriften wird *virtus* nahezu ausschließlich als moralische Norm im Sinne der Tugend verwendet.

Für die Autoren der beiden frühmittelalterlichen Gemeinschaftsspiegel im Korpus ist *virtus* vor allem die Norm und damit die Handlungsmaxime, an die sich die angesprochenen Herrscher zu halten haben. In der Reduktion des Bedeutungsspektrums zeigt sich eine Veränderung in der Wahrscheinlichkeit der Sinnzuschreibung. Wo das spätantike Werk des Augustinus den sprachlichen Aufwand der Christianisierung zeigt, zeigen die frühmittelalterlichen Spiegel, dass auf Variation und Selektion eine Phase der Stabilisierung folgt.²¹ Der Aspekt der moralischen Norm im Sinne christlicher Tugend ist die dominante Bedeutung von *virtus*. Die Verwendungsmuster der anderen Aspekte werden nicht bedient oder überschrieben. Hier zeigt sich, dass bestimmte Bedeutungen eines Wortes nur in bestimmten syntaktischen Kombinationen mit anderen Wörtern zum Tragen kommen. Das Verzeichnen der Gebrauchssituationen allein reicht auch für eine semasiologisch angelegte Untersuchung nicht aus. Besonders deutlich wird das, wenn man die diskursive Integrationskraft des Begriffs betrachtet. Die Christianisierung von *virtus* ist eine Geschichte, die sich in logischer Folge erzählen ließe und die auch so erzählt worden ist. Die Texte von Gregor und Smaragd zeigen, dass sich der Impuls seit der Spätantike durchgesetzt hat. *Virtus* als christliche Tugend

20 Vgl. Schwandt, Semantik, S. 88 mit Anm. 262.

21 Für die Verwendung der Evolutionstheorie zur Beschreibung von semantischen Veränderungen vgl. Niklas Luhmann, Gesellschaftsstruktur und Semantik. Studien zur Wissenssoziologie der modernen Gesellschaft, Bd. 1, Frankfurt 1980, S. 413–594.

hat sich in die gepflegte Semantik eingeschrieben. Dass *virtus* aber ein zentrales Wort auch der Moralphilosophie bleibt, zeigt der „Policraticus“. An diesem Text fällt Folgendes auf: Der Gottesbezug von *virtus* wird erstens kaum expliziert und der philosophische Aspekt zweitens erneut aufgenommen und sprachlich hervorgehoben. Im Werk des Johannes von Salisbury wird *virtus* in erster Linie zur Bezeichnung von Handlungen und Normen verwendet. Einzeln sind diese Befunde durchaus Teil der Forschungsmeinung zu den untersuchten Fürstenspiegeln. Der Vergleich des musterhaften Sprachgebrauchs über das gesamte Korpus führt zu spannenden Nuancierungen bekannter Geschichten. *Virtus* ist eine Vokabel, die durch die beschriebene Polysemie Integrationskraft entwickelt und über Diskursgrenzen hinweg verwendet wird.

An dieser Stelle soll der Blick auf *virtus* in den Fürstenspiegeln genügen. Die Quantifizierung der Ergebnisse und ihre Gegenüberstellung als Muster ermöglicht es, sich von Vorannahmen in Bezug auf ein Wort wie *virtus* zu trennen und den Blick zu objektivieren. Dabei bleiben die Rahmenbedingungen und Vorannahmen der wissenschaftlichen Fragestellung bestehen und leiten den Blick. Stets muss der erste Impuls, den die Anwendung des Computers liefert, mit einer genauen Lektüre begleitet werden, damit die Befunde interpretiert werden können. Uns begleitet also die Frage, was die sprachlichen Beobachtungen suggerieren, das gegebenenfalls die bisherigen Interpretationen ergänzt, hinterfragt oder bestätigt.

II. Grenzen der Übertragbarkeit – Begriffe, Gattungen und Karrieren

Herausforderungen der Begriffsgeschichte entstehen mithin dort, wo Konzepte umstritten sind, oder dort, wo Begriffe angestammte Kontexte verlassen und anders verwendet werden. Übertragung meint dabei nicht notwendigerweise den Wechsel einer Sprache wie den Wechsel vom Griechischen ins Lateinische, sondern eher den Wechsel von einer Diskurssprache in eine andere.²² Daran schließen sich Fragen nach der Übersetzbarkeit von Konzepten an, die häufig an der Unterschiedlichkeit der Referenzrahmen scheitern.²³

22 Vgl. bspw. das Projekt „Übertragungswissen – Wissensübertragungen. Zur Geschichte und Aktualität des Transfers zwischen Lebens- und Geisteswissenschaften (1930/1970/2010)“, das bis 2013 am Zentrum für Literatur- und Kulturforschung von der DFG gefördert wurde, <http://www.zfl-berlin.org/projekt/uebertragungswissen.html>.

23 Mit Thomas S. Kuhn spricht die Wissenschaftsgeschichte von Inkommensurabilität. Vgl. Thomas S. Kuhn, *Commensurability, Comparability, Communicability*, in: James Conant (Hg.), *The Road since Structure. Philosophical Essays, 1970–1993*, with an Autobiographical Interview, Chicago 2002, S. 33–57.

Inzwischen kommen immer mehr Beiträge aus den Translationswissenschaften, die sich mit Übersetzungen und Übersetzbarkeit beschäftigen.²⁴

Der zweite Schritt des vorliegenden Versuchs widmet sich dem Problem der Übertragbarkeit aus einem spezifischen Blickwinkel. Was geschieht mit den Verwendungsmustern von *virtus*, wenn man sie in anderen Gattungen verfolgt? Rechtstexte, in diesem Fall Urkunden, weisen durch ihren formelhaften Sprachgebrauch häufig ein ganz eigenes Vokabular auf. Daher bieten sie sich als Vergleichskorpus an.

Für die Rechtssprache ist bereits die These formuliert worden, dass *virtus* sich formalisiert und in einer bestimmten Form dominiert: im Ablativ Singular als *virtute*. Die Verwendung dieses Kasus zusammen mit einem Genitiv entspricht der deutschen Präposition „kraft“ und ist in juristischen Kontexten – wie etwa in Dokumenten aus der englischen Peasants’ Revolt – zu finden.²⁵ Sie aktualisiert den Bedeutungsaspekt von *virtus* als Wirkmacht. Diese Wortverwendung wird auch im „*Mediae latinitatis lexicon minus*“ von Jan F. Niermeyer dokumentiert: „11. [...] (in Bezug auf ein Schriftstück) Gültigkeit, Rechtsgültigkeit“.²⁶

Wie sieht es nun mit den Verwendungsmustern in einem Vergleichskorpus wie den Urkunden aus dem Kloster Cluny, deren Edition ebenfalls als digitalisiertes Korpus verfügbar ist, aus?²⁷ Analog zu der oben beschriebenen Vorgehensweise werden die Urkunden nach der Verwendungshäufigkeit von *virtus* und ihren Kookkurrenten befragt. Dies geschieht ebenfalls unter Verwendung des eHumanities Desktop.²⁸

In den Urkunden kommt *virtus* 83 mal vor. Die häufigsten Kasus sind der Ablativ Singular und der Genitiv Plural (vgl. Abb. 2). Gemäß der Kriterien für die Bedeutungsaspekte weist diese Kasusverteilung auf zwei Dinge hin. Zum einen ist es durchaus möglich, dass die hohe Frequenz des Ablativs Singular ein Merkmal der Rechtssprache ist. Zum anderen deutet das deutliche Übergewicht der Pluralformen auf den Gebrauch von *virtus* im Sinne der

24 Vgl. zum Einstieg in aktuelle Diskussionen Lavinia Heller (Hg.), *Kultur und Übersetzung. Studien zu einem begrifflichen Verhältnis*, Bielefeld 2017.

25 So wird beispielsweise in der „Chester Indictment Roll“ vom 23. 7. 1381 die Gültigkeit der Proklamation mit der Formulierung *virtute quarum litterarum* begründet. Vgl. The National Archives, Public Records’ Office, CHES 25/8, m. 12 – 12d: „*virtute quarum litterarum proclamatio facta fuit in pleno comitatu Cestriae*“. Insbesondere die Kombination von *virtute* mit dem Genitiv Plural *litterarum* wird in den Akten immer wieder verwendet. Vgl. Schwandt, *Semantik*, S. 179 u. Anm. 548.

26 Vgl. Jan Frederik Niermeyer u. a. (Hg.), *Mediae latinitatis lexicon minus*, Leiden 2002, S. 1448, linke Spalte.

27 Vgl. Institut für Frühmittelalterforschung der Universität Münster, *cartae cluniacenses electronicae*, <http://www.uni-muenster.de/Fruehmittelalter/Projekte/Cluny/CCE/>.

28 Der eHumanities Desktop ist eine Dienstplattform, die an der Universität Frankfurt entwickelt wird. Im Rahmen des Projekts „Computational Historical Semantics“ werden dort Texte und Auswertungstools zur Verfügung gestellt. Vgl. <https://hudedesktop.hucmpute.org/desktop.jsp>.

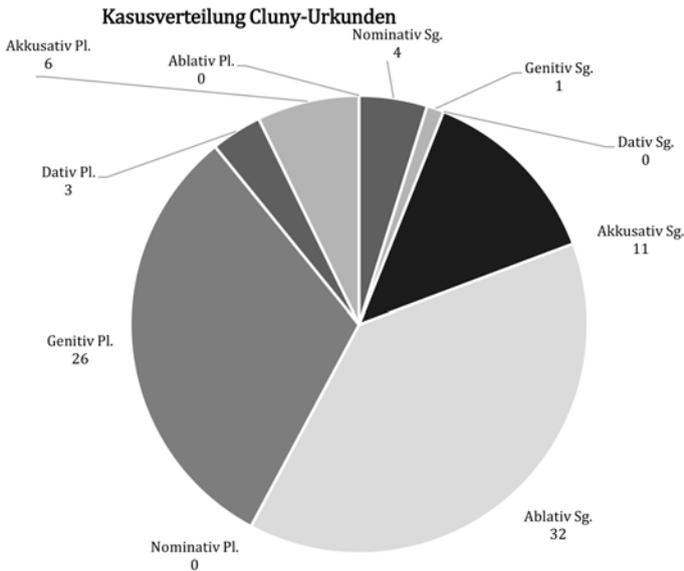


Abbildung 2: Kasusverteilung in den Urkunden von Cluny.

Tugenden hin – eine Beobachtung, die durchaus in unser Bild einer Mönchsgemeinschaft passt, die ihr Leben an christlichen Werten und Normen ausrichtet.²⁹ Ein Blick auf die Kookkurrenten von *virtus* gibt weitere Hinweise auf die Verwendungszusammenhänge. Folgt man der These, dass *virtute* mit einem Genitivattribut verbunden wird, um die Autorität eines Dokumentes oder die Kraft einer Sache auszudrücken, dann lohnt ein Blick auf die mit *virtute* kookkurrierenden Nomina (vgl. Abb. 3).

Auf den ersten Blick fallen all diejenigen Wörter auf, die Personen oder Ämter beziehungsweise Funktionen bezeichnen: *dominus, abbas, frater, prior, monachus, pater, filius, episcopus, conversus, magister, archipresbyter, successor, procurator*. Spürt man diesen Wörtern im Text nach, steht allein *dominus* einmal als Genitivattribut zu *virtute*. Die Formulierung lautet: *virtute Domini Dei*. Wiedergeben ließe sich dies in etwa als „mithilfe der Macht Gottes“. Eine zweite Gruppe von Wörtern, die in der Nähe zu *virtute* auftauchen, repräsentiert noch eher das, was die These zur Präposition „kraft“ stützen würde. Die Kookkurrenten sind: *dictum, scriptum, littera, pars*. Auch hier enttäuscht dann aber der Blick in den Text, denn auch diese Wörter werden nicht in direkter

29 Vgl. zur Geschichte des Klosters Cluny Hans-Werner Goetz, Cluny. Anfänge und Entwicklung eines mittelalterlichen Mönchordens, in: Reinhard Behrens (Hg.), Vorträge 2008–2014. Von der Gründung des Klosters Cluny 910 bis zum „Weimarer Dreieck“, Wettin-Löbejün 2015, S. 15–38.

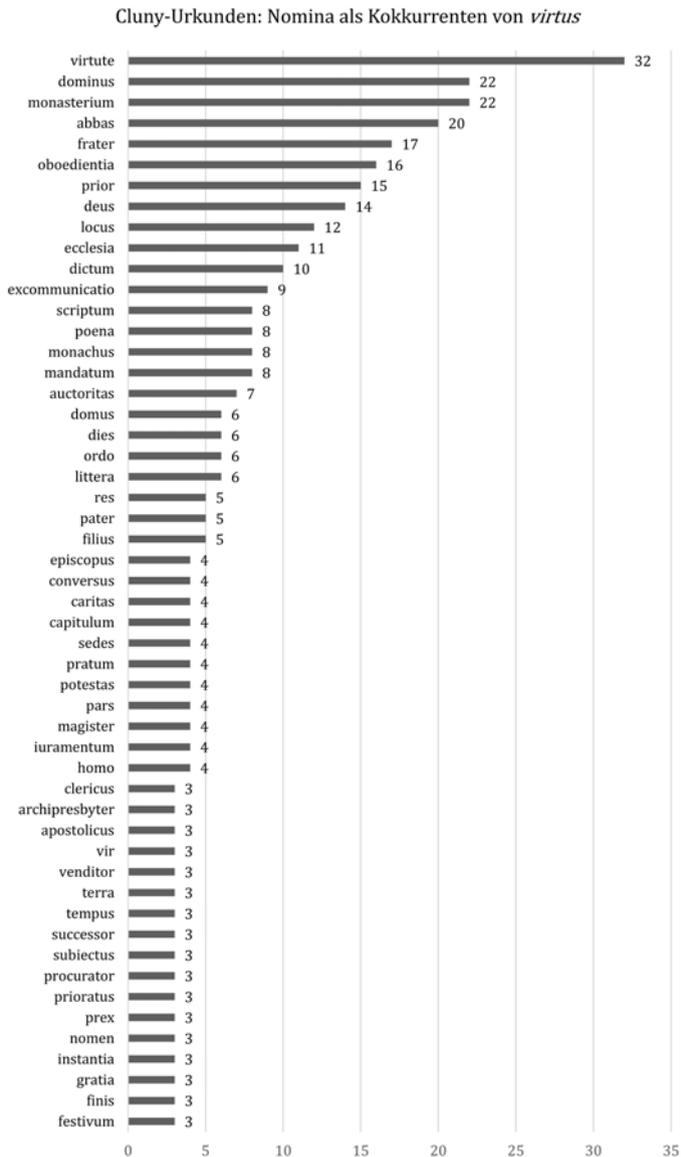


Abbildung 3: Die häufigsten Nomina als Kookkurrenten von *virtute* in den Urkunden von Cluny bis zu einer Mindesthäufigkeit von drei.

syntaktischer Abhängigkeit von *virtus* gebraucht. Ist das schon Indiz genug, um die These zu verwerfen? Es ist in jedem Fall genug, um einen Mangel im

Versuchsaufbau zu identifizieren. Bei den diesen Berechnungen zugrundeliegenden Textversionen fehlt die Möglichkeit, direkt nach den syntaktischen Beziehungen der Suchwörter untereinander zu fragen. Wieder müssen die Erwartungen der Historikerin oder des Historikers, die sich auf digitale Methoden einlassen, angepasst werden. Gerade im Umgang mit digitalisierten Korpora ist es unbedingt notwendig, genau zu prüfen, welche Auszeichnungen im Rahmen der Aufbereitung von Texten vorgenommen worden sind. In diesem Beispiel fehlt die syntaktische Annotation der Daten und diese kann daher vom Computer auch nicht abgerufen und dargestellt werden.

Wir können uns dem Problem aber nähern, indem wir den Kontext verkleinern, den wir für die Kookkurrenten festgelegt haben. Bei der Suche nach direkten Objektbeziehungen ist es auch in einer Sprache wie dem Lateinischen trotz freier Wortstellung sinnvoll, einen kleineren Kontext zu wählen, als den ganzen Satz. Die Liste der Kookkurrenten verändert sich merklich, wenn man nach denjenigen Wörtern sucht, die innerhalb von jeweils fünf Wörtern rechts und links vom Suchwort stehen (vgl. Abb. 4).

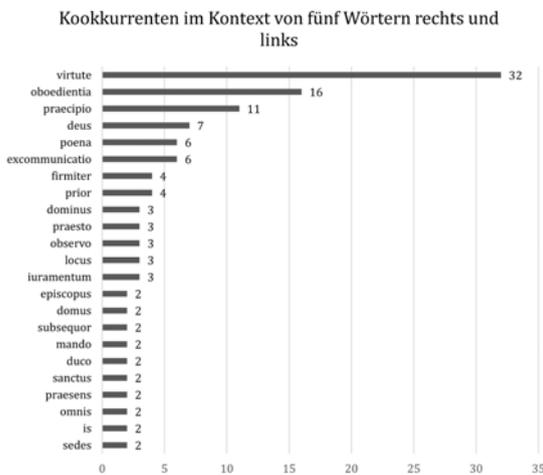


Abbildung 4: Kookkurrenten von *virtute* in den Urkunden von Cluny im Kontext von fünf Wörtern rechts und links bis zu einer Häufigkeit von zwei.

Die Beobachtung, dass signifikant viele Kookkurrenten Personen oder Funktionen bezeichnen, ist nicht mehr aufrecht zu halten. Gleiches gilt für die Wörter, die sich vermeintlich auf Dokumente oder Schriftstücke beziehen. Schaut man auch in dieser Liste zunächst nach den Nomina, dann ist *oboedientia* (Gehorsam) mit Abstand der häufigste Kookkurrent, gefolgt von *deus*, *poena* (Strafe) und *excommunicatio*. Die Veränderung des Kontextes bringt massive Veränderungen des Ergebnisses hervor, die der Forscherin

oder dem Forscher vollkommen andere Zusammenhänge suggerieren. Dabei soll nicht gegen die Methode argumentiert werden, sondern für eine genaue Kontrolle der Daten und der Korpora. Der Computer ersetzt den interpretierenden Menschen nicht, lässt ihn oder sie aber Fragen präzisieren und historische Ausdrucksformen besser verstehen.

In den Urkunden aus Cluny, die sich über mehrere Jahrhunderte erstrecken (die Urkunden stammen aus den Jahren 802 bis 1300), lässt sich die Dominanz des Sprachmusters *virtute* als „kraft“ nicht nachvollziehen. Die Kookkurrenzen suggerieren eher eine Verwendung von *virtute* als Ausdruck der Wirkmächtigkeit der Tugend des Gehorsams.

Das Korpus der Urkunden aus Cluny als Ganzes enthält 1.249.695 Wörter. *Deus* ist mit 6.835 (0,993 %) eines der häufigsten Wörter, *virtus* als Lemma (also alle Wortformen zusammen) hat eine Frequenz von 83. 29 dieser Treffer stehen allein im sechsten Band der Urkundenedition, der die Urkunden aus dem dreizehnten Jahrhundert enthält. Die oben dargestellten Okkurrenzen von *oboedentia* stehen ebenfalls alle in diesen Urkunden. Eine solche Differenzierung des Korpus zeigt erneut, wie wichtig auch die Differenzierung der Beobachtungen und ihrer Kontexte ist, um zu einer historisch belastbaren These zu gelangen. Welche qualitativen Aussagen lassen sich an diese Beobachtung knüpfen? Da es sich bei den Cluny-Urkunden um ein diachron strukturiertes Korpus handelt, könnte man fragen, ob die Formalisierung des Sprachgebrauchs, der an anderer Stelle für die Peasants' Revolt in England Ende des vierzehnten Jahrhunderts beobachtet wurde, vielleicht erst im späteren Mittelalter zu einem stabilen Muster geworden ist. Um solche Thesen aufzustellen und zu verifizieren, müsste man weitere Vergleichskorpora heranziehen. Auch Barbara Rosenwein hat bereits 1989 in ihrem Buch „To be the Neighbor of St. Peter. The Social Meaning of Cluny's Property, 909 – 1049“ die Urkunden des Klosters als sozialhistorische Quelle ausgewertet und dies mithilfe quantifizierender Methoden getan.³⁰ Sie kommt dabei zu dem Schluss, dass man durchaus aus der Wortverwendung auf die Schenkungspraxis des Klosters und darüber auf die Zusammensetzung des sozialen Netzwerks schließen kann, in welches das Kloster eingebunden war.³¹

In unserem Fall ist die These der sprachlichen Formalisierung von *virtute* als Präposition wie das deutsche „kraft“ für die Urkunden aus Cluny nicht zu halten. Es wird deutlich, dass es bei einer Wortgebrauchsgeschichte zunächst um semantische Momentaufnahmen geht, deren Übertragbarkeit und Grenzen für eine Begriffsgeschichte überprüft werden müssen.

30 Vgl. Barbara Rosenwein, *To Be the Neighbor of St. Peter. The Social Meaning of Cluny's Property, 909 – 1049*, New York 1989.

31 Vgl. ebd., S. 135 – 142.

III. Von Titeln und Themen – *virtus* als Repräsentantin eines Diskurses?

Gerade auch als politisch-soziale Grundbegriffe stehen Einzelwörter häufig für einen ganzen Diskussionszusammenhang, eine politische Kontroverse, eine bestimmte Summe von Ausdrucksformen. Als Schlüsselwörter repräsentieren sie einen Diskurs. Fragt man nach *virtus* als Repräsentantin eines Diskurses, dann lohnt es sich, den Blick von den kookkurrierenden Wörtern abzuwenden und nach einem anderen methodischen Zugang zu suchen. Es stellt sich, wie wir gesehen haben, neben der Frage nach der Korpuskontrolle auch immer wieder die Frage nach der Relevanz oder Repräsentativität eines Begriffs für einen bestimmten Diskurs oder auch nur für ein bestimmtes Thema. Mithilfe digitaler Methoden lassen sich Korpora nicht nur nach Wörtern durchsuchen. Es gibt auch Methoden, die mit dem Text oder Korpus als Ganzem arbeiten und nach ihrer Dokumentstruktur und dem Verhältnis von Wörtern untereinander fragen. Solche Methoden eignen sich beispielsweise für die Frage nach dem Stellenwert eines Begriffes innerhalb des Werks eines Autors – wie bei dem Beispiel von *virtus* als *ordo amoris* bei Augustinus. Welchen Stellenwert hat ein Wort wie *virtus* in der „Gottesstadt“ oder im Œuvre des Augustinus? Und worum geht es in den Fürstenspiegeln, die eingangs untersucht worden sind?

1. Onomasiologische Suchbewegungen

Der erste Schritt, um sich anhand der vorgestellten Methoden dieser Frage zu nähern, ist die Berechnung einer Wortliste auf der Basis von „De civitate Dei“. Auf diese Weise erkennt man, dass bei einer Gesamtlänge des Textes von 344.928 Wörtern *virtus* eine Frequenz von 297 hat. Das entspricht etwa 0,184 %.³² *Virtus* ist damit unter den 66 häufigsten Wörtern in der „Gottesstadt“. Auf das Gesamtwerk des Augustinus gerechnet hat *virtus* eine Frequenz von 3.082, das entspricht 0,106 %.³³ Diese Häufigkeiten zu deuten bleibt schwierig ohne den genaueren Kontext zu kennen.

Andere Methoden aus dem Bereich des Text Mining helfen eher die onomasiologische Frage nach den Themen eines Textes zu beantworten.³⁴ Die Webschnittstelle Voyant stellt eine Sammlung von Werkzeugen zur automatischen Verarbeitung von Texten zur Verfügung.³⁵ Insbesondere Tools zur

32 Das häufigste Wort ist *deus* mit einer Frequenz von 3.343. Das entspricht etwa 2,07 %.

33 Das Gesamtwerk bezieht sich auf die unter The Computational Historical Semantics Project, <http://comphistsem.org/home.html>, verfügbaren Werke des Augustinus.

34 Für einen Überblick über Methoden aus diesem Bereich vgl. Chris Biemann u. a. (Hg.), Text Mining. From Ontology Learning to Automated Text Processing Applications, Berlin 2014; oder Gabriel Ignatow, Text Mining. A Guidebook for the Social Sciences, Los Angeles 2016.

35 Vgl. Stéfan Sinclair u. Geoffrey Rockwell, Voyant Tools, <https://voyant-tools.org/docs/#!/guide/about>.

Berechnung und Visualisierung von Wort-Clustern oder Häufigkeiten und deren Veränderung innerhalb eines Textes oder Korpus versprechen weiterführende Beobachtungen.

So kann man beispielsweise einen Text wie „De civitate Dei“ nach Phrasen (vgl. Tab. 1) oder *topics* (vgl. Tab. 2) durchsuchen. Als Phrasen werden all diejenigen Wortkombinationen, die öfter als einmal in der gleichen Weise vorkommen, aufgerufen.³⁶

Tabelle 1: Die häufigsten Phrasen für „De civitate Dei“ bei einer maximalen Länge von zehn Wörtern

Term	Count	Length
propter vitam quae post mortem futura est	5	7
dictum est in isaac vocabitur tibi semen	4	7
cum deus erit omnia in omnibus i cor xv 28	3	10
et erit qui superaverit in domo tua veniet adorare ei	3	10
conclisit enim deus omnes in infidelitate ut omnium misereatur	3	9
est suscitavit enim mihi deus semen aliud pro abel	3	9
in semine tuo benedicentur omnes gentes gen xxii 18	3	9
a flumine aegypti usque ad flumen magnum euphratem	3	8
capillus capitis vestri non peribit luc xxi 18	3	8

Auf den ersten Blick sieht man zwei Dinge: Erstens ist die Anzahl der Wiederholungen von Phrasen relativ gering und zweitens zeigt die Form der Phrasen, dass es sich bei diesen mit großer Wahrscheinlichkeit um die Wiedergabe von Bibelziten handelt; Zitate, die sich auf diese Weise gut aufspüren lassen.³⁷ *Topics* hingegen sind Wort-Cluster, die sich aufgrund der Nähe von einzelnen Wörtern zueinander ergeben.³⁸

36 Dies., „Phrases“, Voyant Tools, <http://voyant-tools.org>.

37 Kolleginnen und Kollegen von der Universität Leipzig haben ein vergleichbares Paradigma zur Erforschung von Zitaten und Fragmenten erprobt. Vgl. Charlotte Schubert u. Markus Klank (Hg.), Das Portal eAQUA. Neue Methoden in der geisteswissenschaftlichen Forschung III, Leipzig 2012, <http://dx.doi.org/10.11588/ea.2012>.

38 Stéfán Sinclair u. Geoffrey Rockwell, „Topics“, Voyant Tools, <http://voyant-tools.org>.

Tabelle 2: Topics in „De civitate Dei“.

	Topic
1	anno caput regnum annos reges finem templum regni temporibus meum
2	dei solum hominum unus caput facere hominibus nomine vita angelis
3	peccata misericordia suas cor ira misericordiam corpore praedixit addunt corde
4	genuit noe diluvium filii nomen cain annos hominum adam civitas
5	cur potius libro satis cicero caeteris aeternam suo dicunt bene
6	mortem vita dominus suis morte mors matth omni animae coelum
7	judicium ignis sanctorum iudicio annis cor loco regnum ratio mortuos
8	homo homines deus hominem animam factum vitae carnis genus corporis
9	christi abest corpus suo christus corpora manuscriptis dicere prorsus ignem
10	omnes deum profecto deus domino vitam corporum sancti mundus peccatum
11	suis vers omnes forte manuscriptis libri multi imperium potius aliquot
12	civitas tempore populus pacem fide virtute tempora gentes codices rege

Durchsucht man „De civitate Dei“ nach diesen *topics*, dann sieht man gleich, dass an dieser Stelle die Interpretationsleistung der Forscherin oder des Forschers besonders gefragt ist. Nicht jede Summe von Wörtern repräsentiert ein sinnvolles Cluster, das sich als *topic* bezeichnen ließe. Außerdem basiert die Aufstellung in Tabelle 2 auf nicht lemmatisierten Wortformen. Trotzdem lassen sich Themen ableiten. In Zeile eins geht es um das Reich Gottes und den Tempel, Zeile vier weist beispielsweise auf biblische Gestalten als Urväter der Menschen hin. Eine Form von *virtus* findet sich in Zeile zwölf in Verbindung mit der irdischen Stadt, der *civitas*, die an die Zeitlichkeit gebunden ist. Jemand, der oder die das Werk des Augustinus kennt, erkennt auch hier die entsprechenden Themen. Wenn digitale Methoden aber mehr leisten sollen, als den Interpretationsprozess zu beschleunigen, indem sie uns mehr Daten zur Verfügung stellen, dann müssen wir bei der Auswahl der Methoden und Werkzeuge kreativer werden und die Möglichkeiten der Digitalisierung und Visualisierung ausnutzen.

2. *Zooming in and zooming out*: Visualisierungen als Provokation

Visualisierungen sind Provokationen im eigentlichen Wortsinn: Sie rufen einen Eindruck hervor, der zum Interpretieren anregt. Dies gilt vor allem für Visualisierungen, die anders funktionieren als einfache Balken- oder Kreis-

diagramme, und durch ihre Darstellungsformen spannende Fragen provozieren. Dabei ist es unbedingt notwendig, die Provokation als produktive Irritation zu verstehen – also für überraschende Beobachtungen offen zu bleiben, die bestehende Vorannahmen herausfordern können. Abbildung 5 zeigt die Verteilung der Wortformen *deus* (Linie oben) und *virtu**³⁹ (Linie unten) im Textverlauf von „De Civitate Dei“.



Abbildung 5: „Bubblelines“ zeigen die Verwendungshäufigkeit von *deus* als Wortform (798) und *virtu** (329) im Textverlauf von „De civitate Dei.“

(Vgl. dazu: Stéfán Sinclair and Geoffrey Rockwell, „Bubblelines“, Voyant Tools, <http://voyant-tools.org>.)

Auf den ersten Blick sieht man, dass sich die beiden Wörter oder Wortformen *deus* und *virtu** im Text unterschiedlich verhalten. *Deus* als die am häufigsten verwendete Wortform weist eine kontinuierlich hohe Frequenz auf mit wenigen „Lücken“. Die mit *virtu** erfassten Wortformen von *virtus* haben nicht annähernd eine ähnlich stabile Frequenz. Die Verwendungshäufigkeit schwankt. Interessant sind vor allem die Stellen, an denen *deus* zurückgeht und *virtu** eine besonders hohe Frequenz hat. Was bedeutet das für den Stellenwert des mit einem Wort verbundenen Begriffs jeweils im Verhältnis zum dominanten Wort? Blicken wir zum Vergleich auf einen anderen Text (vgl. Abb. 6).



Abbildung 6: „Bubblelines“ zeigen die Verwendungshäufigkeit von *deus* als Wortform (110) und *virtu** (434) im Textverlauf des „Policraticus“ von Johannes von Salisbury (Vgl. auch Anm. 33.)

Die Wortformen von *virtu** (untere Linie) sind im „Policraticus“ des Johannes von Salisbury deutlich häufiger als *deus* (obere Linie). Und auch wenn man alle Wortformen von *deus* hinzunehmen würde, verändert sich das Verhältnis nur marginal. Entscheidend bleibt der Eindruck, dass sich in der Entwicklung des *virtus*-Konzepts zwischen der Veröffentlichung der „Gottesstadt“ im fünften

39 *Virtu** steht aufgrund der Trunkierung für alle Wortformen von *virtus*. Der Asterix (*) repräsentiert alle möglichen Endungen bzw. Buchstabenfolgen anschließend an den Stamm „virtu“. Auf diese Weise sucht die Anwendung nach allen Deklinationsformen von *virtus*. Trunkierungen sind vor allem wichtig bei nicht lemmatisierten Wortlisten.

Jahrhundert und der des „Policraticus“ im zwölften Jahrhundert einiges verändert hat – und das bezogen auf ein zentrales semantisches Verhältnis zwischen *virtus* und der dominanten christlichen Bezugsgröße. Die Untersuchungen zu *virtus* im „Policraticus“ zeigen, dass dies nicht mit einem Bedeutungsverlust des Christentums gleichzusetzen ist, sondern auf die Etablierung des christlichen Interpretationszusammenhangs schließen lässt, der nicht mehr in gleichem Maße benannt werden muss wie zur Zeit des Augustinus.⁴⁰ Um einen solchen Verdacht zu verfohlen, bietet es sich an, die Beziehungen der Wortformen untereinander zu analysieren.

Abbildung 7 zeigt eine solche Verteilung von Wortformen, die sich aufgrund ihrer Häufigkeit (gezeigt werden die 140 häufigsten Wortformen und die Formen von *virtus*) und ihrer Nähe zueinander im Text gruppieren lassen. Diese Gruppierung beruht nicht auf der Interpretation der Forscherin oder des Forschers, sondern wird vom Computer als solche vorgeschlagen. Es lassen sich drei (im Original farbig) markierte Cluster erkennen. Links außen stehen einige Pluralformen von *deus*, verbunden mit Pluralformen von *daemon* und den Wortformen *mundi* (Welt) und *rerum* (Dinge). Rechts außen stehen *israel* und *abraham*, verbunden mit Wortformen wie *dominus*, *filius*, *gentes*, *regnum* oder *christi*. Mittig oben befinden sich vor allem Wortformen, die mit *anima*, *corpus*, *mors* und *homo* zu tun haben. In der Mitte sind auch die meisten Wortformen von *virtus* verzeichnet. Auch hier lassen sich die Cluster als Themenkomplexe lesen, deren sprachliche Repräsentationen in Wortformen durchaus in mehr als einem Cluster vorkommen können. So zeigen sich Differenzierung und Verbundenheit von Sprachmustern gleichermaßen.

Nimmt man mehr Wortformen hinzu, verändert sich der Blick auf den Text. Auch wenn im Zentrum der Grafik, dort, wo die meisten Wortformen nah beieinander stehen, nicht mehr viel zu erkennen ist, zeigt sich doch deutlich, dass einige Themenkomplexe weit nach außen wandern. Dies betrifft auf der linken Seite Wortformen, die sich mit dem Alten Testament verbinden lassen: *isaac*, *david*, *israel*, *testamentum* oder auch *jerusalem*. Sie sind verbunden mit unter anderem *carnem*, *populus*, *domus* oder *annos*. Auf der rechten Seite der Grafik, also so weit von diesem Komplex entfernt wie möglich, versammeln sich Wortformen wie *imperium*, *romana*, *bellum*, *gloriam* oder *felicitas* und *daemonum*. Hier wird mithilfe der Visualisierung der Eindruck zweier gegenüberliegender Pole einer Argumentation vermittelt. Und dieser Eindruck wird bei einer Lektüre des Textes durchaus bestätigt. Augustinus etabliert in seiner Schrift eine Gegenüberstellung zweier Städte, einer irdisch-römischen, repräsentiert durch die Wortformen links, und einer himmlisch-christlichen, repräsentiert durch die Wortformen rechts. Die Grafik bildet also erwartbare Strukturen ab. Dies kann als Beleg für die Aussagekraft des Tools gelesen

40 Vgl. Schwandt, Semantik, S. 97 f.

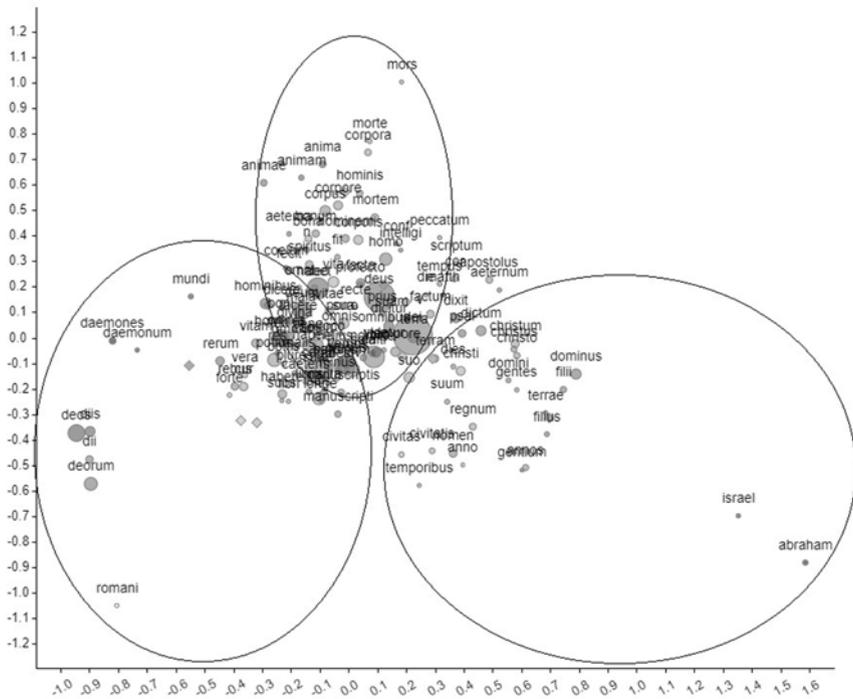


Abbildung 7: „ScatterPlot“ zeigt Cluster von Wortformen und wie sie sich über den Text verteilen.

(Vgl. dazu: Stéfán Sinclair and Geoffrey Rockwell, „ScatterPlot“, Voyant Tools (<http://voyant-tools.org>). Dort ist auch eine genauere Erläuterung zur Berechnungsgrundlage verfügbar. Dargestellt werden die Entfernungen der Wortformen untereinander, die dann übertragen werden in die Entfernung der Punkte zueinander im Koordinatensystem. Daraus ergeben sich dann Cluster von Wörtern, die im Textverlauf häufig eng beieinander stehen.)

werden; es liefert die Ergebnisse, die für die Forschungsfrage relevant sind.

Die Aussagekraft der Visualisierungen kann auch anders getestet werden. Wenn es stimmt, dass *virtus* ein Konzept repräsentiert, das sich zwischen den oben geschilderten Bedeutungsaspekten aufspannt, wie verhält es sich mit diesem Konzept in Texten, die das Wort im Titel tragen? Dem folgenden Versuch liegen als Korpus all jene Werke aus der „Patrologia Latina“ zugrunde, in deren Titel *virtus* vorkommt.⁴¹ Es handelt sich um 17 Texte aus dem fünften

41 Auch hier wurden die Versionen der Texte aus der Datenbank des Frankfurter Projekts Computational Historical Semantics, <http://comhistsem.org> verwendet. Dabei handelt es sich um die folgenden Texte: (1) Anonymus, De vitiorum virtutumque conflictu, 4. Jh. (2) Anonymus, De conflictu vitiorum et virtutum, 4./5. Jh. (3) Augustinus, De quatuor

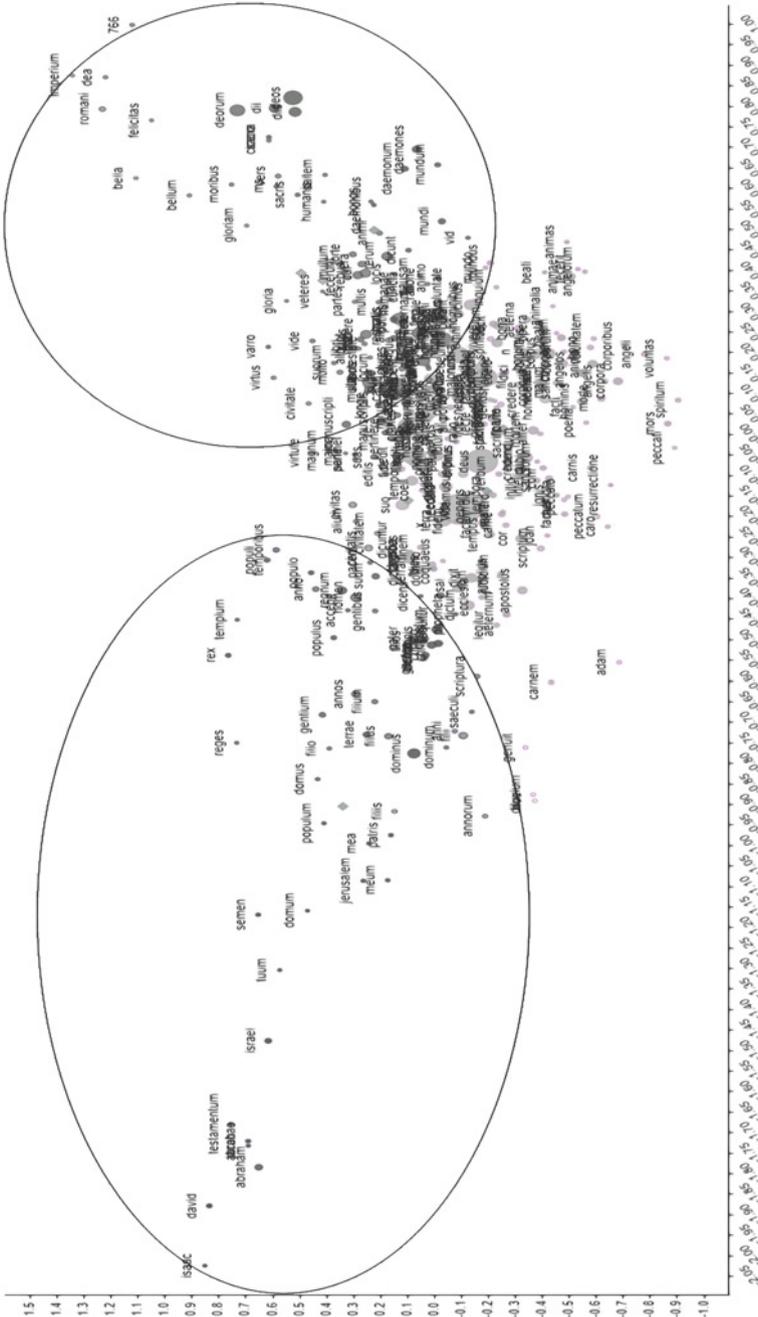


Abbildung 8: „ScatterPlot“ für die 360 häufigsten Wortformen in „De civitate Dei“.

bis fünfzehnten Jahrhundert. Laut Titel geht es in all diesen Werken um *virtus*, aber um welchen Bedeutungsaspekt? Lassen sich Unterschiede in der diachronen Perspektive ausmachen? Ändert sich das Vokabular? Die Annahme bei der Zusammenstellung des Korpus ist also, dass Texte, die das gleiche Wort im Titel haben, einander in Bezug auf Themen und Vokabular ähneln. Diese Ähnlichkeit soll mit dem Versuch überprüft werden.

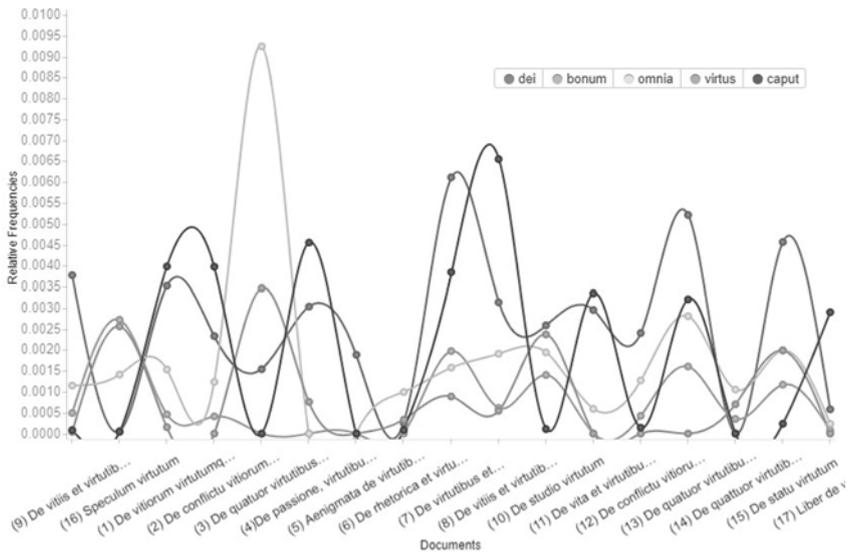


Abbildung 9: „Trends“ zeigt die Verteilung der fünf häufigsten Wortformen im Korpus. (Vgl. Stéfan Sinclair u. Geoffrey Rockwell, „Trends“, Voyant Tools, <http://voyant-tools.org>.)

Virtus als Wortform ist unter den fünf häufigsten Wortformen im Korpus (vgl. Abb. 9). Die häufigsten Wörter sind *dei*, *bonum*, *omnia*, *virtus* und *caput*. *Virtus* als Nominativ Singular wird in jedem der Texte verwendet – allerdings mit deutlichen Schwankungen. Klarer wird der Eindruck, wenn man *virtus* alleine über das Korpus verfolgt (vgl. Abb. 10).

virtutibus charitatis, 5. Jh. (4) Gregor von Tours (zugeschrieben), De passione, virtutibus et gloria Juliani Martyris, 6. Jh. (5) Bonifatius von Mainz (zugeschrieben), Aenigmata de virtutibus, 8. Jh. (6) Alkuin, De rhetorica et virtutibus, 8. Jh. (7) Alkuin, De virtutibus et vitis, 8. Jh. (8) Halitgar von Cambrai, De vitis et virtutibus, 9. Jh. (9) Hrabanus Maurus, De vitis et virtutibus, 9. Jh. (10) Adalgar, De studio virtutum, 9. Jh. (11) Meginfried von Magdeburg, De vita et virtutibus Emmerammi, 11. Jh. (12) Leo IX, De conflictu vitiorum atque virtutum, 11. Jh. (13) Eadmer von Canterbury, De quatuor virtutibus B. Mariae, 12. Jh. (14) Hildebert von Lavardin, De quatuor virtutibus vitae honestae, 12. Jh. (15) Anonymus, De statu virtutum, 12. Jh. (16) Engelbert von Admont, Speculum virtutum, 13. Jh. (17) Jean Germain, Liber de virtutibus, 15. Jh.

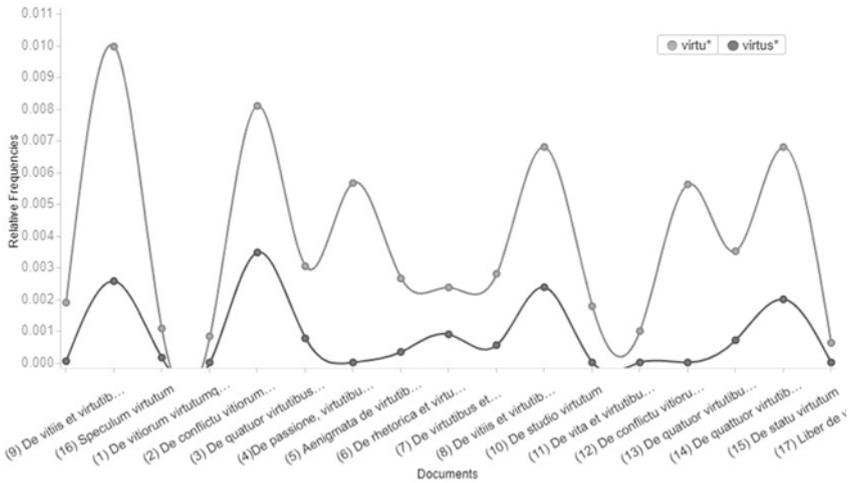


Abbildung 10: „Trends“ für die Verteilung von *virtus* als Wortform und *virtu** als Trunkierung.

Besonders gering ist die Verwendungshäufigkeit von *virtus* in dem anonymen Werk „De vitiorum virtutumque conflictu“ aus dem vierten Jahrhundert und seinem ebenfalls anonymen Pendant „De conflictu vitiorum et virtutum“. Ähnlich gering ist die Frequenz im Werk „De vita et virtutibus Emmerammi“ von Meginfried von Magdeburg aus dem elften Jahrhundert und am Ende der Reihe in Jean Germain's „Liber de virtutibus“ aus dem fünfzehnten Jahrhundert. Die beiden letztgenannten Schriften sind jeweils Bücher mit biografischem Charakter, die die Handlungen und Entscheidungen eines Menschen rühmen, während die erstgenannten eher Abhandlungen über den Kampf der Tugenden gegen die Laster sind. Solche Texte stehen in der Tradition der „Psychomachia“ des Prudentius († 405). Dass alle vier Dokumente eine geringe Frequenz für *virtus* aufweisen, mag dennoch einen ähnlichen Grund haben. Sowohl im allegorischen Kampf der Tugenden und Laster als auch in einer biografischen Evaluierung tauchen öfter noch als *virtus* selbst Wörter zur Bezeichnung der Einzeltugenden auf.

Plottet man nun das Vokabular der verschiedenen Werke mithilfe des Programms ScatterPlot sieht man eine erstaunliche Verteilung. Der Abstand der Werke zueinander repräsentiert die Ähnlichkeit ihres Vokabulars. Vergleichbar mit den Abbildungen 7 und 8 wird die Nähe der Wörter zueinander geplottet und dies dann in einer weiteren Dimension auf die damit verbundene Nähe der Texte zueinander übertragen. Die Visualisierung zeigt also, wie ähnlich das Vokabular der Texte untereinander ist. Je homogener das Vokabular, desto näher stehen die Werke beieinander. Je heterogener das Vokabular, desto größer der Abstand in der Grafik.

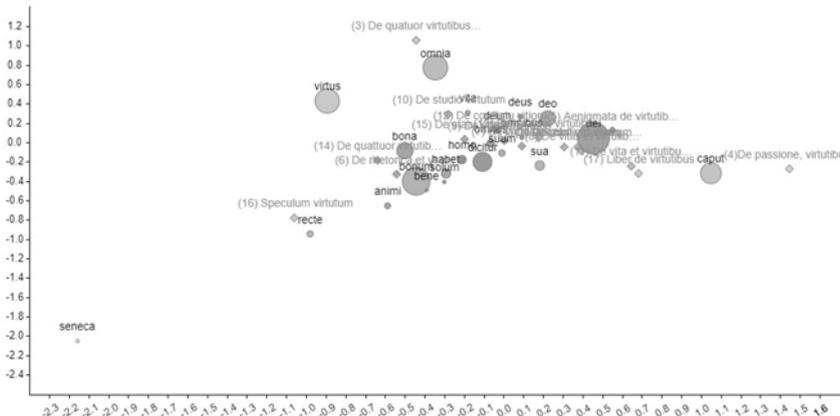


Abbildung 11: „ScatterPlot“ für die vierzig häufigsten Wortformen im Korpus.

Der Blick auf die vierzig häufigsten Wortformen im Korpus zeigt, dass alle 17 Schriften in Bezug auf ihr Vokabular relativ nah beieinander liegen (Abb. 11). Ausreißer sind hier nicht die oben genannten, sondern das „Speculum virtutum“ von Engelbert von Admont aus dem dreizehnten Jahrhundert, „De quatuor virtutibus B. Mariae“ von Eadmer von Canterbury aus dem zwölften Jahrhundert und „De passione, virtutibus et gloria Juliani Martyris“, das Gregor dem Großen und damit dem sechsten Jahrhundert zugeschrieben wird. Bis auf das „Speculum virtutum“ handelt es sich um drei hagiografische Schriften, ähnlich den oben genannten von Jean Germain und Meginfried von Magdeburg. Wenn man jetzt weiter in die Texte hineinzoomt, dann verändert sich auch die Verteilung. Dabei geht es nicht darum, so lange am Objektiv zu spielen, bis man den Ausschnitt zu sehen bekommt, den man gesucht hat, sondern darum, sich auf neue Perspektiven und damit verbundene Interpretationsspielräume einzulassen.

Erhöhen wir die Anzahl der Wortformen auf 160, rücken einzelne Werke wieder näher zusammen und auch das Vokabular selbst liegt weitaus näher beieinander. Werke, die nun weit außen stehen, sind „De passione, virtutibus et gloria Juliani Martyris“, das Gregor von Tours zugeschrieben wird (sechstes Jahrhundert), und die „Aenigmata de virtutibus“ von Bonifatius von Mainz aus dem achten Jahrhundert (Abb. 12).

Was der Vergleich zeigt, ist, dass sich das Vokabular der Texte, also die Summe der in ihnen verwendeten Wörter, und dessen Geschlossenheit unterscheiden. Das, was mit *virtus* im Titel versehen wurde, ist einander nicht immer so ähnlich, wie die Gemeinsamkeit im Titel vermuten lässt. Eine weitere Beobachtung ist, dass sich diese Unterschiede verändern, wenn man den Ausschnitt verändert, den man beobachtet. Je mehr Wörter in Betracht gezogen werden, desto weiter entfernen sich die Werke in der Grafik

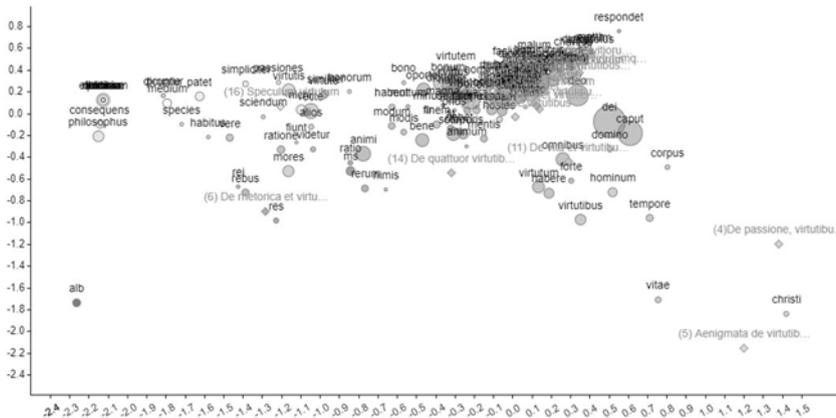


Abbildung 12: „ScatterPlot“ für die 160 häufigsten Wortformen im Korpus.

voneinander – solange, bis der Ausschnitt so groß gewählt wird, dass er die Werke wieder zusammenführt (vgl. Abb. 13).

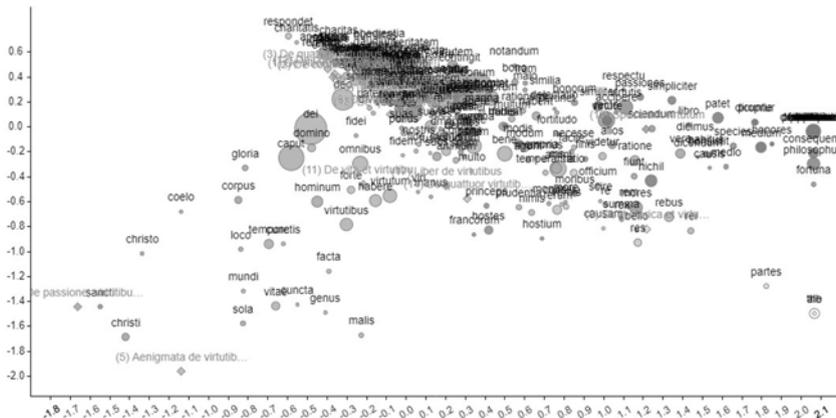


Abbildung 13: „ScatterPlot“ für die 300 häufigsten Wortformen im Korpus.

Die unterstellte Ähnlichkeit eines Textsortenkörpers wird also nicht bestätigt – oder zumindest insoweit infrage gestellt, als Titel offensichtlich nicht das entscheidende Merkmal sind. Vielleicht eignet sich ein solcher Blick auf die strukturellen Eigenschaften von Texten sogar, um Gattungsfragen neu zu stellen, wenn man sich die Geschlossenheit des Vokabulars als ein mögliches Kriterium der Ähnlichkeit vorstellt.

Es werden auf diese Weise auch Diskursausschnitte innerhalb des Korpus sichtbar. Es sind ähnliche Werke, die als Ausreißer erscheinen und sich über die Ähnlichkeit ihres Vokabulars gruppieren lassen. Doch auch hier gilt der Grundsatz, dass es der präzisen Kontrolle der Befunde bedarf, um sie in eine historische Erzählung einzubetten. Der Computer macht Vorschläge, die die Forscherin oder der Forscher einordnen muss.

IV. Die Leistung des Computers und die Erwartung der Historiker – ein Fazit

Die Digital Humanities sind wie die Informationstechnologie mit großen Erwartungen verbunden. Viele davon haben mit der Beschleunigung von Arbeitsprozessen zu tun und werden nicht immer positiv bewertet. Die voranschreitende Digitalisierung von historischen Korpora ist dafür ein gutes Beispiel: Wir wollen große Textmengen, die wir schneller durchsuchen können und die am besten von Zuhause erreichbar sind. Doch wie mit vielen methodischen Paradigmenwechseln, und die Digitalisierung gehört sicher zu diesen, besteht auch bei den Digital Humanities die Herausforderung darin, nicht bei *digital* stehenzubleiben, sondern die methodischen Neuerungen und Angebote in die *humanities* zu tragen. In gewisser Weise versteht sich dieser Aufsatz als ein Versuch, genau dies zu tun. Und am Ende bleibt die Frage: Was können wir vom Computer erwarten?

Als Begriffshistorikerinnen und -historiker sind wir interessiert an der Veränderung, am Wandel eines Begriffs oder seiner Verwendungskontexte. Wir spüren semasiologisch einzelnen sprachlichen Repräsentationen von Begriffen und Konzepten nach oder fragen onomasiologisch nach Argumentationszusammenhängen und Themen. Diese Suchbewegungen kann der Computer beschleunigen, aber auch verändern. Es wird leichter, Irrwege zu identifizieren und Brücken zu finden, weil wir uns von der semantischen Blindheit des Computers provozieren lassen können. Gleichzeitig dürfen wir nicht erwarten, dass der Computer und die verwendeten Werkzeuge neue Zusammenhänge generieren. Der Computer stellt nur Vorhandenes dar. Zahlen allein sind kein Argument, sie können aber Fragen inspirieren, sie sind Beobachtungen, noch keine Ergebnisse. Der obige Versuch hat vor allem auch die Fallstricke beim Umgang mit digitalen Methoden gezeigt. Es ist erstens unbedingt notwendig, dass man genau weiß, was man wissen will. Fragestellungen geben vor, welche Tools von Interesse sein können, und helfen dabei, diese auch weiterzuentwickeln. Zweitens ist jede Methode nur so gut, wie das Material, das ihr zugrunde liegt. Korpuskontrolle ist unerlässlich bei der Arbeit mit digitalen Textsammlungen und Auswertungswerkzeugen. Nur wenn man die Quellenbasis kontrolliert, kann man belastbare Aussagen treffen.

Am Ende bieten digitale Methoden eine Menge neuer Möglichkeiten, um Texte, aber auch andere digitale Objekte zu betrachten. Vor allem die Arbeit mit Visualisierungstools zeigt, dass ein veränderter Blick auf bekanntes Material Interpretationen herausfordern kann. Und dieser neue Blick fordert nicht nur die Begriffsgeschichte oder die Historische Semantik heraus. Er beeinflusst die Art und Weise, wie wir mit historischem Quellenmaterial umgehen und wie wir es wahrnehmen. Das Plädoyer für „distant reading“ von Franco Moretti sollte eben nicht dort, beim fernen Lesen, stehenbleiben.⁴² Es stellt lediglich einen anderen Ausgangspunkt dar für das „close reading“ der Artefakte, die unsere Korpora bilden. Im Wechsel von Vergrößerung und Verkleinerung, von *zooming in* und *zooming out* zeigt sich eine der wesentlichen methodischen Neuerungen der Digital Humanities. Digitale Methoden verändern den Blick auf Texte und die Art, wie wir Texte vergleichen, wie wir Gruppierungen bilden, Zugehörigkeiten, Gemeinsamkeiten und Unterschiede beobachten und beschreiben. Und das geht weit über den Vorteil schnelleren Arbeitens hinaus.

Dr. Silke Schwandt, Universität Bielefeld, Abteilung Geschichte,
Universitätsstraße 25, 33615 Bielefeld
E-Mail: silke.schwandt@uni-bielefeld.de

42 Vgl. Franco Moretti, *Distant Reading*, Konstanz 2016.