Dokumentenarbeit mit hierarchisch strukturierten Texten: Eine historisch vergleichende Analyse von Verfassungen

Knoth, Alexander
alexander.knoth@uni-potsdam.de
Universität Potsdam, Deutschland

Stede, Manfred
stede@uni-potsdam.de
Universität Potsdam, Deutschland

Hägert, Erik
haegert@uni-potsdam.de
Universität Potsdam, Deutschland

1. Einleitung: Verfassungsvergleich als Spiegel staatlichen Wandels?

Staatlich verfasste Gesellschaften sind komplex und differenziert (Mayntz 1997; Schimank 1999). Will man etwas über die „Identität“ von Staaten und deren Wandel erfahren, dann eignen sich Verfassungen, da diese spezfische Dokumentensorte soziologisch als kodifizietre Selbstbeschreibung von Gesellschaften verstanden werden kann (Boli-Bennett und Meyer 1978; Go 2003; Heintz und Schnabel 2006; Boli-Bennett 1979). Moderne Staaten produzieren aber nicht nur Unmengen an amtlichen Dokumenten, sondern sind darüber hinaus durch ihre konstitutionelle sowie rechtsstaatlich-bürokratische Verfasstheit grundsätzlich textlich strukturiert (Weber 1972). Für die historsiche Dokumentenanalyse spielen die Erstellung des Korpus und die Auswahl der Untersuchungsmethoden wichtige Rollen, um sowohl die Textlichkeit, als auch den Kontext angemessen zu berücksichtigen.

In diesem Vorhaben wird mit Verfassungsdokumenten europäischer Staaten gearbeitet, anhand derer staatlicher Wandel von der ersten Verfassungsgebung bis heute sichtbar gemacht wird. Verfassungen beinhalten u.a. Vorstellungen darüber, wie die Gesellschaft beschaffen ist. Konkret wird der historisch-wissenssoziologischen Frage nach der sozialen Konstruktion des (Staats-)Bürgers nachgegangen. Denn historisch betrachtet reflektieren Verfassungen den sukzessiven Umbau von ständisch stratifizierten hin zu souveränen Bürgergesellschaften und damit reflektieren sie ebenso den Wandel des gesellschaftlichen Personals, das es in Form von Personenkategorien und Zugehörigkeitsdimensionen aus dem Material herauszuarbeiten gilt. Was aber genau unter „Verfassung“ verstanden wurde, wie sich die Staaten und ihr ‚Personal’ über dieses Dokument selbst beschreiben und welches Wissen in selbiges eingeht, variiert erheblich (Gosewinkel, Masing und Würschinger 2006; Vorländer 2007). Daher bedarf es eines geeignetne methodisch-analytischen Instrumentariums, um die aufgeworfene Fragen zu beantworten.

2. Dokumentenanalyse im Schnittfeld von historischer Soziologie und Computerlinguistik

Um Verfassungen strukturell und inhaltlich untersuchen zu können, werden Ansätze der historischen Wissenssoziologie (Thelen 1999, 2002; Jepperson 1991) und der Computerlinguistik (z.B. Hausser 2014; Lobin 2010) miteinander verschränkt. Die Entwicklung dieses methodischen Werkzeugs zur „Dokumentenarbeit“ umfasst Verfahrensschritte der Datenerhebung, -aufbereitung und -auswertung, wobei hier vor allem auf methodologische Herausforderungen, d.h. die Korpuserstellung und die semi-automatische Analyse von Dokumentenstrukturen eingegangen wird.

Rechtstexte im Allgemeinen und Verfassungen im Besonderen, weisen eine Dokumentenlogik auf, die stark durch eine formale hierarchische Struktur gekennzeichnet ist. Bei dieser Dokumentenart ist daher davon auszugehen, dass der Struktur eine besonders sinnstiftende Bedeutung zukommt, die es vor allem bei vergleichenden Untersuchungen (synchron wie auch diachron) zu berücksichtigen gilt. Insofern sollte ein computerlinguistisches Verfahren die Strukturinformationen bspw. in welche (Sinn-)Abschnitte sich ein Dokument gliedert für den Vergleich nutzen. Diese Strukturauswertungen können dann wiederum mit statistischen Häufigkeits- und Ähnlichkeitsberechnungen von Worten innerhalb von Fließtexten – wie das u.a. die gängigen Vektorraummodelle (Manning, Raghavan und Schütze 2008; Salton, Wong und Yang 1975) oder insbesondere die derzeit populären „word embedding“ Modell (z.B. Mikolov et al. 2013) machen – kombiniert werden.

Bei der hierarchischen Struktur von Dokumenten anzusetzen bietet einen klaren Ausgangspunkt für die systematische Analyse von großen Textmengen und stiftet zugleich Orientierung im Feld der inhaltsanalytischen Methoden (Kuckartz 2012; Mayring 2015). Diese unterscheiden sich vor allem in Bezug auf ihre Anlage, d.h. entweder Häufigkeiten zählende oder hermeneutisch interpretierende Ausrichtung, und firmieren in den Sozialwissenschaften oftmals unter dem Label „Dokumentenanalyse“. Zwar verbindet alle diese Ansätze,

dass sie sich durch eine ständige Korrespondenz von Forschungsfrage und Arbeit am Material auszeichnen und in der Regel mehrere Iterationen durchlaufen, bevor valide Ergebnisse vorliegen. Dennoch bringt vornehmlich die manuelle Bearbeitung von umfangreichen Textmengen, etwa in Form von Kodier- und Kategorisierschritten der *Grounded Theory* (Strauss und Corbin 1996), Probleme der methodisch kontrollierten Auswertung und damit der Reliabilität der Ergebnisse mit sich.

Außerdem lassen sich über den strukturellen Zugang Fragen erschließen, die über den „reinen" Inhalt hinausgehen, und die die Verwendung wie auch die Art und Weise in den Vordergrund rücken, in der Verfassungen im Zeitverlauf politisch unter Druck geraten, sich also aufgrund wechselnder politischer Machtverhältnisse wandeln. Damit werden die Relation von Dokument und (Entstehungs-)Kontext und besonders die Verfasser von Dokumenten und deren Konstruktion sozialer Wirklichkeit durch die schriftliche Fixierung gesellschaftlichen Wissens (Prior 2011) fokussiert.

Von der Dokumentenstruktur auszugehen heißt, zunächst methodologisch zu fragen, inwieweit sich Dokumente formal wie auch inhaltlich ähneln. Das setzt wiederum voraus, dass sich Dokumente überhaupt vergleichen lassen und so einer Analyse etwaiger struktureller Ähnlichkeiten und spezifischer Unterschiede allererst zugänglich gemacht werden. Hier bieten computergestützte Verfahren einen produktiven Ausgangspunkt, um einerseits bestehende Methoden zu reflektieren und andererseits ein dann auch verallgemeinerbares Werkzeug zur Dokumentenanalyse von Rechtstexten zu entwickeln. Zur Beantwortung der formulierten Frage wird eine innovative, von uns eigens entwickelte Software vorgestellt, mit der sich Verfassungen in ihrer historischen Entwicklung vergleichen lassen. Hierdurch werden Impulse zur Generierung neuer methodischer Ansätze gegeben werden.

3. Arbeitsschritte: Vom Download zur Analysesoftware

Um mit der Auswertung der Dokumente beginnen zu können, muss das Korpus erstellt werden. Ein normales PDF beinhaltet in der Regel kaum explizite Strukturinformationen, lediglich einzelne Worte ließen sich automatisch erfassen, nicht aber die zugrunde liegenden Strukturen abbilden. Hierfür bedürfte es bspw. der Kennzeichnung von Überschriften, Absätzen oder inhaltlich unterscheidbaren Abschnitten. Das Dokument muss also mit weiteren Informationen in seiner Struktur beschrieben werden.

Das konkrete methodische Vorgehen, das hier als „Dokumentenarbeit" zur Korpuserstellung bezeichnet wird, gliedert sich in drei Schritte: das Zusammenstellen der Ausgangsdaten als HTML-Dateien, die Transformation der Ausgangsdaten in das XML-Format sowie die eigentliche Auszeichnung der Ausgangsdaten mit Metadaten zur Modellierung der Struktur.

Entgegen der Annahme, dass derart politisch relevante Dokumente wie Verfassungen als elektronische Ausgangsdaten vorliegen sollten, müssen diese zunächst hergestellt und in ein bearbeitbares Datenformat transformiert werden. Zwar finden sich aktuelle Verfassungen als elektronisch veröffentlichte Ressourcen bspw. in den Rechtsdatenbanken und -portalen der jeweiligen Staaten, im deutschen Fall bspw. „Juris".Es existiert jedoch kein lückenloser, chronologischer Verlauf, aus dem sich alle Änderungen computergestützt entnehmen ließen.

Auf der Seite www.verfassungen.org lassen sich die meisten Verfassungen online (auf Deutsch) abrufen und downloaden. Zudem beinhalten die dortigen Dokumente farblich abgesetzte Änderungen in Textform und nicht etwa als Kommentar oder gesonderte Liste sowie jeweils Totalrevisionen als separate Dokumente. Diese Dokumente werden dann mit offiziell veröffentlichen Verfassungen abgeglichen, um gegebenenfalls inhaltliche Fehler aufzuspüren und zu beheben. Anschließend werden die Daten manuell bereinigt und standardisiert, d.h. nicht benötigte Beschreibungen der Autor*innen der Webseiten oder andere irrelevante Informationen werden entfernt. Diese Dokumente bilden sodann die Grundlage für das Korpus. Die einzelnen Verfassungen beinhalten eine Fülle an textlichen Ergänzungen, Streichungen und anderweitigen textlichen Veränderungen, die einerseits schwer zu identifizieren sind und andererseits nicht chronologisch sortiert, sondern der Texthierarchie folgend vorliegen. Aus diesen Gründen wird zuerst eine Ausgangsverfassung des Jahres 2011, also dem Ende des Untersuchungszeitraums, erstellt, um ausgehend davon jede weitere Änderung als eigenständige, nicht offizielle, „Phantom-Verfassung" zu rekonstruieren. Durch diesen iterativen, historischen Rekonstruktionsprozess wird schließlich die Datengrundlage geschaffen.

Die Verfassungen liegen zunächst als HTML-Dokumente, mit einer sehr flachen Dokumentenstruktur vor. Die zu entwickelnde Analysesoftware benötigt jedoch ein Datenformat, das Metadaten mit Dokumentendaten assoziieren kann. Dem aktuellen Entwicklungsstand entsprechend verwenden wir ein XML-Format (Bubenhofer und Scharloth, 2015).

Deshalb wird im nächsten Schritt der HTML-Code mittels eines XSL-Skripts (Extensible Stylesheet Language) in das technische XML Format (vgl. XML Schema 2001; XQuery 2002; XSLT 1999) überführt und formatiert. Bei XSL bzw. XSLT handelt es sich um eine Programmiersprache zur Transformation (und Formatierung) von XML Derivaten – in unserem Fall die HTML-Dateien – in XML Dokumente. Das XML Format eignet sich in erster Linie dafür, die informationsarmen Ausgangsdaten mit Metadaten (z.B. Attribute, Codes oder Variablen) zur systematischen Beschreibung der Strukturen und der Inhalte anzureichern. Bspw. ließe sich das Ausgangsdatum „Herbert" mit dem Attribut

„Vorname“ verknüpfen und so systematisch alle Vornamen erschließen.

Die Umwandlung von HTML zu XML ist die Grundlage des Mappings der einzelnen Versionen auf einander. Hierfür wurde kein existierender Standard verwendet, sondern das Format so entwickelt, dass es die Struktur der Texte möglichst treu abbildet. Dafür sollen möglichst wenige Elemente verwendet und unnötig tiefe Einbettungen vermieden werden.

Jeder Version wird ein Vorspann vorangestellt (<front>) der den Titel (<docTitle>) und das Datum der jeweiligen Version (<docEdition>) enthält. Diesem Vorfeld folgt der eigentliche Text der Verfassung (<body>). Dieser ist zunächst in Hauptteile gegliedert (<div n="" type="v-teil”>). Diese können wiederum aus Sektionen (<div n="" type="sektion”>) bestehen, welche die Artikel ( ) der Verfassung enthalten. Die Artikel setzen sich aus Sätzen ( ) und gegebenenfalls auch aus listenartigen Aufzählungen (<div n="" type="aufzaehlung”>) zusammen. Letztere bestehen aus einer Reihe von Listenelementen ( ).

Schwesterknoten werden mit eins anfangend durchnummeriert (n). Ein Hauptteil mit n="0" ist eine Präambel. Diese enthalten keine Artikel, sondern eine Reihe von Sätzen ( ) und gegebenenfalls Aufzählungen. Hauptteile, Sektionen und Artikel weisen jeweils ein Element für ihre Überschriften auf. Darüberhinaus enthalten nur Paragraphenelemente ( ) Text. Eine Validierung gegen eine DTD findet nicht statt.

Es ergibt sich folgendes Format:

```
<text>
<front>
<titlePage>
<docTitle></docTitle>
<docEdition></docEdition>
</titlePage>
</front>
<body>
<div n="0" type="v-teil">
<head></head>
<p n="1" type="praeamb-satz"></p>
<div n="1" type="aufzaehlung">
<p n="1" type="aufzaehlung-item"></p>
</div>
<div>
<div n="1" type="v-teil">
<head></head>
<div n="1" type="artikel">
<head></head>
<p n="1" type="satz"></p>
<div n="1" type="aufzaehlung">
<p n="1" type="aufzaehlung-item"></p>
</div>
</div>
<div n="1" type="sektion">
<div n="1" type="artikel">
<head></head>
<p n="1" type="satz"></p>
</div>
</div>
</div>
<body>
<text>
```

An dieser Stelle der Beschreibung der Ausgangsdaten im XML Format setzen wiederum qualitative Dokumentenarbeitsschritte ein, die sich an der analytischen Strategie des Kodierens und Kategorisierens anlehnen. In diesen Vorgang fließen einerseits Kontextinformationen ein, andererseits werden während des Arbeitsschritts wichtige empirische Beobachtungen gemacht, die in Form von Kodiermemos dokumentiert werden. So können die gewonnenen Informationen zu einem späteren Zeitpunkt für die Tiefenstrukturanalyse des Materials oder die Ausdifferenzierung der Analysesoftware genutzt werden.

Mapping als Strukturvergleich

Das Mapping der Strukturelemente der im Vergleich stehenden Versionen aufeinander wird automatisiert vollzogen, indem jedes Element einer strukturellen Ebene (Hauptteil, Sektion, Artikel) mit jeder anderen entsprechenden Ebene der Vergleichsversion abgeglichen wird. Dafür verwenden wir das gängige Cosinus-Maß, das TextÄhnlichkeit durch Modellierung im hochdimensionalen Vektorraum misst. Auf diese Weise wird eine Matrix von Ähnlichkeitswerten aufgebaut.

Währen des Aufbaus der Matrix wird die Anzahl der Berechnungen reduziert, indem, sobald eine Ähnlichkeit vom Wert 1 zu einem Element der Vergleichsversion gefunden wird, also eine perfekte Übereinstimmung vorliegt, das Elementpaar als unveränderte Übereinstimmung abgespeichert und die Elemente aus dem weiteren Vergleich ausgeschlossen werden.

Liegt keine genaue Übereinstimmung vor, wird getestet, ob die beiden zu vergleichenden Texte unterschiedlicher Länge sind. Ist dies der Fall, wird ferner geprüft, ob der kürzere der beiden Texte einen Teilstring des längeren bildet. In solchen Fällen werden die Texte einander als Änderungen (Erweiterungen oder Kürzungen) zugeordnet, abgespeichert und ebenfalls aus der weiteren Berechnung ausgeschlossen.

Schließich bleibt eine Matrix der Ähnlichkeitswerte ausschließlich der Elemente übrig, für die keine Entsprechung gefunden werden konnte.

Die Paare, die die höchsten Cosinusähnlichkeiten (< 1) aufweisen, werden als Änderungen abgespeichert. Die hiernach übrig bleibenden Elemente, denen kein Element der Vergleichsversion zugeordnet werden konnte, sind entweder Tilgungen (keine Entsprechung in der zeitlich späteren Version) oder Hinzufügungen (keine Entsprechung in der früheren Version). Abbildung 1 illustriert den Prozess in tabellarischer Form.

1.

	Art.1	Art2	Art.3	Art.4	Art.5
Art.1	1.0	–	–	–	–
Art.2	–	–	–	–	–
Art.3	–	–	–	–	–
Art.4	–	–	–	–	–

$Art.1 \Rightarrow Art.1$

2.

	Art2	Art.3	Art.4	Art.5
Art.2	0.1	–	–	–
Art.3	–	–	–	–
Art.4	–	–	–	–

$Art.2 \Rightarrow Art.2$

3.

	Art.3	Art.4	Art.5
Art.3	0.8	0.4	0.6
Art.4	0.3	0.7	0.9

4.

$Art.3 \Rightarrow Art.3$
$Art.4 \Rightarrow Art.5$
Neu: Art.4

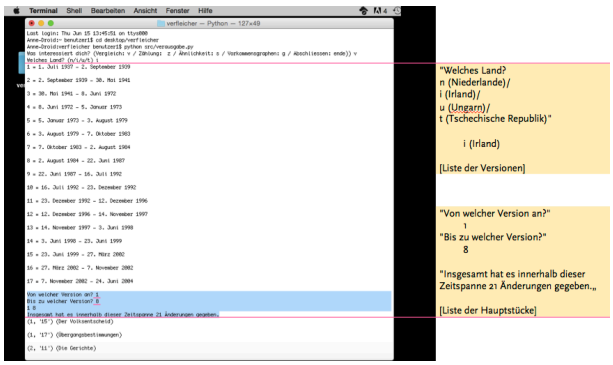
Entwicklungsergebnisse

Die Software zur Verfassungsanalyse ist in der Programmiersprache *Python* geschrieben. Im Zuge der ersten Entwicklungsphase lassen sich rein formal die hierarchische Struktur und die jeweiligen Abschnittslängen der Dokumente vergleichen und auch quantifizieren.

In der vorliegenden Fassung des Werkzeugs können vier verschiedene Operationen ausgeführt werden:

1. Vergleichen

Für den Vergleich wird zunächst ein Land und ein Zeitraum ausgewählt für den die Änderungen ausgegeben werden sollen. Um die Suche weiter einzuschränken, wird zunächst gezeigt, wie viele Änderungen in welchen Hauptstücken in dem angegebenen Zeitraum stattgefunden haben. Nach der Auswahl eines Hauptteils werden die gefundenen Änderungen (in Sektionen und Artikeln), Tilgungen und Hinzufügungen ausgegeben (Abbildung 2).

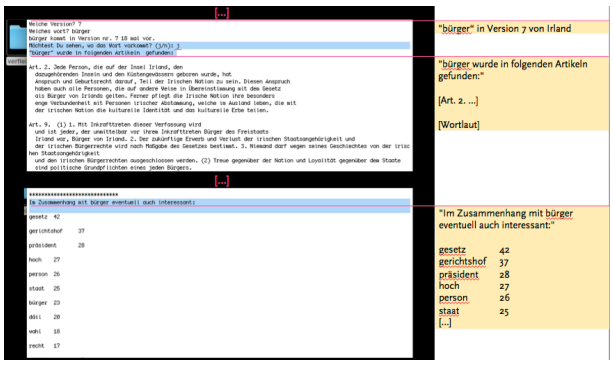


2. Cosinusähnlichkeiten

Mit dieser Funktion (Abbildung 3) lassen sich die Cosinusähnlichkeiten ganzer Versionstexte untereinander berechnen und ausgeben.

3. Wortzählungen und Wortprofile

Der Nutzer kann sich unter Angabe der Version, die von Interesse ist, die Aufkommenshäufigkeiten von Wörtern ausgeben lassen. Zudem lassen sich die Textstellen, die das gezählte Wort enthalten, zusammen mit einer Liste der Wörter ausgeben, die häufiger als ein definierter Schwellenwert (bspw. fünf Mal) in derselben textuellen Umgebung vorkommen (Abbildung 4).



Das Beispiel zeigt die politische Kernkategorie, den Bürger, in Version 7 der irländischen Verfassung. Die textuelle Umgebung ist durch Begriffe wie Gesetz (42 mal), Gerichtshof (37 mal) und Präsident (28 mal), die Hinweise dazu liefern in welchen Sinnzusammenhängen der Bürger theamtisiert wird, gekennzeichnet. Die Begriffe Person (26 mal) und Staat (25 mal) weisen darauf hin, dass es sich beim Bürger offenbar tatsächlich um eine Kategorisierung als Person handelt, die wiederum in irgendeiner Beziehung zum Staat steht. Dieser Zusammenhang, die Beziehung von Bürger und Staat kann nun mithilfe von (historischen) Kontextrecherchen und Literatur basierten Konzepten und Theorien genauer untersucht werden.

Zur Erstellung der Wortprofile, d.h. für die Anreicherung der Daten mit bspw. Lemmata, POS-Taggs und Dependenzrelationen wurde die Pipeline des DARIAH-DKPro-Wrapper [1] des NLP-Toolkits DKPro Core (vgl.

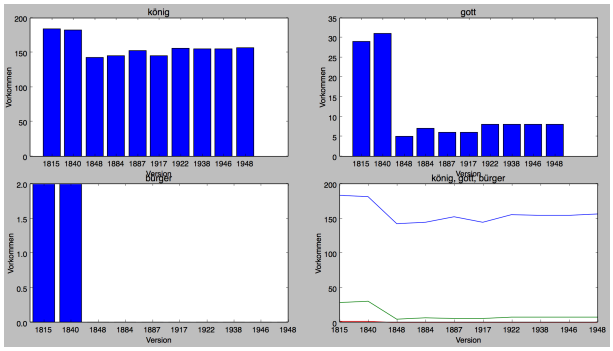
Eckart de Castilho und Gurevych (2014)) benutzt. Das ermöglicht die Ausgabe der syntaktischen Relationen, die das gezählte Wort mit anderen Wörtern eingeht, zusammen mit ihren Häufigkeiten (Abbildung 5).

```
Welche Version? 1
Welches wort? mitglied
mitglied kommt in Version nr. 1 78 mal vor.
SUBJEKT:
'mitglied' kommt 2 mal als Subjekt von 'stimmen' vor.

'mitglied' kommt 2 mal als Subjekt von 'empfangen' vor.

'mitglied' kommt 1 mal als Subjekt von 'stehen' vor.

'mitglied' kommt 1 mal als Subjekt von 'reichen' vor.

'mitglied' kommt 1 mal als Subjekt von 'nehmen' vor.

'mitglied' kommt 1 mal als Subjekt von 'legen' vor.

'mitglied' kommt 1 mal als Subjekt von 'erhalten' vor.

'mitglied' kommt 1 mal als Subjekt von 'bekleiden' vor.

GENITIVATTRIBUT:
'mitglied' kommt 2 mal als Genitivattribut von 'drittel' vor.

'mitglied' kommt 1 mal als Genitivattribut von 'wahl' vor.

'mitglied' kommt 1 mal als Genitivattribut von 'vollmacht' vor.

'mitglied' kommt 1 mal als Genitivattribut von 'teil' vor.

'mitglied' kommt 1 mal als Genitivattribut von 'pension' vor.

'mitglied' kommt 1 mal als Genitivattribut von 'mehrheit' vor.

'mitglied' kommt 1 mal als Genitivattribut von 'hälfte' vor.

'mitglied' kommt 1 mal als Genitivattribut von 'gesamtzahl' vor.

AKKUSATIVOBJEKT:
'mitglied' kommt 1 mal als Akkusativobjekt von 'hinzufügen' vor.

'mitglied' kommt 1 mal als Akkusativobjekt von 'ernennen' vor.
```

Das verwendete Beispiel, die Personenkategorie Mitglied, kommt in der gewählten Verfassungsversion 78 mal vor. Die Informationen, wovon Mitglied das Subjekt ist oder inwiefern Mitglied als Genitivattribut oder Akkusativobjekt von bestimmten Termini vorkommt können u.a. dabei helfen, die Eigenschaften dieser Kategorie oder auch die Prozesse in die diese kategorie eingebunden sein kann, genauer zu bestimmen. So kann ein Mitglied hinzugefügt oder auch ernannt werden. In weiteren Betrachtungen kann dann herausgearbeitet werden, wozu ein Mitglied ernannt oder hinzugefügt werden kann. Das Genitivattribut der Kategorie König gibt bspw. Auskunft darüber, von welcher Bezugsgruppe diese Person überhaupt der König ist. Diese automatisiert verfügbaren Informationen tragen dazu bei, die kategoriale Wissensbestände der Verfassungsstaaten historisch-vergleichend zu untersuchen.

4. Graphische Darstellungen

Nutzende können sich in der aktuellen Version die Auftretensverteilungen von Wörtern in einem anzugebenden Zeitraum als Graph ausgeben lassen. Dabei werden pro eingegebenes Wort ein Balkendiagramm sowie ein Diagramm generiert, das die Kurven aller angegebenen Wörter in einem Graph zugleich darstellt (Abbildung 6).



In dem verwendeten Beispiel werden die Verkommenshäufkeiten von drei Personenkategorien – Gott, König und Bürger – in den Verfassungen der Niederlande für den Zeitraum 1815 bis 1948 dargestellt. Dabei fällt auf, dass der Bürger im Vergleich zum König als Repräsentation des Souveräns eine deutlich untergeordnete Rolle spielt und ab 1848 bis 1948 im Prinzip nicht vorkommt. Die Verfassungen spiegeln das politische System der Monarchie und nicht die Staatsbürgergesellschaft wieder. Ab 1848 nimmt auch das Vorkommen der Kategorie Gott signifikant ab. Während 1840 Gott noch 30 Mal vorkommt, verringert sich die Häufigkeit in den darauffolgenden 100 Jahren auf durchscnittlich sieben. Anahnd dieser Ergebnisse lassen sich ganz verschiedene Interpretationen und tiefergehende Analysen anschließen, um bspw. Erkenntniss über die religiöse Semantiken staatlicher Selbstbeschreibung (Gottesbezog, Gründungsmythen, Vorstellungen der Nation usw.) oder den Wandel des politischen Gemeinwesens und politischer Zugehörigkeit (wer gehört eigentlich dazu?) zu erlangen.

Diese Funktion wird demnächst um weitere bereichert werden, um die Potentiale von Visualisierungen als darstellende Klammer des Strukturvergleichs, der Wortsuche und der syntaktischen Wortrelationen wie auch als analytisches Werkzeug (Lupton 2014) selbst zu sondieren.

4. Zusammenfassung und Ausblick

Derzeit kann die Software alle Änderungen – unabhängig von formalen Totalrevisionen innerhalb der historischen Verfassungsentwicklungen aufzeigen. So lässt sich bspw. feststellen, welche Teile besonders häufig geändert werden oder welche Teile bis heute unangetastet geblieben sind.

Die Vorteile dieser Software gegenüber frei im Internet zugänglichen Versionierungstools (github, gitlab o.ä.) liegen auf der Hand: Zwar bieten solche Programme relativ einfach die Möglichkeit Textänderungen nachzuverfolgen, das gezielte Nachvollziehen von der Änderungshistorie spezifischer Textabschnitte ist ungleich schwieriger. Darüber hinaus bieten die beschrieben Funktionalitäten viel weitgehendere Auswertungsszenarien, als das reine

Mapping, das durch ein Versionierungstool angeboten wird.

Beispielsweise kann mit dem Programm untersucht werden, welche neuen (normativen) Vorgaben Einzug in die Verfassung finden. Diese Änderungen in den Zeitreihen lassen sich dann ihrerseits im Zuge der weitergehenden Untersuchung historisch kontextualisieren und bspw. mit Blick auf staatlichen Wandel oder die Institutionalisierung bzw. Legitimierung neuer Werte, Normen, staatlicher Handlungsverpflichtungen und kultureller Leitideen (Meyer et al. 2005) interpretieren. Das Programm stellt das technische Werkzeug dafür dar, beide Ebenen, Mikro- und Makroebene, gleichermaßen zu betrachten, indem die Änderungshistorie einzelner Abschnitte ins Verhältnis zur Distanzsicht auf die gesamttextlichen Änderungen vieler Verfassungen gesetzt werden können.

Künftig sollen auch Fallvergleiche zwischen verschiedenen europäischen Staaten möglich sein, bspw. indem für frei wählbare Textbereiche die Cosinusähnlichkeit berechnet wird. Dadurch wird u.a. die Herausforderung des Vergleichs verschiedener historischer Kontexte tangiert. Wie kann ein informationstheoretisches Modell aussehen, das verschiedene Vektorräume, die definierte temporäre Sequenzen umfassen, zusammenbringt und miteinander vergleicht? Wie können Okkurrenzen kategorisiert werden, wenn diese bspw. häufiger auftreten?

Die dargestellte Form der Dokumentenarbeit macht Verfassungen nicht nur einer breiten Öffentlichkeit und vielfältigen wissenschaftlichen Erhebungen zugänglich. Vielmehr reflektiert sie Methoden der Dokumentenanalyse, indem sie der spezifischen Dokumentengattung „Verfassung“ besondere Aufmerksamkeit schenkt. Hermeneutisch-interpretative Verfahren versuchen Kontextwissen zumindest zu Beginn der Analyse weitestgehend auszublenden, wohingegen beim skizzierten Vorgehen eben dieses Wissen über die Dokumentenart, deren struktureller Aufbau sowie etwaige kulturell-historische Besonderheiten in die Auszeichung des Textes mit Metadaten für die computerbasierte Bearbeitung einfließt.

Insgesamt leistet die methodische Verschränkung von historischer Wissenssoziologie und Computerlinguistik als Dokumentenarbeit und Entwicklung einer Analysesoftware einen Beitrag zur Untersuchung der Ko-Fabrikation von Sprache und Verfassungsrecht in Europa, indem über einzelene Begriffe und Begriffskombinationen spezifische Wissensbestände und Semantiken in den Blick genommen werden können. Dieses Vorgehen kann dazu beitragen, neue Textkorpora zu erschließen und weitere gesellschaftliche Wissensbestände (bspw. Bibelversionen, Dramen usw.) zu erkunden. Diese Analysen ließen sich mit anderen methodisch ähnlichen, aber gegenständlich anders ausgerichteten Untersuchungen koppeln. Bspw. könnten Verfassungen in Beziehung zu Presseartikeln und den sich darin ablesbaren Diskursen gesetzt und miteinander verglichen werden.

Fußnoten

1. Im Zuge dieses Entwicklungsschrittes wurden u.a. folgende Tagger und Parser verwendet: Open NLP Segmenter, Mate Tools POS-Tagger, Mate Tools Lemmatizer, Open NLP Chunker, Mate Tools Morphological Analyzer, Hyphenation Annotator, CoreNLP Named Entity Recognizer, Mate Tools Dependency Parser.

Bibliographie

Boli-Bennett, John (1979): *The Ideology of Expanding State Authority in National Constitutions, 1870-1970, in: Meyer, John W. / Michael Thomas Hannan (eds.): National development and the world system: educational, economic and political change. Chicago: University of Chicago Press 222-237.*

Boli-Bennett, John / John W. Meyer (1978): The ideology of childhood and the state: Rules distinguishing children in national constitutions, 1870-1970, in: *American Sociological Review 43:* 797–812 .

Bubenhofer, Noah / Joachim Scharloth (2015): Themenheft „Maschinelle Textanalyse“, *in: Zeitschrift für germanistische Linguistik, 43.1.*

Eckart de Castilho, Richard / Gurevych, Iryna (2014): A broad-coverage collection of portable NLP components for building shareable analysis pipelines. In: *Proceedings of the Workshop on Open Infrastructures and Analysis Frameworks for HLT (OIAF4HLT) at COLING 2014*, 1-11, Dublin, Ireland.

Go, Julian (2003): A Globalizing Constitutionalism? Views from the Postcolony, 1945-2000, in: *International Sociology* 18.1: 71-95.

Gosewinkel, Dieter / Johannes Masing / Andreas Würschinger (2006): *Die Verfassungen in Europa 1789-1949.* München: Beck .

Hausser, Roland (2014): *Foundations of Computational Linguistics:* Human-Computer Communication in Natural Language . Berlin: Springer.

Heintz, Bettina / Annette Schnabel (2006): Verfassungen als Spiegel globaler Normen? Eine quantitative Analyse der Gleichberechtigungsartikel in nationalen Verfassungen, in: *Koelner Zeitschrift für Soziologie und Sozialpsychologie*, 58.4, 685-716.

Jepperson, Ronald L (1991): *Institutions, Institutional Effects, and Institutionalism, in: DiMaggio, Paul / Walter W. Powell (eds.): The New Institutionalism in Organizational Analysis. Chicago: University of Chicago Press 143-163.*

Kuckartz, Udo (2012): *Qualitative Inhaltsanalyse.* Methoden, Praxis, Computerunterstützung. Weinheim / Basel: Beltz.

Lobin, Henning (2010): *Computerlinguistik und Texttechnologie.* Paderborn / München: Fink.

Lupton, Deborah (2014): *Digital sociology.* New York: Routledge.

Manning, Christopher D. / Prabhakar Raghavan / Hinrich Schütze (2008): *Introduction to information retrieval.* New York: Cambridge University Press.

Mayntz, Renate (1997): *Soziale Dynamik und politische Steuerung:* theoretische und methodologische Überlegungen. Frankfurt/Main: Campus.

Mayring, Philipp (2015): *Qualitative Inhaltsanalyse.* Grundlagen, Techniken. Weinheim / Basel: Beltz.

Meyer, John W. (2005): *Die Weltgesellschaft und der Nationalstaat, in: ders., Weltkultur: wie die westlichen Prinzipien die Welt durchdringen. Frankfurt/Main: Suhrkamp 85-132.*

Mikolov, Thomas / Wen-tau Yih / Geoffrey Zweig (2013): Linguistic Regularities in Continuous Space Word Representations, in: *Proceedings of the HLT-NAACL conference* 746-752.

Prior, Lindsay (2011): *Using documents in social research.* Los Angeles: Sage.

Salton, Gerard / Andrew Wong / Shungshu Yang (1975): A vector-space model for information retrieval, in: *Journal of the American Society for Information Science* 18: 613-620.

Schimank, Uwe (1999): *Funktionale Differenzierung und Systemintegration der modernen Gesellschaft:* Soziale Integration. Opladen: Westdeutscher Verlag.

Strauss, Anselm L. / Juliet M. Corbin (1996): *Grounded theory:* Grundlagen qualitativer Sozialforschung. Weinheim: Beltz.

Thelen, Kathleen (1999): Historical Institutionalism in Comparative Politics, in: *Annual Review of Political Science* 2.1: 369-404.

Thelen, Kathleen (2002): *The explanatory power of historical institutionalism, in: Mayntz, Renate (eds.): Akteure-Mechanismen–Modelle. Zur Theoriefähigkeit makro-sozialer Analysen. Frankfurt / New York: Campus) 91-107.*

Vorländer, Hans (2007): Europas multiple Konstitutionalismen, in: *Zeitschrift für Staats- und Europawissenschaften* 5.2: 160-180.

Weber, Max (1972): *Wirtschaft und Gesellschaft:* Grundriss der verstehenden Soziologie. Tübingen: Mohr.

„XML Schema". World Wide Web Consortium (W3C) http://www.w3c.org/XML/Schema , [letzter Zugriff] 10.09.2017).

„XQuery 1.0: An XML Query Language". World Wide Web Consortium (W3C) http://www.w3.org/TR/ xquery [letzter Zugriff 10.09.17].

„XSL Transformation Version 1.0". World Wide Web Consortium (W3C) http://www.w3c.org/TR/xslt [letzter Zugriff 10.09.17].