



Computergestützte Textanalysen

Sven-Oliver Proksch

Inhalt

1	Einleitung	818
2	Annahmen, Vorgehensweise und Überblick	818
3	Politikwissenschaftliche Anwendungen	824
4	Methodische Herausforderungen	829
5	Fazit	832
6	Kommentierte Literaturhinweise	833
	Literatur	833

Zusammenfassung

Texte stellen eine der bedeutsamsten Datenquellen in der Politikwissenschaft dar. Mit der computergestützten Textanalyse steht Politikwissenschaftlern ein immer mächtigeres Werkzeug zur Verfügung, um alte und neue Fragen aus verschiedenen Subdisziplinen der Politikwissenschaft zu beantworten. Diese Methoden werden konstant weiterentwickelt und verfeinert, während gleichzeitig immer mehr Textdaten auch elektronisch zur Verfügung stehen. Der vorliegende Beitrag beschreibt die Annahmen und grundsätzlichen Vorgehensweisen und bietet einen Überblick über die wichtigsten computergestützten Textanalyseverfahren von der wörterbuchbasierten Analyse bis hin zu Textskalierung, Textklassifikation und Topic Models. Zudem wird auf geeignete Software verwiesen. Anschließend werden vier politikwissenschaftliche Anwendungsbereiche vorgestellt und mehrere methodische Herausforderungen diskutiert.

S.-O. Proksch (✉)

Cologne Center for Comparative Politics, Wirtschafts- und Sozialwissenschaftliche Fakultät,
Universität zu Köln, Köln, Deutschland

E-Mail: so.proksch@uni-koeln.de

Schlüsselwörter

Automatisierte Textanalyse · Textskalierung · Maschinelles Lernen · Topic-Modelle · Politische Positionen

1 Einleitung

Texte stellen eine der bedeutsamsten Datenquellen in der Politikwissenschaft dar. Kandidaten halten Wahlkampfreden und Abgeordnete nehmen an parlamentarischen Debatten teil. Parteien schreiben Wahlprogramme, Minister dieser Parteien lassen Gesetzesvorlagen erarbeiten. Die meisten großen Zeitungen bieten einen Online-Zugriff auf ihre Archive, und Internetnutzer äußern sich zu politischen Themen auf sozialen Medienplattformen. Mit der digitalen Revolution sind Forscher inzwischen in die Lage versetzt worden, all diese Textdaten (*big data*) computergestützt, d. h. quantitativ, zu untersuchen. Vereinfacht wird dies durch immer mehr frei zugängliche Werkzeuge für die computergestützte Textanalyse. Damit können alte wie neue Forschungsfragen beispielsweise hinsichtlich politischer Repräsentation, Parteienwettbewerb, politischer (online) Partizipation, Politikgestaltung, politischer Konflikte, und historischer politischer Prozesse mit neuen Daten und Analyseverfahren untersucht werden.

Dieses Kapitel beschreibt zunächst die Annahmen und grundsätzlichen Vorgehensweisen und bietet danach einen Überblick über die wichtigsten computergestützten Textanalyseverfahren. Zudem wird auf geeignete *R*-Pakete verwiesen, mit denen Einsteiger und Fortgeschrittene die erwähnten Verfahren durchführen können. Anschließend werden vier politikwissenschaftliche Anwendungsbereiche vorgestellt: die Schätzung von Parteipositionen aus Wahlprogrammen, die sogenannte Sentiment-Analyse in der politischen Kommunikationsforschung, die Identifizierung politischer Themen aus Texten, und die Analyse parlamentarischer Debatten. Abschließend wird auf vier methodische Herausforderungen genauer eingegangen, nämlich die Rolle der Textanalyse in einem politikwissenschaftlichen Forschungsdesign, sowie auf die Notwendigkeit von Robustheitsanalysen, die Berücksichtigung von strategischen Datengenerierungsprozessen und die Herausforderung der Mehrsprachigkeit in vergleichenden Studien.

2 Annahmen, Vorgehensweise und Überblick

Im Folgenden werden häufige Annahmen und Vorgehensweisen bei der computergestützten Textanalyse in der Politikwissenschaft vorgestellt. Besondere Berücksichtigung finden bei dieser Übersicht die klassische Textanalyse mit Wörterbüchern, Skalierungsverfahren für die Schätzung von politischen Positionen, sowie Verfahren zur Klassifikation bzw. Identifizierung von politischen Themen mittels *Topic Models* und maschinellen Lernverfahren. Darüber hinaus wird auf geeignete *R*-Pakete verwiesen, mit denen Einsteiger und Fortgeschrittene die erwähnten Verfahren durchführen können.

2.1 Annahmen und generelle Vorgehensweise

In der Politikwissenschaft gewinnen computergestützte Textanalysen deshalb zunehmend an Beliebtheit, weil politische Texte beständige Produkte politischen Handelns und politischer Kommunikation darstellen und inzwischen in großem Umfang maschinenlesbar vorliegen (z. B. Plenarprotokolle aus Parlamentsarchiven) oder digital generiert werden (z. B. soziale Medien). Eine stark genutzte Textquelle insbesondere in der vergleichenden Politikwissenschaft sind Wahlprogramme politischer Parteien, die als Konsequenz eines innerparteilichen Entscheidungsprozesses hinsichtlich der Positionierung vor einer Wahl betrachtet und entsprechend ausgewertet werden können. Forschungsdesigns in der Politikwissenschaft mit einer computergestützten Textanalysekomponente sollten prinzipiell vier Untersuchungs- und Auswertungsschritte umfassen.

Im ersten Schritt werden Texte durch den Forscher ausgewählt und damit der Textkorpus, d. h. die Textsammlung, definiert. Bereits hierbei ist auf eine sorgfältige und vor allem theoriegeleitete Auswahl zu achten, da insbesondere die Nichtberücksichtigung von relevanten Texten oder die Äquivalenzannahme von Texten sich auf eine Verzerrung der Ergebnisse auswirken kann. Beispielsweise kann es problematisch sein, wenn Gesetzesvorlagen mit Positionspapieren von Interessengruppen verglichen werden sollen, da beide Dokumente unterschiedlichen Textgenerierungsprozessen unterliegen und daher nicht äquivalent zu behandeln sind. Ähnlich problematisch wäre die Auswahl aller parlamentarischen Reden von Abgeordneten ohne zu unterscheiden, ob diese im Rahmen von Debatten zu Regierungs- oder Oppositionsvorhaben stattgefunden haben. Viele öffentliche Datenquellen sind inzwischen durch eine Programmierschnittstelle (API) zugänglich. Oftmals können Textdaten daher mit relativ geringem Aufwand direkt heruntergeladen werden (Munzert et al. 2014; siehe auch den Beitrag von Munzert und Nyhuis in diesem Band).

Im zweiten Schritt wird der Textkorpus in numerische Daten umgewandelt, damit quantitative Auswertungsverfahren eingesetzt werden können. Wenn Dateien dabei nicht im TXT Format vorliegen, werden diese zunächst (beispielsweise aus dem PDF Format) konvertiert. Danach wird die Worteinheit (*Token*) definiert, nach welcher ein Text entsprechend zerlegt werden soll. Dabei kann es sich um ein einzelnes Wort oder auch um Mehrworteinheiten (*N-Gram*) handeln. In vielen – aber nicht allen – computergestützten Textanalysen in der Politikwissenschaft findet die *Bag-of-Words* Annahme Anwendung, bei welcher nur die Häufigkeit der Worteinheiten in einem Text, nicht aber deren Reihenfolge oder grammatikalische Struktur, berücksichtigt wird. Die *Bag-of-Words* Annahme ist zwar stark reduktionistisch, sie erleichtert allerdings in vielen anderen Anwendungsbereichen – z. B. bei der Identifizierung von latenten politischen Themen – wesentlich die statistische Modellierung großer Textkorpora. Ausnahmen bilden aber beispielsweise Ansätze, die Textähnlichkeiten aufgrund der Unterschiede zwischen Textsequenzen bestimmen wie etwa bei der digitalen Plagiatserkennung. Mithilfe dieser sogenannten *text-reuse* Verfahren vergleichen beispielsweise Wilkersen et al. (2015) die im Jahre 2010 verabschiedete US-Gesundheitsreform („Obamacare“) mit 29.000 vorherigen Gesetzesvorlagen im

US-amerikanischen Kongress und stellen aufgrund der Ähnlichkeiten der Textsequenzen fest, dass das Gesetz viele Bestimmungen enthält, die ursprünglich in früheren und gescheiterten Gesetzesvorlagen eingebracht worden waren.

Die entsprechenden Textmerkmale (z. B. Wörter, N-Grams) werden im dritten Schritt ausgezählt und in eine Term-Dokument-Matrix konvertiert. Diese listet die Worteinheiten im Korpus in den Zeilen, die Dokumente in den Spalten und die Häufigkeiten der Worteinheiten im jeweiligen Dokument in den Zellen. Dabei ist zu berücksichtigen, dass es sich im Regelfall um eine dünnbesetzte Matrix (*sparse matrix*) handelt, bei der viele Einträge aus Nullen bestehen. Dies resultiert aus der Tatsache, dass bestimmte Wörter im Sprachgebrauch häufiger auftreten als andere, es also viele Wörter gibt, die in manchen Dokumenten entweder gar nicht oder nur selten auftreten. Wenn Wörter nach ihrer Häufigkeit in einer Rangfolge aufgelistet werden, ist die Häufigkeit eines Wortes umgekehrt proportional zu seiner Rangstelle. Dieser Zusammenhang ist als Zipfsches Gesetz bekannt (Zipf 1935, 1949). Oftmals werden daher auch bei diesem Schritt besonders häufige (aber aus politikwissenschaftlicher Sichtweise nicht relevante) Wörter aus der Datenmatrix entfernt (anhand von Stoppwortlisten) oder von existierenden Algorithmen auf ihren Wortstamm (Stammformreduktion) oder ihr Lemma (Lemmatisierung) zurückgeführt bis eine Minimalanzahl an Silben übrig bleibt. Beide Aspekte dienen der Reduktion der Term-Dokument-Matrix beziehungsweise der Definition von Wortäquivalenzen. Da diese Schritte aber durchaus Folgen für die Analyse selbst haben können, sollten diese transparent gemacht und deren Auswirkungen idealerweise durch Robustheitstests überprüft werden.

Im anschließenden vierten Schritt erfolgt die quantitative Auswertung der Informationen aus der Term-Dokument-Matrix. Die entsprechende Auswertungstechnik ist abhängig von der inhaltlichen Fragestellung: hierbei kann es sich um Textanalysen mit Wörterbüchern, einer Klassifikation von Texten in (bekannte) Kategorien, einer Identifizierung von (unbekannten) Themen, oder einer Schätzung von ideologischen Positionen handeln. Schließlich werden im fünften Schritt die Ergebnisse zusammengefasst, interpretiert, und für eine weiterführende Auswertung (beispielsweise als abhängige oder unabhängige Variable) verwendet.

2.2 Textanalysen mit Wörterbüchern

Die klassische computergestützte Inhaltsanalyse basiert auf dem Einsatz von Wörterbüchern, die eine Brücke zwischen qualitativen und quantitativen Verfahren darstellen (siehe zu letzteren auch die Beiträge von Hutter und Wodak in diesem Band). Ziel ist hierbei die Messung der relativen Häufigkeit vordefinierter Kategorien in den Dokumenten. Auf der einen Seite werden Wörterbücher „qualitativ“ erstellt, indem theoretische Konzepte mit Kategorien und Begriffen assoziiert werden. Dabei spielen Kontextinterpretationen (ist die Verwendung eines Begriffs eindeutig einer Kategorie zuzuschreiben?) eine große Rolle. Begriffe werden jeweils einer Kategorie zugewiesen, was diese Art der Analyse beispielsweise von Topic Models unterscheidet.

Bei der Konstruktion von Wörterbüchern gilt zu beachten, dass immer eine Abwägung getroffen werden muss zwischen zwei Gütekriterien: der Sensitivität des Wörterbuchs auf der einen Seite und der Spezifität auf der anderen. Soll beispielsweise ein Wörterbuch für die Kategorie „Arbeitsmarktpolitik“ erstellt werden, so lassen sich zwei hypothetische Wörterbücher definieren: das erste nur mit dem Begriff „Arbeitsmarktpolitik“ (Wörterbuch A), das andere mit *allen* Begriffen der deutschen Sprache (Wörterbuch B). Wörterbuch A hat eine hohe Spezifität (da das Wort nur im Kontext einer Diskussion zu arbeitsmarktpolitischen Entscheidung auftritt), aber eine geringe Sensitivität (da viele andere Begriffe, die mit Arbeitsmarktpolitik zusammenhängen, ignoriert werden). Umgekehrt hat Wörterbuch B eine sehr hohe Sensitivität (da alle arbeitsmarktpolitischen Begriffe auf jeden Fall abgedeckt werden), aber eine geringe Spezifität (da durch die anderen Wörterbucheinträge nicht mehr sinnvoll arbeitsmarktpolitische Texte identifiziert oder kategorisiert werden können).

Nützliche Wörterbücher müssen daher sinnvoll zwischen diesen Extremen abwägen. In der politischen Kommunikationsforschung erfreut sich die Analyse von Stimmungen oder Empfindungen (*sentiment*) anhand von Wörterbüchern mit positiven und negativen Begriffen großer Beliebtheit (z. B. Soroka 2006, 2012, 2014; Young und Soroka 2012). Auch hierarchische, Politikbereich-spezifische Wörterbücher haben in der politikwissenschaftlichen Forschung Anwendung gefunden (z. B. Laver und Garry 2000). Ein Nachteil von Textanalysen auf Grundlage von Wörterbüchern ist allerdings, dass diese oftmals sehr kontextspezifisch definiert sind. Dies erschwert eine Anwendung auf neue Textkorpora und andere Sprachkontexte.

2.3 Skalierung von Texten

In akteurszentrierten theoretischen Modellen der Politik spielen politische Präferenzen eine zentrale Rolle: Die Präferenzen politischer Akteure wirken sich auf den Parteienwettbewerb, Koalitionsbildungsprozesse, und Gesetzgebungsaktivitäten aus. Allerdings sind politische Präferenzen latent: sie zeigen sich lediglich durch politisches Verhalten, beispielsweise durch das Abstimmungsverhalten oder politische Redebeiträge. Computergestützte Textanalysen bieten daher die Chance, solche latenten Positionen zu messen. Eines der ersten computergestützten Verfahren wurde von Laver et al. (2003) mit *Wordscores* vorgestellt. Die Grundidee dieses Ansatzes ist die Eichung der Dokumentenpositionen in einem eindimensionalen ideologischen Raum anhand von vorgegebenen Werten für Referenzdokumente. Die Auswahl der Referenzdokumente erfolgt durch den Forscher, die Bestimmung der vorgegebenen Werte oftmals auf Grundlage von Experteninterviews. Basierend auf der Beziehung zwischen der relativen Häufigkeit der Wörter in ausgewählten Referenzdokumenten und der relativen Häufigkeit der Wörter in den zu untersuchenden Dokumenten können mit *Wordscores* Positionen in einem eindimensionalen Raum geschätzt werden. Vorgegebene Positionen und Inhalte von Referenzdokumenten verankern also die Dimensionen. Dieses Verfahren kam vielfach in der Schätzung

von Parteipositionen in Wahlprogrammen (z. B. Laver et al. 2003; Proksch und Slapin 2006; Debus 2009), aber auch in parlamentarischen Debatte (z. B. Gianetti und Laver 2005; Hakhverdian 2009; Klemmensen et al. 2007; Bernauer und Bräuninger 2009; Herzog und Benoit 2015) zur Anwendung.

Allerdings hat die Anwendung von *Wordscores* auch einige Probleme aufgeworfen. Ungelöst bleibt beispielsweise die Frage der Referenzdokumentenauswahl bei langen Zeitreihen (Budge und Pennings 2007) und der Transformation von *Wordscores* in Dokumentenwerte, um Positionen der Referenzdokumente mit den geschätzten Positionen der Dokumente vergleichbar zu machen (Lowe 2008; Martin und Vanberg 2008; Benoit und Laver 2008). Schließlich ist zu bedenken, dass die Schätzungen stark von der Auswahl der Referenzdokumente abhängen. Im Extremfall bezieht sich die geschätzte Dimension überhaupt nicht auf die inhaltliche interessierende Dimension (z. B. Umweltpolitik), wenn die Referenztexte (z. B. Wahlprogramme) keine Wörter enthalten, die auf dieser Dimension diskriminieren können, aber durch die Verankerung mit Referenzdokumenten genau dies angenommen wird.

Im Unterschied zu diesem beaufsichtigtem Schätzverfahren (*supervised scaling*) versuchen unbeaufsichtigte Skalierungsverfahren (*unsupervised scaling*) daher, die Ideologie von Akteuren ohne Wörterbücher oder Referenzdokumente zu schätzen (z. B. Slapin und Proksch 2008; Proksch und Slapin 2010; Lo et al. 2016; Lauderdale und Herzog 2016). Das *Wordfish* Modell von Slapin und Proksch (2008) skaliert Positionen ähnlich wie *Wordscores* in einem eindimensionalen Raum, verlässt sich aber ausschließlich auf die relativen Worthäufigkeiten in den zu untersuchenden Dokumenten. Die Technik ist also in Fällen anwendbar, in denen das Auffinden von Referenztexten schwierig oder unmöglich ist. Da der Ansatz unbeaufsichtigt ist und sich auf die verfügbaren Daten stützt, ohne dass Referenzdokumente oder ein Wörterbuch erforderlich sind, sind die Anforderungen an die Datenqualität hoch. Wenn Ideologie aus einem Textkorpus geschätzt werden soll, so sollten die Texte auch tatsächlich mit ideologischen Aussagen zusammenhängen. Die Annahme wird insbesondere schnell verletzt, wenn Dokumente untersucht werden, die nicht primär aufgrund ideologischer Gesichtspunkte generiert werden (z. B. Gesetzesvorlagen). Weiterhin können Agenda-Effekte die Schätzungen beeinflussen, wenn Politiker ähnliches Vokabular benutzen (müssen), um über spezifische Themen zu sprechen. Dies kann zum Beispiel bei einem Vergleich von parlamentarischen Debatten problematisch sein, bei denen das Thema der Debatte vorgegeben ist und sich daher eher die Debatten als Ganzes unterscheiden lassen, nicht aber die Positionen von einzelnen Abgeordneten über Debatten hinweg. Eine Weiterentwicklung des *Wordfish* Modells für diesen Kontext bietet der *Wordshoal* Ansatz von Lauderdale und Herzog (2016), bei welchem die *Wordfish* Schätzungen einzelner Parlamentsdebatten in einem zweiten Schritt noch einmal in einen gemeinsamen Politikraum skaliert werden.

2.4 Klassifikation und Identifizierung politischer Themen

Neben der Schätzung von politischen Positionen ist in vielen politikwissenschaftlichen Forschungsprojekten die Identifizierung von politischen Themen ein wichtiges

Forschungsziel. Generell können hierbei zwei unterschiedliche Vorgehensweisen unterschieden werden: beaufsichtigte (maschinelles Lernen) und unbeaufsichtigte Verfahren (Topic-Modelle). Bei maschinellen Lernverfahren (z. B. *Support Vector Machines*) stellt die Klassifikation von Textdokumenten das Ziel der Untersuchung dar. Zunächst werden für den Trainingsprozess Themenkategorien theoriegeleitet vordefiniert (z. B. Politikfelder oder Sentiment) und ausgewählte Dokumente händisch kodiert bzw. klassifiziert. Danach lernt der maschinelle Klassifikator Wortmerkmale der einzelnen vorkodierten Themenkategorien. Nach Abschluss dieses Lernprozesses kann der Klassifikator unkodierte Textdokumente automatisch klassifizieren. Eine Annahme bei diesen Verfahren ist, dass sich ein Textdokument lediglich einer Kategorie zuordnen lässt. Das Verfahren findet beispielsweise Anwendung bei der Klassifizierung von Politikfeldern von US-Gesetzesvorlagen anhand ihrer Überschriften (Collingwood und Wilkerson 2012).

Sind die Kategorien unbekannt, weil zum Beispiel ein langer Zeitraum mit sich wandelnden Themen untersucht werden soll, oder möchte man die Annahme treffen, dass Dokumente einen Mix an Kategorien enthalten können, so eignen sich sogenannte *Topic-Modelle* (z. B. Blei et al. 2003; Grimmer 2010; Quinn et al. 2010; Roberts et al. 2014). Hierbei sind die Kategorien (*topics*) latent und werden geschätzt. Vordefiniert wird lediglich die Anzahl der zu schätzenden Kategorien. Diese probabilistischen Modelle schätzen anhand der Verteilung von Wörtern über einen Korpus von Dokumenten zugleich, mit welcher Wahrscheinlichkeit ein Wort auf ein bestimmtes Thema/Topic (d. h. eine Kategorie) hindeutet und mit welcher Wahrscheinlichkeit ein Text einem bestimmten Thema zuzuordnen ist. Das zu optimierende Kriterium ist hierbei die bestmögliche Trennung zwischen Texten auf Basis der geschätzten Themenzugehörigkeiten. Für jedes Dokument lassen sich so Topic-Anteile schätzen, beispielsweise zu welchem Anteil ein Dokument Topic A, B, oder C enthält.

2.5 Textanalyse mit R

Für die Anwendung der computergestützten Textanalyse mit politikwissenschaftlichem Bezug bietet sich grundsätzlich die frei zugängliche Software *R* an, die mit einer Bandbreite an Textpaketen die oben erwähnten Modelle umsetzt. Empfehlenswert ist beispielsweise das *R* Paket *Quanteda*, welches viele der gängigen Vorbereitungsschritte (Einlesen der Texte, Aufbereitung, Umwandlung) sowie viele der quantitativen Auswertungsverfahren (Wörterbücher, *Wordscores*, *Wordfish*, *Wordshoal*) umsetzen kann (Benoit 2017). Dokumente und zugehörige Metadaten werden als Korpus-Objekt geladen und gespeichert. Weiterhin besteht eine Schnittstelle zu gängigen Topic-Modellen in *R*, wie beispielsweise den *R*-Paketen *LDA*, *topicmodels* und *STM*. Alternativ eignen sich für die allgemeine Textaufbereitung auch die *R*-Pakete *tm* sowie *tidyr*. Das Paket *preText* bietet die Möglichkeit, die Auswirkungen unterschiedlicher Textbearbeitungsschritte (z. B. Stemming) automatisiert zu untersuchen. Maschinelle Lernverfahren lassen sich mit den Paketen *caret* und *RTextTools* implementieren.

3 Politikwissenschaftliche Anwendungen

Generell sollten bei einem politikwissenschaftlichen Forschungsdesign mit computergestützter Auswertung von Textmaterial einige Grundprinzipien beachtet werden (Grimmer und Stewart 2013). Computergestützte Verfahren bilden offensichtlich nicht die natürliche Sprache der Politik ab, aber oftmals sind Vereinfachungen für das Erreichen des Forschungsziels ausreichend und nützlich. Die Verfahren der maschinellen Sprachverarbeitung erweitern die verfügbaren Ressourcen im Forschungsprozess. Sie sollen (und können) aber den Forscher bei der Analyse nicht ersetzen. Je nach ausgewähltem Verfahren spielt der oder die Forscher am Anfang oder Ende des Textanalyseverfahrens eine wesentliche Rolle: sei es bei der theoriegeleiteten Auswahl von Dokumenten, der Definition von Wörterbüchern, der händischen Klassifikation von Dokumenten, oder der nachträglichen Interpretation und Validierung von geschätzten politischen Dimensionen und Themen. Diese Auflistung zeigt auch, dass es nicht ein dominantes computergestütztes Verfahren gibt, auf welches Forscher unbedingt zurückgreifen sollten, sondern dass je nach Forschungsfrage und Textkorpus ein oder mehrere Verfahren zum Einsatz kommen sollten. Schließlich stehen am Ende eines jeden Prozesses die Kontrolle und Validierung der Ergebnisse. Beispielhaft sollen im Folgenden ausgewählte Anwendungen in der Politikwissenschaft ausführlicher vorgestellt werden, um zum einen die Bandbreite an verfügbaren Alternativen aufzuzeigen und zum anderen häufig eingesetzte Verfahren zu erläutern und einzuordnen.

3.1 Parteipolitische Positionen in Wahlprogrammen: *Wordfish*

Wordfish (Slapin und Proksch 2008) ist ein Textskalierungsverfahren, welches spezifisch für die Schätzung von politischen Positionen in Parteiprogrammen entwickelt wurde. Das Verfahren modelliert die Anzahl an Wörtern in einem Dokument als eine Kombination aus systematischen (Dokument- und Wortposition sowie Fixed Effekte für Wörter und Dokumente) und stochastischen Faktoren (Poisson-Verteilung). Alle zu schätzenden Parameter sind damit latente Variablen und *Wordfish* ein „unbeaufsichtigtes“ Skalierungsverfahren.

Um die Schätzung durchzuführen sind allerdings einige Grundannahmen notwendig. Zunächst muss sichergestellt werden, dass politische Akteure (sprich Parteien) tatsächlich politische Positionen über unterschiedliche Worthäufigkeiten ausdrücken. Da es den Parteien freisteht, die Wahlprogramme hinsichtlich ihrer Länge oder Themenabdeckung gemäß ihren politischen Zielsetzungen auszugestalten, ist diese Annahme realistisch. Weiterhin wird angenommen, dass sich Parteien in einem eindimensionalen politischen Raum verorten. Diese Annahme ist technisch nicht zwingend notwendig. Das *Wordfish*-Modell lässt sich auf weitere Dimensionen erweitern (siehe auch die enge Verwandtschaft des Modells zur Korrespondenzanalyse, Lowe 2017), die Annahme erleichtert allerdings die substanzielle Interpretation der geschätzten Dimensionen. Beispielsweise wäre eine Erwartung, dass Parteien in den meisten demokratischen Parteiensystemen im Gesamtkontext eines Wahl-

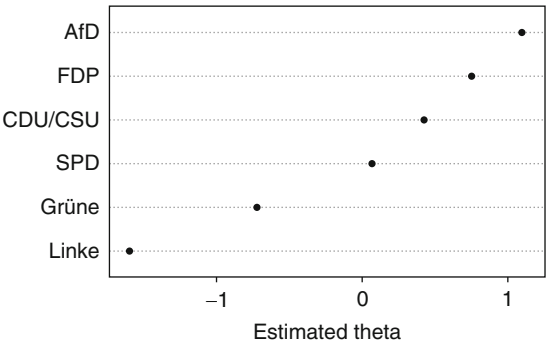
programms zu den meisten politischen Themen Stellung beziehen und sich somit auf einer allgemeinen Links-Rechts-Dimension verorten. Sollten politikfeldspezifische Positionen gewünscht sein, so wäre eine entsprechende Vorauswahl des Korpus (beispielsweise über eine Fokussierung auf wirtschaftspolitische Passagen der Wahlprogramme) empfehlenswert.

Am Beispiel der Wahlprogramme bei der deutschen Bundestagswahl 2017 sollen die wichtigsten Schritte bei der *Wordfish* Schätzung illustriert werden. Tab. 1 zeigt einen Auszug aus der Term-Dokument-Matrix der Wahlprogramme. Dabei wurden die Dokumente mit dem R Paket *Quanteda* vom PDF-Format in Text umgewandelt und die Wörter ausgezählt. So zeigt sich beispielsweise, dass Begriffe wie „Gerechtigkeit“, „Flüchtlinge“ und „Solidarität“ relativ gesehen häufiger von Parteien des linken Spektrums genannt werden, Begriffe wie „Asylbewerber“, „Kriminalität“, und „Zuwanderung“ aber insbesondere in Wahlprogrammen auf dem rechten Spektrum auftauchen. Das Textskalierungsverfahren macht sich die Unterschiede über alle Wörter und Wahlprogramme zunutze und schätzt die Positionen der Dokumente auf einer Dimension. Die Ergebnisse dieser Schätzung sind in Abb. 1 dargestellt. Wahlprogramme, die sich hinsichtlich ihrer relativen Worthäufigkeiten ähnlicher sind, liegen näher beieinander. Ohne dass der Forscher bei der Schätzung Vorgaben hinsichtlich der Dimension oder der Wahlprogramme machen muss, lassen sich die Parteien über die geschätzten Werte von links nach rechts einordnen, mit der Links-

Tab. 1 Auszug aus der Term-Dokument-Matrix deutscher Wahlprogramme bei der Bundestagswahl 2017

Partei	<i>Gerechtigkeit</i>	<i>Flüchtlinge</i>	<i>Solidarität</i>	<i>Asylbewerber</i>	<i>Kriminalität</i>	<i>Zuwanderung</i>	...
<i>AfD</i>	6	2	2	8	9	17	
<i>FDP</i>	0	6	3	0	3	1	
<i>Grüne</i>	79	8	7	4	6	1	
<i>Die Linke</i>	26	13	9	2	2	0	
<i>SPD</i>	24	7	6	1	6	2	
<i>CDU/CSU</i>	1	4	3	1	2	0	

Abb. 1 Wordfish-Positionen deutscher Parteiprogramme bei der Bundestagswahl 2017



partei und der AfD auf den Extremen des Parteienspektrums. Die *Wordfish*-Ergebnisse bieten zudem die Möglichkeit, die geschätzten Wortpositionen zu untersuchen, worüber sich noch mehr über den politischen Sprachgebrauch in Wahlprogrammen aussagen lässt.

Abschließend soll auf eine Weiterentwicklung des *Wordfish*-Modells hingewiesen werden, bei der neben den Parteipositionen auch die Konsistenz von Wahlprogrammen gemessen werden kann (Lo et al. 2016). Vermehrt widmet sich die Parteienforschung der Fragestellung nach der Konsistenz bzw. Inkonsistenz von politischen Positionen. Bei dieser Erweiterung des *Wordfish*-Modells wird ein zusätzlicher Dokumentenparameter (Overdispersion-Parameter) geschätzt, der die unterschiedliche Varianz der Worthäufigkeiten eines Dokuments berücksichtigt. Nach Kontrolle von Textlänge und -position erfasst dieser Parameter Situationen, in denen sich Worthäufigkeiten verändern, weil innerparteiliche Gruppierungen unterschiedliche Themen im Wahlprogramm betonen oder konfliktträchtige Aussagen weggelassen oder mitaufgenommen werden. Dies hat zur Folge, dass Dokumentenvektoren höhere oder niedrigere Worthäufigkeiten enthalten als zu erwarten wäre, wenn die Positionen der Dokumente konsistent formuliert wären. Lo et al. (2016) zeigen anhand von Textsimulationen, dass dieser Schätzer die Konsistenz der Programme ähnlich akkurat messen kann wie menschliche Kodierer. Die Autoren wenden den Schätzer auf Parteiprogramme in Deutschland, Irland, Niederlande und Schweden an und verwenden Positions- und Konsistenzschätzungen als unabhängige Variablen, um Stimmenanteilsveränderungen der Parteien bei Wahlen zu erklären. Unklare Positionen können nach diesen Ergebnissen sogar zu Stimmengewinnen bei Parteien führen, die ihre Positionen mäßigen, aber sie schaden denjenigen Parteien, die extremere Positionen als zuvor einnehmen und dabei in ihrer Position inkonsistenter werden.

3.2 Sentimentanalyse und politische Kommunikation: *Lexicoder*

Die Sentimentanalyse, also die Erkennung ob eine Aussage eher positiv oder negativ ist, nimmt in der politischen Kommunikationsforschung eine zunehmend wichtige Rolle ein (z. B. Atteveldt et al. 2008; Young und Soroka 2012; Soroka et al. 2015a, b; Haselmayer und Jenny 2017). Oft geht es dabei um die Anwendung eines Wörterbuchs zur automatischen Auswertung von Texten, um eine geäußerte Meinung als positiv oder negativ einstufen zu können. Typische Anwendungsgebiete sind hierbei die Analyse von Nachrichteninhalten, politischen Reden, oder Aussagen in den sozialen Medien.

In der englischsprachigen Kommunikationsforschung hat sich mit *Lexicoder* ein Wörterbuch etabliert, welches insbesondere auf politische Nachrichteninhalte hin validiert und optimiert wurde (Young und Soroka 2012). Auf der Basis von drei existierenden Wörterbüchern erstellten die Autoren eine Liste von 4567 positiven und negativen Wörtern. Die Autoren wenden dieses Wörterbuch auf englischsprachige Zeitungen in Kanada während des Wahlkampfes für die Parlamentswahl 2006 an und werten für die beiden wichtigsten Parteien (*Conservatives* und *Liberals*) die „Stimmung“ in Zeitungsartikeln aus. Dabei zählen sie nur positive oder negative

Wörter in den Sätzen, in denen entweder die jeweilige Partei oder der jeweilige Parteivorsitzende erwähnt wird. Schließlich werden für das Sentimentmaß die negativen von den positiven Werten abgezogen. Die Autoren sind mit diesem relativ einfachen Maß in der Lage, die voraussichtliche Wahlabsicht für die beiden Parteien während des Wahlkampfes vorherzusagen. Die Ergebnisse mit händischer Kodierung der Zeitungsartikel sind etwas besser, was darauf schließen lässt, dass die voll automatisierte Sentimentanalyse nicht in jedem Fall die tatsächliche Beurteilung einer Partei in einer Zeitung wiedergeben kann. In der Gesamtsicht schneidet das automatisierte Verfahren zur Vorhersage von Wahlabsichten aber sehr gut ab. Die Sentimentanalyse hat also durchaus noch mehr Anwendungspotenzial bei politikwissenschaftlichen Inhaltsanalysen, insbesondere wenn die Spezifität des Wörterbuchs erhöht werden kann.

3.3 Politische Themensetzung: *Topic Models*

Topic Models sind aus der computergestützten Textanalyse nicht mehr wegzudenken: Sie eignen sich, um latente politische Themensetzungen zu identifizieren und zu messen. Topic-Modelle (oder auch das *Wordfish* Modell) sind generative Modelle des Datengenerierungsprozesses von Worthäufigkeiten. Ein Topic (Thema) wird als eine Mischung aus Wörtern definiert, bei der jedes Wort eine Wahrscheinlichkeit hat, zu einem Topic zu gehören. Ein Dokument ist eine Mischung aus mehreren Topics, d. h. ein einzelnes Dokument kann aus mehreren Themen bestehen. Als solches summieren sich die Themenanteile über alle Themen eines Dokuments hinweg zu eins, ebenso wie die Wortwahrscheinlichkeiten für ein bestimmtes Thema.

Beispielhaft sollen hier Studien von Quinn et al. (2010) und Grimmer (2010) erwähnt werden. Quinn et al. untersuchen die politische Agenda des US-Senats und messen mit ihrem Topic Modell 42 latente Themen in Senatsreden. Im Anschluss untersuchen die Autoren, wie und warum sich die politische Aufmerksamkeit auf diese Themen im Laufe der Zeit verändert. Die Autoren können zeigen, dass Senatoren zwar häufig solche Themen betonen, die mit ihrer Ausschussarbeit zusammenhängen, aber Ausschussvorsitzende nicht häufiger als andere Ausschussmitglieder dazu neigen, solche Themen anzusprechen. Darüber hinaus neigen Senatoren mit ideologisch extremeren Positionen dazu, häufiger im Senat zu sprechen, insbesondere zu kontroversen Themen wie Justiz, Arbeitsmarktpolitik, Steuern und Haushalt.

Topic Models können somit einen spannenden Einblick in parlamentarische und für politische Repräsentation relevante Prozesse liefern. Obwohl Topic Modelle zur Gruppe der „unbeaufsichtigten“ Verfahren zählen, ist eine vorherige Festsetzung der Anzahl der zu schätzenden Topics durch den Forscher notwendig. Wichtige Themen können daher eventuell unentdeckt bleiben, wenn die gewählte Anzahl zu klein ist, und Themen können durch das Topic Modell künstlich erzeugt werden, wenn die Zahl zu hoch eingestellt wird.

In seiner Anwendung des Topic Modells untersucht Grimmer (2010), auf welche Art und Weise US-Senatoren ihren Wählern ihre Arbeit in Washington anhand von

Pressemitteilungen erklären. Das modifizierte Topic Model von Grimmer berücksichtigt die hierarchische Struktur der Pressemitteilungen von Senatoren (Pressemitteilungen bilden die untere Ebene und Senatoren die höhere). Seine Ergebnisse zeigen, dass Ausschussvorsitzende im US-Senat den Themen in ihrem politischen Zuständigkeitsbereich mehr Aufmerksamkeit schenken als der durchschnittliche Senator und dass Senatoren desselben Staates dazu neigen, eine ähnliche Agenda zu haben als dies bei Senatoren aus verschiedenen Staaten der Fall ist. Hingewiesen werden soll an dieser Stelle abschließend auf die Weiterentwicklung des *Structural Topic Models*, welches neben den Themen auch Kovariaten der Dokumente bereits in der Schätzung berücksichtigen kann und somit die Inferenzanalyse wesentlich erleichtert (Roberts et al. 2014).

3.4 Parlamentarische Debatten: Positionierung und Sprachkomplexität

Im Zuge der Digitalisierung von Parlamentsarchiven bietet die computergestützte Textanalyse enorme Potenziale für die vergleichende Parlaments- und Repräsentationsforschung. Zu beachten ist allerdings, dass parlamentarische Reden einem völlig anderen Datengenerierungsprozess unterliegen als beispielsweise Wahlprogramme. Während Parteien in Wahlprogrammen nach Belieben bestimmte Themen betonen oder ignorieren können (und somit ideologische Positionen geschätzt werden können), so unterliegen parlamentarische Debatten anderen Prozessen. Zum einen können aufgrund der begrenzten Debattenzeit niemals alle Abgeordnete im Plenum Position zu einer Gesetzesvorlage beziehen, zum anderen gibt es bei der Auswahl der Redner für die jeweiligen Fraktionen systematische Unterschiede.

Proksch und Slapin (2012, 2015) modellieren den innerparteilichen Prozess der Rednerauswahl in parlamentarischen Debatten und kommen zu dem Schluss, dass in Situationen mit heterogenen Fraktionen oftmals die offizielle Parteilinie besonders stark im Parlament betont wird. Dadurch können Debatten nicht – oder zumindest nicht direkt – dazu verwendet werden, um innerparteiliche Geschlossenheit zu schätzen, da Debatten diese nicht in allen Situationen gleich abbildet. Dennoch bieten sich mehrere Verfahren an, um Debatten computergestützt auszuwerten. Erwähnt werden sollen hier das *Wordshoal*-Verfahren von Lauderdale und Herzog (2016) und eine Anwendung eines Textkomplexitätsmaß durch Spirling (2016).

Das ursprüngliche *Wordfish*-Modell eignet sich nur bedingt für die Analyse von Parlamentsdebatten, da die Schätzwerte über verschiedene Schätzungen hinweg nicht vergleichbar sind (was der Fall für alle latenten Variablenmessungen bzw. Item-Response Modelle ist). Lauderdale und Herzog (2016) fügen daher dem *Wordfish*-Modell eine zweite Skalierungsstufe hinzu. Zunächst werden Redner aus einzelnen Gesetzesdebatten separat skaliert. Da nur wenige Abgeordnete an einzelnen Lesungen bzw. Debatten teilnehmen, schätzen sie in einem zweiten Schritt mithilfe einer Bayesianischen Faktorenanalyse die latenten Positionen der Abgeordneten über das gesamte Parlament hinweg. Die Autoren verfolgen dabei nicht das Ziel, einzelne Debatten auszuwerten, sondern die durchschnittliche ideologische

Positionierung der Abgeordneten über alle Debatten hinweg zu schätzen. Dabei nehmen die Autoren an, dass fehlende Positionen in einer Debatte zufällig verteilt sind und nicht mit der Entscheidung der Abgeordneten, an einer Debatte teilzunehmen, zusammenhängen. Diese Annahme ist angesichts der Ergebnisse von Proksch und Slapin (2015) vermutlich falsch, und eine systematischere Modellierung dieses Aspektes könnte durchaus Auswirkungen auf die geschätzten Werte haben. Die Schätzungen spiegeln zudem nicht die tatsächlichen ideologischen Positionen der Abgeordneten wieder. Vielmehr handelt es sich bei der extrahierten Dimension um den Unterschied zwischen Regierungs- und Oppositionsfraktionen (in Irland), während in einem anderen Kontext eher parteiprogrammatische Unterschiede erkennbar werden (in den USA). Das *Wordshoal*-Modell ist ein sinnvoller Ansatz, um parlamentarische Debatten auszuwerten. Anwender sollten sich jedoch darüber im Klaren sein, dass der besondere parlamentarische Datengenerierungsprozess hierbei die Ergebnisse maßgeblich beeinflusst.

Einen völlig anderen Ansatz – basierend auf einem deskriptiven textanalytischen Verfahren – verfolgt Spirling (2016) in seiner Untersuchung der Auswirkungen der Stimmrechtsausweitung in Großbritannien im 19. Jahrhundert auf das parlamentarische Verhalten von Abgeordneten im britischen *House of Commons*. Spirling geht der Frage nach, ob Minister im Parlament nach der Ausweitung des Stimmrechts weniger komplexe Reden hielten als zuvor, um so zugänglicher für neue Wähler mit geringerem Bildungsniveau zu werden. Hierzu verwendet er einen historischen Textkorpus mit ca. 670.000 Reden von ca. 3600 Abgeordneten zwischen 1832 und 1915. Auf der Grundlage eines verbreiteten und einfach zu berechnenden Lesbarkeitsindex (*Flesch Reading Ease Index*, Flesch 1948) zeigt Spirling, dass dies tatsächlich der Fall war: Minister verwendeten in ihren Reden zum Ende des 19. Jahrhunderts im Schnitt kürzere Sätze und weniger Silben, insbesondere auch im Verhältnis zu Hinterbänklern, bei denen diese Entwicklung nicht im selben Umfang zu beobachten war. Der Aufsatz ist ein gutes Beispiel dafür, wie auch deskriptive computergestützte Verfahren bei der politikwissenschaftlichen Untersuchung historischer Prozesse nützlich für das Testen von Theorien sein können. Kritisch ist allerdings anzumerken, dass sich der verwendete Lesbarkeitsindex nicht konkret an politischer Sprache orientiert, und dass zudem keine Unsicherheitsschätzungen vorliegen. Auch hier ist daher Potenzial für zukünftige Weiterentwicklungen vorhanden.

4 Methodische Herausforderungen

Die computergestützte Textanalyse stellt eine der spannendsten aktuellen methodischen Entwicklungen in der Politikwissenschaft dar. Eines ihrer größten Potenziale liegt dabei in der Möglichkeit, mehrdimensionale Textdaten auf unterschiedlichste Arten und Weisen zu untersuchen, um unbekannte Strukturen zu entdecken und latente theoretische Konzepte der Politikwissenschaft auf innovative Weise zu messen. Nicht zu vernachlässigen ist hierbei auch die produktive Verknüpfung von qualitativer und quantitativer politikwissenschaftlicher Forschung, indem qualitati-

ves Textverständnis, Expertise und händische Kodierungen von Texten mit automatisierten Verfahren kombiniert werden können (siehe wiederum auch die Kapitel von Hutter und Wodak in diesem Band). In vielen nordamerikanischen und europäischen Doktorandenprogrammen in der Politikwissenschaft sowie an *Summer Schools* mit Methodenschwerpunkt (z. B. Essex, ICPSR, Syracuse) zeigt sich, dass die computergestützte Textanalyse als ein zusätzlicher Kernkurs neben Angeboten in Statistik, experimentellen Methoden, kausaler Inferenz, qualitativen Methoden und Forschungsdesign ihren festen Platz im politikwissenschaftlichen Methodencurriculum gefunden hat. Abschließend soll an dieser Stelle auf vier methodische Herausforderungen in der computergestützten Textanalyse in der Politikwissenschaft eingegangen werden.

4.1 Der Platz der Textanalyse im Forschungsdesign: Entdeckung oder kausale Inferenz?

Eine grundsätzliche Überlegung stellt sich bei der Auswahl der Zielsetzung der computergestützten Textanalyse in einem politikwissenschaftlichen Forschungsdesign. Die jüngeren technischen Entwicklungen im Bereich der quantitativen Textanalyse fallen vor allem in den Bereich der unbeaufsichtigten Verfahren, die es erlauben, latente Strukturen zu entdecken und zu messen. Angesichts der unzähligen Möglichkeiten mehrdimensionale Textdaten zu untersuchen, stellen solche dimensionsreduzierende Verfahren (Skalierung oder Topic Models) daher eine sinnvolle Möglichkeit dar, die wesentlichen Inhalte (z. B. Positionen oder Themen) von Textsammlungen zu bestimmen, und dadurch neue Erkenntnisse über politische Prozesse zu gewinnen. Die Entdeckung dieser Strukturen kann daher bereits am Anfang eines Forschungsdesigns eine bedeutende Rolle spielen: die computergestützte Textanalyse kann also für die Theorieformulierung wichtige Impulse liefern.

Es stellt sich allerdings die offensichtliche Frage, inwiefern Ergebnisse einer solchen „Entdeckungs-Analyse“ für das Testen von Theorien und für kausale Inferenzen zulässig sind. Eine Möglichkeit, um mit dieser Herausforderung umzugehen, ist die Unterteilung des Datensatzes in einen Trainings- und Testkorpus (Egami et al. 2017). Weiterhin ist es wichtig, dass sich Politikwissenschaftler vergegenwärtigen, dass sich die Zielsetzungen in der Informatik, aus denen viele der hier vorgestellten Verfahren stammen, grundsätzlich von denen in der Politikwissenschaft unterscheiden (DiMaggio 2015): Während Kausalitätsfragen und Hypothesentests im Vordergrund empirischer politikwissenschaftlicher Studien stehen, so sind Vorgehensweisen in der maschinellen Sprachverarbeitung in der Informatik primär daran interessiert, bestehende Verfahren zu beschleunigen oder zu verbessern.

Hierbei spielt auch das unterschiedliche Vertrauen in menschliche Kodierungen eine Rolle. In überwachten maschinellen Lernverfahren sind menschliche Kodierungen der Goldstandard, ohne die das Lernen überhaupt nicht stattfinden kann. Je komplexer allerdings das zu untersuchende Phänomen, desto unwahrscheinlicher werden Übereinstimmungen menschlicher Kodierungen, was sich unmittelbar auf die Qualität der maschinellen Lernverfahren auswirkt (ibid.). Auf der anderen Seite

sind Validierungen der Ergebnisse unbeaufsichtigter Verfahren durch den Forscher ein wesentlicher Bestandteil des politikwissenschaftlichen Forschungsdesigns. Gerade hier sollte besonders Wert auf eine sorgfältige Kontrolle der Ergebnisse gelegt und die Vorgehensweise der Validierung transparent und nachvollziehbar gemacht werden.

4.2 Die Notwendigkeit von Robustheitsanalysen

Die Anwendung der vorgestellten automatisierten Verfahren ist dank vorprogrammierter Pakete in *R* in den meisten Fällen problemlos möglich, so dass Resultate bereits nach wenigen Augenblicken vorliegen können. Dies bedeutet Segen und Fluch zugleich. Auf der einen Seite können somit große Textsammlungen schnell ausgewertet und Informationen gezielt extrahiert werden. Zum anderen stellt die einfache Durchführung der Verfahren größere Ansprüche an die sorgfältige Zusammenstellung des Textkorpus und ein hohes technisches und statistisches Verständnis des Nutzers. Der Computer, kann die substanzielle Qualität der zur Analyse bereitstehenden Dokumente selbstverständlich nicht beurteilen. Hier sind die Expertise und Einschätzung des Forschers gefragt. Treffen die Annahmen eines Verfahrens überhaupt auf den Textkorpus zu? Welche Formatierungsschritte sollten im Voraus vorgenommen werden (z. B. wie sollen Wörter gezählt werden)? Sind Texte innerhalb des Textkorpus miteinander überhaupt vergleichbar?

Diese Fragen lassen sich vermutlich nur schwer vor Beginn einer Textanalyse beantworten. Allerdings ist es sinnvoll mit Robustheitsanalysen zu demonstrieren, inwiefern bestimmte Entscheidungen bei der Textanalyse (entfernen einzelner Dokumente, unterschiedliches Wörterzählen, Anzahl der zu schätzenden Themen) Auswirkungen auf die Ergebnisse haben können. Eine Möglichkeit, wie diese Art der Robustheitsanalyse systematisch eingesetzt werden kann, zeigen Denny und Spirling (2018) in einem Verfahren, welches automatisiert die Auswirkungen unterschiedlicher Textbearbeitungsschritte auf die Textanalyse auswertet. Dadurch erhalten Anwender die Möglichkeit, die Robustheit ihrer Analysen zu überprüfen und diejenigen Bearbeitungsschritte zu identifizieren, welche für eine signifikante Änderung der Ergebnisse verantwortlich sind.

4.3 Berücksichtigung strategischer Datengenerierungsprozesse

Eine sinnvolle Weiterentwicklung vieler computergestützten Textanalysen in der Politikwissenschaft ist die Berücksichtigung strategischer politischer Prozesse direkt während der Textanalyse (Slapin und Proksch 2014). Politische Sprache findet grundsätzlich in einem strategischen Kontext statt; allerdings berücksichtigen fast keine Textanalysemodelle diesen Aspekt explizit. Politiker sprechen beispielsweise in parlamentarischen Reden andere Zielgruppen an als in Reden auf Wahlkampfveranstaltungen oder in sozialen Medien. Die Zielsetzung von Pressemitteilungen eines Politikers unterscheidet sich vermutlich von der einer parlamentarischen Anfrage an

die Regierung. Besonders sinnvoll erscheint daher eine explizite Modellierung des Textgenerierungsprozesses, der untersucht werden soll. Für die Skalierung parlamentarischer Reden erscheint es beispielsweise sinnvoll die strategische Auswahl der Redner zu berücksichtigen (Proksch und Slapin 2012, 2015; Herzog und Benoit 2015). Vorsicht ist zudem geboten, wenn dieselbe Schätzmethode auf einen Textkorpus mit verschiedenen Dokumententypen angewendet werden soll, insbesondere wenn diese unterschiedlich generiert wurden (Bunea und Ibenskas 2015). Beispielsweise kann der Vergleich von Positionen auf der Grundlage von parlamentarische Reden und Wahlprogrammen schwierig sein, da das Publikum und der Kontext der Botschaften unterschiedlich ausfallen. Wenn diese Unterschiede nicht berücksichtigt werden, kann dies verzerrte Schlussfolgerungen zur Folge haben.

4.4 Vergleichende Politikwissenschaft und Mehrsprachigkeit

Eine nach wie vor große Hürde für die Anwendung computergestützter Textanalysen stellt der Umgang mit mehrsprachigen Dokumenten dar. Daher wird bislang in international-vergleichender Forschung mehrsprachiger Dokumente auf übersetzbare Kodierschemas und menschliche Kodierungen (wie beispielsweise beim *Manifesto Project*) zurückgegriffen. Der Einsatz von vergleichbaren Kodierschemas ist allerdings nicht immer praktikabel oder inhaltlich sinnvoll. Grundsätzlich bieten sich zwei Alternativen an, um auf der Grundlage eines mehrsprachigen Dokumentenkorpus eine Textanalyse durchzuführen. Der eine Weg führt über eine automatisierte Übersetzung aller Dokumente in eine gemeinsame Sprache und der Anwendung eines bestehenden Verfahrens (z. B. Lucas et al. 2015). Alternativ lässt sich ein bestehendes Wörterbuch in andere Sprachen übersetzen und findet somit Anwendung auf die Originaldokumente. Gerade in diesem Bereich gibt es noch reichlich Forschungsbedarf, um die multilingualen computergestützte Textanalyse auch tatsächlich zu einem festen Bestandteil der vergleichenden Politikwissenschaft machen zu können.

5 Fazit

Mit der computergestützten Textanalyse steht Politikwissenschaftlern ein immer mächtigeres Werkzeug zur Verfügung, um alte und neue Fragen aus verschiedenen Subdisziplinen zu beantworten. Zum einen werden die Methoden konstant weiterentwickelt und verfeinert, zum anderen stehen immer mehr Daten elektronisch zur Verfügung. Um die Verfahren effektiv in ein Forschungsdesign integrieren zu können, ist es ratsam, die Annahmen der zur Verfügung stehenden Schätzverfahren und Datenqualität sorgfältig zu prüfen. Viele Verfahren wurden für eine spezielle Art von Textmaterial und einen spezifischen Entstehungskontext entwickelt (z. B. für die Analyse politischer Wahlprogramme). Inwiefern solche Verfahren auf neue Textarten (z. B. Twitter-Daten) angewendet werden können, muss im Einzelnen überprüft und verteidigt werden. Die im Kapitel vorgestellten Verfahren bieten einen Einblick

in ein dynamisches Forschungsgebiet. Der Vergleich über verschiedene Dokumententypen hinweg (politische Reden, Gesetzesvorlagen, soziale Medien, Nachrichten, etc.) mag zwar ein erstrebenswertes Forschungsziel darstellen, aber in der Praxis schwierig umzusetzen sein. Unabhängig von der gewählten Methode sollten Anwender daher a) die Rolle der gewählten Textanalyse im Forschungsdesign klären, b) Robustheitsanalysen durchführen, und c) strategische Datengenerierungsprozesse berücksichtigen.

6 Kommentierte Literaturhinweise

Mehrere Aufsätze eignen sich für einen vertieften Einblick in die unterschiedlichen Textanalyseverfahren in der Sozialwissenschaft (Grimmer und Stewart 2013 für einen allgemeinen Überblick, Slapin und Proksch 2014 für Verfahren im Bereich der vergleichenden Parlaments- und Parteienforschung, DiMaggio 2015 für einen Vergleich der Textanalyse in den Sozialwissenschaften und in der Informatik). Eine Einführung in die automatisierte Datensammlung von Texten bieten Munzert et al. (2014). Für Anwender mit Interesse an der Anwendung von Wörterbüchern in einem politikwissenschaftlichen Kontext bieten sich die Aufsätze von Laver und Garry (2000) sowie Young und Soroka (2012) an. Verschiedene Textskalierungsverfahren werden vorgestellt in Aufsätzen von Laver et al. (2003), Slapin und Proksch (2008), und Lauderdale und Herzog (2016). Für einen Einstieg in Topic Modelle eignen sich Blei et al. (2003) und Roberts et al. (2014). Ein Lösungsansatz für die Anwendung von computergestützten Verfahren in einem multilingualen Kontext in der vergleichenden Politikwissenschaft findet sich in Lucas et al. (2015).

Literatur

- Atteveldt, Van, Jan Kleinnijenhuis Wouter, Nel Ruigrok, und Stefan Schlobach. 2008. Good news or bad news? Conducting sentiment analysis on Dutch text to distinguish between positive and negative relations. *Journal of Information Technology & Politics* 5(1): 73–94.
- Benoit, Kenneth. 2017. quanteda: Quantitative analysis of textual data. <https://doi.org/10.5281/zenodo.1004683>, R package version 0.99.22. <http://quanteda.io>.
- Benoit, Kenneth, und Michael Laver. 2008. Compared to What? A Comment on ‚A Robust Transformation Procedure for Interpreting Political Text‘ by Martin und Vanberg. *Political Analysis* 16(1): 101–111.
- Bernauer, Julian, und Thomas Bräuninger. 2009. Intra-Party Preference Heterogeneity und Faction Membership in the 15th German Bundestag: A Computational Text Analysis of Parliamentary Speeches. *German Politics* 18(3): 385–402.
- Blei, David, Andrew Ng, und Michael Jordan. 2003. Latent dirichlet allocation. *Journal of Machine Learning Research* 3:993–1022.
- Budge, Ian, und Paul Pennings. 2007. Do they work? Validating computerised word frequency estimates against policy series. *Electoral Studies* 26(1): 121–129.
- Bunea, Adriana, und Raimondas Ibenskas. 2015. Quantitative text analysis und the study of EU lobbying und interest groups. *European Union Politics* 16(3): 429–455.
- Collingwood, Loren, und John Wilkerson. 2012. Tradeoffs in Accuracy und Efficiency in Supervised Learning Methods. *Journal of Information Technology & Politics* 9(3): 298–318.

- Debus, Marc. 2009. Pre-Electoral Commitments und Government Formation. *Public Choice* 138(1–2): 45–64.
- Denny, Matthew J., and Arthur Spirling. 2018. Text preprocessing for unsupervised learning: Why it matters, when it misleads, and what to do about it. *Political Analysis*.
- DiMaggio, Paul. 2015. Adapting computational text analysis to social science (and vice versa). *Big Data & Society* 1–5. First published December 1, 2015.
- Egami, Naoki, Christian J. Fong, Justin Grimmer, Margaret E. Roberts, and Brandon M. Stewart. 2017. How to make causal inferences using texts. *Working Paper*.
- Flesch, Rudolf. 1948. A new readability yardstick. *Journal of Applied Psychology* 32(3): 221–233.
- Gianetti, Daniela, and Michael Laver. 2005. Policy Positions und Jobs in the Government. *European Journal of Political Research* 44:91–120.
- Grimmer, Justin. 2010. A Bayesian hierarchical topic model for political texts: Measuring expressed agendas in senate press releases. *Political Analysis* 18(1): 1–35.
- Grimmer, Justin, and Brandon Stewart. 2013. Text as Data: The Promise und Pitfalls of Automatic Content Analysis Methods for Political Texts. *Political Analysis* 21(3): 267–297.
- Hakhverdian, Armén. 2009. Capturing government policy on the left-right scale: Evidence from the United Kingdom, 1956–2006. *Political Studies* 57:720–745.
- Haselmayer, Martin, und Marcelo Jenny. 2017. Sentiment analysis of political communication: Combining a dictionary approach with crowdcoding. *Quality and Quantity* 51:2623–2646.
- Herzog, Alexander, und Kenneth Benoit. 2015. The most unkindest cuts: Speaker selection und expressed government dissent during economic crisis. *Journal of Politics* 77(4): 1157–1175.
- Klemmensen, Robert, Sara B. Hobolt, und Martin E. Hansen. 2007. Estimating policy positions using political texts: An evaluation of the wordscores approach. *Electoral Studies* 26:746–755.
- Lauderdale, Benjamin E., und Alexander Herzog. 2016. Measuring political positions from legislative speech. *Political Analysis* 24(3): 374–394.
- Laver, Michael, und John Garry. 2000. Estimating policy positions from political texts. *American Journal of Political Science* 44(3): 619–634.
- Laver, Michael, Kenneth Benoit, und John Garry. 2003. Extracting policy positions from political texts using words as data. *American Political Science Review* 97(2): 311–332.
- Lo, James, Sven-Oliver Proksch, und Jonathan B. Slapin. 2016. Ideological clarity in multiparty competition: A new measure und test using election manifestos. *British Journal of Political Science* 46(3): 591–610.
- Lowe, Will. 2008. Understanding wordscores. *Political Analysis* 16(4): 356–371.
- Lowe, Will. 2017. Putting it all on the line: Some unified theory for text scaling. *Working Paper*.
- Lucas, Christopher, Richard Nielsen, Margeret Roberts, Brandon Stewart, Alex Storer, und Dustin Tingley. 2015. Computer-assisted text analysis for comparative politics. *Political Analysis* 23(2): 254–277.
- Martin, Lanny W., und Georg Vanberg. 2008. A robust transformation procedure for interpreting political text. *Political Analysis* 16(1): 93–100.
- Munzert, Simon, Christian Rubba, Peter Meißner, und Dominic Nyhuis. 2014. *Automated data collection with R – A practical guide to web scraping und text mining*. Chichester: Wiley.
- Proksch, Sven-Oliver, und Jonathan B. Slapin. 2006. Institutions und Coalition Formation: The German election of 2005. *West European Politics* 29(3): 540–559.
- Proksch, Sven-Oliver, und Jonathan B. Slapin. 2010. Position taking in European Parliament speeches. *British Journal of Political Science* 40(3): 587–611.
- Proksch, Sven-Oliver, und Jonathan B. Slapin. 2012. Institutional foundations of legislative speech. *American Journal of Political Science* 56(3): 520–537.
- Proksch, Sven-Oliver, und Jonathan B. Slapin. 2015. *The Politics of Parliamentary Debate: Parties, Rebels, und Representation*. Cambridge: Cambridge University Press.
- Quinn, Kevin M., Burt L. Monroe, Michael Colaresi, Michael H. Crespin, und Dragomir Radev. 2010. How to analyze political attention with minimal assumptions und costs. *American Journal of Political Science* 54(1): 209–228.

- Roberts, Margeret E., Brandon Stewart, Dustin Tingley, Christopher Lucas, Jetson Leder-Luis, Shana Gadarian, Bethany Albertson, und David Rand. 2014. Structural topic models for open-ended survey responses. *American Journal of Political Science* 58(4): 1064–1082.
- Slapin, Jonathan B., und Sven-Oliver Proksch. 2008. A scaling model for estimating time-series party positions from texts. *American Journal of Political Science* 52(3): 705–722.
- Slapin, Jonathan B., und Sven-Oliver Proksch. 2014. Words as data: Content analysis in legislative studies. In *The Oxford handbook of legislative studies*, Hrsg. S. Martin, T. Saalfeld und K. Strom. Oxford: Oxford University Press.
- Soroka, Stuart. 2006. Good News and Bad News: Asymmetric Responses to Economic Information. *Journal of Politics* 68(2): 372–385.
- Soroka, Stuart. 2012. The Gatekeeping Function: Distributions of Information in Media und the Real World. *Journal of Politics* 74(2): 514–528.
- Soroka, Stuart. 2014. *Negativity in Democratic Politics: Causes and Consequences*. Cambridge Studies in Public Opinion und Political Psychology. New York: Cambridge University Press.
- Soroka, Stuart, Lori Young, und Meital Balmas. 2015a. Bad news or mad news? Sentiment scoring of negativity, fear, and anger in news content. *The Annals of the American Academy of Political and Social Science* 659(1): 108–121.
- Soroka, Stuart, Dominik A. Stecula, und Christopher Wlezien. 2015b. It's (Change in) the (future) economy, stupid: Economic indicators, the media, and public opinion. *American Journal of Political Science* 59(2): 457–474.
- Spirling, Arthur. 2016. Democratization of linguistic complexity: The effect of Franchise extension on parliamentary discourse, 1832–1915. *Journal of Politics* 78(1): 120–136.
- Wilkerson, John, David Smith, und Nicholas Stramp. 2015. Tracing the flow of policy ideas in legislatures: A text reuse approach. *American Journal of Political Science* 59:943–956.
- Young, Lori, und Stuart Soroka. 2012. Affective news: The automated coding of sentiment in political texts. *Political Communication* 29:205–231.
- Zipf, George Kingsley. 1935. *The psycho-biology of language. An introduction to dynamic philology*. Boston: Mifflin.
- Zipf, George Kingsley. 1949. *Human behavior und the principle of least effort. An introduction to human ecology*. Cambridge, MA: Addison-Wesley.