

Bachelorarbeit zur Erlangung des Hochschulgrades Bachelor of Arts (B.
A.) in Sozialwissenschaften

**Wie kann Populismus in politischen Texten von Diktionä-
ren unter Berücksichtigung des Kontextes erkannt wer-
den?**

Entwicklung eines neuen Populismus-Diktionärs

vorgelegt von:

Tom Robin Klotz
Watzmannstr. 87
71067 Sindelfingen
0174/8799379
st172750@stud.uni-stuttgart.de
Matrikel-Nr.: 3531686

Gutachter

Prof. Dr. Raphael Heiko Heiberger
Institut für Sozialwissenschaften
Abteilung Computational Social Science

Abgabedatum: 04.10.23

Inhaltsverzeichnis

Abbildungsverzeichnis	3
1. Einleitung	4
1.1 Vorgehensweise	5
2. Forschungsstand und Methodik der Populismusforschung	6
2.1 Definition.....	6
2.2 Populismus in Textdaten messen	8
2.3 Entwicklung eines Diktionärs.....	10
2.4 Kontextgewichtung.....	11
3. Methodik.....	12
3.1 Forschungsdesign.....	13
3.2 Pre-Processing	14
3.3 Cluster-Algorithmen	16
3.4 K-Dimensionen	17
3.5 Anwendung gewichteter Diktionäre: weightdictR.....	19
4. Ergebnisse	21
4.1 Konstruktvalidität.....	21
4.2 Split-Half Reliabilität.....	26
4.3 Prüfung gewichteter Patterns	28
5. Reflexion des Vorgehens.....	33
6. Fazit	37
Literaturverzeichnis	39

Abbildungsverzeichnis

Abbildung 1 Ellbogen-Kriterium	18
Abbildung 2 Ausschnitt eines gewichteten Diktionärs	19
Abbildung 3 Ausschnitt eines Resultats einer kwic-Funktion.....	20
Abbildung 4 Populismuseinschätzung der Parteien nach Experten	22
Abbildung 5 Anteil populistischer Sätze in Bundestagsreden nach Fraktionen	23
Abbildung 6 Anteil populistischer Sätze in Parteiprogrammen nach Parteien	23
Abbildung 7 Zusammenhänge zwischen Diktionären und Expertenbefragung nach Bundestagsreden	24
Abbildung 8 Zusammenhänge zwischen Diktionären und Expertenbefragung nach Parteiprogrammen.....	25
Abbildung 9 Split-Half Reliabilität auf Basis von Bundestagsreden.....	27
Abbildung 10 Split-Half Reliabilität auf Basis von Parteiprogrammen	27
Abbildung 11 Split-Half Reliabilität gekürzter Diktionäre auf Basis von Bundestagsreden	35
Abbildung 12 Split-Half Reliabilität gekürzter Diktionäre auf Basis von Parteiprogrammen.....	35
Abbildung 13 Konstruktvalidität der Diktionäre mit und ohne seltene Patterns nach Bundestagsreden	36
Abbildung 14 Konstruktvalidität der Diktionäre mit und ohne seltene Patterns nach Parteiprogrammen.....	36

1. Einleitung

Populismus ist seit langem ein Thema, das Gesellschaft, Politik und Wissenschaft beschäftigt. Seitdem Mudde 2004 seine These "A populist Zeitgeist" veröffentlicht hat, sind darauf zahlreiche Forschungsarbeiten gefolgt. Neben diesem umfassenden Forschungskorpus drückt sich die Relevanz des Themas auch in den zahlreichen Wahlerfolgen von populistischen Parteien in Europa aus. Mittlerweile ist die Problematik auch in Deutschland angekommen. Die rechts-populistische AfD schneidet in Sonntagsumfragen mit 20% als zweitstärkste Kraft ab (vgl. Politbarometer 2023). Des Weiteren ist beachtlich, dass trotz des großen Forschungskorpus die Auswirkungen von Populismus auf die Demokratie in der Theorie und Empirie noch nicht vollumfänglich geklärt sind.

In der Theorie konnten sowohl positive Effekte auf die Demokratie wie die Mobilisierung der Bevölkerung und eine verbesserte Responsivität, als auch negative Effekte wie Einschnitte in den Minderheitenschutz und Polarisierung mit Populismus assoziiert werden (vgl. Mudde und Kaltwasser 2017: 83). Die Empirie zeigt ein ähnlich gemischtes Bild. Auf der einen Seite konnten Fälle populistischer Machtübernahmen identifiziert werden, die zu einer demokratischen Erosion geführt haben wie die Wahl von Viktor Orbán zum Ministerpräsidenten 2010 in Ungarn (siehe Batory 2016) oder die Wahl von Hugo Chávez 1998 zum Präsidenten von Venezuela (siehe Hawkins 2016). Auf der anderen Seite konnten auch populistische Machtübernahmen beobachtet werden, die keine negativen Effekte auf die demokratische Stabilität hatten wie in Österreich die FPÖ in den 2000ern (siehe Heinisch und Fallend 2016) oder die Machtübernahme durch Morales in Bolivien 2006 (siehe Anria 2016). Weiter haben Huber und Ruth positive Effekte zwischen Rechts-Populismus und Wahlbeteiligung sowie Links-Populismus und Repräsentation entdeckt (vgl. Huber und Ruth 2017: 476).

Insofern wird erkennbar, dass die Folgen von Populismus für die Demokratie noch nicht ersichtlich sind und weitere Forschung vonnöten ist. Ein Forschungsstrang beschäftigt sich mit der Erfassung von Populismus in Textdaten, um das Phänomen zukünftig besser verstehen zu können. In diesem Strang wurden in den vergangenen Jahren zahlreiche Anstrengungen unternommen.

Eine dieser Anstrengungen ist die Analyse von Populismus über Wortlisten, sogenannte Diktionäre, die das Aufkommen von populistischen Wörtern und Phrasen in Texten zählen. Hierzu haben 2011 Rooduijn und Pauwels einen ersten Meilenstein mit ihrem deutschsprachigen Populismus-Diktionär gesetzt. Auf diesen Vorstoß hin

folgten weitere Entwicklungen, wie etwa Pauwels 2017 oder auch Gründl 2022, der mit seinem Populismus-Diktionär, bestehend aus 238 Einträgen, das bis dato umfangreichste Diktionär entwickelt hat. Gründl konnte in seiner Arbeit beweisen, dass sein Diktionär gegenüber den vorherigen hinsichtlich Validität und Reliabilität überlegen ist.

Dennoch besteht weiterhin ein zentrales Problem von Diktionären: Sie sind nicht in der Lage, den Kontext einer Aussage zu berücksichtigen. Aufgrund dieser ungelösten Problematik von Diktionären haben Grimmer und Stewart 2013 davon abgeraten, Analysen auf Grundlage von Diktionären durchzuführen, weil diese auf "shaky foundations" beruhen (Grimmer und Stewart 2013: 275). Daher werden in aktuellen Forschungsarbeiten andere Methoden gegenüber diktionsärbasierten Messmethoden bevorzugt (siehe Klamm et al. 2023: 1229).

Die Problematik hinter den Diktionären ist, dass diese nicht berücksichtigen können, ob ein Wort wirklich in einem populistischen Kontext steht. Gängige Verzerrungen sind Zitate, Verneinungen und thematische Kontexte. Entsprechend widmet sich diese Arbeit der Lösung dieser Problematik mit folgender Forschungsfrage:

Wie kann Populismus in politischen Texten von Diktionären unter Berücksichtigung des Kontextes erkannt werden?

Als Antwort auf das "Wie" entwickelt und testet diese Arbeit eine Methode, um langfristig die Messung von Populismus in Textdaten zu verbessern. Das Ziel besteht darin, Diktionäre wieder attraktiver für Forschungsarbeiten zu machen, da sie in der Anwendung wesentlich ressourceneffizienter und nachvollziehbarer sind als die aktuellen Methoden. Das entwickelte Diktionär dieser Arbeit soll zukünftige Forschung zum Thema Populismus unterstützen und durch eine verbesserte Genauigkeit dazu beitragen, dass die Populismusforschung zukünftig weitere Fortschritte in der Analyse der Auswirkungen des Phänomens erzielen wird.

1.1 Vorgehensweise

Zur Beantwortung der Forschungsfrage beschäftigt sich diese Arbeit zunächst mit den theoretischen Grundlagen von Populismus und Diktionären. Dem Thema Populismus wird sich über die theoretische Aufarbeitung der verschiedenen Theoriestränge genähert. Darauf folgt die theoretische Aufarbeitung der Analyse von Textdaten. Mit der Darstellung verschiedener Methoden zur Analyse von Textdaten sollen im Vergleich die Vor- und Nachteile von Diktionären herausgearbeitet werden. Diese Erkenntnisse dienen der Relevanzbegründung der Methode. Der theoretische Teil dieser Arbeit

wird mit der Ausführung der Idee der Kontextgewichtung von Diktionären abgeschlossen.

Der folgende Methodenteil dieser Arbeit dient der Erklärung der Umsetzung der Kontextgewichtung von Diktionären. Es sollen Entscheidungen im Forschungsprozess zu Pre-Processing, Algorithmen und Daten offengelegt werden, um die Entwicklung der kontextgewichteten Diktionäre so transparent wie möglich darzustellen.

Der anschließende Ergebnisteil evaluiert die Leistung der entwickelten Methode. Über die Konstruktvalidität, Reliabilität und einem Robustheitstest wird die Leistung der gewichteten Diktionäre im Vergleich zum ungewichteten Original erfasst. In dem Teil der Arbeit soll mit diesem Vergleich bewiesen werden, dass die entwickelte Methode eine valide Antwort auf die Frage nach dem “Wie“ in der Forschungsfrage ist.

Die Ergebnisse der Arbeit werden im Diskussionsteil reflektiert. Es sollen jegliche potenziell aufgetretenen Verzerrungen transparent dargelegt werden, um die Aussagekraft der Ergebnisse beurteilen zu können. Des Weiteren sollen Ideen für aufbauende Forschungsarbeiten aufgezeigt werden, die die Methode weiterentwickeln und identifizierten Verzerrungen entgegenwirken.

Das erklärte Ziel dieser Arbeit ist somit die Beantwortung der Forschungsfrage über die Entwicklung, Testung und Evaluation der Kontextgewichtung vorzunehmen. Das entwickelte Diktionär wird über mein eigens erstelltes R-Paket *weightdictR* zur einfachen Anwendung auf politische Texte aus der DACH-Region zur Verfügung gestellt.

2. Forschungsstand und Methodik der Populismusforschung

2.1 Definition

Nicht ohne Grund wird Populismus in zahlreichen Arbeiten als das umstrittenste Konzept der Sozialwissenschaften bezeichnet. In den letzten Jahren wurden zahlreiche Perspektiven diskutiert, um die “richtige” Definition von Populismus zu finden. Diese Perspektiven haben sich in der Literatur zu drei verschiedenen Ansätzen verdichtet: Dem sozio-kulturellen (siehe Ostiguy 2017), dem politisch-strategischen (siehe Weyland 2001) und dem ideellen Ansatz (siehe Mudde 2004). Diese Theoriestränge haben allesamt unterschiedliche Blickwinkel auf die Analyse von populistischer Kommunikation entwickelt (vgl. Engesser et al. 2017: 1280).

Der sozio-kulturelle Ansatz erfasst Populismus als einen Stil und fragt danach, wie Populisten sich selbst gegenüber der Bevölkerung präsentieren (vgl. Engesser et al. 2017: 1280). In dieser Perspektive rückt der Stil der Kommunikation, also Aspekte wie Körpersprache, Wortwahl und auch der individuelle Kleidungsstil in den

Vordergrund der empirischen Analyse (vgl. Ostiguy 2017: 77). Neben dieser sozio-kulturellen Dimension ist die politisch-kulturelle Komponente die zweite Dimension des sozio-kulturellen Ansatzes. Untersuchungsgegenstand dieser Dimension sind sowohl Führungsstile als auch die Art und Weise, wie politische Entscheidungen herbeigeführt werden (vgl. Ostiguy 2017: 81). Die Verwendung eines bestimmten Stils spricht bestimmte Teile der Gesellschaft an und andere weniger. Dies führt zu einer Mobilisierung und Spaltung von Teilen der Gesellschaft nach der Logik „Ich will diese Leute nicht in der Regierung haben“ und „Ja, ich kann mich mit [*] identifizieren“ (vgl. Ostiguy 2017: 84).

Der politisch-strategische Ansatz versteht Populismus als ein Mittel zum Zweck, um Regierungsmacht zu erreichen und auszubauen. Kern dieser Perspektive sind Methoden und Instrumente zur Machtübernahme und -ausübung (vgl. Weyland 2001: 12). Die Auswahl und Möglichkeiten dieser Methoden und Instrumente hängt davon ab, welcher politische Akteur (eine Person, eine informelle Gruppe oder eine formale Organisation) im Zentrum der Handlung steht und welche Machtressourcen wie Wählerstimmen oder militärischer Einfluss diesem Akteur zur Verfügung stehen (vgl. Weyland 2017: 55). Der Untersuchungsgegenstand dieser Perspektive sind Motive und Ziele populistischer Kommunikation. Es geht darum, zu analysieren, „Warum“ und mit welchem Ziel bestimmte Kommunikationsstrategien angewendet werden (vgl. Engesser et al. 2017: 1280).

Nach dem ideellen Ansatz ist Populismus eine Ideologie, die das „reine Volk“ und die „korrupte Elite“ als homogene Gruppen in einem antagonistischen Verhältnis betrachtet (vgl. Mudde 2004: 543). Diese Denkweise erzeugt eine klare Abgrenzung durch eine Unterscheidung in „wir“ gegen „die“ bzw. „die da oben“ (vgl. Klamm et al. 2023: 1228). Untersuchungsgegenstand dieses Ansatzes sind die Inhalte der populistischen Ideologie. Es geht darum, „Was“ kommuniziert wird (vgl. Engesser et al. 2017: 1280). Die populistische Ideologie bezeichnet Mudde als „thin-centred“, da sie in ihrem Umfang mit großen Ideologien wie dem Sozialismus und dem Liberalismus nicht auf einer Ebene steht. Dennoch kann Populismus laut Mudde mit anderen Ideologien kombiniert werden (vgl. Mudde 2004: 543f.). Das bedeutet, dass populistische Elemente in größere ideologische Rahmen integriert werden können, um politische Ziele zu verfolgen. Beispielsweise können populistische Tendenzen in der linken Politik mit sozialistischen Ideen oder in der rechten Politik mit konservativen Ideen kombiniert werden (vgl. Mudde 2017: 11).

Der Untersuchungsgegenstand dieser Arbeit sind Textdaten, welche als Limitation nicht das „Warum“ und nur in Teilen das „Wie“ hinter populistischer Kommunikation

abbilden, weshalb vorrausgegangene empirische Arbeiten auf Basis von Diktionären (siehe Rooduijn und Pauwels 2011, Pauwels 2017 und Gründl 2022), Populismus aus der Sicht des ideellen Ansatzes als eine Ideologie definieren. Darüber hinaus geht es in dieser Arbeit darum, herauszufinden, inwiefern bestimmte Wörter populistische Inhalte abbilden, weshalb sich diese Arbeit mit dem, "Was" hinter populistischer Kommunikation beschäftigt und entsprechend in der Tradition des ideellen Ansatzes steht.

An dieser Stelle soll angemerkt werden, dass Muddes Darstellung von Populismus als eine kombinierbare Ideologie in der Erstellung des Diktionärs nicht berücksichtigt wurde, weil das entwickelte Diktionär nicht den Anspruch erhebt zwischen rechtem und linkem Populismus zu unterscheiden.

2.2 Populismus in Textdaten messen

In der Literatur wurden bisher vier verschiedene Ansätze, manuelle Codierung (siehe Jagers und Walgrave 2007), holistische Codierung (siehe Hawkins 2009), Diktionäre (siehe Gründl 2022) und überwachtes maschinelles Lernen (siehe Klamm et al. 2023) angewendet, um Populismus in Textdaten zu erfassen. Diese Ansätze sollen mit ihren Stärken und Schwächen vorgestellt werden.

Manuelle Codierung

Der Standard für Textanalysen in den Sozialwissenschaften ist die manuelle Codierung (Inhaltsanalyse) von Texten (vgl. Quinn et al. 2010: 212). In der Vorarbeit wird für das Forschungsvorhaben ein Codebook erstellt, welches das zu interessierende Konzept über ein oder mehrere Dimensionen erfassen soll. Auf dessen Basis werden Coder trainiert, die dem Codebook entsprechende Codierungen der Dokumente in der Auswertung vornehmen (vgl. Quinn et al. 2010: 212; vgl. Pauwels 2017: 125). Die Auswertung erfolgt über die Aggregation der verteilten Codes auf Satz- oder Paragraphebene der Dokumente (vgl. Pauwels 2017: 125).

Auf der einen Seite erreicht die Methode eine hohe Inhaltsvalidität, wie Rooduijn und Pauwels 2011 anhand von Parteiprogrammen aufzeigen konnten (siehe Rooduijn und Pauwels 2011). Auf der anderen Seite ist diese Methode nicht skalierbar, da mit der zunehmenden Anzahl an Dokumenten die Kosten und der zeitliche Aufwand für die Auswertung der Daten steigen (vgl. Grimmer und Stewart 2013: 274).

Holistische Codierung

Holistische Codierung ist ein ähnlicher Ansatz, bei dem Coder trainiert werden, um ein ganzes Dokument auf Basis von Anker-Beispielen zu klassifizieren. Anders als bei der manuellen Codierung wird nach der holistischen Codierung Populismus auf

Dokumentebene analysiert (vgl. Hawkins 2009: 1049). Diese Vorgehensweise ist nicht nur ähnlich aufwendig wie die manuelle Codierung, sondern auch weniger feingliedrig. Des Weiteren ist diese Methode in der Klassifizierung weniger transparent, weil die Klassifizierungsgründe auf der Dokumentebene oft nicht ersichtlich sind im Vergleich zur manuellen Codierung (vgl. Klamm et al. 2023: 1288f.).

Überwachtes maschinelles Lernen

Überwachtes maschinelles Lernen und Diktionäre gehören zu den computergestützten Methoden. Im Vergleich zu den vorherigen Methoden sind diese wesentlich ressourceneffizienter, weshalb solche Ansätze durch die stetig steigende Verfügbarkeit an digitalen Textdaten an Bedeutung gewonnen haben (vgl. Grimmer und Stewart 2013: 271).

Das Ziel von überwachtem maschinellen Lernen ist die Automatisierung der manuellen Codierung (vgl. Grimmer und Stewart 2013: 279). Für die Anwendung wird der Datensatz in Test- und Trainingsdaten gespalten. Die im Trainingsdatensatz enthaltenen Dokumente werden manuell codiert. Diese Codierungen werden genutzt, um einem Computer die Beziehungen zwischen Wörtern und Kategorien beizubringen (vgl. Grimmer und Stewart 2013: 276f.). Der Computer ist anschließend in der Lage die Codierung auf die restlichen Dokumente skalierbar anzuwenden. Auf diese Art und Weise können Kosten- und Zeitressourcen gespart werden, weil nur noch eine geringere Anzahl an Dokumenten händisch codiert werden muss.

Diktionäre

Diktionärbasierte Messmethoden folgen einem deduktiven Prinzip, nach dem im Voraus Wörter oder Phrasen festgelegt werden, die ein theoretisches Konzept erfassen sollen (vgl. Gründl 2022: 1486; vgl. Quinn et al 2010: 212). Diese Wörter und Phrasen werden in einem Diktionär zusammengefasst und können in der Überprüfung am empirischen Material induktiv weiterentwickelt werden (vgl. Gründl 2022: 1487). Über Diktionäre wird in der Anwendung gezählt, wie viele Matches die Patterns eines Diktionärs in den Textdaten haben, welche flexibel auf verschiedenen Ebenen wie zum Beispiel Parteien oder auf Zeithorizonte je nach Kontext und Forschungsinteresse aggregiert werden können.

Die Anwendung solcher Wortlisten ist die ressourceneffizienteste Methode, da sie die komplette händische Codierung ersetzt und daher auf eine große Menge an Textdaten skalierbar angewendet werden kann (vgl. Quinn et al. 2011: 212). Das größte Problem ist dabei, wie anfangs bereits ausgeführt (siehe 1. Einleitung), dass Diktionäre den Kontext nicht berücksichtigen (vgl. Grimmer und Stewart 2013: 275; vgl.

Klamm et al. 2023: 1229). Daher erzeugen Diktionäre in der Anwendung false-positive Fälle. Zum Beispiel haben die Patterns “Korruption“ und “Monopol“ in Pauwels Populismus-Diktionär bei der Analyse der Parteiprogramme der Grünen und der Piraten zu Verzerrungen geführt, weil diese Parteien die Begriffe nicht im populistischen Kontext gegen die Elite verwendet haben, sondern im Kontext der allgemeinen Bekämpfung von Korruption und Monopolen im Land (vgl. Pauwels 2017: 129). Deshalb raten Grimmer und Stewart dazu, Diktionäre nur für den spezifischen Zweck anzuwenden, für den ein Diktionär erstellt wurde, und nicht auf andere Fälle zu übertragen, weil nicht gewährleistet ist, dass die Kontexte anderer Fälle mit dem ursprünglichen Zweck, für den ein Diktionär erstellt wurde, übereinstimmen (vgl. Grimmer und Stewart 2013: 274f.).

Grundsätzlich erscheinen Diktionäre und überwachtes maschinelles Lernen aufgrund ihrer Skalierbarkeit mit größeren Datenmengen am vielversprechendsten. In der gegenwärtigen Populismus-Literatur wird intensiv an Messmethoden auf Grundlage von überwachtem maschinellern Lernen geforscht. Diese sind theoretisch den Diktionären überlegen, weil sie den Kontext berücksichtigen und damit auf verschiedene Fälle angewendet werden können (vgl. Klamm et al. 2023: 1299f.). Andererseits ist die Entwicklung dieser Methoden sehr rechenintensiv und teilweise sind manuelle Codierungen als Vorarbeit notwendig, weshalb diese Methode für diese studentische Arbeit und vergleichbare Projekte nur begrenzt anwendbar ist. Insofern können Diktionäre, wenn die Verzerrung durch Kontexte reduziert wird, eine pragmatische Alternative für Forschungsvorhaben mit geringen Ressourcen gegenüber dem überwachten maschinellen Lernen darstellen. Um diese Nische auszufüllen und niederschwelligen Arbeiten zum Thema Populismus eine inhaltlich angemessene Methode zur Verfügung zu stellen, entwickelt diese Arbeit ein neues kontextgewichtetes Populismus-Diktionär. Welche theoretischen Überlegungen bei der Entwicklung eines Diktionärs aufkommen, werden im Folgenden ausgeführt.

2.3 Entwicklung eines Diktionärs

Die Entwicklung eines Diktionärs ist ein herausforderndes Unterfangen, da die entwickelte Wortliste sowohl theoretisch als auch empirisch reflektiert werden muss (vgl. Quinn et al. 2010: 212). Bei der Entwicklung von Diktionären müssen zwei Faktoren beachtet werden. Erstens, dass ein Diktionär eine hohe Präzision aufweist und entsprechend wenige nicht-populistische Sätze als Populismus identifiziert (false-positive). Zweitens, dass ein Diktionär alle populistischen Inhalte erfasst und nur wenige populistische Sätze als nicht-populistisch identifiziert (false-negative). Diese zwei Kriterien stehen im Konflikt, da die Erhöhung der Patterns die false-negative Fälle

reduziert aber gleichermaßen einen Anstieg an false-positiven Fällen mit sich zieht und umgekehrt (vgl. Gründl 2022: 1487). Die in dieser Arbeit vorgestellte Methodik entwickelt ein neues Diktionär auf Grundlage von den Patterns aus Gründls Diktionär. Gründl hat in der Erstellung seines Diktionärs darauf geachtet, besonders viele Patterns aufzunehmen, um ein umfassendes Populismus-Diktionär zu erstellen. Im Vergleich zu seinen Vorgängern Rooduijn und Pauwels 2011 mit 20 Einträgen (siehe Rooduijn und Pauwels 2011: 1283) und Pauwels 2017 mit 37 Einträgen (siehe Pauwels 2017: 134) ist Gründl dies mit 238 Einträgen gelungen. Insofern sind false-negative Fälle ein zu vernachlässigendes Problem des hier erstellten Diktionärs. Stattdessen stellen die false-positive Fälle eine Herausforderung dar. Diese Herausforderung soll über die Berücksichtigung des Kontextes durch die Kontextgewichtung bewältigt werden, in dem Patterns die false-positive Fälle verursachen, eine geringe Gewichtung erhalten. Die konkrete Logik dahinter wird im folgenden Kapitel erläutert.

2.4 Kontextgewichtung

Zunächst muss festgestellt werden, dass es ein unmögliches Unterfangen ist, einem Diktionär ein Kontextverständnis beizubringen. Es kann aber eine näherungsweise Berücksichtigung des Kontextes über eine Gewichtung der Patterns in ein Diktionär integriert werden. Im Folgenden werden die theoretischen Überlegungen hinter dieser Idee dargelegt.

Ausgangspunkt der Überlegungen hinter diese Methode ist, dass jedes Pattern eines Diktionärs false-positive Fälle produziert. Die grundlegende Annahme dieser Idee ist, dass diese Anzahl an false-positive Fällen von Pattern zu Pattern variiert.

- 1. Annahme: Es besteht ein großes Gefälle in der Anzahl an false-positive Fällen, die ein Pattern produziert.*

Die zweite Annahme folgt der Logik der gewichteten Sentiment-Diktionäre. Diese bewerten bestimmte Wörter positiver bzw. negativer als andere, weil nicht alle Wörter gleichwertige Indikatoren für positive bzw. negative Sprache sind. Dies lässt sich auch auf das Konzept Populismus übertragen.

- 2. Annahme: Die Patterns eines Diktionärs sind keine inhaltlich gleichwertigen Indikatoren für das theoretische Konzept.*

Zum Beispiel ist das Pattern "Volksverräter" populistischer als das Pattern "direkte Demokratie". Das Pattern "Volksverräter" schafft eine einfache Schwarz-Weiß-Dichotomie, in der diejenigen, die nicht mit der populistischen Agenda übereinstimmen, als Feinde des Volkes dargestellt werden. Der Begriff zielt darauf ab, Emotionen zu schüren, indem er eine starke Abneigung und Misstrauen gegenüber denen erzeugt, die

nicht den vermeintlichen "wahren" Interessen der Bevölkerung folgen. Dahingegen ist das Pattern "direkte Demokratie" weniger populistisch, da es nicht direkt darauf abzielt, politische Gegner zu diskreditieren oder Emotionen zu schüren, sondern vielmehr den Fokus auf die Stärkung der Demokratie und die Förderung einer breiten Beteiligung der Bürger legt. Des Weiteren ist zu erwarten, dass das Wort "direkte Demokratie" mehr false-positive Fälle erzeugt, weil es eine grundsätzlich legitime Forderung ist und in sachlichen Debatten über die Demokratie auftreten kann. Dem entgegen ist nicht zu erwarten, dass das Wort "Volksverräter" in sachlichen Debatten oder ähnlichem auftritt und daher weniger false-positive Fälle erzeugt.

Diese beiden Patterns sind in Gründls Populismus-Diktionär enthalten und werden als gleichwertige Indikatoren für das Konzept Populismus verstanden. Nach der vorherigen Argumentation ist diese Anwendung inhaltlich nicht angemessen und eine Vereinfachung.

Um die Faktoren aus Annahme 1 und 2 zu berücksichtigen, wird eine Gewichtung der Patterns vorgenommen. Zur Umsetzung dieser Gewichtung müssen die Kontexte, in denen die Patterns eines Diktionärs in Texten auftreten, untersucht werden. Damit diese Untersuchungen zu aussagekräftigen Ergebnissen kommen, müssen große Mengen an Textdaten analysiert werden. Entsprechend bieten sich Methoden wie die klassische Inhaltsanalyse oder die Erweiterung durch überwachttes maschinelles Lernen für die Untersuchung nicht an, weil diese in der Umsetzung zu aufwendig für den zur Verfügung stehenden Ressourcenumfang in dieser Arbeit sind. Stattdessen werden Cluster-Algorithmen angewendet, um mit der großen Menge an Daten umzugehen. Diese Cluster-Algorithmen können Textdaten gruppieren, wodurch der Klassifikationsaufwand reduziert wird, weil nur wenige Cluster anstatt zahlreicher Textstellen händisch klassifiziert werden müssen.

Die konkrete Umsetzung der Kontextgewichtung, insbesondere der Anwendung von Cluster-Algorithmen, wird im folgenden Abschnitt zur Methodik dieser Arbeit ausführlich erläutert. Durch diese Methodik kann der Kontext in der Anwendung von Diktionären indirekt miteinfließen, wodurch eines der in der Literatur viel kritisierten Herausforderungen von Diktionären angegangen werden kann.

3. Methodik

In diesem Kapitel wird zunächst das Forschungsdesign dieser Arbeit vorgestellt. Die einzelnen Schritte des Forschungsdesign werden in den darauffolgenden Unterkapitel weiter diskutiert.

3.1 Forschungsdesign

Zur Beantwortung der Forschungsfrage soll über die quantitative Analyse großer Mengen an Textdaten ein gewichtetes Diktionär erstellt werden, welches im Vergleich zu regulären Diktionären besser abschneidet. Das gewichtete Diktionär wird auf Basis des bereits existierenden Diktionärs von Gründl erstellt. Ziel der Gewichtung ist die Erstellung einer Skala von 0 bis 1, die erfasst, wie stark ein bestimmtes Pattern als Indikator für Populismus fungiert. Für die Erstellung dieser Skala wurden Bundestagsreden als Datengrundlage genutzt. Diese sind auf Grund ihrer umfangreichen Menge an verfügbaren Textdaten besonders gut für das Vorhaben dieser Arbeit geeignet.

Zur Erstellung der Diktionäre muss zunächst ein Pre-Processing durchgeführt werden, um zu vernachlässigende Informationen aus den Textdaten zu filtern. Daraufhin wird über die “key-word-in-context” Funktion aus dem Quanteda-Paket¹ der Kontext in Form von fünf Wörtern vor und nach einem populistisch identifizierten Fall erhoben. Diese fünf Wörter werden jeweils zu einem Vektor zusammengefügt und nochmals nach zu vernachlässigenden Informationen gefiltert. Dieser bereinigte Kontextvektor dient als Grundlage für die Anwendung von Cluster-Algorithmen. Es werden zwei Cluster-Algorithmen in zwei verschiedenen Varianten angewendet, da ausgehend von den Daten nicht absehbar ist, welche Variante am vielversprechendsten ist. Daher werden in der Folge vier gewichtete Diktionäre erstellt.

Für die Validierung der erstellten kontextbasierten Diktionäre müssen diese auf Textdaten angewendet werden. Für die Anwendung eines solchen Diktionärs existiert noch keine Methode in R, weshalb diese Arbeit eine Methode über das zu dieser Arbeit entwickelte Paket *weightdictR* zur Verfügung stellt. Die Validität der gewichteten Diktionäre wird anhand der Konstruktvalidität den Zusammenhang mit Expertenbefragungen überprüft. Je höher der Zusammenhang eines Diktionärs mit der Experteneinschätzung, desto besser wird die Performance des Diktionärs eingeordnet.

Zur Überprüfung der kontextübergreifenden Anwendung dieser Diktionäre, werden die Ergebnisse auf Grundlage von Bundestagsreden und Parteiprogrammen erhoben und verglichen. Über den Vergleich soll beantwortet werden, ob die Berücksichtigung des Kontextes über die Kontextgewichtung zu einer verbesserten Erkennung von Populismus in Textdaten führt. Die Umsetzung der einzelnen Aspekte des Forschungsdesigns werden in den folgenden Kapiteln erörtert.

¹ Quanteda ist ein weitverbreitetes R-Paket für die Durchführung von Textanalysen, auf dessen Grundlage sowohl die Auswertungen von Gründl als auch die Auswertungen dieser Arbeit durchgeführt werden.

3.2 Pre-Processing

Pre-Processing bezeichnet den Prozess der Datenbereinigung für eine anschließende Datenanalyse. Dieser Prozess ist essenziell für die Analyse von Textdaten, da diese in einem unstrukturierten Format vorliegend. Zur Strukturierung werden die Daten über verschiedene Methoden runtergebrochen, sodass die Daten in dem von der Auswertungsmethode geforderten Format vorliegen. Ein klassisches Vorgehen ist dabei das Entfernen von Satzzeichen, Stoppwörtern und eine Vereinfachung der grammatikalischen Wortstrukturen durch Stemming oder Lemmatization. Stoppwörter sind häufig vorkommende Wörter in Texten, die in der Regel keinen inhaltlichen Mehrwert bieten und oft aus grammatikalischen Gründen vorkommen. Diese werden über Stoppwortlisten der jeweiligen Sprachen entfernt (vgl. Manning et al. 2009: 27). Stemming und Lemmatization sind zwei komplexere Verfahren, die eine knappe Erläuterung von Nöten haben. Stemming und Lemmatization sind linguistische Algorithmen, die darauf abzielen Flexionsformen eines Wortes zu einer gemeinsamen Grundform zu vereinfachen. Stemming entfernt Wortendungen, sogenannte "Suffixe", um den Wortstamm zu extrahieren (vgl. Manning et al. 2009: 32). Dabei entfernt Stemming auch "Ableitungssuffixe", wodurch derivativ verwandte Wörter auf eine gemeinsame Grundform zurückgeführt werden (vgl. Manning et al. 2009: 32). Zum Beispiel wird aus "experiment" und "experience" jeweils "experi". Lemmatization hingegen führt Wörter über die Entfernung ihrer Flexionsformen auf ihre lexikalische Form zurück (vgl. Manning et al. 2009: 32). Das bedeutet, dass Lemmatization an den Wörtern "experiment" und "experience" keine Änderungen vornimmt, weil keine Flexionsformen wie ein Plural "s" vorhanden sind.

Als Datengrundlage für die Berechnung der Kontextgewichtung wurden die Bundestagsreden aus den vergangenen vier Wahlperioden der 17., 18., 19. und 20. herangezogen. In den Perioden 17 und 18 ist mit der LINKE und in der 19. und 20. Ist zusätzlich mit der AfD mindestens eine populistische Partei im Bundestag vertreten.

Das Pre-Processing der Bundestagsreden erfolgt über zwei Schritte. Als erstes werden die Daten für die Anwendung von Gründls Diktionär vorbereitet. Hierzu werden alle Satzzeichen entfernt, Großbuchstaben zu Kleinbuchstaben transformiert und in bestimmten Multiword-Phrasen, die eigentlich von Gründls Diktionär erfasst werden sollen, müssen die Leerzeichen durch Unterstriche ersetzt werden. Zum Beispiel

„ja wollen sie jetzt die aktionäre schützen oder die steuerzahler“

Mit dem pattern: „(wollen|(ein)?fordern|möchten|mögen|verlangen|beanspruchen|wünschen)
(.*)?steuerzahler(innen)?“

Damit dieser Satz von dem zugehörigen Pattern als ein Match erkannt wird, müssen das Pattern und der Text entsprechend transformiert werden:

„ja_wollen_sie_jetzt_die_aktionäre_schützen_oder_die_steuerzahler“

Mit dem Pattern: „(wollen|(ein)?fordern|möchten|mögen|verlangen|beanspruchen|wünschen)(.)?steuerzahler(innen)?“*

Ohne diese Transformation würde Gründls Diktionär nicht diese Fälle erfassen. Dies liegt daran, dass das Quanteda-Paket im Hintergrund mit Patterns, die aufgrund von Wildcards `(.*)?` oder Regex mehrere verschiedene Strings matchen, nicht richtig umgeht.

Auf diese Problematik ist auch Gründl in der Anwendung seines Diktionärs gestoßen und hat mit der Veröffentlichung seines R-Paktes `“multidictR“` eine Funktion für diese Transformation zur Verfügung gestellt². Des Weiteren werden Begrüßungsfloskeln in den Reden entfernt, weil diese keine bedeutenden Informationen für einen populistischen Kontext liefern. Eingriffsintensive Verfahren wie Stemming oder Lemmatization oder das Entfernen von Stoppwörtern werden in diesem Schritt nicht angewendet, weil diese die Identifikation von populistischen Fällen über Gründls Diktionär verzerren. Zum Beispiel wurde durch Stemming aus dem Wort `“demokratie“` `“demokrati“`, wodurch Patterns, die auf dem Wort `“Demokratie“` basieren, wie `“direkt(e|er)?demokratie“`, nicht korrekt matchen. Zudem berücksichtigt das Diktionär verschiedene grammatikalische Endungen und Formen über Regexausdrücke, weshalb Stemming oder Lemmatization nicht zu einer verbesserten Leistung des Diktionärs führen.

Nach der Bereinigung und Aggregation der Kontextdaten werden eingriffsintensivere Verfahren angewendet. Ziel dieses Schrittes ist es, die Kontextdaten für die Anwendung von Cluster-Algorithmen vorzubereiten. Durch das Entfernen von Stoppwörtern wird die Berechnung der Nähe beziehungsweise der Distanz der Kontexte verbessert, sodass diese nicht von inhaltlich unbedeutenden Wörtern verzerrt wird. Darauf zielt ebenfalls die Anwendung von Lemmatization ab, indem der Cluster-Algorithmus besser inhaltlich gleiche Wörter entdeckt. Lemmatization wurde für diesen Fall gegenüber Stemming bevorzugt, da Lemmatization grundsätzlich genauere Resultate erzielt, die für die inhaltliche Leistung der Cluster-Algorithmen wichtig sind (vgl. Manning et al. 2009: 32). Die Anwendung der Cluster-Algorithmen wird im folgenden Kapitel ausgeführt.

² Das bis hierhin vorgestellte Vorgehen wurde ebenfalls für das Pre-Processing der Parteiprogramme angewendet.

3.3 Cluster-Algorithmen

Cluster-Analysen sind statistische Verfahren, die Daten gruppieren, sodass die Distanz zwischen den Gruppen maximiert und die Distanz innerhalb der Gruppen minimiert wird (vgl. Ahlquist und Breunig 2012: 95). Um diese Distanz zwischen Textdaten darzustellen, wird eine Distanzmatrix berechnet. Diese Berechnung kann über verschiedene Distanz- bzw. Ungleichheitsmaße umgesetzt werden, deren Auswahl von den anzuwendenden Cluster-Algorithmen und der Datengrundlage abhängt. In dieser Arbeit wurde die euklidische Distanz für die Berechnung der Distanzmatrix genutzt.

Zur Durchführung einer Cluster-Analyse stehen verschiedene Algorithmen zur Verfügung. Abhängig von der Datengrundlage bieten sich bestimmte Algorithmen für die Durchführung der Analyse mehr und andere weniger an. Generell werden Cluster-Methoden vor allem in explorativen Analysen angewendet, weshalb es gängige Forschungspraxis ist auf die gleichen Daten verschiedene Algorithmen anzuwenden (vgl. Kaufmann und Rousseeuw 2005: 37). Im Folgenden werden die behandelten Algorithmen dieser Arbeit vorgestellt und ihre Auswahl begründet.

K-Means

Der K-Means Algorithmus ist der am weitesten verbreitete Cluster-Algorithmus in der Literatur (vgl. Shukla und Naganna 2014: 1848). Das Ziel von K-Means ist die Reduzierung der durchschnittlichen Distanz zwischen Objekten und dem Mean bzw. dem Centroid eines Clusters (vgl. Manning et al. 2009: 360). Bei der Anwendung von K-Means wird zunächst eine Anzahl an K-Centroids bestimmt, die zufällig ausgewählt werden. Die verbleibenden Objekte im Datensatz werden zu den jeweiligen am nächsten gelegenen Centroids zugeordnet. Für die resultierenden Cluster wird die durchschnittliche Distanz berechnet. Dieser Vorgang wird für eine im Voraus festgelegte Anzahl an Iterationen durchgeführt, um die Variation innerhalb der Cluster (total within-cluster variation kurz wss) sukzessiv zu reduzieren. Als Resultat werden die Clusteraufteilungen ausgegeben, bei der die durchschnittliche Distanz am geringsten ausfällt (vgl. Shukla und Naganna 2014: 1848f.; vgl. Manning et al. 361f.). Eine Limitation des Algorithmus ist, dass er durch die Berechnung des Means als Centroid, anfällig gegenüber Ausreißern ist (vgl. Shukla und Naganna 2014: 1853).

CLARA

Clustering Large Applications (kurz Clara), ist eine Weiterentwicklung des Partitioning Around Medoids (kurz PAM) Algorithmus. Der PAM-Algorithmus versucht ebenfalls die durchschnittliche Distanz wie der K-Means Algorithmus zu reduzieren, jedoch nicht die Distanz zum Centroid, sondern die Distanz zum Medoid, dem zentralsten

Objekt eines Clusters (vgl. Kaufmann und Rousseeuw 2005: 40). Ansonsten folgt der PAM-Algorithmus einem ähnlichen Schema wie K-Means. Es werden zunächst zufällig K-Medoids bestimmt, denen die am nächsten gelegenen Objekte zugeordnet werden. Ausgehend von der Reduzierung der durchschnittlichen Distanz der Objekte zu ihrem Medoid wird das Resultat bewertet. Dieser Vorgang wird für eine im Voraus bestimmte Anzahl an Iterationen durchgeführt und als Ergebnis wird die Clusterteilung mit der geringsten wss zurückgegeben (vgl. Kaufmann und Rousseeuw 2005: 41). Ein Nachteil des PAM-Algorithmus ist, dass er nicht für eine große Anzahl an Fällen konzipiert ist. Dieses Problem löst der CLARA-Algorithmus, welcher auf derselben Logik basiert. CLARA berechnet die Cluster mit Stichproben, wodurch nicht der ganze Datensatz in den Speicher geladen werden muss (vgl. Kaufmann und Rousseeuw 2005: 126). Entsprechend bietet sich der CLARA-Algorithmus für das Vorhaben dieser Arbeit an.

Im Vergleich zum K-Means-Algorithmus, sind diese K-Medoid Methoden weniger anfällig gegenüber Ausreißern (vgl. Kaufmann und Rousseeuw 2005: 41). Das kann Vor- und Nachteile für die Analyse mit sich bringen. Zum aktuellen Zeitpunkt ist noch nicht absehbar, inwiefern die Kontextdaten von populistischen bzw. nicht-populistischen Ausreißern geprägt sind, weshalb die Anwendung beider Cluster-Verfahren sich als Konsequenz ergibt.

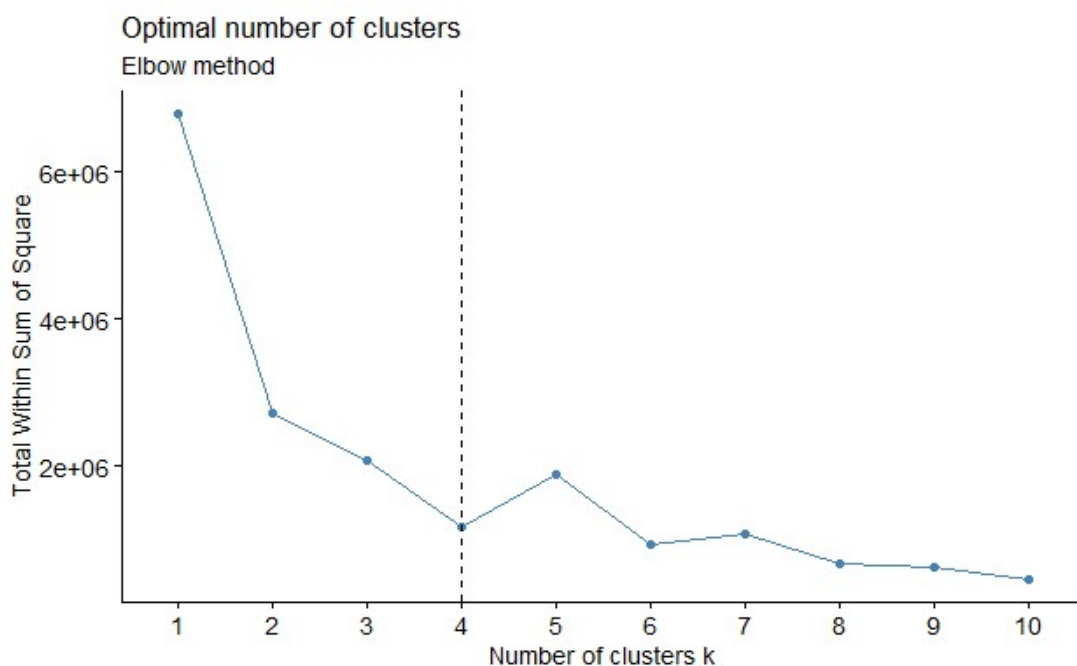
Hierarchische Clusterverfahren wurden nicht in Erwägung gezogen, weil diese für eine große Menge an Daten, wie in dieser Arbeit, nicht vorgesehen sind. Außerdem konnte ein kurzer Test in R beweisen, dass die Datenmenge zu groß ist, sodass ein entsprechender Error als Ergebnis eines hierarchischen Cluster-Algorithmus zurückgegeben wurde.

3.4 K-Dimensionen

Bei der Anwendung von K-Means und CLARA ist zu beachten, dass im Voraus die Anzahl an K-Dimensionen festgelegt werden muss. Diese Entscheidung kann sehr herausfordernd sein, da es kein richtig oder falsch gibt, weil diese Entscheidung von den Daten und dem Ziel der Auswertung abhängig ist. Daher wird hier ähnlich wie bei der Wahl der Cluster-Algorithmen explorativ vorgegangen. Ziel der Auswertung ist es, die Kontexte in populistische und nicht-populistische Kontexte zu klassifizieren, so dass die Populismuskewichtung der Patterns berechnet werden kann. Entsprechend ist naheliegen, die Algorithmen mit einem k von 2 anzuwenden, um ein populistisches und ein nicht-populistisches Cluster zu identifizieren.

Eine weitere Möglichkeit, um k zu bestimmen, ist die Berechnung der Variation innerhalb der Cluster (wss). Anhand der graphischen Darstellung der wss nach der Anzahl k -Cluster soll die Stelle gefunden werden, an der die sukzessive Abnahme der wss auffallend kleiner wird. Demnach ist diese Anzahl an k -Cluster die optimale Anzahl an Clustern (vgl. Manning et al. 2009: 365). Dieses Kriterium wird auch Ellbogen-Kriterium genannt, weil man nach dem Ellbogen in der Grafik sucht. Eine beispielhafte Grafik für die Anwendung von K-Means auf Basis der Kontextdaten von der 19. Wahlperiode ist in Abbildung 1 dargestellt. In Abbildung 1 lässt sich dieser Ellbogenpunkt an der Stelle $k = 4$ feststellen.

Abbildung 1 Ellbogen-Kriterium



Quelle: Eigene Berechnung auf Basis von Open Discourse

Dieses Vorgehen ist die zweite Variante, auf deren Basis die Cluster-Algorithmen angewendet werden. Somit werden vier gewichtete Diktionäre erstellt, K-Means ($k = 2$), K-Means ($k = \text{true } k$), CLARA ($k = 2$) und CLARA ($k = \text{true } k$).

Die Cluster-Algorithmen werden für jede Bundestagsperiode einzeln angewendet, da zu erwarten ist, dass abhängig von der Legislaturperiode Veränderungen in der politischen Kommunikation im Bundestag auftreten, die die Ergebnisse der Clusterung in populistische und nicht-populistische Gruppen verfälschen könnten. Die Bestimmung der Cluster als populistisch bzw. nicht-populistisch erfolgt für $k = 2$ über die Anteile von Kontexten aus populistischen Fraktionen, das bedeutet über die Verteilung der Beiträge der LINKEN und der AfD. Beispielsweise wurde für die 18. Wahlperiode nach K-Means 80% der Beiträge der LINKE Cluster 1 und 20% der Beiträge der LINKE

Cluster 2 zu geordnet. Dementsprechend wurde Cluster 1 als populistisch und Cluster 2 als nicht-populistisch klassifiziert. Für $k = \text{true}$ k wurde zusätzlich neben der Verteilung der Beiträge populistischer Parteien auch die graphische Verteilung der Cluster mitberücksichtigt.

Nach erfolgreichem Abschluss der Klassifizierungsphase werden die Resultate aller Wahlperioden zusammengefügt und auf Ebene der Patterns aggregiert. Die Gewichtung der Patterns erfolgt über den Mean. Als Resultat liegt ein Datensatz vor mit allen Patterns, die auch in den Kontexten vorgekommen sind, und ihre Gewichtung. Dieser Datensatz sieht folgendermaßen aus für das Resultat des K-Means-Algorithmus mit $k = 2$:

Abbildung 2 Ausschnitt eines gewichteten Diktionärs

	pattern	weight
1	\b(a ä)ngst(e)??(de(s r) eine(s r) unsere(s r))_bürger(s innen i...	0.4285714
2	\baberwitzig(e er es en em)??\b	0.6434783
3	\babgehoben(e er es en em)??\b	0.5816993
4	\ban_der_nase_herumführ(t en)\b	0.6785714
5	\bangeblich(e er en)??_[^[:space:]]*(?)partei(en)??\b	0.4615385
6	\banma(ß ss)end(e er es en em)??\b	0.5569620
7	\banständig(e er es en em)_bürger(s n innen in)??\b	0.3333333
8	\banti(-)??demokratisch(e er es en em)??\b	0.6446281
9	arbeitend(e er en)_bevölkerung\b	0.4776119
10	arbeitend(e er en em)_bürger(s n innen in)??\b	0.5714286

Dieser Datensatz repräsentiert ein gewichtetes Diktionär, welches im nächsten Schritt getestet werden soll. Zunächst wird im folgenden Abschnitt die Anwendung solch gewichteter Diktionäre über das R-Paket *weightdictR* und die aufgetretenen Herausforderungen erläutert, da zum aktuellen Zeitpunkt der Arbeit keine Funktion in R für die Anwendung von gewichteten Diktionären mit komplexen Multiword-Patterns existiert.

3.5 Anwendung gewichteter Diktionäre: weightdictR

Bisher wurden gewichtete Diktionäre wie Sentiment-Diktionäre über einen innerjoin der Textdaten als Tokens mit einem Datensatz, der die Patterns und die Gewichtung enthält, angewendet. Dieser Ansatz ist über den inner_join auf einfache Diktionäre ohne Multiword-Phrasen und Regexausdrücke limitiert, weshalb für die Diktionäre dieser Arbeit ein anderes Vorgehen gewählt wird.

Die grundlegende Idee der Anwendung des gewichteten Diktionärs dieser Arbeit ist, dass die Information vorhanden sein muss, welches Pattern in welcher Rede gematched wurde, damit die Gewichtung berücksichtigt werden kann. Diese Informationen sind über die key-word-in-context Funktion (kwic) aus dem Quanteda-Paket

abrufbar. Unter Eingabe von Textdaten und einem Diktionär wird folgender beispielhafter Output von der Funktion erstellt:

Abbildung 3 Ausschnitt eines Results einer kwic-Funktion

docname	from	to	pre	keyword	post	pattern
839752	404	404	die	sogenannten	redispach-kosten	\bso(- _)?genannt(e er es en em)??.b
839218	82	82	keine	sogenannten	postfaktischen	\bso(- _)?genannt(e er es en em)??.b
838986	70	70	der	schande	bezeichnet	\bschandel\b
839076	100	100	journalistischen	kreisen	dass	\bkreisen\b
839224	85	85	einer	realitätsfremden	agrarwelt	\brealitätsfremd(e er es en em)??.b
839173	574	574	uns	täuschen	die	\btäusch(t en)\b
839181	161	161	wir	unseren_bürgerinnen	und	\buns(er er en)??.bürger(n innen)??.b
839248	166	166	die	sogenannte	staatliche	\bso(- _)?genannt(e er es en em)??.b
841221	64	64	news	antidemokratischem	populismus	\bant(- _)?demokratisch(e er es en em)??.b
839291	927	927	der	sogenannten	potas-kommission	\bso(- _)?genannt(e er es en em)??.b

Über die Spalte "docname" können zusätzliche Metainformationen anhand der ID hinzugefügt werden, wie die zugehörige Fraktion des Redners oder die Länge einer Rede. Die Spalten "pre" und "post" wurden im vorherigen Prozess verwendet, um den Kontextvektor zu erstellen. Die Gewichtung wird über die Spalte "pattern" hinzugefügt, indem das gewichtete Wörterbuch mit dem Ergebnis der "kwic"-Funktion gemerged wird. Diese Schritte wurden in die Kernfunktion des Paketes *run_weightdict* zusammengefasst. Voraussetzung für die erfolgreiche Anwendung der Funktion ist, dass im Hintergrund die kwic-Funktion die Patterns des Diktionärs korrekt verarbeitet. Dies ist jedoch nicht der Fall, wie bereits an mehreren Stellen dieser Arbeit angemerkt wurde. Folgender Error liegt der Problematik zugrunde:

„Fehler in h(simpleError(msg, call)) : Fehler bei der Auswertung des Argumentes 'x' bei der Methodenauswahl für Funktion 'which': Incorrectly nested parentheses in regex pattern.

(U_REGEX_MISMATCHED_PAREN, context='(.)?')*

Entscheidend ist dabei die hervorgehobene Zeichenfolge. Dieser Regexausdruck führt dazu, dass ein Pattern wie,

(wollen|(ein)?fordern|möchten|mögen|verlangen|beanspruchen|wünschen) (.)?steuerzahler(innen)?*

unterschiedliche Phrasen in den Textdaten matchen kann. Diese Möglichkeit ist über das Quanteda-Paket nicht abzubilden, weshalb eine reine Anwendung der gewichteten Diktionäre über die kwic-Funktion nicht möglich ist. Diese Limitation ist auch Gründl aufgefallen, weshalb er über sein Paket *multidictR* eine Lösung entwickelt hat (siehe Gründl 2022). Die Kernidee dieser Lösung wurde bereits im Kapitel 3.1 im Pre-Processing angerissen. Über die Funktion *corpus_to_compound_tokens* aus meinem R-Paket wird der Text bereinigt und Leerzeichen in Multiword-Phrasen, die von Gründls Diktionär gematched werden, werden durch Unterstriche ersetzt. Diese

Transformation wird auch für die entsprechenden Patterns in Gründls Diktionär vorgenommen über die Funktion *make_patterns_compound*.

Nach diesen Bereinigungen kann über die Kernfunktion des Pakets *run_weightdict* das gewichtete Diktionär erfolgreich angewendet werden. Diese Bereinigungen können auch standardisiert über den Parameter *prepare_corp = TRUE*, der Funktion *run_weightdict* erfolgen. Davon ist aber abzuraten, da der Prozess *corpus_to_compound_tokens* sehr zeitintensiv ist und sich daher ein Zwischenspeichern der überarbeiteten Textdaten zu empfehlen ist. Des Weiteren können über *run_weightdict* reguläre Diktionäre zum Vergleich angewendet werden, indem der Parameter *add_no_weight_results = TRUE* gesetzt wird.

Weitere Informationen zu *weightdictR* sind in der Dokumentation auf GitHub enthalten (siehe: <https://github.com/TomSOWI/weightdictR>).

4. Ergebnisse

Die Leistung der Diktionäre wird über drei Verfahren geprüft. Zunächst wird die Validität über die Konstruktvalidität geprüft, danach wird die Reliabilität über die Split-Half Reliabilität verglichen und zum Schluss wird die Robustheit getestet, indem ein kleines Sample an Kontexten betrachtet wird.

4.1 Konstruktvalidität

Wie in Kapitel 3.1 erwähnt, dienen die Expertendaten der Validierung der Resultate der Diktionäre. Diese Form der Validierung über den Vergleich mit anderen Messinstrumenten wird als Konstruktvalidität bezeichnet. Zur Berechnung der Konstruktvalidität liegen Expertendaten in den Jahren 2014 und 2019 aus der Chapel Hill Expert Survey (CHES) und 2017 aus der Comparative Study of Electoral Systems (CSES 5) vor.

In der Chapel Hill Expert Survey 2014 wurden Experten zur Position der Parteien zum Verhältnis von Elite und Volk befragt. Die Angaben konnten auf einer 11er Skala angegeben werden von 0 „Elected office holders should make the most important decisions“ bis 10 „The people, not politicians, should make the most important decisions“. Im Jahr 2019 wurden die Experten wieder zu diesem Item und zur Salienz der Parteien zu anti-elitärer Kommunikation und Rhetorik befragt. Die Einschätzungen wurden mit einer 11er Skala von 0 „Not important at all“ bis 10 „Extremely important“ erfasst. Diese Items erfassen die Dimensionen „anti-elitism“ und „people-centrism“ aus der grundlegenden Populismusdefinition, weshalb diese Items zur

Operationalisierung von Populismus genutzt werden.³ Dahingegen wurden 2017 in der 5. Welle der CSES Experten direkt nach Ihrer Populismuseinschätzung der im Bundestag vertretenen Parteien befragt.

Die Einschätzungen der Experten werden in der folgenden Grafik als Durschnitt und z-transformiert dargestellt.

Abbildung 4 Populismuseinschätzung der Parteien nach Experten

Partei	2014	2014_z	2017	2017_z	2019	2019_z
AfD	7.78	1.45	7.0	1.86	9.01	1.89
LINKE	5.40	0.63	4.0	0.43	4.40	0.25
CDU/CSU	1.40	-0.75	1.5	-0.75	1.71	-0.71
Grüne	2.00	-0.55	2.0	-0.51	3.35	-0.12
SPD	1.30	-0.79	2.0	-0.51	1.90	-0.64

Quelle: Eigene Berechnung auf Basis von CHES und CESES 5

Während 2013 mit der Gründung der AfD ihre Ausrichtung noch nicht eindeutig war, hat sie sich mittlerweile als rechts-populistische Partei etabliert (vgl. Fawzi et al. 2017: 114; vgl. Siri 2018: 141f.). Entsprechend dieser Experteneinschätzung schneidet die AfD zu allen Zeitpunkten als populistische Partei ab. Zur Einordnung der LINKE wird noch in der Literatur gerungen. Rooduijn et al. ordnen die LINKE ebenfalls als populistische Partei wie die AfD ein (vgl. Rooduijn et al. 2019). Wohingegen Fawzi et al. die LINKE als Mainstream-Partei und nicht als populistische Partei bewerten (vgl. Fawzi et al. 2017: 114f.). Ausgehend von den Expertendaten liegt die Einordnung der LINKE als Grenzfall nahe, da die Partei nicht die Werte der AfD erreicht, aber überdurchschnittlich hohe Populismuswerte im Vergleich zu den restlichen Parteien erzielt. Die restlichen Parteien werden von den Experten als konstant gering populistisch bewertet. Lediglich 2019 lässt sich für die Grünen eine Steigerung in der Populismuseinordnung der Experten beobachten.

Im nächsten Schritt, vor dem Vergleich der Konstruktvalidität, sollen die deskriptiven Resultate von Gründls Diktionär als Annäherung an die Ergebnisse von Diktionären vorgestellt werden. Dies erfolgt über den Vergleich der Prozentanteil populistischer Sätze nach Parteien zu den Vergleichshorizonten 2014, 2017 und 2019. Wenn die Patterns in Gründls Diktionär gut ausgewählt sind, ist entsprechend zu erwarten, dass die AfD und die LINKE einen höheren Prozentanteile an populistischen Sätzen zu verzeichnen haben als die anderen Parteien gemäß der Einschätzung in der Literatur.

³ 2014 wurde nur "anti-elitism" abgefragt

Eine angemessene Einordnung der Parteien nach Gründl, ist eine grundlegende Voraussetzung für die Leistung der gewichteten Diktionäre.

Abbildung 5 Anteil populistischer Sätze in Bundestagsreden nach Fraktionen

Fraktionen	Prozentanteil populistischer Sätze		
	2014	2017	2019
AfD	-	2.63	2.97
LINKE	1.43	1.46	1.57
CDU/CSU	0.82	0.85	0.91
FDP	-	1.35	0.98
Grüne	0.85	0.93	1.03
SPD	0.93	0.96	1.06

Quelle: Eigene Berechnung auf Basis von Open Discourse

Anhand von Abbildung 5 ist zu erkennen, dass Gründls Diktionär den Populismusanteil der Parteien angemessen in Bundestagsreden erfasst. Die AfD sticht mit einem hohen Anteil populistischer Sätze von über 2 Prozent in den Jahren 2017 und 2019 hervor. Darauf folgt die LINKE, die mit durchschnittlich 1,5 Prozent hohe populistische Werte erzielt. Die anderen Parteien liegen mit ähnlichen Werten um 0,5 Prozentpunkte unter dem Wert der LINKE. Lediglich im Jahr 2017 verzeichnet die FDP einen Ausreißer mit einem Populismusanteil von 1,35 Prozent.

Bei den Parteiprogrammen zeigt sich auf den ersten Blick ein anderes Bild.

Abbildung 6 Anteil populistischer Sätze in Parteiprogrammen nach Parteien

Parteien	Prozentanteil populistischer Sätze		
	2013	2017	2021
AfD	4.17	3.37	2.95
LINKE	0.91	1.12	0.90
CDU/CSU	1.51	1.20	0.75
FDP	1.20	1.81	1.02
Grüne	1.44	1.40	1.30
SPD	1.59	0.84	0.38

Quelle: Eigene Berechnung auf Basis von Manifesto Project

Die AfD schneidet weiterhin als die Partei ab, die über alle Vergleichshorizonte hinweg den höchsten Anteil an populistischer Sprache hat, jedoch schneidet die LINKE nach dieser Darstellung als nicht-populistischer Partei ab (siehe Abbildung 6). Ein Grund dafür ist die Länge des Parteiprogramms, weil in längeren Parteiprogrammen

einzelne populistische Sätze an Gewicht verlieren. Die Parteiprogramme der Linke sind mit ca. 50000 Sätzen doppelt so lang wie die der anderen Parteien, die durchschnittlich etwa 25000 Sätze beinhalten. Des Weiteren ist anzumerken, dass das Parteiprogramm der AfD 2013, als die Partei erst gegründet wurde, aus vier Seiten bestand, weshalb die absolut wenigen populistischen Sätze zu einem hohen relativen Populismusanteil geführt haben. Die restlichen Parteien schneiden ähnlich wie bei den Bundestagsreden nach dieser Darstellung als nicht-populistische Parteien ab. Ebenfalls lässt sich in der Abbildung ein leichter Abwärtstrend in den Prozentanteilen populistischer Sätze beobachten.

Diese Betrachtung liegt nahe, dass die Einordnung der Parteien, bis auf Verzerrungen in den Parteiprogrammen, auf Basis der grundlegenden Patterns aus Gründls Diktionär die Einschätzung der Literatur abbildet. Demgemäß stellt Gründls Diktionär eine gute Grundlage für die Entwicklung der gewichteten Diktionäre dar. Außerdem liegen die Resultate der Diktionäre wie auch die Experten nahe, dass es sich bei der LINKE um einen Grenzfall handelt.

Im nächsten Schritt wird zur Beantwortung der Forschungsfrage überprüft, ob die gewichteten Diktionäre die Parteien genauer einordnen als Gründls Diktionär, indem die Konstruktvalidität verglichen wird.

Die Zusammenhänge zwischen den Einordnungen der Fraktionen bzw. Parteien mit der Einordnung der Experten werden im Folgenden dargestellt. Die Zusammenhänge wurden mithilfe des Pearson-Korrelationskoeffizienten (Pearsons' r) berechnet.

Abbildung 7 Zusammenhänge zwischen Diktionären und Expertenbefragung nach Bundestagsreden

Jahr	gruendl	clara	claraK	kmeans	kmeansK
2014	0.962	0.974	0.959	0.953	0.943
2017	0.964	0.979	0.959	0.948	0.925
2019	0.981	0.977	0.979	0.981	0.981
Durchschnitt	0.969	0.977	0.966	0.961	0.950

Quelle: Eigene Berechnung auf Basis von Open Discourse, CHES und CESES 5

In Abbildung 7 sind die Resultate für die Bundestagsreden dargestellt. Clara und kmeans repräsentieren das entsprechende gewichtete Diktionär für 2 Dimensionen und claraK und kmeansK repräsentieren das gewichtete Diktionär mit der "wahren" Anzahl an Dimensionen. Zunächst ist festzustellen, dass sowohl gruendl als auch die gewichteten Diktionäre sehr hohe Zusammenhänge von größer 0,9 aufweisen. Das

bedeutet, alle Varianten sind in der Lage die, im Bundestag vertretenen Fraktionen angemessen auf einer Populismusskala einzuordnen.

Insgesamt fallen die geringen Unterschiede zwischen gruendl und den gewichteten Diktionären auf. Der größte Unterschied besteht im Jahr 2019 mit einem um 0,039 geringeren Zusammenhang nach kmeansK im Vergleich zu gruendl (siehe Abbildung 7). Am besten schneidet das gewichtete Diktionär nach clara ab. Dieses weist im Schnitt einen um 0,008 stärkeren Zusammenhang mit den Experten im Vergleich zu gruendl auf (siehe Abbildung 7).

Abbildung 8 Zusammenhänge zwischen Diktionären und Expertenbefragung nach Parteiprogrammen

Jahr	gruendl	clara	claraK	kmeans	kmeansK
2013	0.674	0.697	0.676	0.675	0.650
2017	0.824	0.854	0.815	0.807	0.806
2021	0.927	0.933	0.928	0.921	0.914
Durchschnitt	0.809	0.828	0.806	0.801	0.790

Quelle: Eigene Berechnung auf Basis von Manifesto Project, CHES und CESES 5

In Abbildung 8 sind die Ergebnisse für die Parteiprogramme der Parteien dargestellt. Die dargestellte Konstruktvalidität kann zeitliche Verzerrungen beinhalten, weil die Parteiprogramme in den Jahren 2013 und 2021 nicht zum gleichen Zeitpunkt wie die Expertenbefragung von 2014 und 2019 erstellt wurden.

Im Vergleich zu Abbildung 7 fallen die geringeren Zusammenhangswerte auf, insbesondere für die Parteiprogramme aus dem Jahr 2013. Die Ursache wurde bereits in den Erläuterungen zu Abbildung 6 angerissen. Die AfD hatte durch ihr kurzes Parteiprogramm einen überproportionalen hohen Anteil populistischer Wörter auf, während die Parteiprogramme der LINKE von den Diktionären aufgrund ihrer besonders umfangreichen Texte allgemein als weniger populistisch eingeschätzt wurden (siehe Abbildung 6).

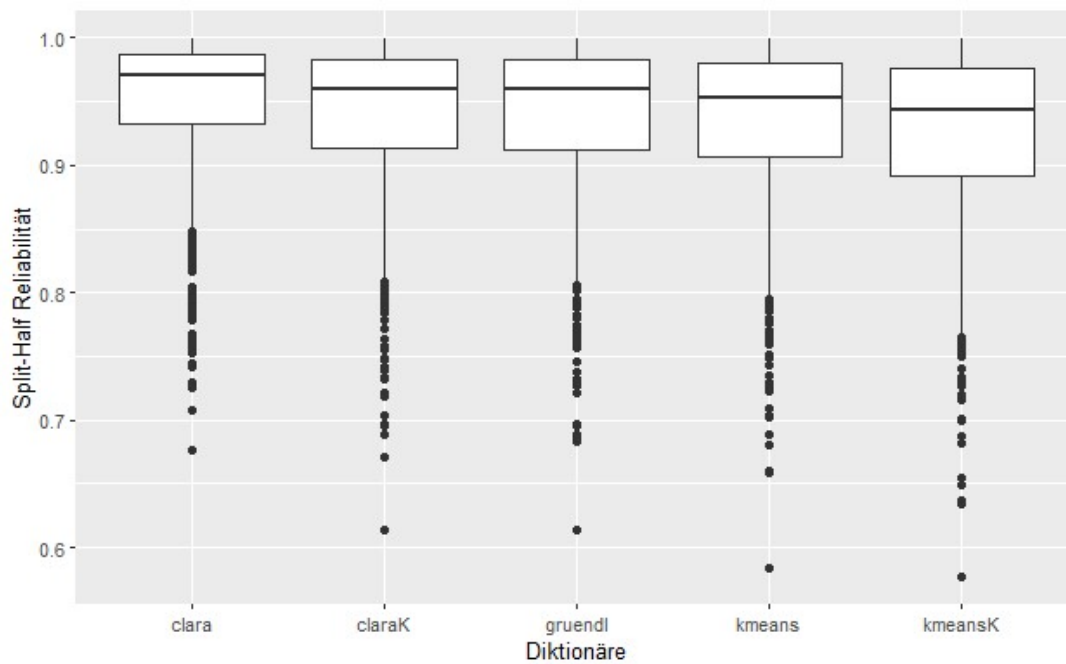
Auch für die Parteiprogramme lassen sich wie für die Bundestagsreden nur geringe Abweichungen zwischen gruendl und den gewichteten Diktionären feststellen. Der größte Unterschied besteht im Jahr 2017 mit einem um 0,03 stärkeren Zusammenhang nach clara (siehe Abbildung 8). Am besten schneidet hier ebenfalls das gewichtete Diktionär nach clara ab. Dieses weist im Schnitt einen um 0,02 stärkeren Zusammenhang mit den Experten im Vergleich zu gruendl auf.

Diese Ergebnisse aus Abbildung 7 und 8 sprechen dafür, dass Diktionäre eine angemessene Methode sind, um Parteien bzw. Fraktionen auf einer Populismusskala einzuordnen. In Hinblick auf die Forschungsfrage lässt sich feststellen, dass die Berücksichtigung des Kontextes über die von mir gewählte Vorgehensweise nur einen leichten Einfluss auf die Konstruktvalidität der Diktionäre hat. Dennoch konnte eine leichte Verbesserung im Vergleich zum Original über die Kontextgewichtung über den Clara-Algorithmus mit zwei Dimensionen erzielt werden. Im nächsten Schritt wird die Split-Half Reliabilität der Diktionäre untersucht.

4.2 Split-Half Reliabilität

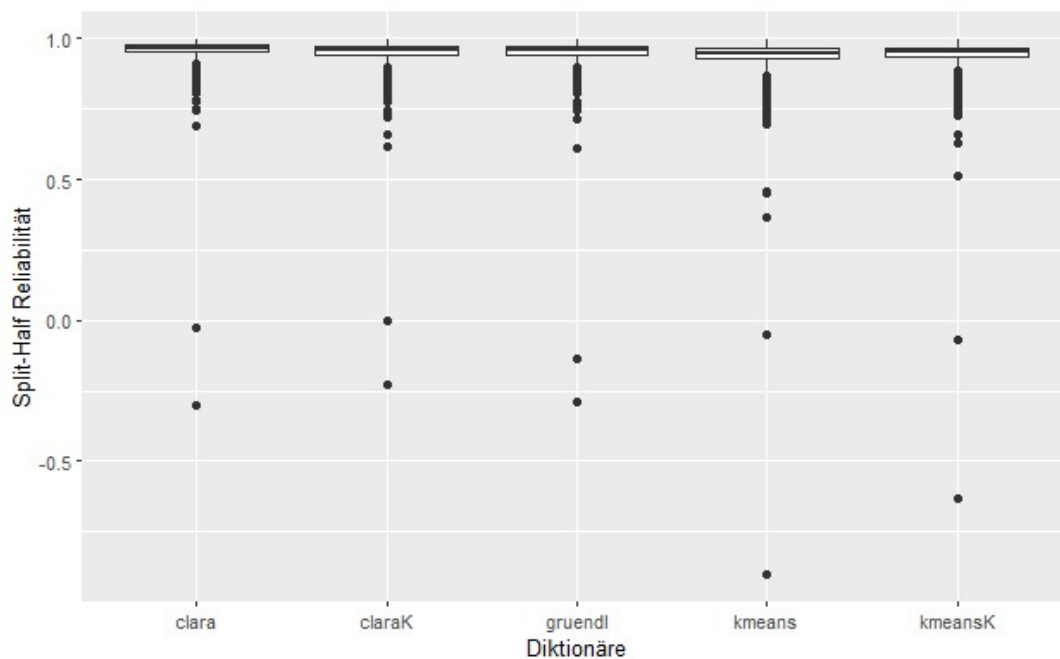
Als weiteres Kriterium, um die Reliabilität von Diktionären zu vergleichen, haben Rooduijn und Pauwels einen Split-Half Reliabilität Test vorgeschlagen (vgl. Rooduijn und Pauwels 2011: 1279). Zur Berechnung wird das Diktionär in zwei Hälften gespalten und der Zusammenhang zwischen den Ergebnissen der Hälften berechnet. Durch die Berechnung soll ermittelt werden, ob die Ergebnisse eines Diktionärs von einzelnen dominanten Einträgen abhängen und ob die Phrasen sowie Wörter dasselbe theoretische Konzept messen (vgl. Gründl 2022: 1493f.). In dem Fall dieser Arbeit basieren die Diktionäre auf denselben Patterns, weshalb die Berechnung zunächst als unbedeutsam erscheint für den Vergleich. Wiederum haben die Gewichtungen einen Effekt auf die Reliabilität, weil zum Beispiel der Einfluss eines Patterns durch eine hohe oder niedrige Gewichtung abgemildert bzw. verstärkt werden kann, wodurch die Reliabilität beeinflusst wird. Aufgrund dessen wird im Folgenden die Split-Half Reliabilität der Diktionäre für Reden und Parteiprogramme verglichen. Die Durchschnittswerte wurden über die Spearman-Brown Formel mit 1000 randomisierten Splits berechnet in Anlehnung an Gründls Modifikation des Vorschlags von Rooduijn und Pauwels (siehe Gründl 2022: 1494).

Abbildung 9 Split-Half Reliabilität auf Basis von Bundestagsreden



Quelle: Eigene Berechnung auf Basis von Open Discourse

Abbildung 10 Split-Half Reliabilität auf Basis von Parteiprogrammen



Quelle: Eigene Berechnung auf Basis von Manifesto Project

Die Abbildungen 9 und 10 stellen die Werte der Split-Half Reliabilität jeweils nach Bundestagsreden und Parteiprogrammen über ein Boxplot dar. Die Boxen in den Abbildungen umschließen die mittleren 50% der Fälle und der Balken repräsentiert den Median. Die einzelnen Punkte repräsentieren einen Zusammenhangswert einer Zufallsaufteilung. Zunächst können sehr befriedigenden Medianwerte der Split-Half

Reliabilität von über 0,9 für alle Diktionäre in beiden Abbildungen festgestellt werden. Nach den Bundestagsreden schneidet clara am besten ab, mit einem Mean von 0,95, wobei selbst der schlechteste Fall mit einem Wert von 0,67 einen soliden Split-Half Reliabilitätswert erzielt. Die anderen Diktionäre claraK (Mean = 0,94) (Min = 0,61), gruendl (Mean = 0,94) (Min = 0,61), kmeans (Mean = 0,93) (Min = 0,58) und kmeansK (Mean = 0,92) (Min = 0,58) schneiden etwas schlechter ab. Clara schneidet auch bei den Parteiprogrammen mit einem Durchschnittswert von 0,96 besser als, claraK (Mean = 0,95), gruendl (Mean = 0,95), kmeans (Mean = 0,93) und kmeansK (0,94) ab. Die Betrachtung der einzelnen Zusammenhänge der Zufallsaufteilungen offenbart ein unerwartetes Ergebnis. Es lassen sich über alle Diktionäre hinweg zwei Ausreißer identifizieren, die teils negative Korrelationen aufweisen. Dies ist darauf zurückzuführen, dass die Parteiprogramme weniger umfangreich sind als die Bundestagsreden, wodurch es zu weniger Matches kommt und teilweise bestimmte Parteien in den Zufallsaufteilungen auch null Matches haben. Insofern, als es sich hierbei nur um 2 Fälle gegenüber 998 Fällen mit ähnlich hohen Werten handelt, können diese Fälle vernachlässigt werden.

Insgesamt lassen sich von den Resultaten der Konstruktvalidität und der Split-Half Reliabilität feststellen, dass das clara-Diktionär am besten abschneidet. Um eine abschließende Schlussfolgerung zu treffen, wird im Folgenden die Stichhaltigkeit der Verbesserung von clara gegenüber gruendl untersucht.

4.3 Prüfung gewichteter Patterns

Über die bereits bekannte kwic-Funktion können die Kontexte der einzelnen Patterns überprüft werden. Um die Stichhaltigkeit der Verbesserung durch die Gewichtung nach clara zu testen, sollen die Kontexte von den 5 Patterns mit der jeweils höchsten und niedrigsten Gewichtung untersucht werden. Bei einer stichhaltigen Verbesserung wäre zu erwarten, dass die Kontexte der Patterns mit der niedrigsten Gewichtung weniger populistisch sind als die Kontexte mit der höchsten Gewichtung. Für die Bestimmung der zu untersuchenden Patterns wurde nach Pattern gefiltert, die mindestens 10-mal in den Textdaten vorkommen.

Für eine hohe Gewichtung wurden die Patterns "Nomenklatura"⁴, "sagen dürfen", "Frechheit", "Schande", "bürgerfern" und für eine niedrige Gewichtung die Patterns "zum Schaden der", "zulasten der deutschen", "wahnwitzig", "auf Kosten der

⁴ Der Begriff wurde in Deutschland von der SED geprägt zur Bezeichnung von Führungspersonal, welches besonders parteitreu ist. Heute wird der Begriff auch als Synonym für die Elite verwendet.

Allgemeinheit“, und “auf Kosten der deutschen“ untersucht. Zunächst werden die Patterns für hohe populistische Gewichtung betrachtet.

Das Pattern “**Nomenklatura**“ kommt häufig in Kontext von Gegenreden zur LINKE vor, wie zum Beispiel:

„Zweitens lasse ich mir nicht von jemandem, der zur Nomenklatura der SED gehört hat, hier Zweiklassenmedizin vorwerfen.“

Außerdem kommt das Wort “Nomenklatura” in seiner Bedeutung als Synonym für die Elite vor:

„Ich würde mir wünschen, dass sich der Deutsche Bundestag hier zum Bewahrer der deutschen Sprache erklärt, sich als Hüter der verbalen Kommunikation mit dem Volk versteht und wir uns nicht mit überbordendem Snobismus einer Sprache einer abgehobenen Nomenklatura befleißigen. Ihr wildes Herumgegendere hat, mit Verlaub, nichts, aber wirklich absolut gar nichts mit Gleichberechtigung der Geschlechter zu tun.“

Des Weiteren wird in dem Zitat die Nähe zum Volk betont, indem eine volksnahe Kommunikation gefordert wird. Entsprechend dieser antielitären und auch volkszentrierten Kontexte erscheint die hohe Gewichtung von 0,92 für “Nomenklatura” gerechtfertigt.

Das Pattern “**sagen dürfen**“ kommt selten in populistischen Kontexten vor. Einer dieser Kontexte ist zum Beispiel folgendes Zitat:

„Nur links-grüne Meinungen in einem engen totalitären Korsett gelten als akzeptabel und zulässig, alles andere wird als Hass und Hetze diffamiert. 71 Prozent der Menschen haben zum Beispiel das Gefühl, dass sie ihre wahre Meinung zur Flüchtlingsfrage nicht sagen dürfen – 71 Prozent! Ein Klima der Angst, in dem keiner sagt, was er meint, ist Totalitarismus, das ist vor allem auch Sozialismus, das ist DDR 2.0

Hier wird die Gefahr der Unterdrückung von Meinungen durch eine “links-grüne” Elite adressiert. Ansonsten sind es populismusfreie Kontexte, in denen das Pattern “sagen dürfen” auftritt, beispielsweise:

„Sie wollen einen falschen Weg einschlagen das wird man im deutschen Bundestag doch wohl noch sagen dürfen.“

Oder

„Zu der Frage, was die Alternative sein soll, muss man hier zwei drei Sätze sagen dürfen (...)“

Oder

„(...) die Ungeimpften tragen derzeit die Verantwortung dafür, dass wir nicht weiterkommen, dass muss man sagen dürfen (...)“

Der Ausdruck “sagen dürfen” wird mehr als Stilmittel benutzt, um der eigenen inhaltlichen Position einen stärkeren Ausdruck zu verleihen. Die Anwendung als Ausdruck für die Unterdrückung von Meinungen durch eine Elite kommt unerwartet selten vor. Insofern erscheint die hohe Gewichtung von 0,91 eher fraglich an dieser Stelle.

Zum Pattern “**Schande**” lassen sich über 600 Kontexte extrahieren. Entsprechend kommt das Wort in diversen Kontexten vor. Der dominanteste populistische Kontext ist dabei die Verunglimpfung des politischen Gegners bzw. einer Elite. Zum Beispiel:

„Ich rede von der Schande unseres Landes, und die heißt Schwarz-Gelb.“

Außerdem wird die Volkszentriertheit betont:

„Der Eingriff in die Persönlichkeitsrechte von Arbeitslosen ist eine Schande für den deutschen Rechtsstaat.“

Dabei wird weniger von “einem Volk” oder “den Menschen” gesprochen, sondern von einzelnen Gruppen, wie hier die Arbeitslosen, die in Mitleidenschaft durch eine Handlung bzw. Entscheidung gezogen werden. Wenige nicht-populistische Kontexte, wie die Zitierung der AfD-Parole “*Mahnmal der Schande*”, können unter der hohen Anzahl an populistischen Kontexten vernachlässigt werden, weshalb die hohe Gewichtung von 0.89 gerechtfertigt erscheint.

Das Vorkommen des Patterns “**Frechheit**” in seinen über 200 Fällen gleicht einem ähnlichen Muster wie dem des Patterns “Schande”. Es lassen sich Verunglimpfungen finden wie:

„Trotzdem besaß Minister Niebel Anfang des Jahres, also ganze 40 Jahre später, die Frechheit dieses Ziel als „sportlich“ zu bezeichnen.“

und Volkszentriertheit wie:

„Was Sie mit diesem Haushalt vorlegen, ist eine beispiellose Frechheit gegenüber den Menschen in diesem Land.“

Demgemäß ist eine hohe Gewichtung von 0,88 angemessen. Allgemein kann festgestellt werden, dass “Frechheit” stärker Volkszentriertheit betont, weil häufiger von “den Menschen” bzw. einem “Volk” in dem Kontext gesprochen wird, während “Schande” stärker im Kontext antielitäre Haltungen auftritt.

Das Pattern “**bürgerfern**” tritt fast ausschließlich in Debatten über die Europäische Union auf. Es verbergen sich populistische und teilweise auch deskriptive Kontexte

hinter dem Ausdruck. Folgendes Beispiel für einen deskriptiven Kontext ist interessant:

„Zurzeit wird die europäische Politik der Bürgerinnen und Bürgern der EU oft als technokratisch und bürgerfern empfunden und dem muss entgegengewirkt werden“

Dieses zeigt eine potenzielle positive Entwicklung von Populismus, da erkannt wird, dass sich eine Elite, in dem Fall die EU, zu weit von den Bürgern entfernt hat. Insofern steht dieses Zitat einer Zustandsbeschreibung ebenfalls mit Populismus in Verbindung. Die populistischen Kontexte wie das folgende populistische Beispiel treten aber am häufigsten auf:

„Sie wollen per EU-Verordnung Ihre Vorstellung von Sozialpolitik in allen 28 EU-Ländern durchpeitschen, egal ob die Menschen das wollen oder nicht. Das ist das grün-sozialistische Europa: Es ist zentralistisch, es ist bürgerfern, und es ist vor allen Dingen undemokratisch.“

Dieses Zitat bildet eine Vielzahl populistischer Elemente ab, wie die Zuschreibung von Bürgerferne und undemokratisches Verhalten. Entsprechend dieser Betrachtung erscheint die hohe Gewichtung von 0.83 für „bürgerfern“ gerechtfertigt.

Nach dieser Untersuchung der hochgewichteten Patterns folgt die Betrachtung der niedriggewichteten Patterns.

Das Pattern **“zum Schaden der“** bezieht sich in der Regel auf eine Personengruppe, häufig ist auch ganz allgemein die Bevölkerung gemeint:

„Für die Menschen im Land ist diese Laufzeitverlängerung [der Atomkraftwerke] ein weiterer Schlag ins Gesicht. Die Bundesregierung setzt ihre Interessen ohne Rücksicht auf Verluste zum Schaden der Menschen jetzt und zum Schaden der Menschen, die in hunderttausend Jahren leben, durch.“

Damit wird das populistische Element bedient, dass die Bundesregierung die Elite an der Bevölkerung vorbeiregiert. Da in der Regel das Pattern genutzt wurde, um auszudrücken, dass eine Elite sich zum Schaden des Volkes bereichert oder das Volk ignoriert, ist die geringe Gewichtung von 0,38 inhaltlich nicht angemessen.

Das Pattern **“zulasten der deutschen“** tritt in den Textdaten auf, um auf ein vermeintliches Fehlverhalten oder eine Fehlentscheidung hinzuweisen:

„Sie verkaufen hier schlichtweg 60 Prozent Staatsverschuldung im Verhältnis zum BIP – ein Maastricht-Kriterium, das auch Sie von der CDU/CSU schon längst über Bord geschmissen haben –, um die Party- und Dolce-Vita-Staaten des Südens zu beglücken, die trotz zehnjähriger Niedrigzinspolitik es nicht auf die Kette gekriegt haben, was übrigens auch zulasten der

deutschen Steuerzahler, der deutschen Sparer, der deutschen Versicherungen, der deutschen Pensionskassen ging.“

und

„Die CDU-geführte Bundesregierung zerstört die deutsche Automobilindustrie durch Kaufprämien und die einseitige Fokussierung auf batteriebetriebene Elektromobilität. Das ist ein Arbeitsplatzvernichtungsprogramm zulasten der deutschen Wirtschaft, des Wohlstandes und der Steuerzahler.“

Diese Fehlentscheidungen sollen angeblich bestimmten Gruppen der Bevölkerung schaden. Jedoch wird hier nicht eine Elite beschuldigt, sich vom Volk zu weit entfernt zu haben, wie bei den Kontexten von “zum Schaden der“. Insofern ist an dieser Stelle die niedrige Gewichtung von 0,39 für “zulasten der deutschen“ gerechtfertigt.

Das Pattern “**wahnwitzig**“ tritt als Adjektiv anders im Text auf als die vorherigen Patterns. Es wird in Textdaten als Ausdruck für einen surrealen Zustand genutzt, wie zum Beispiel hier:

„5000 Menschen kommen pro Jahr ins Gefängnis, weil sie kein Ticket für ein paar Euro gekauft haben. Nur 350 Menschen kommen pro Jahr wegen Steuerhinterziehung in den Knast, obwohl sie dem Staat Milliarden Euro rauben. Das ist wahnwitzig.“

Wie das vorherige Pattern beziehen sich diese Kontexte weniger auf eine Elite, sondern auf Fehler bzw. in dem Fall hier auf Missstände, weshalb eine geringe Gewichtung von 0,41 angemessen ist.

Das Pattern “**auf Kosten der Allgemeinheit**“ beschreibt inhaltlich einen ähnlichen Sachverhalt wie “zulasten der deutschen“ und “zum Schaden der“. Die Kontexte ähneln dabei eher dem Pattern “zulasten der deutschen“, wie folgendes Zitat zeigt:

„Dann werden Sie den Einspeisevorrang für erneuerbare Energien abschaffen müssen. Ich unterstelle Ihnen nicht einmal, dass Sie das wollen. Aber das ist die zwanghafte Konsequenz Ihres verantwortungslosen Energiekonzepts. Sie betreiben volkswirtschaftliches Schiffeversenken auf Kosten der Allgemeinheit.“

Wie in dem Beispiel dargestellt, beziehen sich die Kontexte auf Fehler der Elite. Die geringe Gewichtung von 0,43 ist angemessen, weil die Elite nicht aus populistischen Motiven, sondern inhaltlich kritisiert wird wie bei dem Pattern “zulasten der deutschen“.

Dies gilt ebenfalls für das Pattern “**auf Kosten der deutschen**“, was folgender Kontext untermauert:

„Frau Ministerin, wir haben bereits eine Bund-Länder-Vereinbarung zur Unterstützung der Forschung an Fachhochschulen. Diese gilt es nach ihrem Auslaufen im nächsten Jahr weiter zu stärken und auszubauen. Das wäre der richtige Weg. Wir brauchen keinen Aufbau einer Doppelstruktur auf Kosten des deutschen Steuerzahlers.“

Es ist kein klarer populistischer Kontext, weshalb die niedrige Gewichtung von 0,44 gerechtfertigt ist.

Diese stichprobenartige Untersuchung konnte aufzeigen, dass die Gewichtung, bis auf zwei Abweichungen, grundsätzlich zutreffend ist. Insgesamt beziehen sich die Kontexte der Patterns mit geringer Gewichtung auf Kritik an vermeintlichen Fehlentscheidungen, während die Kontexte der hochgewichteten Patterns in der Regel populistische Elemente beinhaltet. Auf Grundlage dieser Erkenntnis kann vorläufig davon ausgegangen werden, dass die Verbesserung durch die Gewichtung über die clara-Variante stichhaltig ist. Im nächsten Abschnitt wird die Reichweite und Limitation dieser Erkenntnis beurteilt.

Die vorgestellten Kontexte können über das Markdown-Dokument “3.3. Statistiken gewichteter Kontexte“ eigenständig eingesehen und exploriert werden (siehe: <https://github.com/TomSOWI/BA-Arbeit>).

5. Reflexion des Vorgehens

Bevor eine abschließende Einordnung der Ergebnisse dieser Arbeit getroffen werden kann, muss das methodische Vorgehen mit den einhergehenden Limitationen reflektiert werden.

Zunächst handelt es sich bei dem von mir erstellten Verfahren zur Kontextgewichtung von Diktionären um ein exploratives Vorgehen. In dieser Arbeit wurden zwei Cluster-Algorithmen mit zwei unterschiedlichen Vorgehensweisen zur Bestimmung von k angewendet. Weitere Algorithmen und Vorgehensweisen konnten nicht getestet werden, da die Berechnung einer Variante sehr zeitintensiv ist. Entsprechend könnten weitere Explorationen ein besseres kontextgewichtetes Diktionär hervorbringen als das im Rahmen dieser Bachelorarbeit entwickelte Diktionär auf Basis von clara. Entsprechend darf dieses Diktionär nicht als absolut angesehen werden. Weitere Explorationen zukünftiger Arbeiten könnten die Berechnung eines weiteren Cluster-Algorithmus oder der Entwicklung eines weiteren Vorgehens zur Bestimmung von populistischen bzw. nicht-populistischen Clustern beinhalten. Zweiteres erscheint am vielversprechendsten, da die Klassifizierung von populistischen Clustern über die Verteilung von Redebeiträgen populistischer Parteien in der Anwendung teilweise problematisch war. Die Klassifizierung bei zwei Clustern war praktikabel, da sich

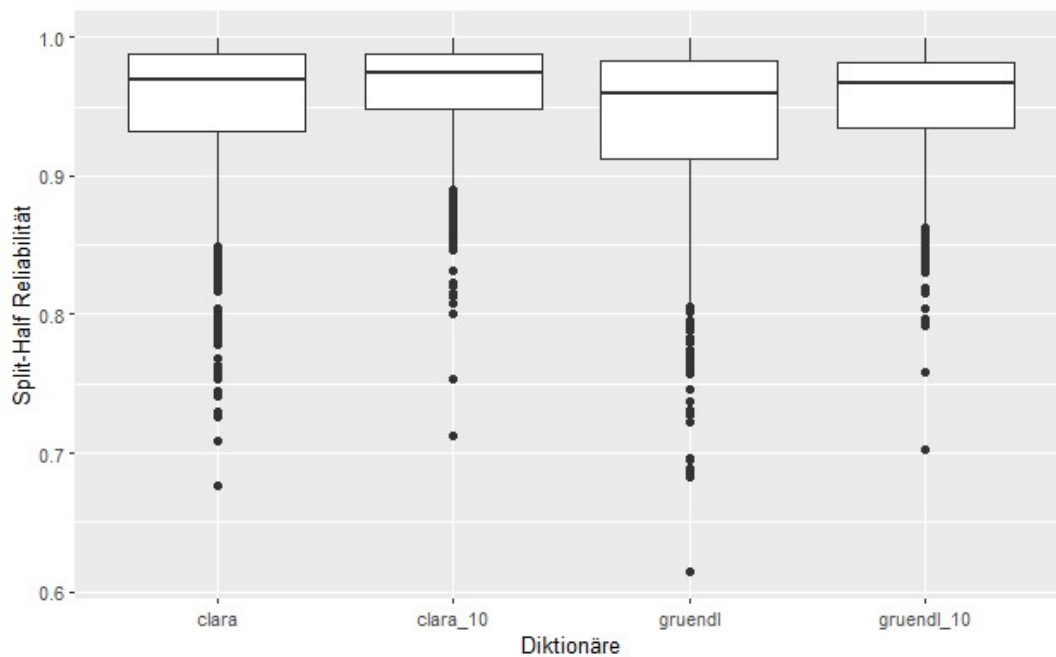
überwiegend die Anteile an Redebeiträgen von populistischen Parteien zwischen den Clustern stark unterschieden haben. Bei dem Vorgehen über die wahre Anzahl an Clustern mussten teilweise sechs Cluster klassifiziert werden. Hierbei gab es häufig nur geringe Unterschiede in den Anteilen der Redebeiträge populistischer Parteien zwischen den Clustern, weshalb diese Einteilung fehleranfällig war. Dies ist auch anhand des schlechten Abschneidens von *kmeansK* und *claraK* in der Konstruktvalidität zu erkennen (siehe Abbildung 7 und 8). Eine weitere Limitation dieser Klassifizierung der Cluster ist, dass das Vorgehen voraussetzt, dass eine populistische Partei in den Textdaten anwesend ist.

Die Validierung über die Konstruktvalidität benötigt ebenfalls eine kritische Betrachtung. Die allgemein hohen Zusammenhangswerte basieren auf einer Einteilung von 6 bzw. 4 Parteien im Jahr 2014 für die Bundestagsreden. Aufgrund dieser geringen Fallzahl dürfen die hohen Zusammenhänge nicht überbewertet werden. Entsprechend bietet es sich an, in einer folgenden Arbeit diese Diktionäre auf das österreichische und schweizerische politische Spektrum anzuwenden, um die Fallzahl an einzuordnenden Parteien zu erhöhen. In diesem Zusammenhang ist zu erwähnen, dass bestimmte landesspezifische Patterns wie, *"zu Lasten der Österreicher"* und *"zu Lasten der Schweizer"* in den deutschen Bundestagsreden nicht vorgekommen sind. Diesen werden die Gewichtungen der korrespondierenden Patterns, hier *"zu Lasten der deutschen"*, hinzugefügt, damit das Diktionär auch im österreichischen und schweizerischen politischen Spektrum angewendet werden kann. Dieser Diktionär ist in dem R-Paket *weightdictR* unter dem Namen *klotz_clara_DACH* abrufbar.

Ein weiteres Problem einer geringen Fallzahl ist die Gewichtung von Patterns, die sehr selten in den Textdaten vorgekommen sind. Für 47 Patterns von 238 basiert diese Gewichtung auf weniger als 10 Fällen. Einerseits bedeutet das, dass die Gewichtung bestimmter Patterns nicht robust und damit fehleranfällig ist, andererseits kann davon ausgegangen werden, dass diese Patterns selten in Texten vorkommen und daher ihr Einfluss zu vernachlässigen ist. Die vorliegenden Abbildungen untersuchen diesen Fall.

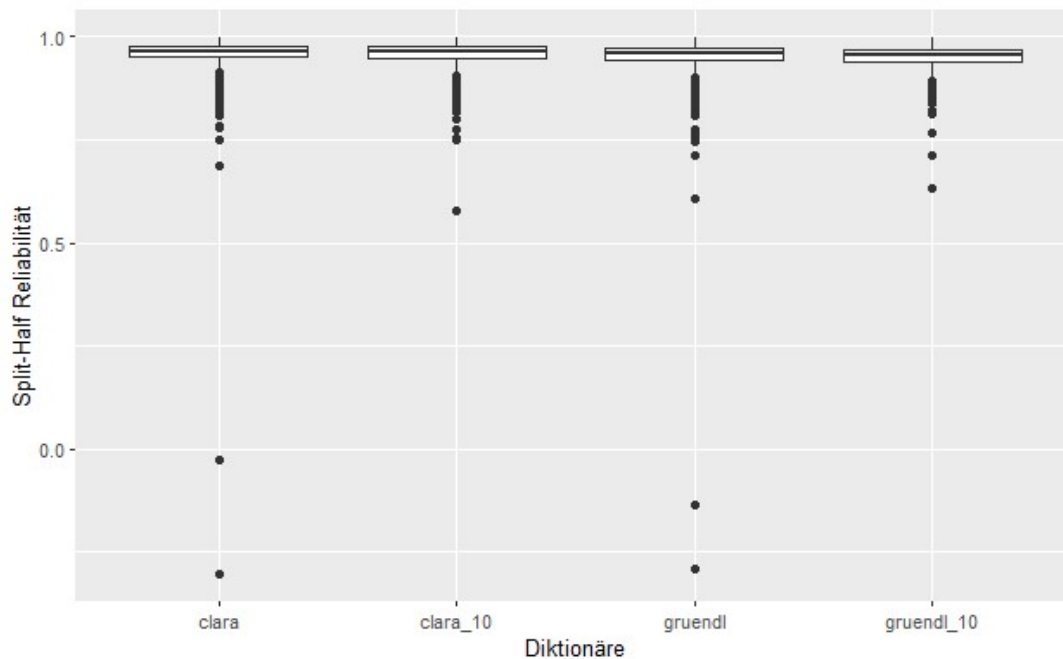
Ein kurzer Blick auf die Split-Half Reliabilität in Abbildung 11 und 12 bestätigt, dass die gekürzten Diktionäre bei der Split-Half Reliabilität besser abschneiden.

Abbildung 11 Split-Half Reliabilität gekürzter Diktionäre auf Basis von Bundestagsreden



Quelle: Eigene Berechnung auf Basis von Open Discourse

Abbildung 12 Split-Half Reliabilität gekürzter Diktionäre auf Basis von Parteiprogrammen



Quelle: Eigene Berechnung auf Basis von Manifesto Project

Das gekürzte gruendl-Diktionär schneidet im Schnitt mit 0,95 besser als das Original mit 0,93 ab. Das gekürzte clara-Diktionär schneidet ebenfalls mit 0,96 besser als das Original mit 0,95 ab. Insgesamt sind das nur marginale Verbesserungen. Ähnlich sieht es bei der Konstruktvalidität aus.

Abbildung 13 Konstruktvalidität der Diktionäre mit und ohne seltene Patterns nach Bundestagsreden

Jahr	gruendl	gruendl10	clara	clara10
2014	0.962	0.965	0.974	0.976
2017	0.964	0.963	0.979	0.981
2019	0.981	0.981	0.977	0.978
Durchschnitt	0.969	0.970	0.977	0.978

Quelle: Eigene Berechnung auf Basis von Open Discourse

Abbildung 14 Konstruktvalidität der Diktionäre mit und ohne seltene Patterns nach Parteiprogrammen

Jahr	gruendl	gruendl10	clara	clara10
2013	0.674	0.681	0.697	0.701
2017	0.824	0.824	0.854	0.854
2021	0.927	0.925	0.933	0.932
Durchschnitt	0.809	0.810	0.828	0.829

Quelle: Eigene Berechnung auf Basis von Manifesto Project

Abbildung 13 und 14 vergleichen die Konstruktvalidität zwischen clara und gruendl mit allen Patterns und ohne selten auftretende Patterns (gruendl10 und clara10). Sowohl für gruendl als auch für clara lässt sich im Durchschnitt eine marginale Verbesserung um einen um 0,01 stärkeren Zusammenhang für die Bundestagsreden als auch für die Parteiprogramme feststellen. Bis auf die Unterschiede der Konstruktvalidität für Parteiprogrammen von 2013 mit 0,07 und 0,04 lassen sich keine größeren Unterschiede beobachten. Insofern kann der Einfluss der gewichteten Patterns mit geringer Fallzahl als geringfügig eingeschätzt werden. Da im Schnitt die Diktionäre ohne den Ausschluss bestimmter Patterns nach der Konstruktvalidität und Reliabilität nur marginal schlechter performen, wird dies nicht als Anlass genommen, weitreichende Änderungen an den Diktionären vorzunehmen.

Zur Verbesserung der Robustheit können stattdessen mehr Daten in Betracht gezogen werden. Dabei muss beachtet werden, dass eine Erweiterung des Erhebungszeitraums dazu führt, dass Verschiebungen im Sprachgebrauch über die Zeit die Ergebnisse verzerren. Daher sollten aufbauende Arbeiten nicht einfach die Anzahl an Bundestagsperioden erhöhen, sondern weitere Quellen im gleichen Zeitraum von 2009 bis 2020 wie zum Beispiel Twitter-Daten hinzuziehen.

Für weitere Explorationen in zukünftigen Anwendungen des Diktionärs, ist in meinem R-Paket ein Dataframe hinterlegt, der das N jeder Gewichtung transparent darlegt (*klotz_clara_N*). Falls bestimmte Patterns beziehungsweise das Diktionär als Ganzes unerwartete Ergebnisse erzielt, könnte ein Lösungsansatz sein, bestimmte Patterns von geringer Robustheit zu entfernen.

6. Fazit

Diese Arbeit hat sich mit der Erfassung von Populismus in Bundestagsreden und Parteiprogrammen auseinandergesetzt, um über ein verbessertes Populismus-Diktionär zukünftige Forschungsarbeiten über Auswirkungen und Folgen von Populismus zu unterstützen. Hierzu wurde in dieser Arbeit eine Methode vorgestellt, um das Messen von Populismus über Diktionäre attraktiver zu machen. Dazu hat sich diese Arbeit mit der Kritik an Diktionären auseinandergesetzt und mit der Kontextgewichtung eine Lösung vorgestellt, um das Kontextproblem anzugehen. Als Grundlage wurde das Diktionär von Gründl genutzt, auf dessen Basis vier gewichtete Diktionäre über die Kontextgewichtung erstellt wurden. Die Resultate der Methode wurden über die Konstruktvalidität, Split-Half Reliabilität und einen Robustheitstest evaluiert. Es konnte gezeigt werden, dass das ursprüngliche Diktionär von Gründl und die Weiterentwicklungen über die Kontextgewichtung sehr befriedigende Werte in der Konstruktvalidität und der Split-Half Reliabilität erzielt haben. Dabei hat das Diktionär clara über alle Vergleichshorizonte hinweg besser abgeschnitten als Gründls Diktionär. Der Robustheitstest konnte bestätigen, dass die Verbesserungen durch die Kontextgewichtung auf Basis des CLARA-Algorithmus mit zwei Dimensionen zu einer stichhaltigen Verbesserung des ursprünglichen Diktionärs von Gründl geführt haben. Die Forschungsfrage kann insofern beantwortet werden, als kontextgewichtete Diktionäre über ihre Gewichtung den Kontext in der Auswertung mitberücksichtigen können und entsprechend bessere Resultate liefern.

Dies ist eine vorläufige Erkenntnis, da aufgrund der erwähnten Limitation in Kapitel 5 weitere Forschung und Tests notwendig sind, um die Erkenntnisse dieser Arbeit zu bestätigen. Zudem wurde die Kontextgewichtung nur auf ein Thema, Populismus im politischen Kontext, angewendet. Um die Stichhaltigkeit der Methode generell zu prüfen, sollten weitere diktionärbasierte Forschungsvorhaben in anderen Kontexten angestoßen werden. Beispielsweise könnte man Populismus weiter in nicht-politischen Kontexten wie Zeitungsartikeln untersuchen, wie Gründl bereits vorgeschlagen hat (vgl. Gründl 2022: 1496). Zudem bieten sich Forschungsvorhaben in anderen Themenbereichen und Sprachen an.

Das clara-Diktionär sowie weitere Methoden zur Exploration und Anwendung gewichteter Diktionäre stehen zur Anwendung in dem R-Paket *weightdictR* zur Verfügung. Die konkrete Umsetzung der Kontextgewichtung dieser Arbeit ist im GitHub unter <https://github.com/TomSOWI/BA-Arbeit> einsehbar. Insgesamt konnte die Arbeit über das entwickelte Paket eine kontextberücksichtigende Analysemethode von Populismus entwickeln, die ressourcenärmeren Forschungsvorhaben ein angemessenes methodisches Werkzeug zur Verfügung stellt.

Literaturverzeichnis

- Ahlquist, John S. und Christian Breunig. 2012. Model-Based Clustering and Typologies in the Social Sciences. *Political Analysis* 20(1): 92–112.
- Anria, Santiago. 2016. More Inclusion, Less Liberalism in Bolivia. *Journal of Democracy* 27(3): 99–108.
- Batory, Agnes. 2016. Populists in Government? Hungary's 'System of National Cooperation'. *Democratization* 23(2): 283–303.
- Chapel Hill Expert Survey. 2019. *Chapel Hill Expert Survey trend file, 1999–2019*. doi: 10.1016/j.electstud.2021.102420.
- Engesser, Sven, Nayla Fawzi, und Anders Olof Larsson. 2017. Populist Online Communication: Introduction to the Special Issue. *Information, Communication & Society* 20(9): 1279–1292.
- Fazwi, Nayla. Magdalena Obermaier. Carsten Reinemann. 2017. Is the Populism Laggard Catching up? In *Populist political communication in Europe*, hrsg. Toril Aalberg, Frank Esser, Carsten Reinemann, Jesper Stromback, Claes De Vreese, 111–126. New York: Routledge.
- Grimmer, Justin und Brandon M. Stewart. 2013. Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. *Political Analysis* 21(3): 267–297.
- Gründl, Johann. 2022. Populist Ideas on Social Media: A Dictionary-Based Measurement of Populist Communication. *New Media & Society* 24(6): 1481–1499.
- Hawkins, Kirk. A. 2009. Is Chávez Populist?: Measuring Populist Discourse in Comparative Perspective. *Comparative Political Studies* 42(8): 1040–1067.
- Hawkins, Kirk A. 2016. Responding to radical populism: Chavismo in Venezuela. *Democratization* 23(2): 242–262.
- Heinisch, Reinhard und Franz Fallend. 2016. Collaboration as Successful Strategy against Right-Wing Populism? The Case of the Centre-Right Coalition in Austria, 2000–2007. *Democratization* 23(2): 324–344.
- Huber, Robert A. und Saskia P. Ruth. 2017. Mind the Gap! Populism, Representation, and Participation in Europe. *Swiss Political Science Review* 23(4): 462–484.
- Jagers, Jan and Stefaan Walgrave. 2007. Populism as Political Communication Style: An Empirical Study of Political Parties Discourse in Belgium. *European Journal of Political Research* 46(3): 319–345.
- Kaufman, Leonard, und Peter J. Rousseeuw. 2005. *Finding Groups in Data: An Introduction to Cluster Analysis*. New York: Wiley.
- Klamm, Christopher, Ines Rehbein, und Simone Paolo Ponzetto. 2023. Our kind of people? Detecting populist references in political debates. In *Findings of the Association for Computational Linguistics: EACL 2023*, 1227–1243. Dubrovnik: Association for Computational Linguistics.
- Manning, Christopher D., Prabhakar Raghavan und Hinrich Schütze. 2009. *Introduction to Information Retrieval*. Cambridge: Cambridge University Press.
- Mudde, Cas. 2004. The populist zeitgeist. *Government and Opposition* 39(4): 542–563.

- Mudde, Cas. 2017. Populism: An Ideational Approach. In *The Oxford Handbook of Populism*, hrsg. Cristóbal Rovira Kaltwasser, Paul Taggart, Paulina Ochoa Espejo und Pierre Ostiguy, 27–47. Oxford: Oxford University Press
- Mudde, Cas und Cristóbal Rovira Kaltwasser. 2017. *Populism: A Very Short Introduction*. New York: Oxford University Press.
- Open Discourse. 2020. *Open Discourse V3 – Harvard Dataverse*. doi: 10.7910/DVN/FIKIBO.
- Ostiguy, Pierre. 2017. Populism: A Socio-Cultural Approach. In *The Oxford Handbook of Populism*, hrsg. Cristóbal Rovira Kaltwasser, Paul Taggart, Paulina Ochoa Espejo, und Pierre Ostiguy, 73–98. Oxford: Oxford University Press.
- Pauwels, Teun. 2017. Measuring populism: a review of current approaches. In *Political Populism: A Handbook. International Studies on Populism*, hrsg. Heinisch Reinhard, Christina Holtz-Bacha und Oscar Mazzoleni, 123–136. Baden-Baden: Nomos.
- Politbarometer. 2023. Politbarometer vom 14. Juli 2023. <https://www.zdf.de/politik/politbarometer/230714-politbarometer-video-2-100.html>. Zugegriffen: 29.09.23.
- Quinn, Kevin M., Burt L. Monroe, Michael Colaresi, Michael H. Crespin und Dragomir R. Radev. 2010. How to Analyze Political Attention with Minimal Assumptions and Costs. *American Journal of Political Science* 54(1): 209–228.
- Rooduijn, Matthijs, und Teun Pauwels. 2011. Measuring Populism: Comparing Two Methods of Content Analysis. *West European Politics* 34(6): 1272–1283.
- Shukla, Shraddha und Naganna S.. 2014. A Review ON K-Means DATA Clustering APPROACH. *International Journal of Information & Computation Technology*. 4(17): 1847-1860.
- Siri, Jasmin. 2018. The Alternative for Germany after the 2017 election. *German Politics* 27(1): 141-145.
- The Comparative Study of Electoral Systems. 2022. *CSES Module 5 Advance Release 4. GESIS, Köln. ZA7557 Datenfile Version 4.0.0*. doi: 10.7804/cses.module5.2022-03-01.
- The Manifesto Data Collection. *Manifesto Project (MRG/CMP/MARPOR). Version 2023a*. Berlin: Wissenschaftszentrum Berlin für Sozialforschung (WZB) / Göttingen: Institut für Demokratieforschung (IfDem). doi: 10.25522/manifesto.mpds.2023a.
- Weyland, Kurt. 2001. Clarifying a Contested Concept: Populism in the Study of Latin American Politics. *Comparative Politics* 34(1): 1–22.
- Weyland, Kurt. 2021. Populism as a Political Strategy: An Approach's Enduring — and Increasing — Advantages. *Political Studies* 69(2): 185–189.

Erklärung über die Eigenständigkeit

Ich erkläre,

dass diese Arbeit selbständig verfasst wurde,

dass keine anderen als die angegebenen Quellen benutzt und alle wörtlich oder
sinngemäß aus anderen Werken übernommenen Aussagen als solche gekennzeichnet wurden,

dass die eingereichte Arbeit weder vollständig noch in wesentlichen Teilen Gegenstand eines anderen Prüfungsverfahrens gewesen ist,

dass die Arbeit weder vollständig noch in Teilen bereits veröffentlicht wurde und

dass das elektronische Exemplar mit den gedruckten Exemplaren übereinstimmt.

Sindelfingen, 04. Oktober 2023

Tom Klotz

A handwritten signature in blue ink, consisting of a stylized 'T' followed by a cursive flourish.