

Universität Stuttgart

Institut für Literaturwissenschaft,
Abteilung Digital Humanities

**Vergleich digitaler Methoden
der Textanalyse zur Identifizierung
rechtspopulistischer Sprache:
Ein Praxistest anhand
ausgewählter politischer Texte
des 20. & 21. Jahrhunderts**

Masterarbeit

An der Universität Stuttgart vorgelegt von
Kevin Stevens

Im Studiengang Master of Arts *Digital Humanities*

Prüfer/Betreuer: Herr Prof. Dr. Anselm Küsters,
Frau Dr. Nora Ketschik

Eingereicht am 01.02.2026

Inhaltsverzeichnis

INHALTSVERZEICHNIS	II
ABBILDUNGSVERZEICHNIS	IV
TABELLENVERZEICHNIS	VI
ABKÜRZUNGSVERZEICHNIS	VII
ZUSAMMENFASSUNG	VIII
ABSTRACT	X
1 EINLEITUNG	11
2 AKTUELLER FORSCHUNGSSTAND	17
3 KORPUSGENERIERUNG	24
4 POLITIKWISSENSCHAFTLICHE GRUNDLAGEN: VORSTELLUNG & EINORDNUNG DER AUSGEWÄHLTEN PARTEIEN AUF DEM POLITISCHEN SPEKTRUM	27
4.1 Alternative für Deutschland (AfD).....	31
4.2 Christlich Demokratische Union (CDU).....	33
4.3 Demokratische Partei der USA (DEM).....	35
4.4 Freie Demokratische Partei (FDP).....	36
4.5 Republikanische Partei der USA (GOP).....	39
4.6 Nationalsozialistische Deutsche Arbeiterpartei (NSDAP)	41
4.7 Sozialdemokratische Partei Deutschlands (SPD)	43
5 AUFSTELLUNG VON HYPOTHESEN BEZÜGLICH DER ÄHNLICHKEIT DER VERSCHIEDENEN PARTEIEN	45
5.1 Hypothese 1: Hohe Übereinstimmung zwischen AfD & Republikanern	45
5.2 Hypothese 2: Besonders hohe Übereinstimmung zwischen AfD & NSDAP	45
5.3 Hypothese 3: Hohe Übereinstimmung von CDU, SPD & FDP mit Demokraten	46
5.4 Hypothese 4: Niedrigste Übereinstimmung von Demokraten, SPD & FDP mit AfD & NSDAP	47
6 VORSTELLUNG & IMPLEMENTIERUNG DER AUSGEWÄHLTEN DIGITALEN METHODEN	48
6.1 Bag-of-Words.....	49

6.1.1	Methodische Grundlagen & Einsatzgebiete	49
6.1.2	Implementierung	51
6.1.3	Analyse der Ergebnisse & Überprüfung der Hypothesen.....	52
6.2	Term Frequency – Inverse Document Frequency.....	58
6.2.1	Methodische Grundlagen & Einsatzgebiete	58
6.2.2	Implementierung	60
6.2.3	Analyse der Ergebnisse & Überprüfung der Hypothesen.....	60
6.3	Stilometrische Analyse.....	62
6.3.1	Methodische Grundlagen & Einsatzgebiete	62
6.3.2	Implementierung	64
6.3.3	Analyse der Ergebnisse & Überprüfung der Hypothesen.....	65
6.4	Wordfish for Ideological Scaling.....	73
6.4.1	Methodische Grundlagen & Einsatzgebiete	73
6.4.2	Implementierung	76
6.4.3	Analyse der Ergebnisse & Überprüfung der Hypothesen.....	77
6.5	Word2Vec	83
6.5.1	Methodische Grundlagen & Einsatzgebiete	83
6.5.2	Implementierung	85
6.5.3	Analyse der Ergebnisse & Überprüfung der Hypothesen.....	86
6.6	Sentence Transformers.....	91
6.6.1	Methodische Grundlagen & Einsatzgebiete	91
6.6.2	Implementierung	94
6.6.3	Analyse der Ergebnisse & Überprüfung der Hypothesen.....	94
7	FAZIT: VERGLEICH DER ERGEBNISSE & REFLEXION DER VERWENDETEN METHODEN.....	99
8	AUSBLICK.....	105
	QUELLEN- & LITERATURVERZEICHNIS	107
	Internetquellen	107
	Sekundärliteratur	115
	Tools & Code Sources	118
	ANHANG	119
	EIGENSTÄNDIGKEITSERKLÄRUNG.....	121

Abbildungsverzeichnis

Abbildung 1: Historisches Konfliktlinienmodell der Parteiensysteme nach Lipset/Rokkan	28
Abbildung 2: Zweidimensionales Konfliktlinienmodell des heutigen deutschen Parteiensystems	29
Abbildung 3: Links-Rechts-Einstufung der Parteien	30
Abbildung 4: Links-Rechts-Verortung der drei großen ideologischen Strömungen des 19. Jahrhunderts	37
Abbildung 5: Visualisierung des bereinigten AfD-Korpus mit <i>wordclouds.com</i>	52
Abbildung 6: Visualisierung des bereinigten CDU-Korpus mit <i>wordclouds.com</i>	53
Abbildung 7: Visualisierung des bereinigten Demokratischen Korpus mit <i>wordclouds.com</i>	54
Abbildung 8: Visualisierung des bereinigten FDP-Korpus mit <i>wordclouds.com</i>	54
Abbildung 9: Visualisierung des bereinigten Republikanischen Korpus mit <i>wordclouds.com</i>	55
Abbildung 10: Visualisierung des bereinigten NSDAP-Korpus mit <i>wordclouds.com</i>	56
Abbildung 11: Visualisierung des bereinigten SPD-Korpus mit <i>wordclouds.com</i> ...	56
Abbildung 12: Dendrogramm der Clusteranalyse der unbereinigten Parteienkorpora, visualisiert mit <i>Stylo</i>	66
Abbildung 13: Dendrogramm der Clusteranalyse der bereinigten Parteienkorpora, visualisiert mit <i>Stylo</i>	68
Abbildung 14: Principal Components Analysis der unbereinigten Parteienkorpora in Form einer Kovarianzmatrix, visualisiert mit <i>Stylo</i>	69
Abbildung 15: Principal Components Analysis der bereinigten Parteienkorpora in Form einer Kovarianzmatrix, visualisiert mit <i>Stylo</i>	71
Abbildung 16: Geschätzte Dokumentpositionen der jeweiligen Parteienkorpora in Wordfish, sortiert in absteigender Reihenfolge	77

Abbildung 17: Estimated Feature Scores aus dem Wordfish-Modell für die ersten 30 Begriffe des bereinigten AfD-Korpus	78
Abbildung 18: Visualisierung der von Wordfish geschätzten Dokumentpositionen der jeweiligen Parteienkorpora im zweidimensionalen Koordinatensystem	80
Abbildung 19: Clusterplot der verschiedenen Parteien, basierend auf deren jeweiligen Wordfish-Positionen	81
Abbildung 20: Darstellung der Ähnlichkeitswerte aller Parteien mit Cosine Similarity in Matrizenform	87
Abbildung 21: Liste der Top-Ähnlichkeiten zwischen den Parteien in absteigender Reihenfolge.....	87
Abbildung 22: Überarbeitete Cosine Similarity Matrix	89
Abbildung 23: Überarbeitete Liste der Top-Ähnlichkeiten zwischen den Parteien in absteigender Reihenfolge	90
Abbildung 24: Heatmap der Cosine Similarities zwischen allen Parteien, basierend auf Sentence Transformers	95
Abbildung 25: Ergebnisse der PCA aller Partei-Embeddings in 2D, basierend auf Sentence Transformers	97
Abbildung 26: Übersicht über die Gesamtmenge der eingetretenen Hypothesen (in %)	100

Tabellenverzeichnis

Tabelle 1: Hierarchische Sortierung der <i>exact matches</i> der Top 10 <i>mfw</i> aller Parteien, paarweise	57
Tabelle 2: Ergebnisse der <i>Tf-Idf</i> -Implementierung	60
Tabelle 3: Hierarchische Sortierung der <i>exact matches</i> der Top 10 <i>mfw</i> aller Parteien, paarweise (<i>Tf-Idf</i> -bereinigt)	61
Tabelle 4: Übersicht der überprüften Hypothesen anhand der verwendeten Methoden	99
Tabelle 5: Übersicht der größten und geringsten Ähnlichkeiten zwischen den Parteien pro Methode, (falls vorhanden) basierend auf <i>Cosine Similarity</i>	100
Tabelle 6: Übersicht der Parteien mit den höchsten Ähnlichkeitswerten zur NSDAP pro Methode, (falls vorhanden) basierend auf <i>Cosine Similarity</i> ..	101
Tabelle 7: <i>Ungefilterte</i> Ergebnisse der <i>Tf-Idf</i> -Implementierung	119

Abkürzungsverzeichnis

AfD	A lternative für D eutschland
BoW	B ag of W ords
CDU	C hristlich D emokratische U nion
CA	C luster A nalysen
CT	C onsensus T ree
DEM	D emokratische Partei der USA
DH	D igital H umanities
FDP	F reie D emokratische P artei
GOP	G rand O ld P arty (Republikanische Partei der USA)
Idf	I nverse d ocument f requency
mfw	m ost f requent w ords
NLP	N atural L anguage P rocessing
NSDAP	N ationalsozialistische D eutsche A rbeiterpartei
PCA	P rincipal C omponent A nalysis
PCV	P rincipal C omponent V isualization
SPD	S ozialdemokratische P artei D eutschlands
Tf	T erm F requency
Tf-Idf	T erm f requency – I nverse d ocument f requency
w2v	w ord 2 v ec
WD	W orking D irectory
WP	W ahlperiode

Zusammenfassung

Welche Methoden eignen sich am besten, um rechtspopulistische Sprache in politischen Texten zu identifizieren? Zur Beantwortung dieser Frage werden verschiedene digitale Methoden der Textanalyse vorgestellt, um sie anschließend beim Vergleich verschiedener politischer Texte einzusetzen. Die Analyse wird anhand eigens zu diesem Zweck ausgewählter sowohl konservativer, als auch liberaler deutscher und amerikanischer Parteien und deren Parteiprogrammen sowie stichprobenartig ausgewählter politischer Reden durchgeführt.

Ergänzt wird diese Analyse unter anderem durch Visualisierungen mit *Most Frequent Words* und verschiedenen Metriken zur Ähnlichkeitsmessung (*Cosine Similarity* und *Principal Component Analysis*). Durch diese Verfahren sollen vor allem Gemeinsamkeiten und Unterschiede der politischen Sprache der ausgewählten Parteien untereinander aufgezeigt und untersucht werden. Darüber hinaus wird der rechtspopulistische Standard der jeweiligen Parteien an deren Ähnlichkeit zur Nationalsozialistischen Deutschen Arbeiterpartei (NSDAP) gemessen und beurteilt.

Abhängig von der jeweiligen Methode wurde die Implementierung entweder in *Python* oder *R* durchgeführt. Hierbei wurden die Korpora der verschiedenen Parteien, bestehend aus dem Wahlprogramm der jeweiligen Partei sowie zehn stichprobenartig ausgewählten politischen Reden, miteinander verglichen und die Ähnlichkeit der verschiedenen Parteien mit Hilfe von Distanzmaßen ausgewertet. Dies geschah nicht nur in Bezug auf die Identifizierung rechtspopulistischer Sprache, welche in dieser Arbeit durch besondere Nähe zur NSDAP definiert wurde, sondern darüber hinaus auch auf der Grundlage zuvor aufgestellter Hypothesen, welche als zusätzliche „Kontrollinstanz“ die Eignung der ausgewählten Methoden zusätzlich reflektieren sollen.

Es stellte sich heraus, dass –gemessen an der Anzahl der zutreffenden Hypothesen– sich hauptsächlich die *Sentence Transformers*, die *Stilometrische Analyse* und *Word2Vec*, sowie in Teilen auch *Wordfish for Ideological Scaling*, als die erfolgreichsten der verwendeten Methoden erwiesen haben. Und auch in Bezug auf die Identifizierung rechtspopulistischer Sprache bedeutet dies, dass sich diese vier Methoden gleichermaßen hierzu eignen, da *Sentence Transformers* und *Word2Vec*, sowie die *Stilometrische Analyse* und *Wordfish for Ideological Scaling* untereinander jeweils nahezu identische Übereinstimmungen in ihren Ergebnissen bezüglich der semantischen Ähnlichkeit zur NSDAP aufweisen.

Schlüsselwörter: *Digital Humanities, Politikwissenschaft, Parteien, Parteiprogramme, politische Reden, Rechtspopulismus, Ähnlichkeitsmaße, digitale Methoden, Bag-of-Words, Tf-Idf, Stilometrie, Word2Vec, Wordfish for Ideological Scaling, Sentence*

Transformers, Embeddings, Most Frequent Words, Cosine Similarity, Principal Component Analysis.

Abstract

What are the best suited methods for identifying right-wing populist language in political texts? To answer this question, various digital methods of text analysis are presented and then used to compare different political texts. The analysis is carried out using conservative and liberal German and American parties and their party programs, which were selected specifically for this purpose, as well as representatively selected political speeches.

This analysis is supplemented by visualizations using most frequent words and various metrics for similarity measurement (cosine similarity and principal component analysis). These methods are primarily intended to highlight and examine similarities and differences in the political language of the selected parties. In addition, the right-wing populist standard of the respective parties is measured and assessed based on their similarity to the National Socialist German Workers' Party (NSDAP).

Depending on the method used, the implementation was carried out in either Python or R. The corpora of the various parties, consisting of each party's election program and ten randomly selected political speeches, were compared with each other and the similarity of the various parties was evaluated using distance measures. This was done not only in relation to the identification of right-wing populist language, which was defined in this work as being particularly close to the NSDAP, but also based on previously established hypotheses, which were intended to reflect the suitability of the selected methods as an additional "control instance."

It turned out that, based on the number of correct hypotheses, *Sentence Transformers*, *Stylometric Analysis*, and *Word2Vec*, as well as *Wordfish for Ideological Scaling* to some extent, proved to be the most successful of the methods used. In terms of identifying right-wing populist language, this means that these four methods are equally suitable for this purpose, as *Sentence Transformers* and *Word2Vec*, as well as *Stylometric Analysis* and *Wordfish for Ideological Scaling*, respectively show almost identical results in terms of semantic similarity to the NSDAP among each other.

Key Words: *Digital Humanities, political science, parties, party programs, political speeches, right-wing populism, similarity measures, digital methods, Bag-of-Words, Tf-Idf, Stylometry, Word2Vec, Wordfish for Ideological Scaling, Sentence Transformers, Embeddings, Most Frequent Words, Cosine Similarity, Principal Component Analysis.*

1 Einleitung

„Das AfD-Programm basiert auf Fremdenfeindlichkeit, Islamfeindlichkeit und Intoleranz. Die Partei knüpft an die Naziideologie an.“¹ So formulierte es der ehemalige deutsche Bundesinnenminister (FDP) und Rechtsanwalt Gerhart Baum in seinem letzten Buch „Besinnt Euch!“, das posthum im Juni 2025 erschienen ist. Denn die Erstarkung rechtspopulistischer und rechter Parteien in den letzten Jahren stellt nicht nur unsere deutsche Demokratie vor große Herausforderungen, sondern auch die europäischen Grundwerte und transatlantischen Beziehungen grundsätzlich infrage. Was eine rechtskonservative bis rechtspopulistische Mehrheitsregierung in Deutschland bedeuten würde, wird aktuell am Beispiel der Vereinigten Staaten von Amerika sehr anschaulich ersichtlich: Wirtschaftliche Abschottung (Strafzölle), politische Isolation (NATO) und nicht zuletzt die gezielte Dämonisierung, Verfolgung und Ausweisung bestimmter Bevölkerungsgruppen (darunter vor allem irreguläre Einwanderer) wären in dieser oder ähnlicher Form auch in Deutschland an der Tagesordnung.

Um diese Parallelität nachzuweisen und die Ähnlichkeit sowohl deutscher als auch amerikanischer rechtspopulistischer Rhetorik, Sprachmuster und politischer Zielsetzungen aufzuzeigen, soll in dieser Forschungsarbeit ein Vergleich verschiedener ausgewählter deutscher und amerikanischer Parteien durchgeführt werden, um auf der Basis der jeweiligen Parteiprogramme sowie mehrerer politischer Reden der wichtigsten Vertreterinnen und Vertreter² dieser Parteien sowohl aus der aktuellen als auch den letzten drei Regierungsperioden³ einen parteiübergreifenden Vergleich durchzuführen. Dieser Vergleich ermöglicht anschließend, die Ähnlichkeit der verschiedenen Parteien untereinander aufgrund ihrer spezifischen Sprachmuster einzuordnen.

Für diese Auswertung werden diverse digitale Methoden aus dem Bereich der *Digital Humanities*⁴ eingesetzt, welche die Ähnlichkeit der politischen Texte der jeweiligen

¹ Baum, Gerhart: Besinnt Euch! Das Vermächtnis eines großen Liberalen. Berlin 2025.

² Wenn es nicht möglich war, in ausreichender Anzahl ausschließlich politische Reden der wichtigsten Vertreterinnen und Vertreter einer bestimmten Partei in transkribierter Form zu bekommen, wie dies beispielsweise bei der AfD der Fall war, so wurden ergänzend dazu ebenso zufällig ermittelte Reden beliebiger Parteimitglieder für diese Untersuchung herangezogen.

³ Dies entspricht dem Zeitraum von 2013-2025. Hierbei sind die deutschen Wahlperioden (WP) mit den Regierungszeiträumen der amerikanischen Präsidenten von den Jahreszahlen her identisch, sodass die folgenden Regierungsperioden entstehen: WP 21/ 60. Präsidentschaftswahl: seit 2025, WP 20/ 59. Präsidentschaftswahl: 2021-2025, WP 19/ 58. Präsidentschaftswahl: 2017-2021 sowie WP 18/ 57. Präsidentschaftswahl: 2013-2017.

⁴ Die *Digital Humanities*, zu Deutsch auch „Digitale Geisteswissenschaften“, sind ein interdisziplinäres Forschungsfeld, welches geistes- (darunter v.a. literaturwissenschaftliche) Ansätze mit digitalen

Parteien maschinell auswerten. Auf diese Weise wird eine objektive Beurteilung gewährleistet sowie eine neutrale Auswertung sichergestellt. Zu diesen digitalen Methoden zählen unter anderem *Bag-of-Words*, *Tf-Idf*, eine *stilometrische Analyse* mit Hilfe des R-Packages „*Stylo*“, *Word2Vec*, *Wordfish for Ideological Scaling*, sowie ein *Sentence Transformer* der Open-Source Plattform „Hugging Face“.⁵ Die Ähnlichkeit der verschiedenen Texte bzw. Korpora jeder Partei soll dabei jeweils anhand der beiden Distanzmaße *Cosine Similarity* und *Principal Component Analysis* gemessen werden.

Die Motivation für diese spezifische Thematik stammt noch aus dem Wintersemester 2023/24, als unser Jahrgang am Ende des ersten Semesters im Seminar „Computerlinguistische Methoden der Digital Humanities“ bei Fr. Dr. Melanie Andresen für ein Forschungsprojekt in kleine Gruppen eingeteilt wurde. Die Gruppe, welcher ich angehörte, untersuchte die Verwendung nationalsozialistisch geprägter Begriffe oder Begriffe aus der Zeit des Nationalsozialismus (darunter bspw. „entartet“, „Lebensraum“ usw.) im Kontext heutiger politischer Reden der AfD Bundestagsfraktion. Untersucht wurden dabei vor allem die Bundestags-Plenarprotokolle, welche auf der offiziellen Webseite des Deutschen Bundestages frei zugänglich einsehbar sind.

Die Auswertung erfolgte in Form einer Kollokationsanalyse, da uns insbesondere der spezifische Kontext, in dem die gesuchten Begriffe verwendet wurden, interessierte. Durchgeführt wurde diese Kollokationsanalyse mit Hilfe des Open-Source Tools *AntConc*. Neben falsch-positiven Treffern, wie beispielsweise der Verwendung des Begriffes „Lebensraum“ im natur- und tierschutzrechtlichen Kontext, ergab die Analyse aber auch überraschend viele *true positives*. Hierunter zählten unter anderem die Verwendung der Begriffe „entartet“ in Bezug auf die doppelte Staatsbürgerschaft, welche von der AfD vehement abgelehnt wird, sowie der Verwendung des Begriffes „Lebensraum“ in Bezug auf bezahlbaren Wohnraum für deutsche Familien.⁶ Während in diesem Forschungsprojekt also bereits Untersuchungen hinsichtlich der Verwendung nationalsozialistisch geprägter Begriffe durch aktuelle AfD-Politikerinnen und Politiker vorgenommen wurden, sollen in dieser konkreten Forschungsarbeit nicht nur die Kookkurrenzen und Korrelationen einzelner, wörterbuchartig vordefinierter

Methoden aus dem Bereich der Computerlinguistik, Informatik sowie der Maschinellen Sprachverarbeitung (*Natural Language Processing*, *NLP*) verknüpft. Das Ziel ist hierbei, große Mengen an Text nicht mehr selbst lesen zu müssen, sondern diese stattdessen maschinell auswerten zu lassen, um durch größere Datenmengen eine höhere Generalisierbarkeit zu ermöglichen.

⁵ Eine Übersicht der verschiedenen Methoden ist in Kapitel 6 Vorstellung & Implementierung der ausgewählten digitalen Methoden zu finden, wo diese jeweils ausführlich vorgestellt und beschrieben werden.

⁶ Bei Interesse finden Sie alle Ergebnisse inklusive Auswertung des damaligen Forschungsprojekts auf meinem persönlichen GitHub-Repository: [Präsentation Forschungsprojekt.pptx](https://github.com/Stevens29081997/Masterarbeit_Kevin_Stevens/blob/main/Pr%C3%A4sentation_Forschungsprojekt.pptx) (https://github.com/Stevens29081997/Masterarbeit_Kevin_Stevens/blob/main/Pr%C3%A4sentation_Forschungsprojekt.pptx).

Begriffe, sondern die Ähnlichkeit ganzer politischer Reden und Parteiprogramme verschiedener, ausgewählter Parteien untersucht und maschinell ausgewertet sowie quantitativ messbar gemacht werden.

Die zentrale Forschungsfrage dieser Arbeit lautet somit: Inwiefern können unterschiedliche digitale Methoden der Textanalyse zur Identifizierung rechtspopulistischer Sprache in ausgewählten politischen Texten des 20. & 21. Jahrhunderts eingesetzt werden? Hierbei soll das Parteiprogramm sowie mehrere politische Reden der wichtigsten Vertreter der NSDAP als „Baseline“ für rechtspopulistische Sprache angesehen werden. Mit diesem Korpus sollen dann mehrere deutsche und US-amerikanische Parteien verglichen werden, um die Ähnlichkeiten dieser Parteien sowohl untereinander, als auch mit der NSDAP objektiv mit Hilfe digitaler Methoden messbar zu machen. Hierbei gilt: Je näher eine der ausgewählten Parteien in Bezug auf die Werte der Ähnlichkeitsmaße an die NSDAP herankommt, als desto rechtspopulistischer wird ihre Sprache angesehen. Auf dieser Grundlage sollen in den nachfolgenden Kapiteln mehrere Hypothesen bezüglich der Ähnlichkeiten der jeweiligen Sprachmuster aufgestellt werden, welche anschließend durch eine entsprechende Auswertung der Ergebnisse der verschiedenen digitalen Methoden entweder verifiziert oder aber widerlegt werden sollen.

Bevor der obigen Forschungsfrage jedoch nachgegangen werden kann, müssen zunächst die Begriffe des Populismus und vor allem des Rechtspopulismus geklärt werden, welche dieser zentralen Fragestellung zugrunde liegen. In der wissenschaftlichen Literatur wird Populismus häufig als „dünne Ideologie“ definiert, da er, im Gegensatz zum Sozialismus oder Liberalismus, nicht über denselben Grad an Schärfe und Einheitlichkeit verfügt, wodurch er sich nur sehr schwer von anderen Ideologien abgrenzen lässt.⁷ Das Kernkonzept des Populismus besteht hierbei aus dem Antagonismus zwischen dem einfachen „Volk“ und der sog. „Elite“, die als dessen Antagonist fungiert. Dabei kann der Populismus ohne Weiteres mit einer Vielzahl verschiedenster Ideologien – darunter bspw. dem Kommunismus, Ökologismus, Nationalismus und Sozialismus – kombiniert werden. Darüber hinaus weist der Populismus in der Regel eher moralistische als programmatisch orientierte Züge auf, da für den Diskurs von Populisten die normative Unterscheidung zwischen „der Elite“ und „dem Volk“ von wesentlicher Bedeutung ist, nicht aber der empirische Unterschied im Verhalten oder bestimmten Einstellungen. Populismus zeichnet sich demnach durch eine Sichtweise aus, die eine deutliche Unterscheidung zwischen „Freund“ und

⁷ Mudde, Cas: The Populist Zeitgeist. In: Hertner, Isabelle und Jones, Erik (Hgg.): *Government and Opposition. An International Journal of Comparative Politics*, Band 39, Ausgabe 4. Oxford/Malden 2004, S. 544 (<https://www.cambridge.org/core/journals/government-and-opposition/article/populist-zeitgeist/2CD34F8B25C4FFF4F322316833DB94B7>, letzter Zugriff am 21.08.2025), Übersetzung des Autors. Gilt nachfolgend bis anders angegeben.

„Feind“ vorgibt, was einem klassischen „Schwarz-Weiß-Schema“ mit klarem Feindbild entspricht. Populismus ist in diesem Zusammenhang also als eine Art „Argumentationsstrategie“ zu verstehen, die von diversen Ideologien eingesetzt wird, um ihre jeweiligen Kommunikationsziele zu erreichen.⁸ Ganz konkret bedeutet dies zum einen die Darlegung vermeintlich einfacher Lösungen für komplexe Probleme, zum anderen aber ebenso die entscheidende Bedeutung einer zentralen Führungspersönlichkeit und eines gemeinsamen Identitätsbegriffes.⁹

Dies impliziert, dass sowohl Rechts- als auch Linkspopulismus als mögliche Phänomene des politischen Spektrums identifiziert werden können. Eine Gemeinsamkeit dieser beiden Formen ist, wie bereits beschrieben, die Konstruktion des „einen, wahren Volkes“ als Antithese zu einer als „entfremdet“ oder „feindlich“ beschriebenen Elite, kurz gesagt, die Erschaffung eines dichotomen Weltbildes bestehend aus dem „guten Volk“ und den „bösen Eliten“.¹⁰ Meist fokussieren sich Rechtspopulisten dabei vor allem auf kulturelle Themen (wie beispielsweise Migration), während Linkspopulisten vorwiegend ökonomische Themen (zum Beispiel soziale Ungleichheit) in den Mittelpunkt rücken.¹¹ Dabei richtet sich der Rechtspopulismus allerdings nicht nur gegen Eliten, sondern auch gegen ganze Bevölkerungsgruppen mit anderer ethnischer Herkunft.¹² Demnach gehören schutzsuchende Geflüchtete oder ethnische Minderheiten nicht zum „Volk“.¹³ Dieser für den Rechtspopulismus spezifische Anti-Pluralismus beschränkt sich jedoch nicht nur auf die Exklusion bestimmter Bevölkerungsgruppen aufgrund ihrer Herkunft, sondern ebenso auf Basis ihrer Religion und Weltanschauung.¹⁴ Während der „klassische“ Populismus also einen Konflikt „unten gegen oben“ (sprich „Volk“ vs. „Elite“) suggeriert, ergänzt der Rechtspopulismus diesen Antagonismus um den Konflikt „innen gegen außen“ (also „Einheimische“ vs. „Fremde“), der nicht nur auf nationalistischem, sondern ebenso auch auf völkischem Gedankengut basiert.¹⁵

⁸ Scheufele, Bertram: Warum Frames und Ideologien keine Synonyme sind und Populismus keines von beiden ist – Konzeptionelle Brückenschläge zwischen Frames und Ideologien und Einwände zur aktuellen Populismus-Forschung. In: Tepe, Peter und Oehm, Stefan (Hgg.): *Mythos-Magazin: Politisches Framing 1* (2019), S. 22 (https://mythos-magazin.de/politisches-framing/bs_warum_frames_und_ideologien.pdf, letzter Zugriff am 21.08.2025).

⁹ Bücker, Sinah; Schade, Sarah Maria und Wiegerling, Ulrike: Die AfD: Woher sie kommt, wie sie funktioniert, wer sie unterstützt. In: Walther, Eva und Isemann, Simon D. (Hgg.): *Die AfD – psychologisch betrachtet*. Wiesbaden 2019, S. 37.

¹⁰ Ebd.

¹¹ Rooduijn, Matthijs und Akkerman, Tjitske: Flank attacks: Populism and left-right radicalism in Western Europe. In: *Party Politics*, Band 23, Ausgabe 3, S. 200 (<https://journals.sagepub.com/doi/full/10.1177/1354068815596514>, letzter Zugriff am 21.08.2025).

¹² Hillje, Johannes: Das „Wir“ der AfD. Kommunikation und kollektive Identität im Rechtspopulismus. Frankfurt am Main 2022, S. 15.

¹³ Ebd.

¹⁴ Ebd.

¹⁵ Ebd.

Phillip Becher definiert Rechtspopulismus als eine

sich in das Kontinuum des Rechtsextremismus einfügende und in Wechselwirkung mit anderen rechten Strömungen stehende Bewegung [...], die sich anschickt, eine ‚parlamentsfähige Massenbasis für administrativ-autoritäre Politik‘ im Spannungsfeld ‚zwischen Partei und Sammlungsbewegung‘ herzustellen.¹⁶

Diese Definition kombiniert sehr anschaulich sowohl die politisch-parlamentarischen (Parteien und Öffentlichkeitsarbeit) als auch die außerparlamentarischen Elemente (Strömungen, Bewegungen) des Rechtspopulismus. Ergänzend dazu bemerkt Kolja Möller, dass der Rechtspopulismus sich „im Rahmen des demokratischen Paradoxes“ bewege, die Rechtsbindung öffentlicher Herrschaftsausübung nicht vollständig in Frage stelle und sich auf die Mechanismen des politischen Systems einlasse.¹⁷ Auch wenn der Rechtspopulismus eine „illiberale Transformation der Verfassung“ avisiere, berufe er sich im Gegensatz zu (neo-)faschistischen Bewegungen nicht auf die „befreiende Kraft einer dauerhaften physischen Gewaltanwendung“. ¹⁸ Dennoch betont Möller, dass der autoritäre Populismus durchaus mit der Gefahr einer faschistischen Umwälzung einhergehe.¹⁹

Da diese Arbeit explizit nach rechtspopulistischen Sprachmustern sucht, wird der Fokus in der nachfolgenden Analyse auf den Aspekt des Rechtspopulismus gelegt, wobei dieser als möglichst große semantische Nähe zur Nationalsozialistischen Deutschen Arbeiterpartei (NSDAP) verstanden wird. Dieser gelang es im Deutschland der 1930er Jahre erfolgreich, die oben beschriebene Unterscheidung zwischen Volk und Elite durch die Unterscheidung Volk/die Anderen zu ersetzen. Dadurch fand sich der deutsche „Volkskörper“ in einer Dauerbedrohung durch mögliche Eindringlinge (gemeint sind in diesem Fall „die Juden“) wieder, was zur Folge hatte, dass gerade nicht Ruhe, Ordnung und Sicherheit, sondern eine stetige Suche nach potentiellen „Gefährdern“ beginnt.²⁰ Diese populistische Dimension wird umso deutlicher, da der Nationalsozialismus in Deutschland eine „plebejische, rebellische, von antikapitalistischen Stimmungen erfüllte Bewegung“ gegen „die Eliten“ wendete.²¹ Der praktische Ausdruck dieser Elitenkritik äußerte sich hierbei in der Form des Antisemitismus, der eine Bewegung in Gang setzte, welche ständig neue Feinde

¹⁶ Becher, Phillip: Rechtspopulismus. Köln 2013, S. 25-26.

¹⁷ Möller, Kolja: Volk und Elite. Eine Gesellschaftstheorie des Populismus. 2. Auflage, Berlin 2024, S. 105.

¹⁸ Ebd.

¹⁹ Ebd.

²⁰ Ebd., S. 102.

²¹ Ebd., S. 104.

identifizierte, die „mit möglichst unmittelbaren Maßnahmen durch den politischen Apparat auszulöschen waren.“²²

Nach dieser Klärung der Begriffe und Definition von Rechtspopulismus soll im folgenden Kapitel nun zunächst der Forschungsstand der aktuellen Forschung etwas näher skizziert werden.

²² Ebd.

2 Aktueller Forschungsstand

Der Ansatz, politische Texte mithilfe digitaler Methoden und Ähnlichkeitsmaße, wie beispielsweise der *Cosine Similarity*, zu vergleichen, ist jedoch keinesfalls neu, denn in der bisherigen Forschung wurden ähnliche Versuche wie in dieser Arbeit bereits mehrmals durchgeführt. So hat beispielsweise Peer Trilcke im Jahr 2019 eine digitale Analyse der Wahlprogramme der an der brandenburgischen Landtagswahl teilnehmenden Parteien durchgeführt.²³ Bei dieser Analyse stützt sich Trilcke vornehmlich auf statistische Sprachanalysen der Wahlprogramme der Grünen, SPD, FDP, CDU, Linke und AfD. Hierbei interessiert sich Trilcke besonders für die häufigsten Wörter (*Most Frequent Words*), besonders charakteristische bzw. überrepräsentierte Wörter (*Keywords*) und uncharakteristische bzw. unterrepräsentierte Wörter (*Negative Keywords*) im Korpus. Nach denen für eine solche Untersuchung üblichen Schritten des *Preprocessings*²⁴ wurden dann die diversen Datenanalysen durchgeführt. Mithilfe frei verfügbarer Computerprogramme²⁵ wurden daraufhin zum einen nach Wortarten differenzierte Häufigkeitsanalysen (Ermittlung der „Most Frequent Words“) und zum anderen nach Wortarten differenzierte Schlüsselwortanalysen (Ermittlung der „Keywords“) durchgeführt. Dabei wurden sowohl über- als auch unterrepräsentierte Wörter ermittelt. Die Ergebnisse dieser Analysen wurden schließlich für eine zusammenfassende Präsentation auf insgesamt zwölf Postern (inklusive der dazugehörigen Datensätze) aufbereitet. Inhaltlich werden diese jedoch nicht weiter analysiert oder kommentiert, da Trilcke diese Aufgabe vorwiegend im Bereich des politischen Journalismus und nicht bei sich selbst verortet.

Einen ähnlichen Ansatz wählten Jacobs & Kinder, welche für die Bundestagswahl 2021²⁶ die Wahlprogramme sechs verschiedener deutscher Parteien mit diversen

²³ Trilcke, Peer: Wortwelten der Wahlprogramme: Digitale Analysen der Programme der Parteien zur brandenburgischen Landtagswahl 2019. In: *Netzwerk Digitale Geisteswissenschaften* (<https://www.uni-potsdam.de/de/digital-humanities/blog/wahlanalyse2019>, letzter Zugriff am 07.08.2025).

²⁴ Hierunter fallen vor allem die für eine solche Analyse wesentlichen Schritte der Extraktion und Lemmatisierung der Texte sowie Part-of-Speech-Tagging und Datenbereinigung/-homogenisierung. Siehe hierzu: Trilcke: Wortwelten der Wahlprogramme (<https://www.uni-potsdam.de/de/digital-humanities/blog/wahlanalyse2019>).

²⁵ Ebd.

²⁶ Beispiele aus den sozialen Medien belegen ebenfalls ein deutliches Interesse an der Analyse und dem Vergleich der Wahlprogramme verschiedener Parteien mithilfe des Einsatzes digitaler Methoden. Dies zeigt sich unter anderem auf *Reddit*, wo für die Bundestagswahl 2021 beispielsweise eine Analyse der häufigsten Substantive veröffentlicht wurde. Siehe hierzu: https://www.reddit.com/r/de/comments/pujwxr/analyse_der_wahlprogramme_was_sind_die_h%C3%A4ufigsten/#lightbox, letzter Zugriff am 27.09.2025.

Textähnlichkeitsmethoden verglichen.²⁷ Hierzu zählen vor allem die *Jaccard-Bag-Similarity*, *Latent Semantic Analysis*, *doc2vec* und *Sentence Bidirectional Encoder Representations from Transformers* (kurz *SBERT*).²⁸ Darüber hinaus kombiniert die Arbeit *k-means-Clustering* von SBERT-Embeddings mit dem *Fowlkes-Mallows-Score* zur Dokumentenähnlichkeit. Durch diese Methoden sollen vor allem die Gemeinsamkeiten und Unterschiede der Wahlprogramme der verschiedenen Parteien in Bezug auf deren Länge, Hauptaussagen, Verständlichkeit, Sympathie und semantische Komplexität beleuchtet werden. So ergibt Jacobs' & Kinders Analyse unter anderem beispielsweise, dass die Wahlprogramme der SPD und CDU die besten Chancen besitzen, am verständlichsten *und* sympathischsten zu sein und ähnliche Werte im Hinblick auf deren Länge und semantische Komplexität aufweisen. Jacobs & Kinder betonen darüber hinaus, dass, obwohl derartige (maschinelle) Auswertungen qualitative Analysen oder ein *Deep Reading* der Texte nicht ersetzen können, dennoch Prognosen liefern, welche dann wiederum in empirischen Studien überprüft werden können. Sie könnten zudem als Motivation dienen, bestimmte Aspekte zukünftiger Wahlprogramme zu ändern, um sie dadurch verständlicher und/oder sympathischer zu formulieren.

Stefan Hartmann hat ebenfalls die Wahlprogramme der im Bundestag vertretenen Parteien für die Bundestagswahl 2021 mit Hilfe der Programmiersprache *R* analysiert.²⁹ Hartmann legt dabei besonderen Fokus auf Keywords (sprich welche Wörter für die Wahlprogramme besonders charakteristisch sind, also deutlich häufiger auftreten als bspw. in einem Vergleichsdatensatz), Wortwolken, N-Gramme (Sequenzen aus einer beliebigen Anzahl von Wörtern) sowie Wortwolken mit Trigrammen (Wortsequenzen bestehend aus drei Wörtern) für die jeweiligen Parteien. Hierfür nutzt Hartmann verschiedene Assoziationsmaße, darunter vor allem die *Log-Likelihood Ratio* und den *Sørensen–Dice-Koeffizienten*. Ergänzend zur Verwendung dieser Assoziationsmaße nutzt Hartmann darüber hinaus auch das Tf-Idf Maß, das Term Frequency (Tf) mit Inverse Document Frequency (Idf) kombiniert.³⁰ Besonders beachtlich hierbei ist, dass Hartmann eine detaillierte Schritt-für-Schritt-Anleitung für dieses Vorgehen in *R* bietet, inklusive vordefinierten, kommentierten Codes sowie aller

²⁷ Vgl. hierzu Jacobs, Arthur M. und Kinder, Anette: Electoral Programs of German Parties 2021: A Computational Analysis Of Their Comprehensibility and Likeability Based On 'SentiArt' (https://arxiv.org/abs/2109.12500?utm_source=arxiv, letzter Zugriff am 09.08.2025).

²⁸ Ebd., S. 2. Gilt nachfolgend bis anders angegeben.

²⁹ Hartmann, Stefan: Quantitative Analyse von Wahlprogrammen (<https://empirical-linguistics.github.io/wahlprogramme2021/#Pakete>, letzter Zugriff am 13.09.2025).

³⁰ Ersteres bezieht sich auf die Frequenz eines Wortes (bzw. in Hartmanns konkretem Anwendungsfall: eines Trigramms) in einem Dokument, während letzteres dafür sorgt, dass sehr geläufige und damit frequente Wörter bzw. N-Gramme weniger stark gewichtet werden. Für weitere Informationen zu Tf-Idf siehe auch Kapitel „6.2 Term Frequency – Inverse Document Frequency (Tf-Idf)“.

Ergebnisse der Untersuchung, was eine maximale Transparenz und Reproduzierbarkeit dieser Studie im Sinne der *Digital Humanities* gewährleistet.

Eine Studie der Universität Hohenheim, veröffentlicht im August 2021, untersucht ebenfalls die Wahlprogramme zur Bundestagswahl 2021, jedoch im Vergleich aller Bundestagswahlen seit 1949.³¹ Wie schon zuvor bei Jacobs & Kinder stehen hier insbesondere die Länge, Verständlichkeit, sowie Themen und Begriffe der Wahlprogramme im Vordergrund der Analyse. Es wird also untersucht, wie sich die Wahlprogramme der verschiedenen Parteien im Laufe der Zeit verändert haben. Ergänzt wird diese Studie jedoch zusätzlich noch durch eine Populismus- sowie Sentiment-Analyse. Zu den verwendeten digitalen Methoden dieser Arbeit zählen vor allem Part-of-Speech-Tagging, Lemmatisierung der Texte, die Erstellung von Wordclouds, Tf-Idf sowie wörterbuchbasierte Sentiment-Analyse.³² Die Studie gelangt zu dem Ergebnis, dass Wahlprogramme über die Jahre hinweg zwar insgesamt immer länger und umfangreicher werden, es aber durchaus bei der Länge der Wahlprogramme große Unterschiede zwischen den einzelnen Parteien gibt.³³ Dabei stehen die Länge der Wahlprogramme und ihre formale Verständlichkeit jedoch in keinerlei Korrelation zueinander, da bei der Bundestagswahl 2021 die Linkspartei zwar das längste, aber gleichzeitig auch das verständlichste Wahlprogramm lieferte.³⁴ Darüber hinaus sind im Vergleich zu früheren Wahlen Populismus und Negativismus bei der Bundestagswahl 2021 eher weniger ausgeprägt.³⁵ Am populistischsten schreibt dabei die AfD, welche zusammen mit der Linkspartei auch am häufigsten negative Begriffe verwendet. Außerdem hat die Themen- und Begriffsanalyse gezeigt, dass die Sozialpolitik sprachlich zu den zentralen Themen der Bundestagswahl 2021 gehört, dicht gefolgt von der Außen- und Umweltpolitik, die ebenfalls häufig unter den fünf meistgenutzten Themen zu finden sind. Ferner sind die Wahlprogramme 2021 im Vergleich zu früheren Wahlen sprachlich eher positiv, wobei die negativste bzw. am wenigsten positive Sprache die Wahlprogramme der AfD und der Linken aufweisen.³⁶ Dies betrifft jedoch nicht nur die untersuchte Wahl, sondern ist auch im langjährigen Vergleich erkennbar.³⁷

Ceron et al. haben für die Bundestagswahl 2021 ebenfalls untersucht, inwiefern sich unterschiedliche Textrepräsentationen (sowohl strukturell als auch heuristisch) für den

³¹ Brettschneider, Frank und Thoms, Claudia: Die Wahlprogramme zur Bundestagswahl 2021 im Vergleich aller Bundestagswahlen seit 1949. Eine Studie der Universität Hohenheim (https://www.uni-hohenheim.de/uploads/media/Wahlprogramm-Check_2021_Bundestagswahl.pdf, letzter Zugriff am 10.08.2025).

³² Ebd., S. 39 f. & 54.

³³ Ebd., S. 5.

³⁴ Ebd.

³⁵ Ebd., S. 6. Gilt nachfolgend bis anders angegeben.

³⁶ Ebd., S. 55.

³⁷ Ebd.

Vergleich von Parteiprogrammen eignen, also inwieweit Darstellungen von Texten die Nähe von Parteien erfassen können und wie ebendiese Darstellungen wiederum optimiert werden können.³⁸ Das Ziel der Analyse bestand demnach also darin, die Frage zu beantworten, welche Embedding- und Normalisierungsstrategien besonders gut dazu geeignet sind, zuverlässige Ähnlichkeitsbewertungen für den Vergleich von Parteiprogrammen zu liefern. Durchgeführt wurde diese Studie hauptsächlich mit Hilfe des vortrainierten Sprachmodells *Sentence-BERT* (SBERT) aufgrund von dessen Fokus auf Satzähnlichkeit und Rechenleistung, wobei dieses Modell zunächst lediglich als Ausgangsbasis diente und im weiteren Verlauf der Analyse noch weiter „finegetuned“ wurde.³⁹ Ceron et al. gelangen zu dem Ergebnis, dass die in der Politikwissenschaft als relevant für diese Art der Analyse identifizierten Aspekte, nämlich beispielsweise die Annotation verschiedener Themenfelder und Behauptungen, offenbar weniger entscheidend sind.⁴⁰ Stattdessen besteht die Möglichkeit, diese aufwendige Handarbeit durch computergestützte Verfahren der Sprachverarbeitung (*NLP-Tools*) zu ersetzen. Im Rahmen der vorliegenden Studie wurde nämlich ein automatisiertes Klassifikationsverfahren eingesetzt, um die in den Aussagen enthaltenen Informationen zu erfassen. Darüber hinaus präsentierten Ceron et al. eine Methode namens *Twin Matching*, die wöchentlich evaluiert wird und eine approximative Abschätzung der inhaltlichen Ähnlichkeit von Parteien ermöglicht. Hierbei ist besonders auffällig, dass eines ihrer Modelle, das vollständig ohne manuelle Annotationen auskam, im Vergleich zu den anderen Modellen eine herausragende Leistung zeigte. Die Untersuchung ergab außerdem, dass die Vorgehensweise bei der Vorbereitung und nachträglichen Verfeinerung von Textdaten einen entscheidenden Einfluss auf die Qualität von Vorhersagen hat. Die manuelle Annotation von Inhalten wurde hierbei als weniger relevant eingestuft.

Ebenfalls von Ceron et al. stammt eine Arbeit zu einer politikbereichsbezogenen Analyse zum besseren Verständnis der Positionierung von Parteien.⁴¹ Die vorliegende Studie präsentiert eine neuartige Analyse-Methodik für Parteiprogramme, die auf der additiven Manifestzerlegung basiert. Dieses Projekt nimmt die Berechnung von Ähnlichkeiten pro Politikbereich in den Fokus, wodurch die Möglichkeit besteht, die Übereinstimmung zwischen zwei Parteien in Bezug auf bestimmte Themenfelder zu analysieren und zu bestimmen, in welchen Bereichen jeweils eine stärkere oder schwächere Übereinstimmung vorliegt. Dieses Verfahren nutzt moderne Sprachmodelle, um Politikbereiche zu definieren, Prognosen zu erstellen und

³⁸ Ceron, Tanise; Blokker, Nico und Padó, Sebastian: Optimizing text representations to capture (dis)similarity between political parties (<https://arxiv.org/abs/2210.11989>, letzter Zugriff am 10.08.2025).

³⁹ Ebd., S. 5.

⁴⁰ Ebd., S. 9. Gilt nachfolgend bis anders angegeben.

⁴¹ Ceron, Tanise; Nikolaev, Dmitry und Padó, Sebastian: Additive manifesto decomposition: A policy domain aware method for understanding party positioning (<https://arxiv.org/abs/2305.10136>, letzter Zugriff am 13.08.2025).

Ähnlichkeiten zwischen Parteien zu quantifizieren. Die Methode verbindet technische Fortschritte in der Textdarstellung mit inhaltlicher Analyse und wurde erfolgreich auf die deutsche Parteienlandschaft angewandt. Ceron et al. setzen bei dieser Analyse vorwiegend auf Textrepräsentationen durch *Sentence Embeddings*, um auf diese Weise die *Cosinus-Distanz* zwischen allen Paaren von Sätzen zu berechnen, die zu zwei bestimmten Kategorien gehören.⁴² Der Mittelwert dieser beiden Sätze wird dann wiederum als Maß für die thematische Kohärenz zwischen den beiden Kategorien verwendet.⁴³ Die Ergebnisse zeigen eine hohe Übereinstimmung mit Expert:inneneinschätzungen, ermöglichen die Rekonstruktion klassischer politischer Skalen wie der Rechts-Links-Einteilung und weisen teilweise überraschende inhaltliche Übereinstimmungen zwischen politischen Lagern auf.⁴⁴ Automatisierte Verfahren ersetzen dabei oft aufwendige manuelle Annotationen, ohne dass dabei Qualitätseinbußen zu verzeichnen sind. Darüber hinaus kann dieser Ansatz ebenso auf andere Sprachen, Länder und deren politische Parteien angewendet werden, falls eine ausreichende Textmenge zur Verfügung steht.

Rheault und Cochrane stellen ein weiteres, spannendes Fallbeispiel dar, da sie einen ähnlichen Anwendungsfall wie den in dieser Arbeit untersucht haben. Sie analysieren in ihrer Studie ebenfalls die semantische Nähe verschiedener Parteien über politische Reden.⁴⁵ Dafür trainieren sie *Word2Vec*-Modelle auf umfangreiche parlamentarische Korpora aus Großbritannien, Kanada und den Vereinigten Staaten von Amerika.⁴⁶ *Word Embeddings* –also Vektoren, die durch neuronale Netze kodiert werden, um die semantische Ähnlichkeit numerisch beschreiben zu können– werden in diesem Kontext speziell zur Analyse politischer Sprache eingesetzt. Rheault und Cochrane erweitern klassische Embeddings, indem sie die Parteizugehörigkeit als zusätzliche Variable mit einbeziehen und daraus sogenannte „Partei-Embeddings“ bilden. Diese Modelle wurden dann mit großen parlamentarischen Korpora aus Großbritannien, Kanada und den USA trainiert. Rheault und Cochrane gelangen daraufhin zu dem Ergebnis, dass Partei-Embeddings unterschwellige Konzepte wie Ideologie erfolgreich erfassen können und dass dieser Ansatz Forschenden einen geeigneten Rahmen für die Analyse politischer Sprache bietet.⁴⁷ Sie merken außerdem an, dass *Word Embeddings* eine leistungsfähige Grundlage zur Analyse politischer Texte und zur

⁴² Ebd., S. 4.

⁴³ Ebd.

⁴⁴ Ebd., S. 9. Gilt nachfolgend bis anders angegeben.

⁴⁵ Vgl. hierzu Rheault, Ludovic und Cochrane, Christopher: Word Embeddings for the Analysis of Ideological Placement in Parliamentary Corpora. In: *Political Analysis*, Vol. 28(1), 2019/20, S. 112–133 (<https://www.cambridge.org/core/journals/political-analysis/article/word-embeddings-for-the-analysis-of-ideological-placement-in-parliamentary-corpora/017F0CEA9B3DB6E1B94AC36A509A8A7B>, letzter Zugriff am 20.10.2025).

⁴⁶ Ebd., S. 112. Übersetzung des Autors; gilt nachfolgend bis anders angegeben.

⁴⁷ Ebd., S. 130. Übersetzung des Autors; gilt nachfolgend bis anders angegeben.

Visualisierung ideologischer Strukturen bieten, da semantische Methoden im Gegensatz zu reinen Wortzählmethoden, die lediglich die Anzahl der Wörter erfassen, dazu in der Lage sind, die Bedeutung von Wörtern und die Veränderungen im Sprachgebrauch über die Zeit hinweg zu erfassen. Ihre speziell angepasste Modellversion –mit den sogenannten „Partei-Embeddings“, welche die Parteizugehörigkeit mit einbeziehen– erlaubt darüber hinaus eine zuverlässige Positionierung politischer Akteure im ideologischen Raum. Die Ergebnisse zeigen eine hohe Übereinstimmung mit den Expertenschätzungen, den manuellen Kodierungen und den Abstimmungsdaten. Gleichzeitig weisen die Autoren jedoch auch auf noch offene Herausforderungen hin: So verlangt die Methode etwa große Textkorpora, ist rechenintensiv und derzeit noch ohne ausgereifte Unsicherheitsabschätzungen. Für zukünftige Arbeiten könnten demnach Bayes'sche neuronale Netze oder Variationsinferenz eine geeignete Methode darstellen. Rheault und Cochrane gelangen außerdem zu dem Schluss, dass sich das Verfahren als robust, objektiv und mit geringfügig subjektiven Entscheidungen anwenden lässt, was einen signifikanten Fortschritt für quantitative Analysen politischer Sprache darstellt.

Einen etwas anderen Ansatz verfolgten He et al., die mittels *Topic Embeddings* keine Wahlprogramme oder politischen Reden, sondern vielmehr die thematische Polarisierung von Nachrichtenartikeln mittels *Cosine Distance* im *Embedding Space* untersucht haben.⁴⁸ Dieses Beispiel verdeutlicht sehr anschaulich die vielfältigen Einsatzmöglichkeiten und Anwendungsfälle digitaler Methoden über den in dieser Arbeit vorwiegend untersuchten Gegenstand der politischen Texte hinaus.

Weitere Anwendungsbereiche zeigt unter anderem Christian Rauh auf, der Pressemitteilungen der Europäischen Kommission auf deren Komplexität in Bezug auf die darin verwendete Sprache untersuchte.⁴⁹ Für die Analyse nutzte Rauh ein umfangreiches Korpus von fast 45.000 englischsprachigen Pressemitteilungen der Europäischen Kommission aus dem Zeitraum von 1985–2020. Die Komplexität der verwendeten Sprache machte Rauh anhand der folgenden drei Parameter fest: nämlich der sprachlichen Zugänglichkeit, der Verwendung von Fachsprache und dem Handlungsbezug der politischen Kommunikation, also ob die Europäische Kommission verdeutlicht, was tatsächlich politisch entschieden und umgesetzt wird.⁵⁰ Die Messung der sprachlichen Zugänglichkeit basiert auf dem *Flesch-Kincaid Reading Ease Score*, welcher die Satzlänge (in Wörtern) und die durchschnittliche Wortlänge

⁴⁸ He, Zihao; Mokherian, Negar; Câmara, António; Abeliuk, Andrés und Lerman, Kristina: Detecting Polarized Topics Using Partisanship-aware Contextualized Topic Embeddings (<https://arxiv.org/abs/2104.07814>, letzter Zugriff am 12.08.2025).

⁴⁹ Rauh, Christian: Clear messages to the European public? The language of European Commission press releases 1985–2020. In: Journal of European Integration, Vol. 45, Nr. 4, S. 683–701 (<https://doi.org/10.1080/07036337.2022.2134860>, letzter Zugriff am 14.08.2025).

⁵⁰ Ebd., S. 691–92. Gilt nachfolgend bis anders angegeben.

(in Silben) kombiniert. Die Verwendung von Fachsprache wird anhand eines Vergleichs mit Wortfrequenzen aus dem Google Books Korpus ermittelt, während der Handlungsbezug mithilfe von *Part-of-Speech-Tagging* bestimmt wird. Als Vergleichskorpora dienen nationale Regierungs-Pressemitteilungen, wissenschaftliche Abstracts und Zeitungstexte, welche wiederum mit *Topic Modeling* und *Topical Inverse Regression Matching* zur inhaltlichen Vergleichbarkeit von Themen zwischen nationalen und supranationalen Akteuren eingesetzt werden. Die Studie gelangt zu dem Schluss, dass die Pressemitteilungen der Kommission im Vergleich zu denen nationaler Regierungen oder Medien eine geringere Verständlichkeit aufweisen.⁵¹ Sie sind durch eine komplexere Grammatik, einen stärkeren Fachjargon und einen abstrakteren, Substantiv-lastigeren Stil gekennzeichnet. Dieser technokratische Sprachstil ähnelt dabei eher wissenschaftlichen Texten als öffentlicher politischer Kommunikation. Trotz der gestiegenen politischen Kompetenzen der Kommission bei gleichzeitig zunehmender antieuropäischer Politisierung der EU ist über einen Zeitraum von 35 Jahren eine kaum wahrnehmbare Verbesserung in Bezug auf Lesbarkeit und Stil zu verzeichnen, schließt Rauh.

Nachdem in diesem Kapitel der bisherige Forschungsstand inklusive der verwendeten Methodik skizziert wurde, soll im nächsten Kapitel nun vor allem die Korpusgenerierung, also die Auswahl und Zusammensetzung des Korpus etwas näher beleuchtet und vor allem auch begründet werden.

⁵¹ Ebd., S. 695-96. Gilt nachfolgend bis anders angegeben.

3 Korpusgenerierung

Da für diesen speziellen Anwendungsfall (noch) kein existierendes Korpus aus Parteiprogrammen und politischen Reden der ausgewählten Parteien vorhanden war, auf das zurückgegriffen hätte werden können, musste für diese Arbeit das Korpus selbst zusammengestellt werden.⁵² Im Folgenden sollen daher die Auswahl der im Korpus repräsentierten Parteien sowie die Zusammensetzung des Korpus selbst etwas näher erläutert werden.

Da das Ziel dieser Arbeit ein Vergleich der sprachlichen Ähnlichkeit verschiedener deutscher und amerikanischer Parteien zur Identifizierung rechtspopulistischer Sprache⁵³ in politischen Texten mit Hilfe digitaler Methoden ist, wird zunächst eine rechtspopulistische Baseline benötigt, anhand derer dann wiederum der rechtspopulistische Grad der anderen politischen Texte gemessen werden kann. Als dieser Standard bietet sich wiederum die Nationalsozialistische Deutsche Arbeiterpartei (NSDAP) hervorragend an, da sie in ihren politischen Zielen und deren Formulierungen durchaus als rechtspopulistische Partei gelten kann.⁵⁴

Ausgehend von dieser Baseline werden daraufhin die anderen Parteien auf Ähnlichkeit mit der NSDAP geprüft, wobei insbesondere die Sprache in deren politischen Texten untersucht werden soll. Bei hohen Übereinstimmungswerten mit der NSDAP kann somit von einem hohen Anteil an Rechtspopulismus in den politischen Texten und damit einhergehend auch der Sprache der jeweiligen Partei ausgegangen werden. Hierbei ist vor allem die Ähnlichkeit aktueller rechtskonservativer bis rechter Parteien mit der NSDAP von besonderem Interesse, um den aktuellen Rechtsruck in Politik und Gesellschaft in Deutschland und den Vereinigten Staaten von Amerika anhand dieser politischen Texte zu überprüfen und, wenn möglich, diese Ähnlichkeit dann auch nachweisen zu können.

Zu diesem Zweck wurden auf amerikanischer Seite die Republikanische Partei (engl. *Grand Old Party*, kurz *GOP*) und auf der deutschen Seite die beiden Parteien

⁵² Ungeachtet dessen existieren aber durchaus vordefinierte Korpora, welche bereits annotierte Sammlungen zahlreicher Wahlprogramme (auch internationaler Parteien) enthalten. Ein solches Korpus wäre bspw. das sog. *Manifesto Project*, welches aktuell aus beinahe 3.000 maschinenlesbaren Wahlprogrammen aus über 60 verschiedenen Ländern in fast 40 Sprachen besteht. Vgl. hierzu: <https://manifesto-project.wzb.eu/information/documents/corpus>, letzter Zugriff am 20.08.2025. Da für diese Arbeit aber weniger eine möglichst breite Vielfalt verschiedener, internationaler Wahlprogramme, sondern vielmehr die Kombination aus Wahlprogrammen und politischen Reden einiger weniger, ausgewählter Parteien von Interesse ist, konnte das Manifesto Corpus zu diesem Zweck leider nicht genutzt oder anderweitig gewinnbringend eingesetzt werden.

⁵³ Unter „rechtspopulistischer Sprache“ versteht der Autor vor allem politische Texte, welche der in Kapitel 1 „Einleitung“ thematisierten Definition von Rechtspopulismus entsprechen.

⁵⁴ Näheres hierzu s. Kapitel 4.6 Nationalsozialistische Deutsche Arbeiterpartei (NSDAP).

Alternative für Deutschland (AfD) und die Christlich Demokratische Union (CDU) ausgewählt. Für das mittlere Spektrum wurde repräsentativ die Freie Demokratische Partei (FDP) ausgewählt, während die Sozialdemokratische Partei Deutschlands (SPD) sowie die Demokratische Partei der USA (DEM) die linke Mitte des politischen Spektrums repräsentieren.⁵⁵ Für jede Partei wurde wiederum das Parteiprogramm sowie zehn politische Reden von mindestens zweien der wichtigsten Vertreterinnen und Vertreter der jeweiligen Partei für die Analyse stichprobenartig ausgewählt. Hierbei wurde der Zeitraum von der aktuellen⁵⁶ bis maximal zu den letzten drei Regierungsperioden⁵⁷ abgedeckt, wobei der Fokus aus Gründen der Aktualität eher auf den letzten drei bis fünf Jahren liegt.

Für die Parteien der Opposition im Deutschen Bundestag wurden hierfür neben Nachrichtenartikeln, welche die Reden im Wortlaut transkribiert hatten, vor allem die offiziellen Plenarprotokolle des Deutschen Bundestages⁵⁸ verwendet, um aus diesen die gewünschte Anzahl an politischen Reden der jeweiligen Parteien zu extrahieren. Bei den Parteien der aktuellen Bundesregierung (also CDU und SPD) war dies sogar direkt auf der offiziellen Website der Deutschen Bundesregierung⁵⁹ möglich, wo sämtliche Reden des amtierenden Bundespräsidenten und Bundeskanzlers sowie der aktuellen Bundesministerinnen und Bundesminister nicht nur veröffentlicht, sondern darüber hinaus auch archiviert werden.

Im Falle der NSDAP gestaltete sich dies allerdings schon deutlich schwieriger, da die Gründe für die schwere öffentliche Zugänglichkeit nationalsozialistischer Reden im Internet natürlich auch historisch bedingt und daher durchaus gerechtfertigt sind. Hinzu kommt das Problem, dass bis heute keine vollständige Edition aller Reden Adolf Hitlers vorhanden ist.⁶⁰ Aus diesem Grund musste für die Analyse der politischen Reden

⁵⁵ Hierbei darf jedoch nicht außer Acht gelassen werden, dass die „Links-Rechts-Zuordnung“ in Deutschland und den USA nicht deckungsgleich ist. So werden beispielsweise demokratische Politikerinnen oder Politiker, welche sich für eine öffentlich zugängliche, gesetzliche Krankenversicherung in den USA einsetzen, von Republikanern oftmals als „kommunistisch“ und „links“ bezeichnet, während dieses System in Deutschland dem Normalfall entspricht und selbst von „rechten“ Parteien mitgetragen wird. Auch die demokratische Präsidentschaftskandidatin Kamala Harris, welche in den USA als (für amerikanische Verhältnisse) „links“ angesehen wird, hätte in Deutschland aller Wahrscheinlichkeit nach eher für die CDU als für Die Linke kandidiert.

⁵⁶ Seit 2025.

⁵⁷ 2021-2025, 2017-2021 & 2013-2017; im Fall der NSDAP deckt der Zeitraum ebenfalls die zwölf Regierungsjahre 1933-45 ab, mit der Ausnahme einer einzelnen Rede aus dem Jahr 1920.

⁵⁸ Die offiziellen Plenarprotokolle des Deutschen Bundestages sind über dessen Website frei zugänglich und können als pdf-Datei heruntergeladen werden: <https://www.bundestag.de/dokumente/protokolle/plenarprotokolle>, letzter Zugriff am 27.08.2025.

⁵⁹ Siehe hierzu <https://www.bundesregierung.de/breg-de/service/newsletter-und-abos/bulletin>, letzter Zugriff am 23.08.2025.

⁶⁰ Sich diesem Problem annehmend, arbeitet seit Januar 2024 ein Team aus Wissenschaftlerinnen und Wissenschaftlern der Goethe Universität Frankfurt am Main unter der Leitung Professor Cornelißens an einer neuen, vollständigen Edition aller gesammelten Hitlerreden aus der Regierungszeit der NSDAP (1933-45). Vgl. hierzu: Sauter, Anke: Neue Grundlage für die Hitler-Forschung. In: Philosophie und

aufgrund der stark eingeschränkten Zugänglichkeit auf diejenigen Reden zurückgegriffen werden, welche im Internet als transkribierte Digitalisate frei verfügbar waren. Diese wurden überwiegend von verschiedenen Online-Archiven bereitgestellt.⁶¹

Für die beiden US-amerikanischen Parteien wurden die politischen Reden Teils der offiziellen Website des Weißen Hauses, Teils der jeweiligen Websites der Parteien selbst, und Teils Nachrichtenartikeln, welche die Reden im Wortlaut transkribiert hatten, entnommen. Um eine einheitliche Vergleichbarkeit der verschiedenen Reden untereinander zu gewährleisten, wurden die Reden und Parteiprogramme der beiden US-amerikanischen Parteien mit Hilfe des kostenlosen Übersetzungstools *DeepL*⁶² ins Deutsche übersetzt. Dadurch ist eine zuverlässige und aussagekräftige Analyse durch die digitalen Methoden und gute Vergleichbarkeit mit den deutschen Parteien sichergestellt.

Alle in dieser Arbeit analysierten Reden enthalten in der Kopfzeile einen Internetlink, welcher deren Quelle samt letztem Zugriffsdatum angibt, und können bei Interesse im eigens hierfür angelegten [Ordner](#) im GitHub-Repository⁶³ im txt-Format eingesehen werden. Die Parteiprogramme wurden jeweils der offiziellen Website der jeweiligen Partei entnommen und sind im PDF-Format ebenfalls über das GitHub-Repository im entsprechenden [Ordner](#) einsehbar.

Geschichtswissenschaften, UniReport 2.24 (<https://aktuelles.uni-frankfurt.de/unireport/neue-grundlage-fuer-die-hitler-forschung/>, letzter Zugriff am 13.09.2025).

⁶¹ Hierzu zählen vor allem das Institut für Zeitgeschichte München/Berlin, die Forschungsstelle für Zeitgeschichte in Hamburg und die Universitätsbibliothek Paderborn, denen an dieser Stelle mein besonderer Dank für die Digitalisierung sowie für die öffentliche Bereitstellung dieser Dokumente gebührt.

⁶² Siehe hierzu <https://www.deepl.com/de/translator>.

⁶³ Siehe hierzu https://github.com/Stevens29081997/Masterarbeit_Kevin_Stevens.

4 Politikwissenschaftliche Grundlagen: Vorstellung & Einordnung der ausgewählten Parteien auf dem politischen Spektrum

Parteien sind ein wichtiger Bestandteil der Demokratie und dienen als Schnittstelle zwischen Staat und Gesellschaft.⁶⁴ Sie ermöglichen es Bürgerinnen und Bürgern, politische Repräsentanten ihrer Interessen zu wählen und somit indirekt am politischen Geschehen eines freien Staates mitzuwirken.⁶⁵ Für die ideologische und programmatische Einordnung politischer Parteien in ein bestimmtes Spektrum ergeben sich indes mehrere Möglichkeiten. Hierunter zählen vor allem das klassische „Links-Rechts-Schema“, aber auch etwas neuere Modelle, wie beispielsweise das von Seymour Martin Lipset und Stein Rokkan Ende der 1960er Jahre entwickelte Cleavage-Konzept.

Um eine Partei in ein sogenanntes „Links-Rechts-Schema“ einzuordnen, muss zunächst die Positionierung der jeweiligen Partei bezüglich der zwei gesellschaftlichen Grundkonflikte, nämlich dem sozioökonomischen Verteilungskonflikt auf der einen und dem soziokulturellen Wertekonflikt auf der anderen Seite, analysiert werden.⁶⁶ Der sozioökonomische Verteilungskonflikt besteht hierbei aus den sich gegenüberstehenden Grundpositionen der Marktfreiheit (als rechte Position) und der sozialen Gerechtigkeit (als linke Position).⁶⁷ Im Falle des soziokulturellen Wertekonflikts stehen sich dahingegen konservativ-autoritäre und libertäre Werthaltungen gegenüber.⁶⁸

Als Alternative zum klassischen „Links-Rechts-Schema“ kann dabei auch das sogenannte *Cleavage-Konzept* angesehen werden, welches politische Parteien nicht eindimensional nach ihrer ideologischen Position im Spektrum von progressiv/gleichheitsorientiert bis konservativ/marktorientiert anordnet, sondern Parteien vielmehr als Ausdruck gesellschaftlicher Konfliktlinien (z. B. Arbeit–Kapital,

⁶⁴ Gschliesser, Nathalie: Der Gebrauch von Sprache in Deutschen Wahlkampagnen : Langage utilisé pendant les campagnes électorales en Allemagne. Masterarbeit an der Faculté de philosophie, arts et lettres, Université catholique de Louvain, 2018 (<https://thesis.dial.uclouvain.be/entities/masterthesis/38a91b1e-5334-4408-a337-45ee560b4241>, letzter Zugriff am 27.09.2025).

⁶⁵ Ebd.

⁶⁶ Bückner: Die AfD, S. 36.

⁶⁷ Ebd., S. 36-37.

⁶⁸ Ebd., S. 37.

Kirche–Staat, Stadt–Land, Zentrum–Peripherie) versteht, die ihre Entstehung und Ausrichtung prägen.⁶⁹



Abbildung 1: Historisches Konfliktlinienmodell der Parteiensysteme nach Lipset/Rokkan⁷⁰

Aufgrund seiner erhöhten Komplexität wird in dieser Arbeit jedoch auf die Anwendung des *Cleavage-Konzepts* auf die verschiedenen Parteien verzichtet und der Fokus stattdessen ausschließlich auf das klassische „Links-Rechts-Schema“ gerichtet. Anhand der beiden gesellschaftlichen Grundkonflikte dieses Konzepts, nämlich dem sozioökonomischen Verteilungskonflikt und dem soziokulturellen Wertekonflikt, soll im folgenden Kapitel nun die Einordnung der verschiedenen ausgewählten politischen Parteien erfolgen.

Diese beiden gesellschaftlichen Grundkonflikte werden im folgenden Schaubild sehr anschaulich visualisiert: Abbildung 2 zeigt ein zweidimensionales Konfliktlinienmodell des heutigen deutschen Parteiensystems.

⁶⁹ Vgl. hierzu auch Decker, Frank: *Jenseits von links und rechts. Lassen sich Parteien noch klassifizieren?* (<https://www.bpb.de/shop/zeitschriften/apuz/279819/jenseits-von-links-und-rechts/>, letzter Zugriff am 17.09.2025).

⁷⁰ Quelle: Stein Rokkan: *Eine Familie von Modellen für die vergleichende Geschichte Europas*. In: *Zeitschrift für Soziologie* 2/1980, S. 121, zitiert nach Decker, Frank: *Jenseits von links und rechts. Lassen sich Parteien noch klassifizieren?* (<https://www.bpb.de/shop/zeitschriften/apuz/279819/jenseits-von-links-und-rechts/>, letzter Zugriff am 17.09.2025).

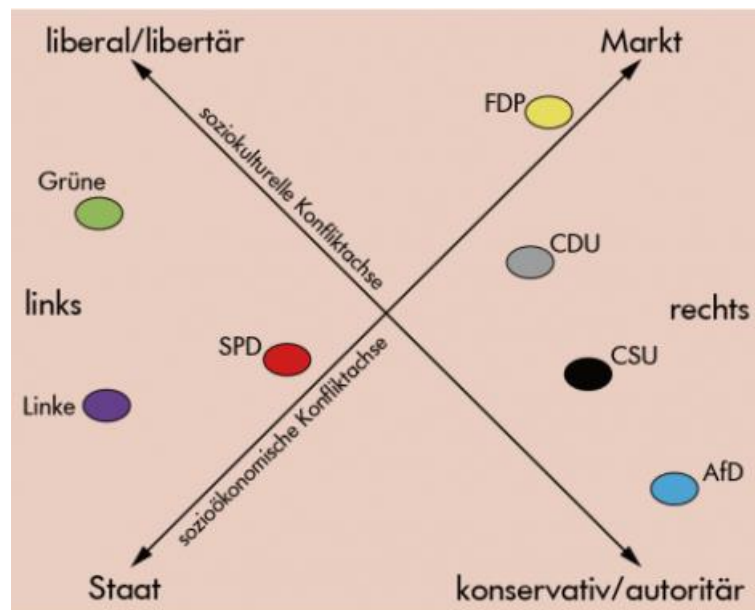


Abbildung 2: Zweidimensionales Konfliktlinienmodell des heutigen deutschen Parteiensystems⁷¹

Abbildung 2 zeigt, dass die Parteien FDP, CDU/CSU und AfD allesamt dem „rechten“ politischen Spektrum zuzuordnen sind, da sie, politisch betrachtet, deutlich rechts der Mitte stehen. Von denen in dieser Arbeit untersuchten deutschen Parteien ist somit einzig und allein die SPD dem „linken“ Spektrum zuzuordnen, da sie links der politischen Mitte steht. Wie die verschiedenen Parteien jedoch in der breiten deutschen Öffentlichkeit wahrgenommen werden, zeigt wiederum eine Links-Rechts-Einstufung der Parteien (Abbildung 3):

⁷¹ Quelle: Bundeszentrale für Politische Bildung (bpb, <https://www.bpb.de/shop/zeitschriften/apuz/279819/jenseits-von-links-und-rechts/>, letzter Zugriff am 17.09.2025).

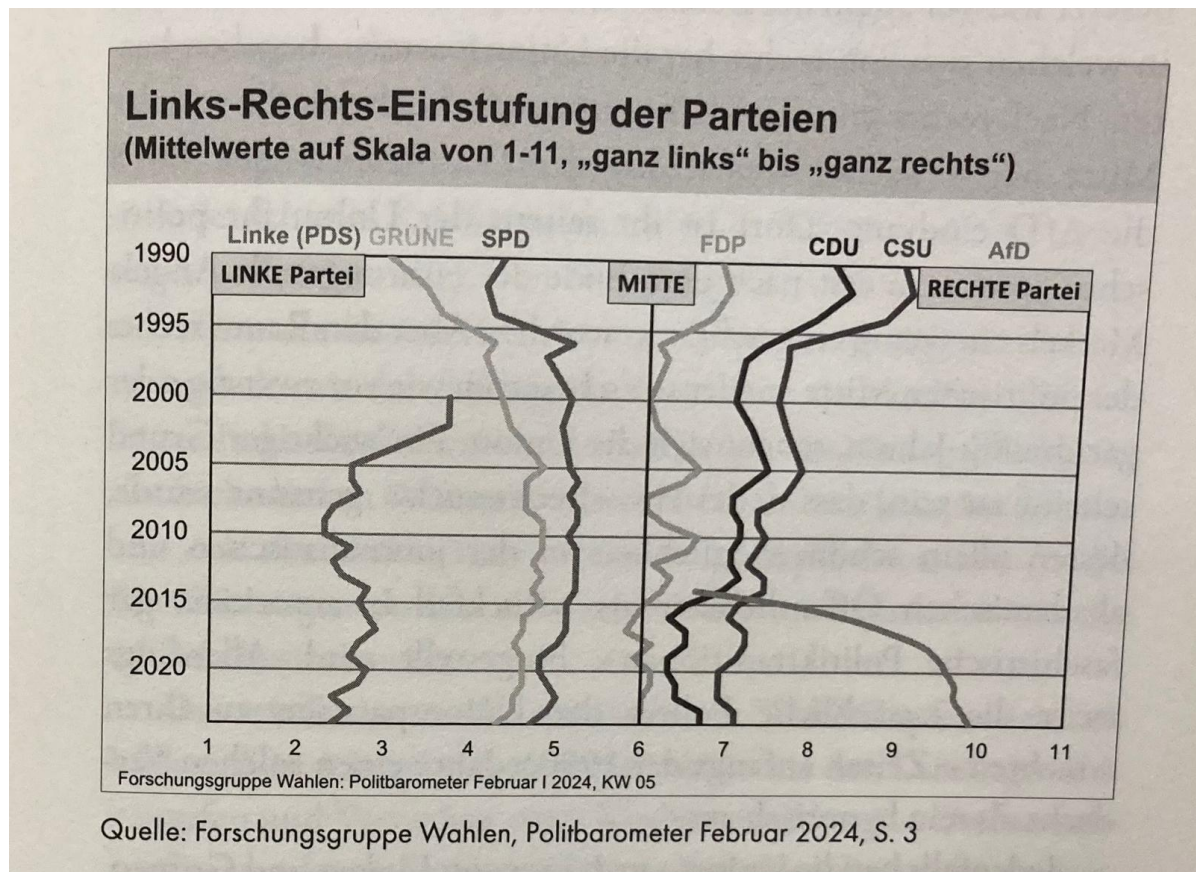


Abbildung 3: Links-Rechts-Einstufung der Parteien⁷²

Besonders auffällig ist hierbei der vermeintliche „Linksrutsch“ der Unionsparteien seit den 1990er Jahren, wie auch Werner Patzelt anmerkt.⁷³ Auch die als überwiegend links wahrgenommenen Grünen und Sozialdemokraten haben sich nach dem Eindruck der deutschen Öffentlichkeit seit 2015 erneut weiter Richtung links verschoben.⁷⁴ Aktuell stehen sie sogar noch weiter links der Mitte, als die als CDU und CSU rechts der Mitte wahrgenommen werden.⁷⁵ Dadurch war vor dem Aufkommen der AfD vorwiegend der linke und mittlere Bereich des politischen Spektrums abgedeckt, vor allem auch der Bereich um die politische Mitte herum.⁷⁶ Die Lücke hin nach rechts, welche durch die Unionsparteien frei geworden war, wurde seit 2015 systematisch und kontinuierlich von der AfD geschlossen,⁷⁷ welche sich selbst immer weiter radikalisierte

⁷² Quelle: Forschungsgruppe Wahlen, Politbarometer Februar 2024, S. 3, zitiert nach Patzelt, Werner J.: Deutschlands blaues Wunder. Die AfD und der Populismus. München 2025, S. 41.

⁷³ Patzelt, Werner J.: Deutschlands blaues Wunder. Die AfD und der Populismus. München 2025, S. 41.

⁷⁴ Ebd.

⁷⁵ Ebd.

⁷⁶ Ebd., S. 41-42.

⁷⁷ Ebd., S. 42.

und dadurch auch in der öffentlichen Wahrnehmung immer weiter rechts auf dem politischen Spektrum eingestuft wurde.

Basierend auf dem obigen Schema (Abbildung 2) sollen im nachfolgenden Kapitel nun die verschiedenen politischen Parteien zunächst im Einzelnen einmal kurz vorgestellt und auf der politischen Skala eingeordnet werden.

4.1 Alternative für Deutschland (AfD)

Die „Alternative für Deutschland“ (kurz AfD) wurde Anfang des Jahres 2013 zunächst als eurokritische Partei gegründet, woher sich auch der Name der Partei ableitet.⁷⁸ Bereits ein Jahr später zog sie ins Europäische Parlament ein und ist mittlerweile in allen Parlamenten der Bundesrepublik vertreten.⁷⁹ Darüber hinaus ist sie seit der Bundestagswahl 2017 die drittstärkste Kraft und größte Oppositionspartei im Deutschen Bundestag,⁸⁰ seit der Bundestagswahl im Februar 2025 sogar die zweitstärkste Kraft.

Dabei zeichnet sich die AfD nicht nur durch die „rasante Geschwindigkeit ihres Aufstiegs“ aus, sondern vor allem auch durch die Tatsache, dass es sich bei der AfD um die erste rechte Partei handelt, der es gelang, sich in der Bundesrepublik Deutschland tatsächlich parlamentarisch zu etablieren,⁸¹ und noch dazu über einen längeren Zeitraum hinweg. Trotz ihrer zu Beginn noch vorrangig konservativ-wirtschaftsliberalen Ausrichtung entwickelte sich die AfD innerhalb kürzester Zeit zu einer rechtspopulistischen Partei, deren Entwicklung vor allem durch mehrere parteipolitische Richtungswechsel geprägt wurde.⁸² Diese waren nicht nur mit diversen parteiinternen Streitigkeiten und anschließenden Parteiaustritten verbunden; die Partei macht darüber hinaus auch regelmäßig mit Skandalen aus dem rechtsextremen Lager Schlagzeilen,⁸³ ohne sich dabei jedoch auf klare und eindeutige Weise von diesen zu distanzieren. Bereits vor September 2013, also kurz nach der Gründung der Partei, waren Mitglieder extrem rechter und rechtspopulistischer Kleinstparteien vermehrt der AfD beigetreten.⁸⁴ Hierzu zählte vor allem die rechtspopulistische Anti-Islam-Partei „Die Freiheit“, deren ehemalige Mitglieder insbesondere in den Landesverbänden

⁷⁸ Der Name der Partei ist hierbei als direkter Bezug auf die finanziellen Hilfen der Bundesregierung an Griechenland zu verstehen, welche die damalige Bundeskanzlerin Angela Merkel im Jahre 2010 als „alternativlos“ bezeichnete. Vgl. hierzu Bücken: Die AfD, S. 29.

⁷⁹ Ebd., S. 28.

⁸⁰ Ebd.

⁸¹ Ebd.

⁸² Ebd.

⁸³ Ebd.

⁸⁴ Friedrich, Sebastian: Die AfD. Analysen – Hintergründe – Kontroversen. Berlin 2017, S. 53.

Brandenburg und Mecklenburg-Vorpommern einflussreiche Positionen innerhalb der AfD besetzten.⁸⁵

Nach der Eurokrise gilt allgemein die sogenannte europäische „Flüchtlingskrise“⁸⁶ als zweites wichtiges populistisches Momentum der AfD, da die Debatte um den Zuzug geflüchteter Menschen nach Deutschland nicht nur EU-kritische und fremdenfeindliche Ansichten innerhalb der deutschen Gesellschaft begünstigte, sondern darüber hinaus auch zu großer Unzufriedenheit in Teilen der Bevölkerung in Bezug auf die Asylpolitik der damaligen Bundesregierung führte.⁸⁷

Da ihre Programmatik ebenso marktliberale wie sozialprotektionistische Ansätze beinhaltet, ist die Zuordnung der AfD bezüglich des oben beschriebenen sozioökonomischen Verteilungskonflikts nicht eindeutig möglich.⁸⁸ In Bezug auf den soziokulturellen Wertekonflikt ist die AfD hingegen deutlich rechts von CDU und CSU zu verorten, was unter anderem an bestimmten Aussagen im Bundestagswahlprogramm nachgewiesen werden kann, demnach Deutschland einen „Paradigmenwechsel hin zu einer nationalen Bevölkerungspolitik“ brauche.⁸⁹

Dennoch ist bis heute nicht eindeutig zu beantworten, wie weit rechts die AfD tatsächlich steht, da die Partei bisher immer in einen wirtschaftsliberalen, einen nationalkonservativen und einen rechtspopulistischen Flügel untergliedert wurde.⁹⁰ Mittlerweile hat sich aber die Bezeichnung der Partei als „rechtspopulistisch“ weitgehend durchgesetzt.⁹¹ Dies liegt hauptsächlich daran, dass mit dem vermehrten Zuzug Schutzsuchender nach Deutschland seit dem Sommer 2015 die Eurokritik in

⁸⁵ Ebd.

⁸⁶ Im Zuge des vermehrten Zuzugs Schutzsuchender ab dem Jahr 2015 entfachte in Deutschland eine heftige Debatte zur Asylpolitik. Begriffe wie *Flüchtlingswelle*, *-strom* oder *-krise*, die damals häufig in der Berichterstattung genutzt wurden und heutzutage teilweise immer noch werden, gelten dabei als problematisch für die Bezeichnung des Zuzugs geflüchteter Menschen und der damit einhergehenden Debatte zur Asylpolitik, da sie ein „Ausgeliefertsein“ Deutschlands vermitteln und Schutzsuchenden die Verantwortung für asylpolitische Probleme geben. Daher wird empfohlen, vorwiegend neutralere Begriffe wie *Asyldebatte*, *Asylpolitik* oder *Zuzug Geflüchteter* zu verwenden. Vgl. hierzu Bücker: Die AfD, S. 34, zit. n. Vassiliou-Enz et al. 2017.

⁸⁷ Bücker: Die AfD, S. 33. Für eine nähere Analyse der Korrelation zwischen Migration und Wahlverhalten in Deutschland siehe auch Angeli, Oliviero und Otteni, Cyrill: Migration und Wahlverhalten in Deutschland. In: Brinkmann, Heinz Ulrich und Reuband, Karl-Heinz (Hgg.): *Rechtspopulismus in Deutschland. Wahlverhalten in Zeiten politischer Polarisierung*. Wiesbaden 2022, S. 371-393, sowie Kühnel, Steffen und Leibold, Jürgen: Die Auswirkungen von Migration und Integration auf das Wahlverhalten in Deutschland. In: Brinkmann, Heinz Ulrich und Reuband, Karl-Heinz (Hgg.): *Rechtspopulismus in Deutschland. Wahlverhalten in Zeiten politischer Polarisierung*. Wiesbaden 2022, S. 395-421.

⁸⁸ Ebd., S. 37.

⁸⁹ Ebd.

⁹⁰ Ebd.

⁹¹ Ebd.

den Hintergrund und die Asyl- und Migrationspolitik in den Vordergrund trat.⁹² Mit dieser Verschiebung ging auch eine fundamentale Änderung der Positionierung der Partei einher, und zwar von ordoliberalen wirtschaftlichen Ansichten hin zu rechts- bis nationalkonservativen.⁹³ Diese „Rechtsverschiebung“ ist nicht nur in dem 2016 verabschiedeten Grundsatzprogramm, sondern auch im Wahlprogramm für die Bundestagswahl 2017 deutlich erkennbar.⁹⁴ Diese Betrachtung der Zeitspanne von der Gründung der Partei im Februar 2013 bis zur Bundestagswahl 2017 belegt somit eine fundamentale Verschiebung der politischen Ausrichtung der AfD von zu Beginn noch euroskeptisch zu rechts- und nationalkonservativ.⁹⁵

Armin Pfahl-Traughber geht in einer politikwissenschaftlichen Analyse sogar so weit, die AfD nach einer sorgfältigen Abwägung der Argumente für und wider einer Einschätzung als rechtsextremistisch nicht nur als rechtspopulistisch, sondern sogar als rechtsextremistische Partei einzustufen: „In der Bilanz bedeutet dies, dass man es bei der AfD mittlerweile selbst mit einer rechtsextremistischen Partei zu tun hat“, welcher es gelungen sei, die politische Repräsentationslücke rechts der Unionsparteien CDU und CSU erfolgreich zu schließen,⁹⁶ wie auch in Abbildung 3 ersichtlich wird.

4.2 Christlich Demokratische Union (CDU)

Die Christlich Demokratische Union⁹⁷ (kurz CDU) entstand in der Nachkriegszeit⁹⁸ zunächst als politisches Novum, das sich vorwiegend der Überwindung des Nationalsozialismus samt christlich fundiertem Neuanfang verschrieben hatte.⁹⁹ Seither hat die Union (bestehend aus CDU und ihrer bayerischen Schwesterpartei CSU) die politischen Entscheidungen der Bundesrepublik Deutschland so stark geprägt wie keine andere Partei, da sie seit 1949 bis ins Jahr 2016 insgesamt 47 Jahre lang Regierungspartei war, wodurch sie auch den Bundeskanzler bzw. die

⁹² Ebd., S. 38.

⁹³ Ebd., S. 38-39.

⁹⁴ Ebd., S. 39.

⁹⁵ Ebd.

⁹⁶ Pfahl-Traughber, Armin: Die AfD und der Rechtsextremismus. Eine Analyse aus politikwissenschaftlicher Perspektive. Wiesbaden 2019, S. 33-41, Zitat auf S. 41.

⁹⁷ Der Name „Union“ ist hierbei bewusst gewählt, um das Wesen der Partei als Vereinigung verschiedener religiöser Konfessionen, sozialer Schichten sowie unterschiedlicher Interessen zu einer gemeinsamen politischen Kraft zu betonen. Vgl. hierzu Hemmelmann, Petra: Der Kompass der CDU. Analyse der Grundsatz- und Wahlprogramme von Adenauer bis Merkel. Wiesbaden 2017, S. 145.

⁹⁸ Bis zur Gründung der Bundespartei vergingen jedoch noch einige Jahre. Erst im Oktober 1950 schlossen sich die selbstständigen Landesparteien (mit Ausnahme der CSU) auf dem ersten Parteitag in Goslar zusammen. Vgl. hierzu Hemmelmann: Kompass der CDU, S. 146-147.

⁹⁹ Hemmelmann: Kompass der CDU, S. 143. Gilt nachfolgend bis anders angegeben.

Bundeskanzlerin stellte. Dadurch zählt die Union trotz ihres vergleichsweise jungen Alters bereits zu den erfolgreichsten Parteien Europas.

Seit jeher sieht sich die CDU dabei als „Volkspartei der Mitte“ an, während sie die SPD hingegen ins linke Lager verbannte.¹⁰⁰ Dabei vermied die Union es ganz bewusst, sich als „rechte“ Partei zu bezeichnen, um den nach der Zeit des Nationalsozialismus in der öffentlichen Wahrnehmung stark diskreditierten national-konservativen Raum weitläufig zu umschließen, wobei gleichzeitig jedoch eine klare begriffliche Abgrenzung nach Rechtsaußen geschaffen werden sollte.¹⁰¹ Dieses Konzept der „schichten- und konfessionsübergreifenden Sammlungspartei der Mitte“¹⁰² hatte durchaus Erfolg, sodass die Union nicht nur aus der Bundestagswahl 1949 sozusagen „aus dem Stand“ als stärkste Kraft hervorging, wodurch sie mit Konrad Adenauer auch den ersten Bundeskanzler der Bundesrepublik stellte, sondern darüber hinaus auch bis 1969 als Regierungspartei zwanzig Jahre lang ununterbrochen an der Macht blieb, wodurch sie die maßgeblichen Richtungsentscheidungen¹⁰³ der noch sehr jungen Republik aktiv (mit)gestaltete.¹⁰⁴

In der jüngeren Vergangenheit, vor allem seit dem Amtsantritt Angela Merkels als erster Bundeskanzlerin der Bundesrepublik Deutschland im Jahr 2005 und dann nochmals zehn Jahre später mit dem Beginn der europäischen Asyldebatte 2015/16, lässt sich in der öffentlichen Wahrnehmung ein deutlicher „Linksruck“ der politischen Ausrichtung der CDU verzeichnen (s. Abbildung 3). Dies führte unter anderem zu dem Vorwurf, die CDU habe unter Angela Merkel, die das Wesen des Konservativen als „Die Wirklichkeit annehmen, fähig zur Erneuerung sein und zugleich Wertvolles bewahren“¹⁰⁵ definiert hatte, ihre Identität und ihre konservative „DNA“ verloren.¹⁰⁶ Diesen Vorwurf entkräftigt jedoch Petra Hemmelmann, indem sie sämtliche

¹⁰⁰ Walter, Franz; Werwath, Christian und D'Antonio, Oliver: Die CDU. Entstehung und Verfall christdemokratischer Geschlossenheit (=Die politischen Parteien der Bundesrepublik Deutschland). 1. Auflage, Baden-Baden 2011, S. 19.

¹⁰¹ Bereits Konrad Adenauer wehrte sich vehement gegen innerparteiliche Versuche, die CDU als rechte Partei zu bezeichnen. Vgl. hierzu Walter: Die CDU, S. 19.

¹⁰² Hemmelmann: Kompass der CDU, S. 147.

¹⁰³ Hierzu zählen vor allem außenpolitisch die Wiedererlangung der staatlichen Souveränität, die Wiederbewaffnung, die Westbindung, die Aussöhnung mit Frankreich und Israel, die europäische Einigung sowie innenpolitisch der Aufbau der Sozialen Marktwirtschaft, die Forcierung des sozialen Wohnungsbaus, die Eingliederung von Vertriebenen und die Einführung des Kindergelds. Vgl. hierzu Hemmelmann: Kompass der CDU, S. 148.

¹⁰⁴ Hemmelmann: Kompass der CDU, S. 147.

¹⁰⁵ Brechenmacher, Thomas: Die CDU unter Angela Merkel (2000-2018). In: Lammert, Norbert: *Christlich Demokratische Union. Beiträge und Positionen zur Geschichte der CDU*. 1. Auflage, München 2020, S. 81-135, Zitat auf S. 88.

¹⁰⁶ Von medialer Seite wurde Angela Merkel immer wieder vorgeworfen, die CDU zu „entkernen“. Vgl. hierzu Hemmelmann: Kompass der CDU, S. 427.

Wahlprogramme der CDU von Adenauer bis Merkel analysiert. Hemmelmann gelangt zu dem Schluss,

*[...] dass die Markenidentität der CDU, wie sie in den Parteiprogrammen ihren Ausdruck findet, nicht entkernt wurde. [...] Eine inhaltliche Entkernung mit einhergehender Fokussierung auf die Person Merkel ist in den Programmen nicht feststellbar.*¹⁰⁷

Ganz im Gegenteil sei auf der Basis der vorliegenden Erkenntnisse unter Merkel sogar vielmehr von einer „rhetorischen Profilschärfung entsprechend dem langfristigen Markenkern der Partei“ zu sprechen.¹⁰⁸

4.3 Demokratische Partei der USA (DEM)

Das Parteiensystem der USA wird seit jeher von zwei großen Parteien dominiert, nämlich der Demokratischen Partei (DEM) auf der einen und der Republikanischen Partei (*Grand Old Party*, kurz *GOP*) auf der anderen Seite.¹⁰⁹ Diese beiden Parteien haben sich im Laufe ihrer Geschichte jedoch mehrmals gewandelt, sodass ihr Stellenwert im politischen System der USA in der ersten Hälfte des 21. Jahrhunderts keineswegs klar war.¹¹⁰ Im Gegensatz zu europäischen Ländern handelt es sich bei den amerikanischen Parteien nämlich nicht um straff organisierte, homogene Organisationen mit einem inhaltlich fixierten, polarisierenden Parteiprogramm.¹¹¹

Der Beginn der amerikanischen Parteien lässt sich auf das Jahr 1800 datieren, als die Organisationen zweier Lager parteiähnliche Funktionen übernahmen, die bis in die heutige Zeit nachwirken.¹¹² Hierbei handelte es sich einerseits um die *Democratic-Republicans*, die Vorläufer der heutigen Demokraten, und die *National Republicans*, welche als Vorgänger der heutigen Republikaner gelten.¹¹³ 1848 gründeten die aus den *Democratic-Republicans* hervorgegangenen Demokraten schließlich ihr erstes *National Committee*, welches den Beginn einer nationalen Parteiorganisation darstellte.¹¹⁴

Letztendlich war es jedoch der demokratische Präsident Woodrow Wilson (1913-21), welcher unter anderem den Kriegseintritt der USA in den Ersten Weltkrieg durchsetzte

¹⁰⁷ Hemmelmann: Kompass der CDU, S. 435.

¹⁰⁸ Ebd.

¹⁰⁹ Gellner Winand und Kleiber, Martin: Das Regierungssystem der USA. Eine Einführung. 1. Auflage, Baden-Baden 2007, S. 142.

¹¹⁰ Ebd., S. 143.

¹¹¹ Ebd.

¹¹² Ebd., S. 144.

¹¹³ Ebd., S. 144-145.

¹¹⁴ Ebd., S. 147.

und heute meist wegen seines Engagements für den Völkerbund erinnert wird, der zahlreiche Reformen anregte und den Wandlungsprozess der Demokratischen Partei hin zu einer -im amerikanischen Sinne- „liberalen“, also moderat-linken Partei, begann, indem er beispielsweise eine progressive Einkommenssteuer einführte.¹¹⁵

Darüber hinaus ist die Demokratische Partei der USA eng mit der amerikanischen Arbeiterbewegung (*labor movement*) verbunden, was spätestens seit der Etablierung des Gewerkschaftsbundes *Congress of Industrial Organizations (CIO)* Mitte der 1930er Jahre evident wurde.¹¹⁶ Diese Allianz zwischen Arbeiterschaft und Demokraten hat seither die grundlegende Ausrichtung der Partei bestimmt und langfristigen Einfluss jenseits der Durchschnittswähler ausgeübt. Dies gilt ebenso für das Verhältnis der christlichen Rechten zur Republikanischen Partei der USA. Noch wichtiger ist jedoch, dass diese Bündnisse sich deutlich von den Allianzen der anderen großen sozialen Bewegungen unterscheiden, denen es nicht gelang, einen festen Platz innerhalb der politischen Parteien zu finden und zu behaupten.

Während der amerikanischen Bürgerrechtsbewegung (*Civil Rights Movement*) der 1950er und 60er Jahre wurde die Demokratische Partei zunehmend mit den Bürgerrechten für Schwarze identifiziert, wohingegen weiße Wählerinnen und Wähler aller Bevölkerungsschichten mit einer auf „Familienwerten“ basierenden Botschaft angesprochen werden konnten, wie sie vor allem durch die Republikaner vermittelt wurde.¹¹⁷ Hinzu kommt, dass sich Demokraten und Republikaner im Laufe der Zeit nicht nur programmatisch immer weiter voneinander entfernt haben, sondern darüber hinaus auch sehr unterschiedliche Sektoren der amerikanischen Wählerschaft ansprechen, ganz besonders in Bezug auf Herkunft, ethnischer Zugehörigkeit, Geschlecht, Religion und Wohnort.¹¹⁸

4.4 Freie Demokratische Partei (FDP)

Die Freie Demokratische Partei, kurz FDP, wurde im Dezember 1948 in Heppenheim gegründet und entstand aus dem Zusammenschluss der liberalen Parteien aller vier Besatzungszonen.¹¹⁹ Trotz eines regional teilweise sehr unterschiedlichen Liberalismus-Verständnisses konnte sich die FDP im Wahlkampf zur ersten deutschen Bundestagswahl (August 1949) als antisozialistische, nationale und teils antiklerikale

¹¹⁵ Greven, Thomas: Die Republikaner. Anatomie einer amerikanischen Partei. München 2004, S. 59.

¹¹⁶ Tarrow, Sidney: *Movements and Parties. Critical Connections in American Political Development*. Cambridge et al. 2021, S. 14. Übersetzung des Autors, gilt nachfolgend bis zum Ende des Absatzes.

¹¹⁷ Greven: Die Republikaner, S. 69.

¹¹⁸ Tarrow: *Movements and Parties*, S. 22, Übersetzung des Autors.

¹¹⁹ Treibel, Jan: Die FDP. Prozesse innerparteilicher Führung 2000-2012. 1. Auflage, Baden-Baden 2014, S. 50.

Alternative zu CDU und SPD mit fast 12% der Gesamtstimmen behaupten.¹²⁰ Seither ist es der FDP gelungen, als Koalitionspartner immer wieder an verschiedenen Regierungen und Koalitionen der Bundesrepublik aktiv als Regierungspartei mitzuwirken. Dies war zuletzt in der sogenannten „Ampel-Koalition“ (2021-25), bestehend aus den Parteien SPD, FDP und Grüne der Fall. Die Koalition zerbrach jedoch vorzeitig, was zu Neuwahlen führte. In der aktuellen Legislaturperiode (seit Februar 2025) ist die FDP jedoch an der 5%-Hürde knapp gescheitert und daher im aktuellen Deutschen Bundestag nicht mehr vertreten.

Vom politischen Selbstverständnis her begreift sich die FDP als den organisatorischen Träger des Liberalismus, der sich neben Konservatismus und Sozialismus als dritte politische Ideologie des 18. Und 19. Jahrhunderts herausbildete.¹²¹ Dies wird im folgenden Venn-Diagramm ersichtlich, auf dem eine Verortung dieser drei politischen Ideologien auch in Bezug auf das zu Beginn bereits erwähnte „Links-Rechts-Schema“ erfolgt:

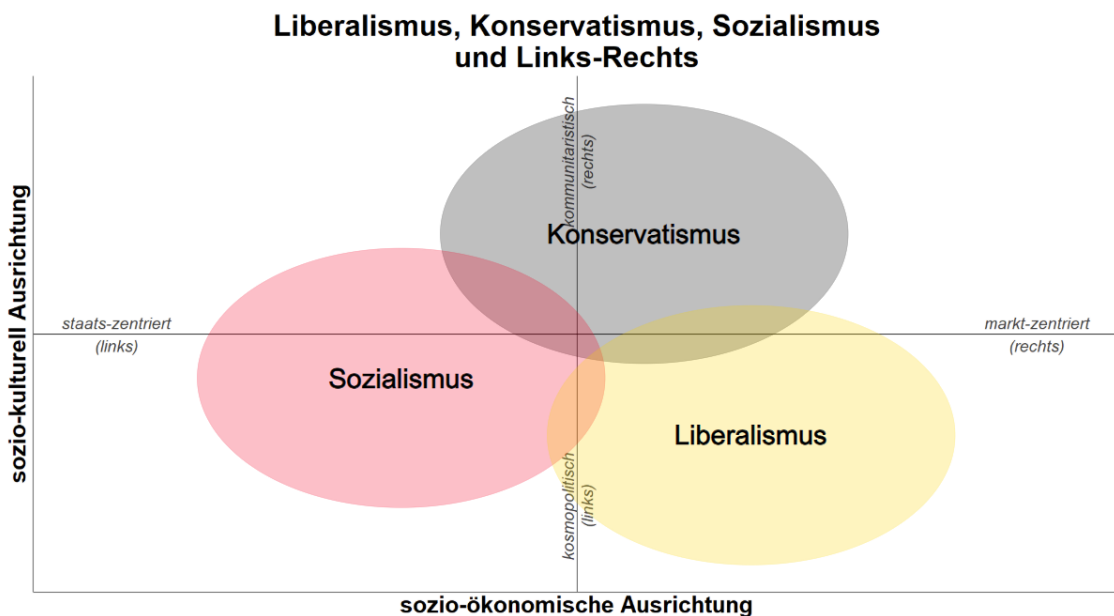


Abbildung 4: Links-Rechts-Verortung der drei großen ideologischen Strömungen des 19. Jahrhunderts¹²²

Der Liberalismus kann demzufolge also sowohl als kosmopolitisch als auch markt-zentriert eingeordnet werden. Zentrale Grundprinzipien dieser politischen Strömung sind unter anderem Meinungsfreiheit, Rechtsstaatlichkeit, eine repräsentative

¹²⁰ Ebd., S. 51.

¹²¹ Ebd., S. 45.

¹²² Quelle: Kleiser, Thomas: Wer und was ist Links oder Rechts? Teil I: Bedeutung von Links und Rechts (<https://socialsciencemeetsdatascience.de/analysis/article1.1.html>, letzter Zugriff am 24.09.2025).

Demokratie und freie Marktwirtschaft.¹²³ Ökonomisch setzen die Liberalen auf die freien Kräfte des Marktes und des Wettbewerbs durch die Mechanismen von Angebot und Nachfrage.¹²⁴ Bis heute können mindestens zwei Hauptströmungen des Liberalismus auf dem sozioökonomischen Grundkonflikt unterschieden werden, die beide auch innerhalb der FDP zu finden sind.¹²⁵ Hierbei handelt es sich um den Wirtschaftsliberalismus¹²⁶ auf der einen und den Sozialliberalismus¹²⁷ auf der anderen Seite.

Ferner definiert die FDP ihr inhaltlich-programmatisches Profil gemäß den im Jahr 2012 auf dem Bundesparteitag in Karlsruhe beschlossenen „Karlsruher Freiheitsthesen“, in denen die „Freiheit des Einzelnen“ als „Grund und Grenze liberaler Politik“ definiert wird.¹²⁸ Eine gewisse grundlegende Skepsis gegenüber jeglicher Art von staatlichen Eingriffen in Wirtschaft oder Gesellschaft war dabei seit jeher ein wesentlicher Kernbestandteil der Programmatik der FDP, ganz getreu dem Motto „Privat vor Staat“. In der Wirtschaftspolitik bevorzugt die FDP demnach eine zurückhaltende staatliche Ordnungspolitik sowie Deregulierung. Gesellschafts- und familienpolitisch hingegen zeigt sich die FDP durchaus offen für Liberalisierungsschritte, während sie innenpolitisch gegen umfassende Überwachungsrechte des Staates kämpft. Verkehrspolitik neigt sie eher zu einer vorsichtigen beziehungsweise kritischen Haltung gegenüber einer Verkehrswende hin zur Elektromobilität.

¹²³ Ebd., S. 46.

¹²⁴ Ebd.

¹²⁵ Ebd.

¹²⁶ Oberstes Ziel des Wirtschaftsliberalismus ist die Entfaltung des freien Individuums auf einem freien Markt innerhalb eines freien Wettbewerbs. Die einzige Aufgabe des Staates ist die Wahrung der Sicherheit, weshalb seine Rolle möglichst passiv, mit minimalen Eingriffen in den Markt und die freie Wirtschaft, auszufallen hat. Diese liberale wirtschaftspolitische Ausrichtung fand in der Praxis beispielsweise Ende des 20. Jahrhunderts in Großbritannien (unter Margret Thatcher) und den USA (unter Ronald Reagan) Anwendung. Vgl. hierzu Treibel: Die FDP, S. 46.

¹²⁷ Nach sozialliberaler Ansicht kann die freie Entfaltung des Individuums jedoch nicht durch einen komplett inaktiven Staat erreicht werden, da dieser gegebenenfalls auf Marktversagen reagieren können und daher die Rahmenbedingungen für individuelles Handeln so ausgestalten muss, dass Chancengleichheit für alle Individuen herrscht. Demnach ist ein begrenztes staatliches Eingreifen in Markt und Wettbewerb nötig. Ein liberaler Wohlfahrtsstaat und eine nach klaren Regeln begrenzte Marktwirtschaft sind ebenso unabdingbar mit der Ideologie des Sozialliberalismus verbunden wie öffentliche Aufgaben des Staates, zu denen beispielsweise die Unterhaltung der öffentlichen Infrastruktur oder die soziale Existenzsicherung zählen. Vgl. hierzu Treibel: Die FDP, S. 47.

¹²⁸ Lempp, Jakob; Serfling, Oliver und Rolf, Jan Niklas: Parteienhängerschaft in Deutschland. Eine Analyse der Parteien und ihrer Anhängerschaften in Bund und Ländern. Wiesbaden 2023, S. 119-120. Gilt nachfolgend bis zum Ende des Absatzes.

4.5 Republikanische Partei der USA (GOP)

Die moderne Republikanische Partei (*Grand Old Party*, kurz *GOP*) ging im Jahre 1854 aus den Resten der *National-Republicans* hervor und bildet seit dieser Zeit, zusammen mit den Demokraten, das Zwei-Parteiensystem der USA, das wir heute kennen.¹²⁹ Ursprünglich vorwiegend für ein Ende der Sklaverei eintretend, stellte sie mit Abraham Lincoln ihren ersten Präsidenten (1861-65).¹³⁰ Über die nächsten 50 Jahre hinweg wandelte sich die Partei jedoch und wurde nun zu einer eindeutig konservativen und entschlossen wirtschaftsfreundlichen Partei.¹³¹ In der zweiten Hälfte des 20. Jahrhunderts entwickelte sich innerhalb der Republikanischen Partei darüber hinaus auch eine konservative Bewegung gegen den eigenen damaligen Präsidenten Dwight D. Eisenhower (1953-61) und dessen sogenannten „Modernen Republikanismus“. ¹³² Neben dem Antikommunismus war vor allem eine verfassungspolitisch begründete Ablehnung zu großer Eingriffe des Bundes in die Angelegenheiten der Einzelstaaten ein verbindendes Element dieser neuartigen Bewegung.¹³³

In den 1980er Jahren gelang es Ronald Reagan, dem ehemaligen Gouverneur von Kalifornien (1967-75) und späteren US-Präsidenten (1981-89), der zuvor von der Demokratischen zur Republikanischen Partei gewechselt war, sowohl die elitär-intellektuellen Neokonservativen, als auch die populistischen Neuen Rechten und die Christlichen Rechten auf republikanischer Seite zu vereinen, wodurch er eine regelrechte „konservative Revolution“ hervorbrachte.¹³⁴ Diese Entwicklung brachte unweigerlich aber ebenso einen Rechtsruck der Partei mit sich, nicht zuletzt unter dem Einfluss evangelikaler Christen und des amerikanischen Südens.¹³⁵ Denn während die Demokraten ihrer Bandbreite von Interessengruppen neue Wählergruppen hinzufügten, absorbierte die Republikanische Partei eine Reihe rechter Bewegungen, stieß ihren gemäßigten Flügel ab und wurde im Zuge dessen zu einer ideologisch geprägten Partei.¹³⁶

Zusätzlich verfügt die Republikanische Partei im Gegensatz zu den Demokraten in der heutigen Zeit über eine höhere Wählerkohäsion, da es den Republikanern im Rahmen der Amtszeit von Präsident George W. Bush (2001-09) gelang, weite Teile der Bevölkerung, darunter vor allem religiös-konservative Wählerinnen und Wähler nicht nur zu mobilisieren, sondern auch parteipolitisch langfristig und dauerhaft an sich zu

¹²⁹ Gellner/Kleiber: Regierungssystem der USA, S. 147.

¹³⁰ Greven: Die Republikaner, S. 47.

¹³¹ Ebd., S. 59.

¹³² Ebd., S. 67.

¹³³ Ebd.

¹³⁴ Ebd., S. 72-73.

¹³⁵ Ebd., S. 212.

¹³⁶ Tarrow: Movements and Parties, S. 25, Übersetzung des Autors.

binden.¹³⁷ Zudem haben die Republikaner nach den Terroranschlägen vom 11. September 2001 unumstritten das Thema „Nationale Sicherheit“ für sich in Beschlag genommen. Hinzu kommt, dass die Republikanische Partei vor allem durch das schnell gewachsene, von Rupert Murdoch aufgebaute konservative Sendernetzwerk *Fox News* auf medialer Ebene sehr stark unterstützt wird, da es der Partei „eine exzellente Möglichkeit zur öffentlichen Darstellung auf nationaler Ebene“ bietet.

Torben Lütjen geht sogar so weit, die Geschichte der Republikanischen Partei als eine Geschichte der „permanenten Steigerung eines populistischen Politikstils“ zu bezeichnen.¹³⁸ Dieser populistischen Erzählung zufolge ist Amerika in zwei unterschiedliche Sphären unterteilt: Das sogenannte *Heartland*, wo die Menschen ehrlich, bodenständig, hart arbeitend, patriotisch und gottesfürchtig sind, und dem anderen Teil des Landes, der sich aus einer „urbanen Elite“ zusammensetzt, die in Washington D.C, der New Yorker Wall Street, den Hollywood Hills und San Francisco lebt. Nach diesem -auch medial sehr populären- Narrativ sind die Menschen, die dort wohnen, in jedem Fall das Gegenteil von dem, was die Bewohner des *Heartland* ausmacht. Darüber hinaus ist die „urbane Elite“ arrogant, abgehoben und schaut dieser Erzählung zufolge auf die hart arbeitenden Menschen des *Heartland* herab.

Ferner bewegten sich die Republikaner im 21. Jahrhundert hin zu dem Modell einer Bewegungspartei, indem sie einerseits einen Großteil der Energie der Tea-Party-Bewegung¹³⁹ aufgriffen -man könnte sogar sagen, ab 2010 *war* die *Tea Party* die Republikanische Partei, da ihr rechter Flügel die gesamte Partei übernommen hatte,¹⁴⁰- und andererseits von der Trump-Bewegung übernommen wurden.¹⁴¹ Dieser Kurswechsel steigerte die ideologische Bedeutung innerhalb der Partei und machte die Parteiführung zunehmend zurückhaltender, mit ihren politischen Gegnern zu verhandeln.¹⁴² Vor allem unter Donald Trump lässt sich ab dem Jahr 2016 eine „beispiellose Radikalisierung und sprachliche Enthemmung“ innerhalb der Republikanischen Partei beobachten, wie Torben Lütjen anmerkt.¹⁴³ So habe Trump nicht nur die populistische Mentalität der Republikaner radikalisiert, sondern die Partei darüber hinaus auch näher an den Politikstil des europäischen Rechtspopulismus herangeführt. Als Gemeinsamkeiten nennt Lütjen unter anderem die extreme

¹³⁷ Gellner/Kleiber: Regierungssystem der USA, S. 154. Gilt nachfolgend bis zum Ende des Absatzes.

¹³⁸ Lütjen, Torben: Partei der Extreme: Die Republikaner. Über die Implosion des amerikanischen Konservatismus. Bielefeld 2016, S. 99. Gilt nachfolgend bis zum Ende des Absatzes.

¹³⁹ Die sogenannte *Tea-Party* ist eine Bewegung, die im Jahr 2009, nach dem Übergang von Bush zu Obama, innerhalb der Republikanischen Partei entstand und seither als ihr rechter Flügel angesehen wird. Zu den wichtigsten Unterstützern der *Tea-Party* zählen hauptsächlich konservative Lobby- und Interessensgruppen. Vgl. hierzu Lütjen: Partei der Extreme, S. 113-114.

¹⁴⁰ Lütjen: Partei der Extreme, S. 120-121.

¹⁴¹ Tarrow: Movements and Parties, S. 15, Übersetzung des Autors.

¹⁴² Ebd.

¹⁴³ Lütjen: Partei der Extreme, S. 129-130. Gilt nachfolgend bis zum Ende des Absatzes.

Personalisierung, den Tabubruch als Methode, den Hang zum Verschwörungsdenken, den Anti-Intellectualismus sowie die Konzentration auf einen aggressiven Ethno-Nationalismus bei gleichzeitiger Flexibilität auf zahlreichen anderen Politikfeldern.

Zwar hatte die Republikanische Partei bereits vor Donald Trumps erster Amtszeit seit mehr als zwei Jahrzehnten fundamentalistische Christen, libertäre Staatsskeptiker und rechtsradikale Weiße hofiert, stand aber inhaltlich weiter für Freihandel, solide Haushalte, Big Business und robusten Internationalismus.¹⁴⁴ Mit seiner ersten Amtszeit (2017-21) richtete Trump die Partei mit seinem protektionistischen, immigrationsfeindlichen und isolationistischen Kurs jedoch neu aus. Entscheidendes Merkmal seiner Politik waren Identitätsfragen wie ethnische Herkunft, Nationalität, Religion und Geschlecht, was er, mehr als jeder andere Politiker, auch entsprechend ins Zentrum seiner Reden und Handlungen stellte. Nach dem Gewinn der Kandidatur und seinem Sieg bei den Präsidentschafts- und Kongresswahlen 2016 schnitt Donald Trump die Republikanische Partei völlig auf seine Person zu, radikalisierte sie weiter und besetzte Schlüsselpositionen mit seinen Gefolgsleuten, sodass die Partei mehr und mehr zum Vehikel von Trumps Ambitionen und Impulsen verkam und Loyalität zu seiner Person zum entscheidenden Kriterium für eine Karriere bei den Republikanern wurde.

4.6 Nationalsozialistische Deutsche Arbeiterpartei (NSDAP)

Die Nationalsozialistische Deutsche Arbeiterpartei, kurz NSDAP, ist verantwortlich für die Entstehung des Zweiten Weltkrieges, in dessen Verlauf über 50 Millionen Menschen den Tod fanden, sowie an der systematischen Ausgrenzung, Verfolgung und Ermordung von insgesamt 6 Millionen europäischen Jüdinnen und Juden. Ohne diese schrecklichen Gräueltaten außer Acht zu lassen, soll der Fokus dieser Arbeit jedoch auf den innen- und außenpolitischen Zielsetzungen der NSDAP liegen. Nachdem diese ausreichend skizziert und analysiert wurden, soll auf diese Weise, auf der Basis des NSDAP Wahlprogramms sowie der zu diesem Zweck exemplarisch ausgewählten Reden der wichtigsten politischen Vertreter der Partei, ein Vergleich mit den anderen Parteien ermöglicht werden. Hierbei soll die NSDAP als rechtspopulistische „Baseline“ angesehen werden. Dies bedeutet, dass die anderen Parteien, je „näher“ beziehungsweise ähnlicher sie der NSDAP sind, infolgedessen automatisch auch als umso „rechter“ beziehungsweise rechtspopulistischer angesehen werden können.

¹⁴⁴ Bierling, Stephan: Die Unvereinigten Staaten. Das politische System der USA und die Zukunft der Demokratie. 2. Auflage, München 2024, S. 91-92. Gilt nachfolgend bis zum Ende des Absatzes.

Die NSDAP wurde im Jahr 1919, also kurz nach dem Ende des Ersten Weltkrieges, als neue Partei gegründet und trug zunächst noch den Namen „Deutsche Arbeiterpartei“ (DAP).¹⁴⁵ Dabei war die DAP anfangs nur eine unter vielen völkisch gesinnten und rechtsradikalen Gruppierungen, die sich alle in ihrer Abneigung gegen die neue Republik, ihren nationalistischen Forderungen, ihrem Militarismus und ihren antisemitischen Ansichten ähnelten. Kurze Zeit später, am 24. Februar 1920, wurde das Parteiprogramm, aufgrund seiner 25 zentralen Thesen auch „25-Punkte-Programm“ genannt, im Münchner Hofbräuhaus offiziell vorgestellt und die Partei fortan in „Nationalsozialistische Deutsche Arbeiterpartei“ umbenannt. Zentrale Forderungen dieses Programms waren für die damalige völkische Denkwelt typische Programmpunkte, darunter unter anderem beispielsweise die Errichtung eines „Groß-Deutschlands“, die Durchsetzung einer Pressezensur, die Einführung eines deutschen „Gemein-Rechts“, die „Hebung der Volksgesundheit“ und die Aufstellung eines durch Wehrpflicht gespeisten „Volksheeres“. Auch zahlreiche antisemitische Maßnahmen waren im nationalsozialistischen Wahlprogramm enthalten: So sollten Jüdinnen und Juden keine deutschen „Volksgenossen“ und damit auch keine deutschen Staatsbürger mehr sein, und „Nicht-Deutsche“ hatten das Reich zu verlassen. Weitere Absätze sahen außerdem den Aufbau einer mittelständisch ausgerichteten, nicht-kapitalistischen Volkswirtschaft vor. Auch eine vehemente Ablehnung des Versailler Vertrags, der nicht nur in völkischen, sondern auch in nationalkonservativen und antidemokratischen Kreisen gemeinhin als „Schandfrieden“ oder „Diktatfrieden“ betitelt wurde, mit all seinen Bestimmungen (Gebietsabtretungen, Reparationszahlungen, alleinige Kriegsschuld Deutschlands) fand im Parteiprogramm der NSDAP Ausdruck.¹⁴⁶

Nach einem gescheiterten Putschversuch in München am 9. November 1923 wurden die führenden Parteimitglieder inhaftiert und die Partei zunächst reichsweit verboten.¹⁴⁷ Nach der Neugründung der Partei 1925 wuchs diese in der zweiten Hälfte der 1920er Jahre langsam, aber beständig, und auch die Organisationsstrukturen und Strategien der Partei veränderten sich.¹⁴⁸ So versuchte die NSDAP seit 1928 ganz gezielt, neue Wählergruppen jenseits der radikal-völkischen Kreise ihres bisherigen Klientels anzusprechen.¹⁴⁹ Hierzu zählte nicht nur ein mittelständisches Publikum, sondern ebenso Studenten oder Beamte. Und auch wenn die Partei sich fortan auf demokratischem Wege um Wahlerfolge und Sitze im Parlament bemühte, so blieb die

¹⁴⁵ Ahlheim, Hannah: Der Nationalsozialismus 1933-1939. Paderborn 2025, S. 15. Gilt nachfolgend bis anders angegeben.

¹⁴⁶ Ebd., S. 16.

¹⁴⁷ Ebd., S. 17.

¹⁴⁸ Ebd., S. 17-18.

¹⁴⁹ Ebd., S. 18. Gilt nachfolgend bis anders angegeben.

Zerschlagung der Republik samt all ihrer demokratischen Institutionen weiterhin ein zentrales und zugleich vor allem offen formuliertes Ziel.

Die Weltwirtschaftskrise Ende Oktober 1929 mit anschließender hoher Arbeitslosigkeit im Deutschen Reich (Höchstwert von 16,3% im Februar 1932) beflügelten den weiteren Aufstieg der NSDAP zusätzlich, sodass diese bei den darauffolgenden Reichstagswahlen im September 1930 ganze 18,3% der Stimmen für sich verbuchen konnte, wodurch sie zur zweitstärksten Kraft im Deutschen Reichstag wurde.¹⁵⁰ Nur zwei Jahre später, bei der Reichstagswahl am 31. Juli 1932, wählten ganze 37,3% der Deutschen die Nationalsozialisten, was das beste Ergebnis darstellte, das jemals eine Partei bei einer Reichstagswahl vor 1933 erringen konnte.¹⁵¹ Somit war die NSDAP nun zu einer Volkspartei geworden, deren Unterstützer den unterschiedlichsten Schichten und Milieus der deutschen Gesellschaft entstammten.

4.7 Sozialdemokratische Partei Deutschlands (SPD)

Die Sozialdemokratische Partei Deutschlands, kurz SPD, blickt auf eine lange Geschichte zurück und gilt als die älteste Partei Deutschlands.¹⁵² Ihre Ursprünge finden sich in den Auseinandersetzungen um die soziale Frage zwischen den von Karl Marx und Friedrich Engels inspirierten August Bebel und Wilhelm Liebknecht auf der einen und dem eher reformorientierten Ferdinand Lassalle auf der anderen Seite und reichen bis in die zweite Hälfte des 19. Jahrhunderts zurück. Während der Zeit des Nationalsozialismus verboten, entstand die SPD in den drei westlichen Besatzungszonen noch vor der Gründung der Bundesrepublik neu. Seither prägte sie die neu gegründete Bundesrepublik zusammen mit der CDU maßgeblich als eine der beiden großen Volksparteien und stellte über insgesamt mehr als 20 Jahre lang insgesamt vier Bundeskanzler und drei Bundespräsidenten.

Die SPD positionierte sich dabei spätestens seit dem sogenannten „Godesberger Programm“ aus dem Jahre 1959 als sozialdemokratische Volkspartei im Rahmen der Parteienlandschaft der Bundesrepublik.¹⁵³ Dieses Grundsatzprogramm beeinflusst die Weichenstellungen der Partei bis in die heutige Zeit: Noch das 2007 verabschiedete „Hamburger Programm“ benennt als gleichrangige Grundwerte der SPD die bereits 1959 als „Grundwerte des Sozialismus“ definierten Normen Freiheit, Gerechtigkeit und Solidarität. Hierbei steht vor allem die soziale Gerechtigkeit als inhaltlich-programmatischer Kernbegriff nach wie vor im Zentrum der deutschen Sozialdemokratie. Darunter versteht die SPD -ganz im Gegensatz zur FDP- einen

¹⁵⁰ Ebd., S. 19-20.

¹⁵¹ Ebd., S. 22. Gilt nachfolgend bis anders angegeben.

¹⁵² Lempp: Parteienherrschaft in Deutschland, S. 78-79. Gilt nachfolgend bis zum Ende des Absatzes.

¹⁵³ Ebd., S. 77-78. Gilt nachfolgend bis zum Ende des Absatzes.

starken und aktiven Staat als Treiber sozialpolitischer Reformen, welcher in Kombination mit der besonderen Bedeutung der Sozialpolitik auch als „Sozialstaat“ bezeichnet wird. Dieser Sozialstaat gilt daher nicht nur als „Kernkompetenz“, sondern darüber hinaus auch als der eigentliche „Markenkern“ der SPD.

Aber auch um diesen Markenkern herum ist die SPD heute als moderne Volkspartei breit aufgestellt, was sowohl die abgedeckten Politikfelder als auch das repräsentierte Meinungsspektrum betrifft.¹⁵⁴ Dadurch kommt es auch immer wieder zu innerparteilichen Debatten, wie beispielsweise um die sozialpolitischen Reformen der frühen 2000er Jahre („Hartz IV“) unter Bundeskanzler Gerhard Schröder, welche eine große Kontroverse auslösten und unter anderem eine Abwanderung bisher treuer SPD-Wähler zur Linkspartei und später auch zur AfD zur Folge hatte.

¹⁵⁴ Ebd., S. 78. Gilt nachfolgend bis zum Ende des Absatzes.

5 Aufstellung von Hypothesen bezüglich der Ähnlichkeit der verschiedenen Parteien

Nachdem im vorangegangenen Kapitel die ausgewählten Parteien überblickshaft vorgestellt und auf dem politischen Spektrum grob eingeordnet wurden, sollen in diesem Kapitel nun mehrere Hypothesen bezüglich der zu erwartenden semantischen Ähnlichkeit ebendieser Parteien aufgestellt werden. Diese zu erwartende Ähnlichkeit (*expected similarity*) ergibt sich aus den jeweiligen Verortungen der einzelnen Parteien auf dem politischen Spektrum, die wiederum auf den (sowohl innen- als auch außenpolitischen) Zielsetzungen und generellen Werten und Überzeugungen der einzelnen Parteien beruhen.

5.1 Hypothese 1: Hohe Übereinstimmung zwischen AfD & Republikanern

Da es sich sowohl bei der AfD als auch der Republikanischen Partei der USA (besonders seit diese unter dem Einfluss Donald Trumps steht) um sehr nationalkonservative Parteien mit ähnlichen politischen Zielsetzungen (internationaler Isolationismus, radikale Asylpolitik, Besinnung auf „traditionelle“ Werte) handelt, liegt hier eine entsprechend hohe Ähnlichkeit besonders nahe. Gestützt wird diese Hypothese zusätzlich durch die offen kommunizierte, beidseitige Sympathie und Kooperation bei gemeinsamen Veranstaltungen, wie beispielsweise dem virtuellen Gastauftritt Elon Musks auf dem Parteitag der AfD Ende Januar 2025.¹⁵⁵ Zwar ist und war Musk selbst zu diesem Zeitpunkt kein Mitglied der Republikanischen Partei, nahm aber dennoch als Leiter des *Department of Government Efficiency* (DOGE) und enger Berater des amerikanischen Präsidenten Donald Trump starken Einfluss vor allem auf dessen innenpolitische Entscheidungen.

5.2 Hypothese 2: Besonders hohe Übereinstimmung zwischen AfD & NSDAP

Immer wieder wird ein Verbot der AfD öffentlichkeitswirksam diskutiert. Zugrunde liegen dieser Diskussion vorwiegend die Einschätzungen des Bundesamts für Verfassungsschutz (BfV), welches die Partei seit längerem als Verdachtsfall beobachtet. Nach einzelnen Landesgruppen und -verbänden erfolgte Anfang Mai 2025 schließlich die offizielle, bundesweite Einstufung der gesamten Partei als „gesichert

¹⁵⁵ Siehe hierzu <https://www.tagesschau.de/inland/bundestagswahl/parteien/musk-afd-wahlkampfshow-100.html>, letzter Zugriff am 03.10.2025.

rechtsextremistisch“.¹⁵⁶ Obwohl diese Einschätzung nach Klage der AfD zwischenzeitlich bis auf weiteres ausgesetzt wurde, stellt sich doch die Frage nach den Gründen für diese Einstufung. Nach Angaben des Verfassungsschutzes lagen diese vorwiegend in einem „völkischen Nationalismus“, der durchaus „in der Tradition des Nationalsozialismus“ stehe.¹⁵⁷ Darüber hinaus legt nicht zuletzt die immer wieder medial präsente Verwendung äußerst kritischer Begriffe aus dem NS-Vokabular durch offizielle Vertreterinnen und Vertreter der AfD¹⁵⁸ eine besonders hohe semantische Ähnlichkeit zwischen diesen beiden Parteien nahe.

In Bezug auf die Republikanische Partei der USA stellt sich nun wiederum die Frage: Wenn in Hypothese 1 von einer „hohen Übereinstimmung“ zwischen AfD & Republikanern ausgegangen wird und in Hypothese 2 von einer „besonders hohen Übereinstimmung“ zwischen AfD & NSDAP, müsste dies im Umkehrschluss wiederum auch eine relativ hohe Übereinstimmung zwischen NSDAP und den Republikanern bedeuten. Teil dieser Überlegung ist allerdings auch, dass die AfD trotzdem etwas näher bei der NSDAP als den Republikanern zu verorten sein wird, da die GOP keine spezifisch deutschen Themen behandelt und daher möglicherweise etwas geringere Ähnlichkeitswerte mit den beiden deutschen Parteien aufweist als diese jeweils untereinander.

5.3 Hypothese 3: Hohe Übereinstimmung von CDU, SPD & FDP mit Demokraten

Basierend vor allem auf der wirtschaftspolitischen Ausrichtung sollte sich eine hohe Übereinstimmung der FDP, welche für möglichst geringe staatliche Eingriff in die Wirtschaft bei gleichzeitigem maximalen Schutz der individuellen Freiheits- und Persönlichkeitsrechte der Staatsbürger steht, mit der Demokratischen Partei der USA ergeben. Aus gesellschaftspolitischer Sichtweise sind die Demokraten aber auch der CDU und SPD sehr ähnlich, da sie zwar auf die gemäßigt konservative bis liberale amerikanische Mittelschicht ausgerichtet sind, aber ebenso eine stark ausgeprägte Verbindung zur klassischen „Arbeiterklasse“ und Gewerkschaften besitzen.

¹⁵⁶ Siehe hierzu Grasnick, Belinda und Kumpfmüller, Konstantin: Einstufung des Verfassungsschutzes: Warum die AfD "gesichert rechtsextremistisch" ist. In: Tagesschau.de (<https://www.tagesschau.de/inland/innenpolitik/afd-verfassungsschutz-rechtsextremismus-100.html>, letzter Zugriff am 03.10.2025).

¹⁵⁷ Ebd.

¹⁵⁸ Siehe hierzu Kapitel „1 Einleitung“ dieser Arbeit.

5.4 Hypothese 4: Niedrigste Übereinstimmung von Demokraten, SPD & FDP mit AfD & NSDAP

Auf dem politischen Spektrum, wo sich sowohl NSDAP als auch AfD als deutlich rechts einordnen lassen, ist keine Überschneidung mit der Demokratischen Partei der USA, der SPD oder auch der FDP zu erwarten, da sich diese Parteien eher im Bereich Mitte bis Links verorten lassen. Falls sich diese Vermutung auch auf die politischen Texte dieser Parteien übertragen lässt, wäre in diesem Bereich mit den niedrigsten Werten für eine semantische Übereinstimmung zu rechnen.

Basierend auf diesen vier Hypothesen ist folglich also mit der Bildung zweier großer „Hauptlager“ zu rechnen: Einem eher rechten Lager, bestehend aus NSDAP, AfD und Republikanern auf der einen Seite, sowie einem eher liberalen bzw. moderaten Lager, bestehend aus CDU, SPD, FDP und den Demokraten auf der anderen Seite. Hierbei wird die Rolle der CDU besonders spannend zu beobachten sein, da sie sich selbst als konservative Partei versteht und daher möglicherweise keinem der beiden Lager eindeutig zugeordnet werden kann. Die CDU könnte somit als Bindeglied zwischen den beiden Lagern fungieren.

Nachdem in diesem Kapitel nun die verschiedenen Hypothesen bezüglich der semantischen Ähnlichkeit unter den politischen Texten der ausgewählten Parteien aufgestellt und diskutiert wurden, sollen im nächsten Kapitel zunächst die methodischen Grundlagen und deren Einsatzgebiete in den Digitalen Geisteswissenschaften vorgestellt werden, mit deren Hilfe die semantische Auswertung der politischen Texte erfolgen soll. Nicht zuletzt auf den Ergebnissen ebendieser Analyse basieren letztendlich die Gesamtaussagen dieser Arbeit, weshalb auf den methodischen Grundlagen, welche dieser Analyse zugrunde liegen, ein ganz besonderer Fokus liegen soll.

6 Vorstellung & Implementierung der ausgewählten digitalen Methoden

Über die Jahre hinweg haben sich in den *Digital Humanities* seit deren Entstehung verschiedene digitale Methoden zur computergestützten Auswertung großer Textmengen etabliert. Eine Auswahl einiger dieser Methoden soll in diesem Kapitel nun vorgestellt, auf das Korpus angewendet und die daraus resultierenden Ergebnisse ausgewertet und miteinander verglichen werden. Zu den in dieser Arbeit verwendeten Methoden zählen sowohl frequenzbasierte als auch vektorbasierte Methoden, welche anhand unterschiedlicher Ähnlichkeits- und Distanzmaße verglichen und evaluiert werden sollen. Bei diesen Methoden handelt es sich namentlich um *Bag-of-Words* (BoW), *Term Frequency – Inverse Document Frequency* (Tf-Idf) und die *Stilometrische Analyse*, sowie *Wordfish for Ideological Scaling*, *Word2Vec* und *Sentence Transformers*.

Bei *Bow* und *Tf-Idf* handelt es sich um reine frequenzbasierte Methoden, welche die absoluten Wortfrequenzen eines Textes bzw. Korpus auswerten, wobei *Tf-Idf* zusätzlich noch die Relation der für einen einzelnen Text besonders charakteristischen Begriffe im Verhältnis zu allen andern Begriffen angibt. Diese Methoden eignen sich also besonders gut dafür, einen ersten Überblick über die Themen zu bekommen, welche besonders prominent in einem bestimmten Text oder Korpus diskutiert werden. Die *Stilometrische Analyse* wurde als Methode ausgewählt, weil sie es ermöglicht, den charakteristischen, ganz eigenen Stil eines Textes zu erfassen und mit anderen Texten zu vergleichen, was insbesondere für die Vergleichbarkeit von politischen Reden interessant sein kann. Die Methode *Wordfish for Ideological Scaling* bietet sich an, da sie den innerhalb eines Textes verwendeten Begriffen Werte (*theta-scores*) auf einer ideologischen Skala zuweist, aufgrund dessen der gesamte Text dann „bewertet“ und eingeordnet wird. Ein *Word2Vec-Modell* bildet die Wörter eines Textes als Vektoren ab, deren Ähnlichkeit zueinander mit bestimmten Distanzmaßen (wie beispielsweise der Cosine-Similarity) gemessen werden kann, was auf übergeordneter Ebene ebenso für ganze Texte bzw. Korpora gilt. Sogenannte *Sentence Transformers* stellen dabei eine Weiterentwicklung von *Word2Vec* dar, welche erst in den letzten Jahren an Einfluss gewonnen hat und spätestens seit der Einführung großer Sprachmodelle (*LLMs*) die bei weitem populärste und modernste dieser Methoden darstellt.

Durch diese vielseitige Auswahl aus sowohl „klassischen“ älteren, als auch modernsten „State-of-the-Art“ Methoden, sollen möglichst vielfältige Erkenntnisse gewonnen werden, auf deren Basis dann auf übergeordneter Ebene wiederum Rückschlüsse ermöglicht werden sollen, welche dieser Methoden sich besonders gut für den konkreten Anwendungsfall dieser Arbeit eignen.

6.1 Bag-of-Words

6.1.1 Methodische Grundlagen & Einsatzgebiete

Frequenzbasierte Methoden arbeiten mit Wortfrequenzen, was bedeutet, dass sie die Häufigkeit der Wörter in einem bestimmten Text oder Korpus analysieren. Zu den populärsten Methoden dieser Art zählen *Bag-of-Words* (*BoW*) und deren Weiterentwicklung *Term Frequency – Inverse Document Frequency* (*Tf-Idf*, s. Kapitel 6.2 *Term Frequency – Inverse Document Frequency*). Bei *BoW* handelt es sich um eine Feature-Extraktionsmethode, welche Textdaten zur weiteren Analyse und Verarbeitung (beispielsweise in Algorithmen oder für maschinelles Lernen) extrahiert.¹⁵⁹ Dabei sind *BoW*-Modelle eine unstrukturierte Zusammenstellung aller Wörter eines Textdokuments, die ausschließlich nach ihrer Häufigkeit definiert werden. Wortreihenfolge, Syntax oder Kontext spielen dabei keine Rolle. In vielen Text-Mining-Pipelines ist die Erstellung von *BoW*-Modellen der erste von mehreren Schritten zur Verarbeitung natürlicher Sprache (*NLP*).

Im Rahmen einer Textklassifizierung werden Begriffe, die in einem Dokument besonders häufig auftreten, als Hauptgedanken des Dokuments interpretiert. Lässt sich in einem Dokument beispielsweise eine hohe Frequenz der Verwendung der Begriffe „Präsident“, „Wähler“ und „Wahl“ feststellen, ist die Wahrscheinlichkeit sehr hoch, dass es sich um einen politischen Text handelt, der sich spezifisch mit einer Präsidentschaftswahl befasst. Im Rahmen der Textklassifizierung mit *BoW* wird dementsprechend eine Hochrechnung vorgenommen, die auf der Annahme basiert, dass Dokumente mit ähnlichem Inhalt auch vom Typ her ähnlich sind.

Im konkreten Anwendungsfall bedeutet dies, dass im Rahmen des *BoW*-Ansatzes eine Transformation jedes einzelnen Wortes eines Textes zu einer separaten Dimension (oder Achse) des Vektorraums erfolgt. Der Wert jeder Dimension des resultierenden Vektorraums entspricht also lediglich der Häufigkeit eines Wortes in einem Korpus, wobei jedes Wort als Punkt in diesem Vektorraum dargestellt werden kann. Die Position eines Punktes entlang einer bestimmten Dimension wird wiederum durch die Häufigkeit bestimmt, mit der das Wort dieser Dimension im Dokument des Punktes vorkommt.

Nachteile von *BoW*-Modellen sind unter anderem die Wortkorrelation, da der *BoW*-Ansatz davon ausgeht, dass Wörter in einem Dokument oder Korpus voneinander unabhängig sind und daher bei der Messung der Häufigkeit einzelner Begriffe keine Korrelationen in der Verwendung zwischen Wörtern berücksichtigt. Ein weiteres

¹⁵⁹ Murel, Jacob und Kavlakoglu, Eda: What is bag of words? (<https://www.ibm.com/think/topics/bag-of-words>, letzter Zugriff am 13.10.2025). Übersetzung des Autors; gilt nachfolgend bis anders angegeben.

Problem bilden zusammengesetzte Phrasen, in denen zwei oder mehr Wörter als eine semantische Einheit fungieren. Ein einfaches *BoW*-Modell könnte beispielsweise „Mr. President“ als zwei unabhängige Wörter darstellen, obwohl sie zusammengehören. Hinzu kommt, dass mehrdeutige Begriffe, wie beispielsweise „Bank“ (Sitzgelegenheit vs. Kreditinstitut) eventuell nicht richtig zugeordnet werden können. Da *Bag of Words* bei der Modellierung von Wörtern weder Kontext noch Bedeutung berücksichtigt, fasst es all diese unterschiedlichen Bedeutungen unter einem Wort zusammen und lässt dadurch potenziell wichtige Informationen zum Thema eines Textes (und damit zur möglichen Klassifizierung) außer Acht.

In der Praxis werden *BoW*-Ansätze beispielsweise eingesetzt, um *Hate Speech* auf Social-Media-Plattformen zu identifizieren. Dies geschieht jedoch mit unterschiedlichem Erfolg. Während Salminen et al. *Bag-of-Words* mit *Word2Vec* und Deep-Learning-Klassifikatoren wie *BERT* vergleichen und zu dem Schluss gelangen, dass *BERT Bag-of-Words* übertrifft und dass *Tf-Idf* die Vorhersagen von *BoW*-Modellen nicht wesentlich verbessert,¹⁶⁰ präsentieren im Gegensatz dazu Pandey et al. einen Algorithmus, der *Bag-of-Words* und *Naïve Bayes* zur Erkennung von *Hate Speech* mit einer Genauigkeit von etwa 99 % verwendet.¹⁶¹ Diese unterschiedlichen Ergebnisse könnten beispielsweise durch Unterschiede in der Datengröße und Stichprobenauswahl sowie durch unterschiedliche Vorgehensweisen beim Preprocessing erklärt werden. Eine weitere Möglichkeit wäre auch, dass die vergleichbare Leistung zwischen *BERT* und Klassifikatoren, die *BoW* verwenden, von der Größe der Klassifizierungskategorien des Datensatzes abhängt. Besonders in den *Digital Humanities* werden große Mengen literarischer Texte oftmals über Wortfrequenzen miteinander verglichen, um auf diese Weise beispielsweise Stil, Genre oder zeitliche Trends zu identifizieren.¹⁶² Hierbei handelt es sich häufig um

¹⁶⁰ Vgl. hierzu Salminen, Joni; Hopf, Maximilian; Chowdhury, Shammur A.; Jung, Soon-gyo; Almerexhi, Hind, und Jansen, Bernard J.: „Developing an online hate classifier for multiple social media platforms“. In: *Human-centric Computing and Information Sciences*, Vol. 10, 2020 (<https://hcis-journal.springeropen.com/articles/10.1186/s13673-019-0205-6>, letzter Zugriff am 13.10.2025). Zit. n. Murel: What is bag of words? (<https://www.ibm.com/think/topics/bag-of-words>, letzter Zugriff am 13.10.2025). Übersetzung des Autors.

¹⁶¹ Vgl. hierzu Pandey, Yogesh; Sharma, Monika; Kashaf Siddiqui, Mohammad und Singh Yadav, Sudeept: „Hate Speech Detection Model Using Bag of Words and Naïve Bayes“. In: *Advances in Data and Information Sciences*, S. 457–470, 2020 (https://link.springer.com/chapter/10.1007/978-981-16-5689-7_40, letzter Zugriff am 13.10.2025). Zit. n. Murel: What is bag of words? (<https://www.ibm.com/think/topics/bag-of-words>, letzter Zugriff am 13.10.2025). Übersetzung des Autors.

¹⁶² An dieser Stelle sei exemplarisch auf Allison et al. verwiesen, die in einer frühen Stanford-Literary-Lab Studie Wortfrequenzen zur automatisierten Genre- bzw. Stil-Erkennung und historischen Vergleichen über tausende Romane einsetzten. Siehe hierzu Allison, Sarah; Heuser, Ryan; Jockers, Matthew; Moretti, Franco; Witmore, Michael: Quantitative Formalism: An Experiment. In: *Pamphlets of the Stanford Literary Lab* (Pamphlet 1). Stanford Literary Lab 2011 (<https://d-nb.info/1164077457/34>, letzter Zugriff am 13.10.2025).

einen ersten, explorativen Schritt, der dann als Ausgangspunkt für weitere Analysen dient.¹⁶³

6.1.2 Implementierung

Da bei *BoW* der jeweilige Text auf Wortebene in seine einzelnen Bestandteile aufgeteilt wird und diese Wörter getrennt voneinander und kontextunabhängig analysiert werden, ist es wichtig, die jeweiligen Texte zunächst zu tokenisieren, zu lemmatisieren und unnötige Stoppwörter zu entfernen (also beispielsweise sämtliche Artikel, Präpositionen und Konjunktionen, da diese nichts über den eigentlichen Inhalt des Textes aussagen). Das bedeutet, dass der bzw. die Texte zunächst von allen Satzzeichen (Punktuations-, Anführungszeichen etc.) bereinigt und alle Wörter auf ihre Grundformen (Lemmas) zurückgeführt werden müssen. Dies ist besonders wichtig, damit bei der späteren Analyse der Wortfrequenzen beispielsweise nicht „bringt“ und „bringen“ als zwei unterschiedliche Wörter getrennt voneinander gewertet werden, und geschieht mithilfe der beiden Python-Packages *NLTK* und *spaCy*¹⁶⁴. Diese eignen sich sowohl zur Tokenisierung, als auch für die anschließende Lemmatisierung bestens. Für *spaCy* muss hierfür zunächst jedoch noch ein Modell für die deutsche Sprache ausgewählt und hinzugefügt werden. Dies ist über die offizielle *spaCy*-Website möglich.¹⁶⁵ Die Parteiprogramme der verschiedenen Parteien müssen zusätzlich noch vom pdf- ins txt-Format übertragen werden, um diese bei der weiteren Analyse berücksichtigen zu können.¹⁶⁶ Dies geschieht durch den Einsatz des Python-Packages *pdfminer*, welches frei verfügbar ist und direkt in Python über die Bibliothek integriert werden kann.

Nachdem die Tokenisierung und Lemmatisierung mit *NLTK* und *spaCy* erfolgt ist und auch alle Stoppwörter herausgefiltert sind, werden als nächstes alle Zahlen, Sonderzeichen etc. mit Hilfe von *Regular Expressions* entfernt. Was nun übrig bleibt, ist das Korpus jeder Partei (also sämtliche Reden sowie das Parteiprogramm) komprimiert in einem einzigen String. Dieser wird als txt-Datei automatisch im Ordner der jeweiligen Partei gespeichert. Auf diese bereinigte und vorverarbeitete Datei wird

¹⁶³ So nutzt beispielsweise Matthew Jockers *BoW*-Features (im Sinne von Wortfrequenzen und *TF-IDF*) als zentrale Repräsentation für *Clustering*, *Principal Component Analysis (PCA)* und Klassifikationsanalysen für große Textmengen. Siehe hierzu Jockers, Matthew L.: *Macroanalysis: Digital Methods and Literary History*. Urbana/Chicago 2013.

¹⁶⁴ Der Python-Code, auf dem das für diese Arbeit generierte *BoW*-Modell basiert, ist samt Anleitung zur Erstellung eines lokalen *BoW*-Modells auf der Website „datacamp.com“ zu finden. Vgl. hierzu https://www.datacamp.com/tutorial/python-bag-of-words-model?dc_referrer=https%3A%2F%2Fwww.google.com%2F, letzter Zugriff am 11.11.2025.

¹⁶⁵ Ausgewählt wurde für den in dieser Arbeit konkreten Anwendungsfall das Modell „*de_core_news_lg*“, da es sowohl Morphologizer, Parser, Lemmatizer und NER beinhaltet. Siehe. hierzu <https://spacy.io/models/de>, letzter Zugriff am 10.11.2025.

¹⁶⁶ Dies geschah mit Hilfe des Skripts „*pdf2txt.py*“ (s. [GitHub-Repository](#) im Ordner „Code“; gilt nachfolgend für alle Skripte).

6.1.3 Analyse der Ergebnisse & Überprüfung der Hypothesen

Diese Grafik bestätigt nicht nur die Ergebnisse des *BoW*-Modells, sie geht darüber hinaus sogar noch weiter und erweitert die Perspektive auf das Korpus, da das Visualisierungstool hier nicht auf die 10 *mfw* beschränkt ist, sondern alle Ergebnisse zeigt. Dadurch werden zusätzliche Erkenntnisse gewonnen, wie unter anderem, dass neben den zehn bereits genannten Begriffen auch „Staat“, „Familien“, „Kinder“ und „Menschen“, aber auch Themen wie „Einwanderung“ und „Sicherheit“ eine wichtige

52

Rolle im politischen Jargon der AfD spielen. Auch die europakritische bzw. antieuropäische Ausrichtung der AfD spiegelt sich in den Worten „Euro“ und „EU“, die beide in den 10 *mfw* vertreten sind, wider.

Bei der CDU bestehen die 10 häufigsten Wörter aus 'mehr' (1028), 'müssen' (1020), 'deutschland' (927), 'land' (870), 'stärken' (680), 'menschen' (615), 'dafür' (556), 'setzen' (556), 'europa' (490) und 'deshalb' (490). Da diese zehn Begriffe bisher noch nicht allzu aussagekräftig erscheinen, soll an dieser Stelle ebenfalls eine Visualisierung des kompletten CDU-Korpus vorgenommen werden. Das Ergebnis sieht folgendermaßen aus:



Abbildung 6: Visualisierung des bereinigten CDU-Korpus mit *wordclouds.com*

Nach der Erstellung der Wortwolke fällt auf, dass im Wortschatz der CDU auch Begriffe wie „leben“, „Freiheit“, „Wissen“, „Frieden“, „Sicherheit“ und „Unternehmen“ eine zentrale Rolle in der Rhetorik einnehmen, auch wenn diese nicht unter den 10 *mfw* vertreten sind.

Im Falle der Demokraten handelt es sich bei den 10 häufigsten Wörtern um die Begriffe 'präsident' (2015), 'mehr' (1530), 'biden' (1450), 'amerika' (1430), 'menschen' (1210), 'demokraten' (1080), 'dollar' (1020), 'regierung' (970), 'unserer' (940) und 'amerikaner' (885). Die entsprechende Visualisierung sieht folgendermaßen aus:



Abbildung 7: Visualisierung des bereinigten Demokratischen Korpus mit *wordclouds.com*

Bei der Visualisierung des demokratischen Korpus fällt auf, dass neben den bereits genannten 10 *mfw* ebenso den Begriffen „Frauen“, „Pandemie“, „Zukunft“ und „helfen“ eine große Bedeutung zukommt. Auch der Name des US-Präsidenten Donald Trump wird offenbar sehr häufig erwähnt, wodurch diesem ebenfalls eine große Bedeutung beigemessen werden kann.

Bei der FDP handelt es sich im Falle der 10 *mfw* um die Begriffe 'mehr' (885), 'deutschland' (875), 'müssen' (695), 'demokraten' (675), 'freie' (620), 'menschen' (485), 'deshalb' (450), 'dafür' (410), 'land' (375) und 'unserer' (350). Eine Visualisierung mit WordClouds liefert darüber hinaus die folgende Grafik:



Abbildung 8: Visualisierung des bereinigten FDP-Korpus mit *wordclouds.com*

Die 10 *mfw* der Republikanischen Partei der USA bestehen aus den Begriffen 'trump' (990), 'land' (965), 'amerika' (840), 'präsident' (745), 'menschen' (725), 'mehr' (635), 'heute' (595), 'tun' (575), 'donald' (545), und 'amerikanischen' (540). Eine Visualisierung des republikanischen Korpus sieht entsprechend folgendermaßen aus:



Aus dieser Abbildung geht hervor, dass Begriffe wie „Welt“, „Dollar“, „Nation“, „Biden“ und „Volk“ ebenfalls sehr häufig genannt werden und daher auch sehr wichtig für den politischen Diskurs der Republikanischen Partei zu sein scheinen.



Aus dieser Visualisierung geht hervor, dass außerdem die Begriffe „Führung“, „Führer“, „Welt“, „Kraft“ und „Kampf“ eine entscheidende Rolle in der Rhetorik der NSDAP spielen.

Bei der SPD handelt es sich bei den 10 *mfw* um die Begriffe 'mehr' (1120), 'deutschland' (1095), 'müssen' (720), 'dafür' (655), 'besser' (605), 'menschen' (570), 'stärken' (540), 'spd' (490), 'europa' (470) und 'setzen' (460).



Abbildung 11: Visualisierung des bereinigten SPD-Korpus mit *wordclouds.com*

Aus dieser Abbildung wird ersichtlich, dass darüber hinaus auch die Begriffe „Europa“, „Ukraine“, „Zukunft“, „Arbeit“, „gemeinsam“ und „kämpfen“ häufig im politischen Jargon der SPD verwendet werden.

Vergleicht man diese Ergebnisse nun auf der Basis der Top 10 *mfw*, so sind die Paare mit der höchsten Übereinstimmung, basierend auf *exact matches*, in Tabelle 1 hierarchisch dargestellt.

Tabelle 1: Hierarchische Sortierung der *exact matches* der Top 10 *mfw* aller Parteien, paarweise¹⁶⁸

Parteien	<i>Exact matches</i> (inkl. Anzahl)
CDU – SPD	8: mehr, müssen, deutschland, stärken, menschen, dafür, setzen, europa.
CDU – FDP	7: mehr, müssen, deutschland, menschen, deshalb, dafür, land.
AfD – NSDAP	5: deutschen, deutsche, deutschland, mehr, müssen.
FDP – SPD	5: mehr, deutschland, müssen, menschen, dafür.
DEM – GOP	4: präsident, amerika, menschen, mehr.
FDP – DEM	4: mehr, menschen, demokraten, unserer.
AfD – CDU/FDP/SPD	Jeweils 4: mehr, müssen, deutschland, dafür.
CDU – GOP	3: mehr, menschen, land.
NSDAP – CDU/SPD	Jeweils 3: mehr, müssen, deutschland.
DEM – CDU/SPD	Jeweils 2: mehr, menschen.
NSDAP – GOP	2: mehr, heute.

Die Paare mit der höchsten Übereinstimmung, basierend auf *exact matches*, sind demnach CDU & SPD (8), sowie CDU & FDP (7). AfD & NSDAP kommen auf genauso viele Matches wie FDP & SPD (5). Am wenigsten Matches weisen NSDAP und Republikaner (2) sowie Demokraten mit sowohl CDU als auch SPD (jeweils 2) auf. Aus diesen Ergebnissen wird ersichtlich, dass sich keine der vier zuvor aufgestellten Hypothesen (s. Kapitel 5 Aufstellung von Hypothesen bezüglich der Ähnlichkeit der verschiedenen Parteien) mit *BoW* verifizieren lässt, da weder eine hohe Übereinstimmung zwischen AfD & Republikanern (Hypothese 1), noch eine besonders hohe Übereinstimmung zwischen AfD & NSDAP (Hypothese 2) vorliegt, da diese lediglich 5/10 *exact matches* bei den 10 *mfw* aufweisen. Auch die Hypothesen 3 (Hohe Übereinstimmung von CDU, SPD & FDP mit Demokraten) und 4 (Niedrigste Übereinstimmung von Demokraten, SPD & FDP mit AfD & NSDAP) wurden nicht erfüllt, da CDU & SPD mehr *exact matches* mit der AfD (jeweils 4) als mit den Demokraten (nur jeweils 2) teilen. Darüber hinaus eignet sich *BoW* auch nicht zur Identifizierung rechtspopulistischer Sprache, da die *exact matches* mit der NSDAP bei den *mfw* dafür zu unspezifisch sind (=„deutsche“, „deutschland“, „müssen“).

Im folgenden Kapitel soll die Methode *BoW* noch um die sogenannte *Term Frequency–Inverse Document Frequency* (*Tf-Idf*) erweitert werden. Diese Anpassung ermöglicht

¹⁶⁸ Die Ergebnisse dieser Übersicht wurden mit Hilfe von ChatGPT (<https://chatgpt.com/>) erstellt.

es, nicht mehr nur die reine Frequenz bestimmter Wörter innerhalb eines Textes bzw. Korpus zu zählen, sondern darüber hinaus auch weniger aussagekräftige Wörter in ihrer Bedeutung für den Text abzuschwächen, indem deren Gewichtung relativiert wird. Im Gegenzug werden aussagekräftigere Wörter dafür entsprechend stärker gewertet. Dadurch entsteht eine Glättung der verwendeten Wörter über die reinen Wortfrequenzen hinaus.

6.2 Term Frequency – Inverse Document Frequency

6.2.1 Methodische Grundlagen & Einsatzgebiete

Eine zentrale Fragestellung sowohl im *Text Mining* als auch im *NLP* ist die inhaltliche und semantische Analyse eines Dokuments.¹⁶⁹ Hierfür muss man sich jedoch zunächst die einzelnen Wörter anschauen, aus denen das Dokument besteht. Ein Indikator dafür, wie wichtig ein bestimmtes Wort ist, ist dessen *Term Frequency* (*Tf*, zu Deutsch „Begriffshäufigkeit“). Diese gibt an, wie häufig ein bestimmtes Wort in einem Korpus vorkommt. Es gibt jedoch auch Wörter in einem Dokument, die sehr häufig vorkommen, aber nicht wichtig sind. Hierzu zählen unter anderem Funktionswörter, wie beispielsweise Artikel, Präpositionen oder Konjunktionen. Solche Wörter können einer Stoppwortliste hinzugefügt und dann vor der Analyse entfernt werden, jedoch ist es ebenso möglich, dass einige dieser Wörter in manchen Texten wichtiger sind als in anderen. Daher ist eine Stoppwortliste kein besonders ausgereifter Ansatz, um die Worthäufigkeit oft verwendeter Begriffe anzupassen.

Ein weiterer Ansatz ist, sich die *Inverse Document Frequency* (*Idf*, zu Deutsch „umgekehrte Dokumenthäufigkeit“) eines bestimmten Begriffs anzuschauen, welche die Gewichtung häufig verwendeter Begriffe reduziert und dafür diejenige von Begriffen erhöht, welche in einer Sammlung von Texten nicht besonders häufig vorkommen. Dies kann dann wiederum mit der *Term Frequency* kombiniert werden, um den *Tf-Idf*-Wert eines bestimmten Begriffs zu berechnen. Dabei werden die beiden Werte (also *Term Frequency* und *Inverse Document Frequency*) miteinander multipliziert, wobei die Häufigkeit eines Begriffs angepasst wird, je nachdem, wie oft dieser verwendet wird.

Mit dem *Tf-Idf*-Wert soll demnach also gemessen werden, wie wichtig ein Wort für ein Dokument in einer Sammlung (oder einem Korpus) von Dokumenten ist, beispielsweise für einen Roman in einer Sammlung von Romanen oder eine Website in einer Sammlung von Websites. Es handelt sich hierbei jedoch eher um eine

¹⁶⁹ Silge, Julia und Robinson, David: *Text Mining with R. A Tidy Approach*. 1. Auflage, Beijing et al. 2017 (<https://www.tidytextmining.com/tfidf#tfidf>, letzter Zugriff am 04.10.2025). Übersetzung des Autors; gilt nachfolgend bis anders angegeben.

Faustregel oder heuristische Größe, denn obwohl sich das *Tf-Idf* vor allem beim *Text Mining* oder in Suchmaschinen durchaus als nützlich erwiesen hat, werden seine theoretischen Grundlagen laut Silge und Robinson von Experten der Informationstheorie als nicht ausreichend gesichert angesehen.

Die *Inverse Document Frequency* für einen bestimmten Begriff ist definiert als

$$\text{idf}(\text{term}) = \ln \left(\frac{n_{\text{documents}}}{n_{\text{documents containing term}}} \right)$$

Um daraus die *Tf-Idf* zu berechnen, muss man dann noch die Häufigkeit eines Begriffes (*Term Frequency*) mit der *Idf* multiplizieren. Hieraus ergibt sich dann wiederum die folgende Gesamtformel für *Tf-Idf*:

$$\text{Tf-Idf} = \text{tf} \cdot \text{idf} = \text{tf} \cdot \ln \left(\frac{n_{\text{documents}}}{n_{\text{documents containing term}}} \right)$$

Die *Tf-Idf*-Analyse wird hauptsächlich dazu verwendet, um zu quantifizieren, wie wichtig verschiedene Begriffe in einem Dokument sind, das Teil einer größeren Dokumentsammlung bzw. eines Korpus ist.

In den *Digital Humanities* wird *Tf-Idf* in vielen Arbeitsabläufen meist als Standard-Feature-Repräsentation eingesetzt. Tang et al.¹⁷⁰ verwendeten *Tf-Idf* beispielsweise bei einer bibliometrischen Analyse der DH-Literatur von 1989–2014, um zu untersuchen, wie sich die inhaltliche Kohärenz (*intellectual cohesion*) der Digital-Humanities-Forschung über etwa 25 Jahre hinweg entwickelt hat. Dazu analysierten sie, welche Themen, Schlagwörter und Begriffe in der DH-Forschung häufiger oder seltener gemeinsam auftreten – und wie sich diese Beziehungen im Laufe der Zeit verändern. *Tf-Idf* war hier jedoch kein Selbstzweck, sondern lediglich der erste Schritt einer mehrstufigen Pipeline. Maarten Grootendorst¹⁷¹ nutzt *c-TF-IDF* (class-based *TF-IDF*) als Variante von *Tf-Idf*, die in modernen *Topic-Modeling-Pipelines* für Forschungs- und Kulturdaten eingesetzt wird. *BERTopic* und die *c-TF-IDF*-Repräsentation werden dabei auch in Analysen genutzt, die für DH-Datensätze relevant sind, beispielsweise bei Vergleichen von Topic-Modellen auf literarischen und/oder Korpus-basierten Daten.

¹⁷⁰ Vgl. hierzu Tang, Muh-Chyun; Cheng, Yun Jen und Chen, Kuang Hua: A longitudinal study of intellectual cohesion in digital humanities using bibliometric analyses. In: *Scientometrics*, Volume 113, S. 985–1008, 2017 (<https://link.springer.com/article/10.1007/s11192-017-2496-6>, letzter Zugriff am 14.10.2025).

¹⁷¹ Grootendorst, Maarten: BERTopic: Neural topic modeling with a class-based TF-IDF procedure (<https://arxiv.org/abs/2203.05794>, letzter Zugriff am 14.10.2025).

6.2.2 Implementierung

Obwohl der Python-Code zur Implementierung von *Tf-Idf* größtenteils auf dem von *BoW* beruht, mussten noch einige Anpassungen vorgenommen werden.¹⁷² Hierzu zählen unter anderem die *Term Frequency (Tf)*, die zunächst für die bereinigten Korpora (also die lemmatisierten Dateien) aller Parteien bestimmt werden muss, sowie die *Inverse Document Frequency (Idf)*, welche wiederum auf diese Begriffe angewendet werden muss. Hierbei kommt auch die weiter oben beschriebene Formel zur Berechnung der *Idf* eines bestimmten Begriffs zum Einsatz. Das Ergebnis besteht aus einer Liste der 10 Begriffe mit dem höchsten *Tf-Idf*-Wert einer jeden Partei in absteigender Reihenfolge. Diese Ergebnisse sollen nun im Folgenden visualisiert und diskutiert werden.

6.2.3 Analyse der Ergebnisse & Überprüfung der Hypothesen

Für eine bessere Übersichtlichkeit sind die Ergebnisse dieser Auswertung in der folgenden Tabelle zusammengefasst.¹⁷³

Tabelle 2: Ergebnisse der *Tf-Idf*-Implementierung

Partei	Top 10 <i>Tf-Idf</i> Begriffe (inkl. Wert)
AfD	afd (0.5596), euro (0.2186), deutschen (0.1472), lehnt (0.1435), ezb (0.1305), eeg (0.1109), bundestag (0.1020), pepp (0.1013), masseneinwanderung (0.0811), geburtenrate (0.0811).
CDU	schäuble (0.3273), bundeskanzlerin (0.1546), ddr (0.1171), deutschen (0.0991), bürgerinnen (0.0871), ampel (0.0831), politikwechsel (0.0820), euro (0.0785), forschungseinrichtungen (0.0761), cdu (0.0758).
DEM	dollar (0.6342), biden (0.4027), amerikaner (0.2458), amerika (0.2388), lgbtqi (0.1787), medicare (0.1492), harris (0.1306), medicaid (0.1304), veteranen (0.1275), trump (0.1177).
FDP	bürgerinnen (0.1963), euro (0.1771), lindner (0.1429), demokraten (0.1176), wachstumsinitiative (0.0880), deutschen (0.0836), subventionen (0.0790), ängste (0.0779), habeck (0.0779), unternehmerische (0.0770).
GOP	republikaner (0.3980), joe (0.3733), trump (0.3365), stimmzettel (0.3243), dollar (0.3164), amerika (0.2855), sir (0.2654), heritage (0.2359), biden (0.2205), mike (0.1966).

¹⁷² Diese Anpassungen beruhen auf dem von „datacamp.com“ vorgeschlagenen Muster (https://www.datacamp.com/tutorial/python-bag-of-words-model?dc_referrer=https%3A%2F%2Fwww.google.com%2F, letzter Zugriff am 11.11.2025) und wurden durch eigene Ergänzungen noch weiter ausgebaut. Das finale Skript trägt den Namen „tf-idf.py“ (s. GitHub) und wurde mit Hilfe von ChatGPT (<https://chatgpt.com/>) erstellt.

¹⁷³ Hierbei handelt es sich jedoch um eine bereinigte, nochmals manuell nachgefilterte Ausgabe. Für die ungefilterten Ergebnisse inkl. der jeweiligen Begründungen für die manuelle Nachfilterung der Ergebnisse, siehe Tabelle 7 (Anhang).

- NSDAP jude (0.4147), deutschen (0.2689), führer (0.2431), volke (0.2413), bolschewismus (0.2393), selber (0.1848), reiches (0.1694), judentum (0.1595), volk (0.1542), allmählich (0.1540).
- SPD regierungsprogramm (0.5733), spd (0.2448), bundestagswahl (0.1674), ebert (0.1651), brandt (0.1230), sozialdemokratie (0.1216), bürgerinnen (0.1173), zukunftsinvestitionen (0.0956), sozialdemokratische (0.0869), hannover (0.0869).

Anhand dieser Darstellung der Ergebnisse wird ersichtlich, dass jetzt nur noch diejenigen Begriffe in der Top 10 Trefferliste angezeigt werden, die spezifisch für die jeweilige Partei besonders repräsentativ sind. Da nur noch diese „Alleinstellungsmerkmale“ der jeweiligen Partei abgebildet werden, fallen vor allem solche Begriffe weg, die in allen Korpora vorkommen. Daher werden Wörter wie beispielsweise „Europa“, die in allen Korpora besonders häufig vorkommen, herausgefiltert und gar nicht mehr angezeigt. Ein hoher *Tf-idf*-Wert bedeutet demnach, dass ein Begriff für ein bestimmtes Dokument sehr charakteristisch ist, im restlichen Korpus aber eher selten vorkommt.

Vergleicht man nun wieder die Anzahl der *exact matches* bei den *mfw* der Parteien untereinander, müssten diese also durch die *Tf-idf*-Filterung jetzt deutlich geringer ausfallen, was tatsächlich auch der Fall ist (s. Tabelle 3).

Tabelle 3: Hierarchische Sortierung der *exact matches* der Top 10 *mfw* aller Parteien, paarweise (*Tf-idf*-bereinigt)¹⁷⁴

Parteien	<i>Exact matches</i> (inkl. Anzahl)
DEM – GOP	4: trump, dollar, amerika, biden.
CDU – FDP	3: bürgerinnen, euro, deutschen.
AfD – CDU/FDP	Jeweils 2: deutschen, euro.
NSDAP – AfD/CDU/FDP	Jeweils 1: deutschen.
SPD – CDU/FDP	Jeweils 1: bürgerinnen.
SPD – AfD/DEM/GOP/NSDAP	Keine Übereinstimmung!
DEM – CDU/FDP/SPD/NSDAP	Keine Übereinstimmung!
GOP – CDU/FDP/SPD/NSDAP	Keine Übereinstimmung!

Ganz konkret bedeutet das in diesem Fall, dass nach der erfolgten *Tf-idf*-Bereinigung die ähnlichsten Parteien jetzt nicht mehr CDU & SPD, sondern die beiden amerikanischen Parteien der Demokraten und Republikaner sind (4/10 Matches), dicht gefolgt von CDU & FDP (3/10 Matches). Da es keine Übereinstimmung zwischen AfD und Republikanern sowie keine besonders hohe Übereinstimmung zwischen AfD &

¹⁷⁴ Die Ergebnisse dieser Übersicht wurden mit Hilfe von ChatGPT (<https://chatgpt.com/>) erstellt.

NSDAP (lediglich 1 Match) gibt, treffen weder Hypothese 1 noch Hypothese 2 im Falle von *Tf-Idf* zu. Hypothese 3, die von einer hohen Übereinstimmung von CDU, SPD & FDP mit den Demokraten ausgeht, wird ebenfalls nicht erfüllt. Auch Hypothese 4, die von der niedrigsten Übereinstimmung von Demokraten, SPD & FDP mit AfD & NSDAP ausgeht, trifft in diesem Fall nicht zu, da AfD & NSDAP mit CDU & FDP jeweils 1 bzw. 2 Matches aufweisen.

Damit wird keine der vier zuvor aufgestellten Hypothesen erfüllt. Dies bedeutet jedoch nicht, dass die Methode *Tf-Idf* an sich fehlerhaft oder ungeeignet wäre. Vielmehr kommt es auf den konkreten Anwendungsfall an, denn *Tf-Idf* eignet sich nicht zur Links-Rechts Klassifikation der jeweiligen Parteien, sondern eher für eine semantische Analyse der verwendeten Begriffe.

6.3 Stilometrische Analyse

6.3.1 Methodische Grundlagen & Einsatzgebiete

Bei der Stilometrischen Analyse bzw. Stilometrie (engl. *Stylometry*) handelt es sich um eine digitale Methode zur quantitativen Textanalyse, welche die stilistischen Merkmale von Texten zu komparativen Zwecken misst.¹⁷⁵ Zu diesen Merkmalen zählen beispielsweise die Wort- bzw. Satzlänge, verschiedene Frequenzen (von Wörtern, Wortformen, Wortlängen etc.), die Reichhaltigkeit des Vokabulars, die Verwendung von Satzzeichen, sowie die Verwendung bestimmter Ausdrücke und Präferenzen für bestimmte Schreibweisen. Dabei spielt auch die statistische Analyse eine wichtige Rolle. Die Stilometrie geht davon aus, dass Schreibstile zwar nicht vollkommen konstant, aber doch ausreichend charakteristisch sind, um über die zuvor beschriebenen, messbaren Merkmale beispielsweise Rückschlüsse auf die Autorschaft, die Entstehungszeit oder das Genre eines bestimmten Textes bzw. Werkes zu ziehen.¹⁷⁶ Die Methode basiert größtenteils auf Zuschreibungsstudien (*Attribution Studies*) und Computerlinguistik, kann aber auch für forensische Analysen verwendet werden.¹⁷⁷ Sie geht außerdem davon aus, dass Individuen einzigartig sind

¹⁷⁵ Digital Humanities Workbench (https://www2.fgw.vu.nl/werkbanken/dighum/data_analysis/text_analysis/stylometry.php, letzter Zugriff am 24.10.2025). Übersetzung des Autors; gilt nachfolgend bis anders angegeben.

¹⁷⁶ Byszuk, Joanna: The Voices of Doctor Who – How Stylometry Can be Useful in Revealing New Information about TV Series. In: *Digital Humanities Quarterly*, Vol. 14, Nr. 4 (2020), Absatz 28 (<https://dhq.digitalhumanities.org/vol/14/4/000499/000499.html>, letzter Zugriff am 24.10.2025). Übersetzung des Autors.

¹⁷⁷ Wermer-Colan, Alex: Stylometry Methods and Practices. A review of various stylometry methods and programs for the digital humanities (<https://guides.temple.edu/stylometryfordh>, letzter Zugriff am 24.10.2025). Übersetzung des Autors; gilt nachfolgend bis anders angegeben.

und dass sich diese Einzigartigkeit im persönlichen Schreibstil des oder der jeweiligen Autorin widerspiegelt.

Wie viele andere Ansätze in den *Digital Humanities* kombiniert die Stilometrie traditionelles *Close Reading* mit *Distant Reading*. Dabei kann sie jedoch nur auf statistische Wahrscheinlichkeiten hinweisen, aber keine endgültigen Aussagen zur Autorschaft treffen. Darüber hinaus wird kritisiert, dass sich der Schreibstil eines Autors oder einer Autorin auch über die Zeit im Laufe des Lebens bzw. der Karriere verändern kann.¹⁷⁸ Außerdem ändert sich der Schreibstil auch abhängig von dem jeweiligen Genre, sodass der genreübergreifende Unterschied der Werke eines einzelnen Autors letztendlich größer sein kann als derjenige zwischen verschiedenen Autorinnen und Autoren innerhalb desselben Genres.

Die Zuordnung von Urheberschaft im Sinne der Stilometrie stellt hierbei eine der ältesten Formen der digitalen Geisteswissenschaften dar.¹⁷⁹ Sie geht bereits auf das 15. Jahrhundert und den Vergleich unterschiedlicher Übersetzungen zurück. Im Jahr 1851 spekulierte der britische Mathematiker August de Morgan, dass die Wortlänge ein einzigartiger Marker für den Stil eines Autors sein könnte. Fast vierzig Jahre später, im Jahr 1890, publizierte der polnische Philosoph Wincenty Lutosławski seine Abhandlung *Principes de stylométrie* und prägte damit den Begriff der „Stilometrie“ bis heute. Zuvor setzte er diese Methode zur Analyse der Werke von Platon ein, wodurch er auch als der erste gilt, der *Most Frequent Words* (MFWs) verglichen hat.¹⁸⁰

Einen sehr populären Anwendungsfall der Stilometrie in den *Digital Humanities* stellt die Bestimmung der Urheberschaft der sogenannten *Federalist Papers* dar, einer Reihe von Artikeln, die 1787–88 mit dem Ziel veröffentlicht wurden, die Ratifizierung der neuen Verfassung der Vereinigten Staaten zu fördern.¹⁸¹ Die Publikation erfolgte durch die Autoren Jay, Hamilton und Madison unter dem Pseudonym „Publius“. Der Verfasser der Artikel war bekannt, während die Urheberschaft anderer Artikel weiterhin Gegenstand von Diskussionen war. In den frühen 1960er Jahren unternahmen Mosteller und Wallace daher den Versuch, diese Ungewissheit mithilfe stilometrischer Methoden zu lösen. Nach der Veröffentlichung ihrer Studie zu den *Federalist Papers*

¹⁷⁸ Eve, Martin Paul: *The Digital Humanities and Literary Studies*. Oxford 2022, S. 28-29. Übersetzung des Autors; gilt nachfolgend bis anders angegeben.

¹⁷⁹ Wermer-Colan, Alex: *Stylometry Methods and Practices. A review of various stylometry methods and programs for the digital humanities* (<https://guides.temple.edu/stylometryfordh>, letzter Zugriff am 24.10.2025). Übersetzung des Autors; gilt nachfolgend bis zum Ende des Absatzes.

¹⁸⁰ Byszuk: *The Voices of Doctor Who*, Absatz 28 (<https://dhq.digitalhumanities.org/vol/14/4/000499/000499.html>, letzter Zugriff am 24.10.2025). Übersetzung des Autors.

¹⁸¹ Digital Humanities Workbench (https://www2.fgw.vu.nl/werkbanken/dighum/data_analysis/text_analysis/stylometry.php, letzter Zugriff am 24.10.2025). Übersetzung des Autors; gilt nachfolgend bis anders angegeben.

im Jahr 1964¹⁸² erfuhr die Stilometrie daraufhin eine signifikante Steigerung ihrer Popularität.¹⁸³ In der Folge entwickelte sie sich methodisch weiter, vor allem auch dank der Verfügbarkeit leistungsfähigerer Computer und der Studie von Burrows¹⁸⁴ zu Jane Austens Werken aus dem Jahr 1987. In dieser Studie präsentierte er das sogenannte Delta-Maß (*Delta measure*)¹⁸⁵, eine Methode zum Vergleich nicht nur einzelner Wörter, sondern ganzer Texte, die sich als deutlich zuverlässiger erwies.

In den *Digital Humanities* wurde eine Stilometrische Analyse ebenfalls dafür eingesetzt, die Autorin J. K. Rowling, bekannt geworden hauptsächlich aufgrund des internationalen Erfolgs der Harry-Potter-Reihe, als Autorin eines Werkes zu identifizieren, welches diese unter einem Pseudonym veröffentlicht hatte.¹⁸⁶ Rowling gab später die Autorschaft zu. Martin Paul Eve weist jedoch darauf hin, dass diese Art der Autorschaftserkennung nicht immer von Erfolg gekrönt ist.¹⁸⁷ So wurden Ende des 20. Jahrhunderts beispielsweise bei einigen Gerichtsverhandlungen stilometrische Techniken angewandt, um dadurch schriftliche Beweise auszuwerten, was allerdings in einem live im Fernsehen übertragenen Misserfolg der Urheberschaftszuweisung unter Verwendung dieser Methode gipfelte. Einen weiteren Fall bietet etwa um dieselbe Zeit das Gedicht „A Funeral Elegy“, welches mithilfe stilometrischer Analysen William Shakespeare zugeschrieben wurde. Nach einigem medialen Aufruhr, gefolgt von weiteren Analysen, stellte sich jedoch bald darauf heraus, dass John Ford ein weitaus wahrscheinlicherer Kandidat für die Autorschaft des Gedichts war.

6.3.2 Implementierung

Die Stilometrische Analyse wurde, ebenso wie *Wordfish*, in der Programmiersprache *R* entwickelt. Die Implementierung geschieht hierbei zunächst mithilfe des *R-Packages* „Stylo“. Dadurch öffnet sich ein Fenster (=Stylo-GUI), in dem alle Analysen durchgeführt werden können. Hier können auch Parameter wie Analyse-Typ,

¹⁸² Vgl. hierzu Mosteller, Frederick und Wallace, David L.: *Inference and disputed authorship: The Federalist*. Reading 1964.

¹⁸³ Byszuk: The Voices of Doctor Who, Absatz 28 (<https://dhq.digitalhumanities.org/vol/14/4/000499/000499.html>, letzter Zugriff am 24.10.2025). Übersetzung des Autors; gilt nachfolgend bis zum Ende des Absatzes.

¹⁸⁴ Vgl. hierzu Burrows, John F.: *Computation into criticism: a study of Jane Austen's novels and an experiment in method*. Oxford/New York: 1987.

¹⁸⁵ Vgl. hierzu Burrows, John F.: 'Delta': a Measure of Stylistic Difference and a Guide to Likely Authorship. In: *Literary and Linguistic Computing*, 17(3), 2002, S. 267–287.

¹⁸⁶ Patrick Juola beschreibt in diesem Zusammenhang die stilometrischen Untersuchungen (u. a. mithilfe der Open-Source-Software „JGAAP“), die im Fall *The Cuckoo's Calling* von Robert Galbraith zur Identifikation von J. K. Rowling als Autorin herangezogen wurden. Siehe hierzu auch Juola, Patrick: *The Rowling Case: A Proposed Standard Analytic Protocol for Authorship Questions*. In: *Digital Scholarship in the Humanities*, Volume 30, Issue 1 (2015), S. 100–113 (<https://doi.org/10.1093/litc/fqv040>, letzter Zugriff am 27.10.2025).

¹⁸⁷ Eve: *The Digital Humanities*, S. 29–31. Übersetzung des Autors; gilt nachfolgend bis zum Ende des Absatzes.

Features, Clustering, etc. eingestellt werden. Da es sich für eine bessere Nachvollziehbarkeit aber empfiehlt, wird die folgende Analyse trotzdem per Code in *R* durchgeführt.

Diese Analyse wird mit zwei verschiedenen Korpora durchgeführt: Einmal mit dem kompletten, unverarbeiteten und nicht-bereinigten Korpus aller Parteien (=“Corpus_combined“), und einmal mit den lemmatisierten und bereinigten Daten, die bereits für die vorherigen Anwendungsfälle genutzt wurden (=“Corpus_lemmatized“). Hierfür ist es jedoch zunächst notwendig, mit Hilfe eines *R*-Skripts¹⁸⁸ die Dateien jedes Parteienkorpus in einer einzigen großen txt-Datei zusammenzufügen. Diese Dateien werden daraufhin in einen eigenen Ordner verschoben, innerhalb dessen sie dann miteinander verglichen werden können.

Dieser Ordner wird dann im eigentlichen *Stylo*-Skript¹⁸⁹ zunächst als Working Directory (WD) gesetzt. Für die darauffolgende Analyse der bereinigten Parteienkorpora muss dann entsprechend der Zielordner der bereinigten Dateien als WD gesetzt werden. Alle Parameter der Analyse, die auch über das zuvor beschriebene „Stylo-GUI“ eingestellt werden können, werden nun im Code definiert. So kann für jede Analyse zwischen verschiedenen Analysearten gewählt werden, darunter beispielsweise ein sogenannter *Consensus Tree* (CT), eine *Cluster Analyse* (CA) oder eine *Principal Component Visualization* (PCV). Damit diese Analyse nicht zu weit ausgreift, wird sich für jedes Korpus jeweils auf eine CA und eine PCV mit einem Limit von 100 *mfw* beschränkt. Als Distanzmaß wird erneut die Cosine Similarity gewählt, um die Einheitlichkeit und Vergleichbarkeit der Ergebnisse zu den vorherigen Methoden zu wahren. Da die Korpora bereits entweder komplett oder überhaupt nicht bereinigt sind, ist zudem auch kein weiteres Sampling in den Voreinstellungen mehr notwendig. Die Ergebnisse dieser Analyse sollen im folgenden Kapitel nun vorgestellt und analysiert werden.

6.3.3 Analyse der Ergebnisse & Überprüfung der Hypothesen

Bei der Analyse der nicht-vorverarbeiteten Parteienkorpora (=“Corpus_combined“) per CA fällt zunächst auf, dass es zwei große Haupt-Strömungen innerhalb der unverarbeiteten Korpora zu geben scheint, nämlich einmal die etablierten deutschen Parteien SPD, FDP und CDU, welche –gemessen an der Cosine Distance– nah beieinander liegen und sich demnach sehr ähnlich sind, sowie dem anderen „Ast“,

¹⁸⁸ Hierbei handelt es sich um das Skript „[combine_party_texts.r](https://m365.cloud.microsoft.com/chat/)“ (s. GitHub), welches mit Hilfe von Microsoft Copilot (<https://m365.cloud.microsoft.com/chat/>) erstellt wurde.

¹⁸⁹ Hierbei handelt es sich um das Skript „[stylo.r](https://m365.cloud.microsoft.com/chat/)“ (s. GitHub), welches mit Hilfe von Microsoft Copilot (<https://m365.cloud.microsoft.com/chat/>) erstellt wurde.

bestehend aus Republikanern, NSDAP, Demokraten und AfD. Diese Ergebnisse werden in Abbildung 12 dargestellt.

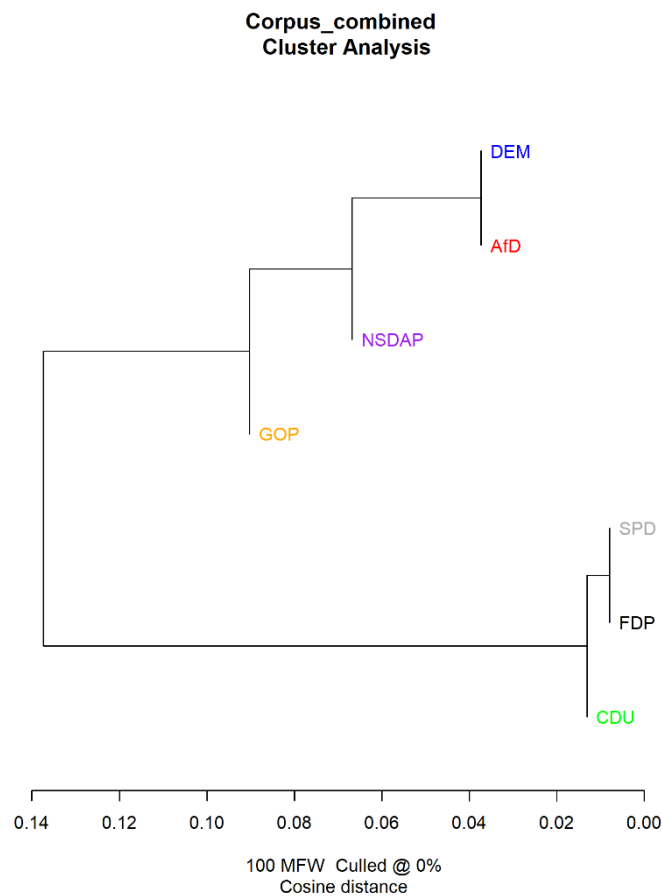


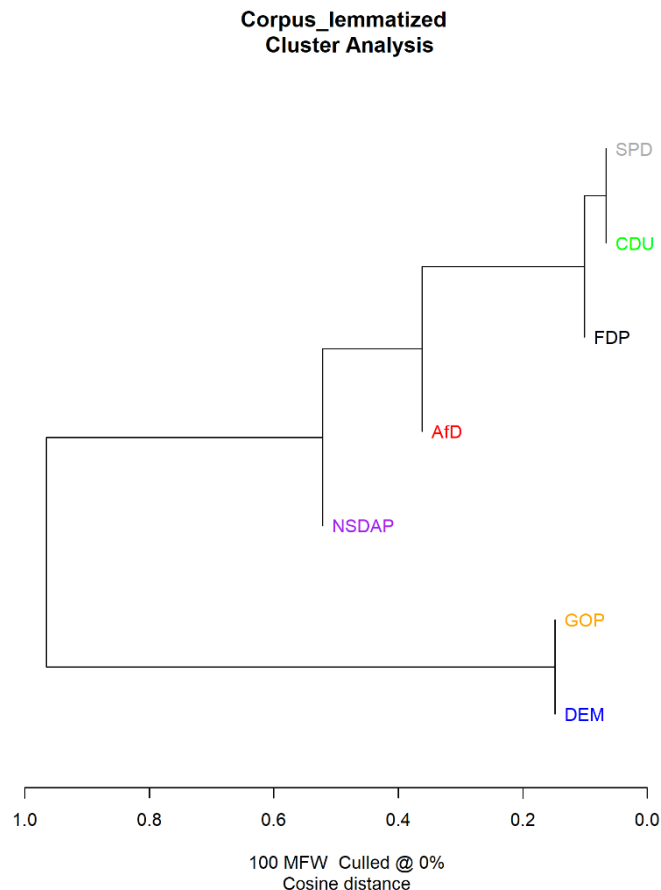
Abbildung 12: Dendrogramm der Clusteranalyse der unbereinigten Parteienkorpora, visualisiert mit Stylo

In dieser hierarchischen Clusteranalyse werden die verschiedenen Parteien anhand der Ähnlichkeit ihrer 100 *mfw* gruppiert. Die *Cosine Distance* misst hierbei als Distanzmaß die Ähnlichkeit der Wortverteilungen: Je kleiner der Wert, desto ähnlicher sind sich die jeweiligen Korpora bzw. Parteien. Auf der x-Achse sind die Distanzwerte dargestellt. Je weiter links die Parteien auf der x-Achse positioniert sind, desto größer sind die Unterschiede zwischen ihnen, je weiter rechts, desto größer ist die Ähnlichkeit. Die Unterschrift „100 MFW Culled @ 0%“ bedeutet wiederum, dass die Analyse auf den 100 häufigsten Wörtern jedes Korpus ohne Entfernung der Ausreißer basiert.

Da die drei Parteien SPD, FDP und CDU sehr nah beieinander liegen (Cosine Distance von unter 0.02), sind sich ihre Sprachmuster sehr ähnlich, das heißt vermutlich eher moderat und formal. AfD und Demokraten bilden ein eigenes Cluster, relativ nah zueinander (Cosine Distance von ca. 0.04). Dies könnte dadurch erklärt werden, dass trotz unterschiedlicher politischer Ausrichtung ihre Texte ähnliche sprachliche

Strukturen aufzeigen, beispielsweise eher populistische oder emotionale Sprache. NSDAP und Republikaner liegen zwischen den beiden Clustern, aber näher an AfD und Demokraten. Demnach teilen NSDAP und Republikaner bestimmte sprachliche Muster, die sie von SPD, FDP und CDU unterscheiden. Die größte Trennung besteht jedoch zwischen SPD, FDP und CDU auf der einen und Demokraten, AfD, NSDAP und Republikanern auf der anderen Seite. Daraus folgt, dass es zwei große Sprachcluster gibt, nämlich erstens das Cluster 1, bestehend aus den etablierten deutschen Parteien SPD, FDP und CDU, sowie Cluster 2 (Demokraten, AfD, NSDAP und Republikaner), welches aufgrund der großen Nähe zur NSDAP eher populistisch geprägt zu sein scheint.

Die Vorteile dieser unverarbeiteten Korpora bestehen darin, dass sie unter anderem stilistische Merkmale und Flexionen (z.B. die Verwendung bestimmter Verbformen) enthalten, die für Stilanalysen relevant sein können, da für stilistische Merkmale die originale Wortform oft stilprägend und daher entscheidend sein kann. Zudem werden Vorverarbeitungsfehler vermieden, da eine Lemmatisierung, besonders bei etwas komplexerer Morphologie, auch Fehler mit sich bringen kann. Um jedoch einen Vergleichswert dieser Analyse zu erhalten, soll im Folgenden nun dieselbe Analyse mit den lemmatisierten und vorverarbeiteten Parteienkorpora durchgeführt werden und die Ergebnisse miteinander verglichen werden. Vorteile von lemmatisierten und vorverarbeiteten Daten sind zum einen eine reduzierte Varianz, da Wörter wie „laufen“, „lief“ oder „gelaufen“ zum Infinitiv „laufen“ zusammengeführt werden, was zu einer deutlich geringeren Streuung führt. Hinzu kommt eine bessere Vergleichbarkeit bei inhaltlichen Analysen, wie bestimmten Themen oder semantischer Nähe, da unnötige Funktionswörter zuvor bereits herausgefiltert werden. Vergleicht man diese Ergebnisse nun mit derselben Analyse der bereinigten Korpora, gelangt man zu dem folgenden Ergebnis (s. Abbildung 13).



**Abbildung 13: Dendrogramm der Clusteranalyse der bereinigten
Parteienkorpora, visualisiert mit Stylo**

Wie aus Abbildung 13 hervorgeht, gibt es nun mehrere wesentliche Änderungen im Vergleich zum nicht-vorverarbeiteten Korpus. Hierunter fällt einerseits die Trennung der amerikanischen und deutschen Parteien. So nehmen jetzt beispielsweise Republikaner und Demokraten bei den bereinigten Korpora die Stelle der etablierten deutschen Parteien im unteren Cluster 1 ein. Auch die Werte der Cosine Distance sind fast identisch. Dies liegt vor allem daran, dass sich die Themen, also die bereinigten *mfw* der Republikaner und Demokraten größtenteils decken, wie bereits aus den vorangegangenen Analysen mit *BoW* und *Tf-Idf* ersichtlich wurde. In Cluster 2, dem „oberen Ast“, liegen SPD, CDU und FDP semantisch erneut nah beieinander, hieran hat sich also nichts geändert, mit der Ausnahme, dass sie nun näher an AfD und NSDAP angrenzen. Bei den deutschen Parteien lässt sich jetzt also eine Art „Reihenfolge“ hin zum Rechtspopulismus erkennen. Hierfür gibt es mehrere Erklärungen. Zum einen ist es auffällig, dass alle deutschen Parteien einem gemeinsamen Zweig angehören, während die amerikanischen Parteien abgeschottet sind. Dies hängt sicherlich mit den Themengebieten zusammen, welche in den bereinigten 100 *mfw* vertreten sind. Dass die amerikanischen Parteien andere (und

untereinander sehr ähnliche) Themen diskutieren als die deutschen Parteien, ist einleuchtend, daher auch diese strikte Trennung innerhalb des Dendrogramms. Auch eine ähnliche Verwendung bestimmter Begriffe wie beispielsweise „Volk“ etc. ruft eine Nähe der AfD zur NSDAP hervor. Da beide Parteien sich in ihren verwendeten Begriffen aber stets auf Deutschland beziehen, sind sie näher an den etablierten deutschen Parteien verortet als an den beiden amerikanischen. Populistische Rhetorik und bestimmte Formulierungen oder Sprachmuster werden hier nicht mehr berücksichtigt.

Eine weitere Visualisierung zeigt das Ergebnis einer *Principal Components Analysis* (PCA) der unbereinigten Parteienkorpora in Form einer Kovarianzmatrix, bei welcher erneut die 100 *mfw* ohne Ausreißerentfernung analysiert wurden (s. Abbildung 14).

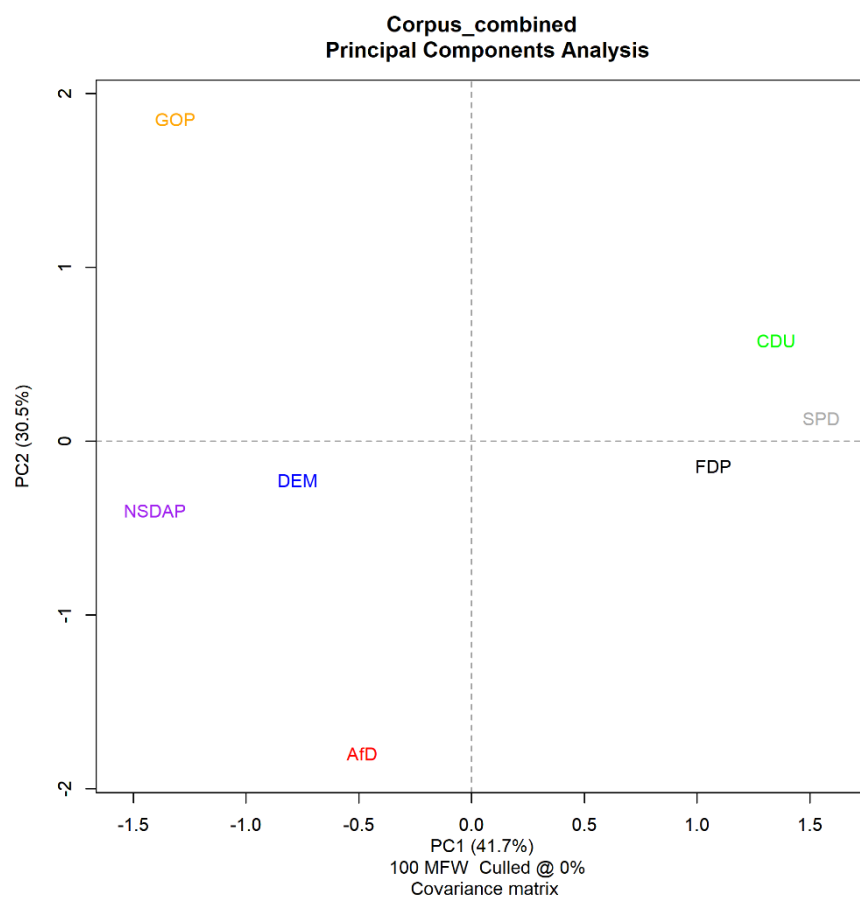


Abbildung 14: Principal Components Analysis der unbereinigten Parteienkorpora in Form einer Kovarianzmatrix, visualisiert mit *Stylo*

Die PCA basiert hierbei auf Kovarianz anstelle von Distanz, wie dies noch zuvor bei den Dendrogrammen der Fall war. Dies bedeutet, dass nun auch die absoluten Unterschiede in den Wortfrequenzen berücksichtigt werden. Daher können Parteien mit längeren Texten oder einer höheren Wortanzahl die Analyse stärker beeinflussen,

da die Hauptkomponenten nicht nur die relative Wortverteilung, sondern auch absolute Frequenzunterschiede widerspiegeln. Dies ist jedoch unproblematisch, da alle Korpora ungefähr die gleiche Größe besitzen. Die Achse „PC1“ bildet hierbei die erste Hauptkomponente und erklärt den größten Teil der Varianz (41,7%), also der Unterschiede zwischen den Parteien. Parteien, die auf der x-Achse weit auseinander liegen, unterscheiden sich stark in den Wortmustern, die PC1 repräsentiert. So können im unbereinigten Korpus beispielsweise deutliche sprachliche Unterschiede zwischen CDU, SPD und FDP (rechts) und AfD, NSDAP, Demokraten und Republikanern (links) festgestellt werden. „PC2“ bildet dementsprechend die zweite Hauptkomponente und erklärt wiederum den zweitgrößten Teil der Varianz (30,5%). Folglich variieren Parteien, die sich auf der y-Achse stark unterscheiden, in einem anderen sprachlichen Muster. Beispielsweise sind die Republikanische Partei (oben) und AfD (unten) in der Matrix beide „links“ auf PC1 verortet, aber unterscheiden sich in PC2 (z. B. in bestimmten inhaltlichen Themen oder dem Gebrauch von Funktionswörtern).

Zusammen decken diese beiden Dimensionen über 70% der Unterschiede in den Textdaten aus den verschiedenen Parteienkorpora ab. Hierbei bildet jede Partei einen Punkt im zweidimensionalen Raum. Die Position basiert auf den 100 *mfw* und deren Verteilung. Je näher zwei Punkte also beieinander liegen, desto ähnlicher sind die Texte dieser Parteien in Bezug auf Wortgebrauch und Stil. Jede Hauptkomponente ist eine lineare Kombination der ursprünglichen Variablen (Wortfrequenzen), gewichtet nach ihrem Beitrag zur Varianz. Die Werte auf den beiden Achsen (z. B. -1.5 bis +1.5) stellen dabei standardisierte Scores der Parteien auf den jeweiligen Hauptkomponenten dar. Sie geben an, wie stark eine Partei in Richtung der jeweiligen Komponente ausgeprägt ist. Ein negativer Wert bedeutet hierbei, dass die Partei Wörter verwendet, die eher negativ geprägt sind. Ein positiver Wert bedeutet wiederum, dass die Partei Wörter verwendet, die eher eine positive Ladung besitzen. Die horizontale Distanz (PC1) steht also für den Hauptunterschied im Sprachstil (z. B. positive oder negative bzw. moderate oder populistische Sprache), während die vertikale Distanz (PC2) als zweitwichtiges Muster zum Beispiel den Themenfokus oder die Satzstruktur repräsentiert. Parteien, die nahe beieinander liegen, sind sich also entsprechend sprachlich ähnlicher.

So liegen beispielsweise CDU, SPD und FDP (rechts oben) wie bereits zuvor im Dendrogramm nahe beieinander, was auf ähnliche sprachliche Muster hinweist. Die AfD ist unten links wiederum weit entfernt von den etablierten deutschen Parteien, was auf eine andere Wortwahl und einen deutlich anderen, höchstwahrscheinlich populistischen Sprachstil hindeutet, da die AfD am nächsten an die NSDAP angrenzt. NSDAP und Demokraten (unten links, nahe AfD) liegen ebenfalls im linken Bereich, was auf Gemeinsamkeiten im Sprachgebrauch hinweist, die sie von den etablierten deutschen Parteien unterscheiden. Die Republikanische Partei (oben links) befindet sich zwar weiter oben, aber immer noch auf der linken Seite, also eher im Cluster mit

AfD, NSDAP und Demokraten, jedoch mit einer anderen, eigenen Ausprägung. Wie bereits zuvor im Dendrogramm gibt es erneut zwei große Gruppen, nämlich Cluster 1 (rechts), bestehend aus den etablierten deutschen Parteien(CDU, SPD und FDP, welche durch ähnliche Sprachmuster miteinander verbunden sind, sowie dem Cluster 2 (links) aus AfD, NSDAP, Demokraten und Republikanern, welche untereinander stärker voneinander abweichen.

Die entsprechende Visualisierung der bereinigten und lemmatisierten Parteienkorpora ist in Abbildung 15 ersichtlich. Hier spiegelt sich die Struktur des entsprechenden Dendrogramms wider, in welcher sich die Parteien an der Nationalitätsgrenze scheiden.

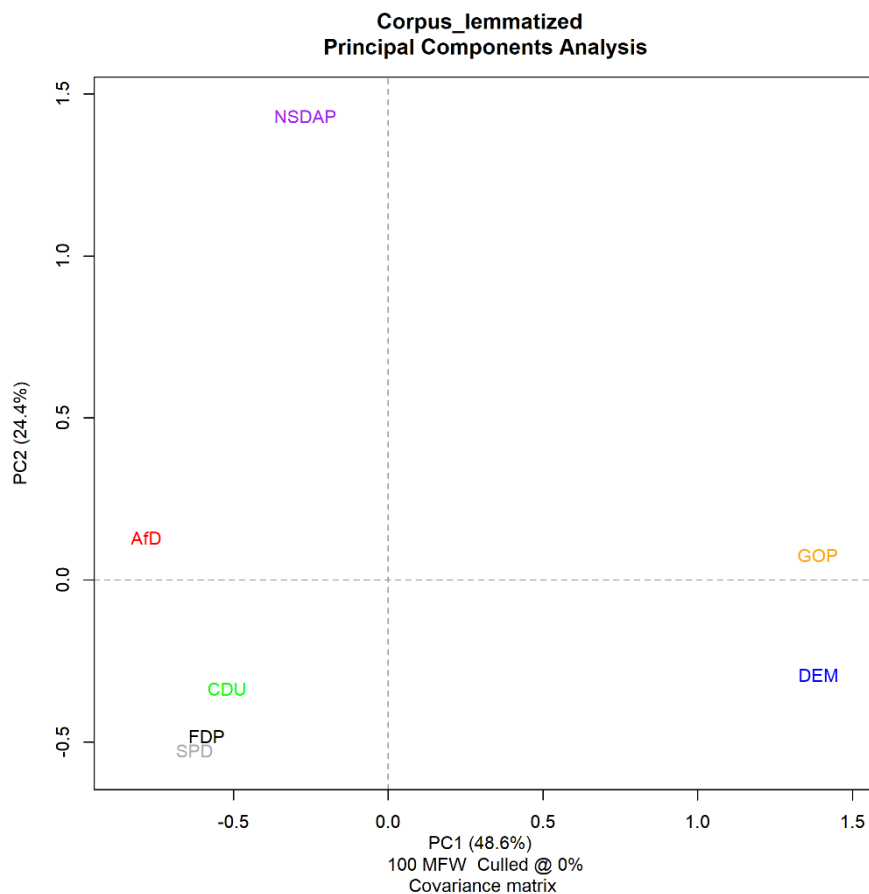


Abbildung 15: Principal Components Analysis der bereinigten Parteienkorpora in Form einer Kovarianzmatrix, visualisiert mit *Stylo*

Besonders auffällig ist hierbei, dass nach der Bereinigung der Korpora (Lemmatisierung, Entfernung der Stoppwörter etc.) die etablierten deutschen Parteien mit den amerikanischen Parteien ihre Positionen auf der x-Achse tauschen. Dies bedeutet, dass die deutschen Parteien allesamt deutlich negativer gewertete Begriffe

verwenden als die amerikanischen. Nur das Verhältnis zwischen AfD und NSDAP bleibt weitestgehend unverändert und wird dadurch nicht berührt.

Zusammenfassend lässt sich also noch einmal sagen: Die Analyse mit unbereinigten Korpora ist für den Stil (populistische Rhetorik und bestimmte Formulierungen oder Sprachmuster) der jeweiligen Parteien entscheidend, wohingegen die Analyse der bereinigten Parteienkorpora mehr die aussagekräftigsten Themengebiete der jeweiligen Parteien in den Vordergrund stellt und diese miteinander vergleicht. Durch die *PCA* wird darüber hinaus noch viel stärker zwischen der x- und y-Achse nach positiven und negativen Werten unterschieden, wobei der Sprachstil gar nicht so sehr im Fokus steht, sondern vielmehr das Vokabular. Im Gegensatz zur *Cluster Analysis*, welche die semantische Nähe von Parteien mit Hilfe der *Cosine Distance* misst, erweitert die *PCA* dieses Feld noch um die Variablen des Vokabulars (z. B. moderate oder populistische Sprache durch positive oder negative Bewertung von Begriffen), sowie den Themenfokus oder die Satzstruktur auf einer weiteren Achse, wodurch eine noch tiefergehende Analyse ermöglicht wird.

Aus diesem Grund soll in Bezug auf die Überprüfung der aufgestellten Hypothesen von den nicht-vorverarbeiteten, kombinierten Parteienkorpora ausgegangen werden, da diese am besten die Rhetorik und den Sprachstil der verschiedenen Parteien wiedergeben. Hier lässt sich einerseits festhalten, dass bei den nicht-vorverarbeiteten Parteienkorpora Hypothese 1 nicht zutrifft, da die AfD eine deutlich höhere Ähnlichkeit zu den Sprachmustern der Demokraten als zu denen der Republikaner aufweist. Dies gilt ebenfalls für die vorverarbeiteten Korpora, da die deutschen Parteien hier ein separates Cluster, getrennt von den amerikanischen Parteien, bilden. Hypothese 2, die von einer besonders hohen Übereinstimmung zwischen AfD & NSDAP ausgeht, wird in beiden (also sowohl vorverarbeiteten als auch nicht- vorverarbeiteten) Korpora erfüllt. Hypothese 3 wird in beiden Korpora nur teilweise erfüllt, da CDU, SPD & FDP untereinander zwar eine sehr hohe Übereinstimmung aufweisen, aber nicht mit der Demokratischen Partei der USA, welche in beiden Korpora ein separates Cluster bildet. Hypothese 4, welche von der niedrigsten Übereinstimmung von Demokraten, SPD & FDP mit AfD & NSDAP ausgeht, wird in beiden Korpora nicht erfüllt, da im unvorverarbeiteten Korpus die Demokraten näher an der AfD liegen als an CDU, FDP & SPD, und im vorverarbeiteten Korpus eine klare Trennung und räumliche Abgrenzung zwischen Demokraten und den etablierten deutschen Parteien erkennbar ist. Im Vergleich zu den vorherigen Methoden ist bei der Stilometrischen Analyse also teilweise eine sehr schöne Trennung gemäß der zuvor aufgestellten Hypothesen sichtbar, auch wenn der Abstand zwischen den einzelnen Clustern nicht exakt dem in den Hypothesen zunächst vermuteten entspricht.

6.4 Wordfish for Ideological Scaling

6.4.1 Methodische Grundlagen & Einsatzgebiete

Bei der digitalen Methode *Wordfish for Ideological Scaling*, kurz *Wordfish*, handelt es sich um ein statistisches Textskalierungsmodell, das dazu dient, ideologische Positionen politischer Akteure (beispielsweise Parteien) aus Texten, insbesondere Wahlprogrammen, zu extrahieren, was diese Methode für den konkreten Anwendungsfall dieser Arbeit besonders interessant macht. Denn das Verfahren dieser Methode ermöglicht die zeitliche Analyse (*Time-Series Scaling*) von Parteipositionen, ohne dass hierbei jedoch auf eigens vordefinierte Referenztexte zurückgegriffen werden muss,¹⁹⁰ was aufgrund der dadurch entfallenden Arbeitsschritte eine wesentliche Zeitersparnis darstellt. *Wordfish* basiert auf der Annahme, dass die relative Frequenz der Verwendung von Wörtern in einem Text Aufschluss über die politische Position des Autors geben kann. Das Programm bedient sich dabei eines statistischen Sprachmodells, das von der *Naïve-Bayes*-Logik und *Poisson*-Verteilungen für Wortfrequenzen ausgeht. Der von Slapin und Proksch vorgestellte Ansatz suggeriert außerdem, dass Wörter unabhängig voneinander auftreten und die Häufigkeit eines Wortes einer *Poisson*-Verteilung folgt.

Anschließend werden die Modellparameter mittels eines *Expectation–Maximization*-(*EM*)-Algorithmus geschätzt.¹⁹¹ Im Rahmen des *Expectation*-Schritts werden zunächst die Erwartungswerte der latenten Variablen (also Wortgewichte, Parteipositionen etc.) berechnet, während im *Maximization*-Schritt die *Log-Likelihood* maximiert wird, um dadurch eine Aktualisierung der Parameter zu bewirken. Diese Schritte werden daraufhin so oft wiederholt, bis die *Likelihood* übereinstimmt. Das sogenannte „Parametrische Bootstrapping“ liefert nach 500 Wiederholungen etwa 95 prozentige Konfidenzintervalle für die geschätzten Parteipositionen.¹⁹² Je länger die jeweiligen Texte sind bzw. je mehr Wörter analysiert werden, desto kleiner werden die Konfidenzintervalle und desto höher die Präzision. Für spezifische Politikfelder, beispielsweise Wirtschaft oder Außenpolitik, müssen die einzelnen Textabschnitte jedoch getrennt voneinander analysiert werden, wobei Wörter mit besonders hohen positiven bzw. negativen Gewichtungen ideologisch trennscharfe Begriffe markieren.¹⁹³

¹⁹⁰ Vgl. hierzu Slapin, Jonathan B. und Proksch, Sven-Oliver: A Scaling Model for Estimating Time-Series Party Positions from Texts. In: American Journal of Political Science, Volume 52, Issue 3 (2008), S. 705-722 (<https://doi.org/10.1111/j.1540-5907.2008.00338.x>, letzter Zugriff am 17.10.2025). Hier S. 708, gilt nachfolgend bis anders angegeben.

¹⁹¹ Ebd., S. 709-710. Gilt nachfolgend bis anders angegeben.

¹⁹² Ebd., S. 710-711. Gilt nachfolgend bis anders angegeben.

¹⁹³ Ebd., S. 712.

Der Wordfish-Algorithmus untersucht die Unterschiede zwischen Dokumenten und identifiziert natürliche Gruppierungen entlang einer dimensional Skala, indem er die Häufigkeit derselben Wörter in verschiedenen Dokumenten miteinander vergleicht.¹⁹⁴ Der Algorithmus identifiziert und gewichtet die Wörter, deren Häufigkeit zwischen den Dokumenten am stärksten variiert, und verwendet diese Muster dann, um Dokumente entlang einer Skala zu clustern. Wenn ein Wordfish-Modell initialisiert wird, sind alle gemessenen Parameter nur eine „erste bestmögliche Schätzung“ der latenten Skalierung von Dokumenten und Wörtern. Je nach Qualität der Textdaten sollten diese Algorithmen dann ihre anfänglichen Vorhersagen iterativ verfeinern und sich schrittweise statistisch robusteren und aufschlussreicheren Modellstrukturen annähern. Darüber hinaus behandelt Wordfish jedes Dokument als reine *Bag of Words*, da der Algorithmus nur die Häufigkeit der Wörter betrachtet und die Wortreihenfolge ignoriert. Es spielt demnach also keine Rolle, wo Wörter in einem bestimmten Dokument vorkommen, sondern nur, welche Wörter in diesem Dokument vorkommen und mit welcher Häufigkeit. Die *BoW*-Modellierung kann bei längeren Texten jedoch problematisch sein, bei denen verschiedene Abschnitte des Inhalts je nach Kontext unterschiedliche Bedeutungen vermitteln können.

Einen klassischen Anwendungsfall für die *Digital Humanities* bieten Slapin und Proksch selbst, welche die *Wordfish*-Methode in ihrer Studie auf deutsche Parteiprogramme von 1990–2005 angewendet haben.¹⁹⁵ Die Ergebnisse belegen realistische politische Verschiebungen, darunter beispielsweise einen SPD-Rechtsruck unter Bundeskanzler Gerhard Schröder. Dabei werden Wörter mit vergleichsweise hohen Referenzwerten (z. B. „Faschismus“, „Einkommenssteuer“ oder „Frauenbewegung“) zur inhaltlichen Validierung der Ergebnisse genutzt. Die Methode weist außerdem relativ hohe Korrelationen (0.8–0.98) mit anderen,

¹⁹⁴ Wermer-Colan, Alex; Lemire-Garlic, Nicole und Antsen, Jeff: Text Mining YouTube Comment Data with Wordfish in R (<https://doi.org/10.46430/phen0120>, letzter Zugriff am 08.12.2025). Übersetzung des Autors; gilt nachfolgend bis zum Ende des Absatzes.

¹⁹⁵ Ebd., S. 712-720. Gilt nachfolgend bis zum Ende des Absatzes.

vergleichbaren Verfahren auf, darunter *CMP*¹⁹⁶, dem Computerprogramm *Wordscores*¹⁹⁷ sowie Expertensurveys.

Ein weiteres Anwendungsbeispiel bildet Andrea Ceronis Studie zur Analyse von Fraktionsanträgen und Parteikongressen in Italien aus dem Jahr 2015.¹⁹⁸ Ceroni verwendet *Wordfish*, um innerparteiliche Debatten zu skalieren und inhaltliche Veränderungen im Laufe der Zeit zu untersuchen. Aydoğan et al. zeigen eine weitere Anwendungsmöglichkeit auf, indem sie *Wordfish* auf kurze Social Media Posts der Plattform *Twitter* bzw. *X* angewendet haben.¹⁹⁹ Innerhalb dieser Studie wurden die Tweets pro Gruppe aggregiert (also je Partei für Eliten und für Unterstützer) und dann die Wortfrequenzmatrix aufgebaut, also je Gruppe ein „Dokument“ auf Twitter-Basis. Anschließend wurden die ideologischen Positionen der Gruppen geschätzt. Daraus ergaben sich Positionen für die Parteien und indirekt auch für deren Anhängergruppen. Dies zeigt, dass *Wordfish* auch für sehr kurze, nicht-akademische Texte genutzt werden kann.

Darüber hinaus stellte Benjamin Riesch unlängst eine Weiterentwicklung von *Wordfish* namens *Wordkrill* vor.²⁰⁰ *Wordkrill* basiert dabei auf dem Modell von *Wordfish* und erweitert dieses, sodass mehrere ideologische oder thematische Dimensionen

¹⁹⁶ Das *Comparative Manifestos Project* hat die Möglichkeiten der Wissenschaft zur Durchführung komparativer Forschung erheblich verbessert, indem es Schätzungen zu Parteipositionen über Länder und Zeiträume hinweg bereitstellt. Die *CMP-Group* hat dabei 56 Themenbereiche erstellt, die sich weiter in sieben Hauptkategorien unterteilen lassen. Die Generierung von Parteipositionen erfolgt durch die Kodierung der Anzahl der Sätze, die zu jedem Themenbereich gehören, und der Division dieser Anzahl durch die Gesamtzahl aller Sätze im Parteiprogramm. Vgl. hierzu Budge, Ian; Robertson, David und Hearl, Derek: *Ideology, Strategy, and Party Change: Spatial Analyses of Post-War Election Programmes in 19 Democracies*. Cambridge 1987, sowie Budge, Ian; Klingemann, Hans-Dieter; Volkens, Andrea; Bara, Judith und Tanenbaum, Eric: *Mapping Policy Preferences: Estimates for Parties, Electors, and Governments 1945-1998*. Oxford 2001. Zit. n. Slapin et al.: *A Scaling Model for Estimating Party Positions*, S. 706, Übersetzung des Autors.

¹⁹⁷ Frühere Computercodierungssysteme basierten auf der Verknüpfung von Texten mit computergestützten Wörterbüchern (*dictionaries*), die Begriffe oder Ausdrücke enthielten, die mit vorab festgelegten politischen Positionen verglichen wurden. Laver, Benoit und Garry weisen jedoch darauf hin, dass diese Methode den menschlichen Aufwand nicht wirklich reduziert, da sie die wesentliche Mitarbeit ganzer Teams von Forschenden erfordert und die manuelle Eingabe umfangreicher, handcodierter Wörterbücher stets das Risiko menschlicher Fehler mit sich bringt. Vgl. hierzu Laver, Michael; Benoit, Kenneth und Garry, John: *Extracting Policy Positions from Political Texts Using Words as Data*. In: *American Political Science Review* 97/2 (2003), S. 311–332. Zit. n. Slapin et al.: *A Scaling Model for Estimating Party Positions*, S. 707, Übersetzung des Autors.

¹⁹⁸ Vgl. hierzu Ceroni, Andrea: *Changing politics, changing language: The effect of institutional and communicative changes on political language measured through content analysis of Italian intra-party debates*. In: *Journal of Language and Politics*, Band 14, Ausgabe 4 (2015), S. 528-551 (<https://air.unimi.it/bitstream/2434/365263/2/changinglanguage.pdf>, letzter Zugriff am 18.10.2025).

¹⁹⁹ Vgl. hierzu Aydoğan, Abdullah; Tuna, Tayfun und Yildirim, A. Kadir: *Ideological Congruence and Social Media Text as Data*. In: *Journal of Representative Democracy*, Volume 55, Issue 2/2019, S. 159-178 (<https://doi.org/10.1080/00344893.2019.1592012>, letzter Zugriff am 18.10.2025).

²⁰⁰ Vgl. hierzu Riesch, Benjamin: *Wordkrill: Extending Wordfish into the multidimensional political space* (<https://www.arxiv.org/pdf/2506.20275>, letzter Zugriff am 18.10.2025).

abgebildet werden können. Diese Erweiterung trägt dazu bei, das Vorgehen realistischer zu gestalten, insbesondere in Fällen, in denen politische Kommunikation nicht eindimensional (links/rechts) verläuft. Gleichzeitig sind mit dieser Erweiterung jedoch auch methodische und datenbezogene Herausforderungen verbunden, wie etwa die Identifikation der Dimensionen, die Interpretation und die Anforderungen an das Text-Korpus, um nur einige Beispiele zu nennen.

Dieses Forschungsfeld hat sich in den letzten Jahren seit Wordfish aber auch weiter entwickelt, sodass es heute zahlreiche neuere, modernere Alternativen zu Wordfish gibt. Hierzu zählen unter anderem etwa technisch verwandte, ähnliche Methoden, wie beispielsweise *Latent Semantic Scaling*²⁰¹. Aufgrund der Referenztexte mit bekannten politischen Positionen würde sich aber auch *Wordscores*²⁰² anbieten, um einen solchen Anwendungsfall zu untersuchen. Für diese Arbeit bot sich aber dennoch die Verwendung von Wordfish aufgrund einer besseren Vergleichbarkeit an, da es in der Vergangenheit bereits für ähnliche Untersuchungen eingesetzt wurde.²⁰³

6.4.2 Implementierung

Das Wordfish-Modell wurde ursprünglich als *R*-Package entwickelt.²⁰⁴ Inzwischen ist es aber auch möglich, Wordfish über eine Python-Implementierung von *Latent Semantic Scaling* zu nutzen.²⁰⁵ Im ursprünglichen Modell von Slapin & Proksch fungiert als Input eine Wortfrequenzmatrix, bestehend aus Parteien und Wörtern.²⁰⁶ Diese Matrix bildet in Form einer Tabelle ab, welches Wort in welchem Dokument des jeweiligen Korpus wie oft vorkommt. Als Output werden dann geschätzte Parteipositionen, Wortgewichte und Konfidenzintervalle erzeugt. Das Modell setzt jedoch voraus, dass Wortbedeutungen über die Zeit konstant bleiben und sich nicht verändern.

Der Wordfish-Algorithmus ist mittlerweile aber auch in dem *R*-Package *quanteda* enthalten, welches die Daten direkt in ein brauchbares Wordfish-Format umwandelt.²⁰⁷ Um das Wordfish-Modell in *quanteda* auszuführen, müssen zunächst jedoch drei Arten von Textdatenobjekten erstellt werden: Ein Korpus, Tokens und eine *Document*

²⁰¹ Siehe hierzu <https://github.com/umanlp/SemScale>, letzter Zugriff am 08.12.2025.

²⁰² Siehe hierzu <https://tutorials.quanteda.io/machine-learning/wordscores/>, letzter Zugriff am 08.12.2025.

²⁰³ Siehe hierzu eine Vorlage der offiziellen *quanteda* Website mit Beispielen für irische Haushaltsreden (<https://tutorials.quanteda.io/machine-learning/wordfish/>, letzter Zugriff am 08.12.2025).

²⁰⁴ Für den ursprünglichen, unveränderten Wordfish-Code für *R* siehe <http://www.wordfish.org/software.html>, letzter Zugriff am 17.10.2025.

²⁰⁵ Siehe hierzu: <https://github.com/umanlp/SemScale>, letzter Zugriff am 08.12.2025.

²⁰⁶ Slapin et al.: A Scaling Model for Estimating Party Positions, S. 711. Gilt nachfolgend bis zum Ende des Absatzes.

²⁰⁷ Wermer-Colan et al.: Text Mining YouTube Comment Data (<https://doi.org/10.46430/phen0120>, letzter Zugriff am 08.12.2025). Übersetzung des Autors; gilt nachfolgend bis zum Ende des Absatzes.

Feature Matrix (DFM). Das Korpus enthält alle Dokumente, die analysiert werden sollen. In *quanteda* sind Tokens eine Liste von Zeichenvektoren, die mit dem Dokument verknüpft sind, aus dem sie stammen. Diese Form ermöglicht eine weitere Bereinigung und Vorverarbeitung des Textes. So können die Tokens nicht nur lemmatisiert, sondern auch zusätzliche Stoppwörter entfernt werden.

Auch wenn eine Lemmatisierung oder die Entfernung von Stoppwörtern aufgrund der bereits zuvor erfolgten Vorverarbeitung der Parteienkorpora nicht mehr notwendig ist, bietet sich aufgrund dieser Vorteile *quanteda* für die Implementierung an. Als Basis des Codes dient eine Vorlage der offiziellen *quanteda* Website mit Beispielen für irische Haushaltsreden.²⁰⁸ Dieser Code musste final noch etwas angepasst werden.²⁰⁹ Nach dem Laden der *quanteda*-Packages wird als Korpus zunächst der Zielordner mit den vorverarbeiteten Parteienkorpora festgelegt. Die Namen der jeweiligen Parteien werden dabei direkt aus den Dateinamen übernommen, sodass keine manuelle Erstellung einer Liste von Parteien mehr notwendig ist. Daraus wird anschließend ein *quanteda*-Corpus gebaut und Wordfish ausgeführt, das die Dokumentpositionen als Parteien-Koordinaten ausgibt. Dabei werden die Wordfish-Dimensionen für jede Partei (*theta*) erzeugt und in *RStudio* geplottet. Zusätzlich werden die verschiedenen Parteien in einer Tabelle nach ihrer jeweiligen Wordfish-Position sortiert und in einem Clusterplot, in diesem Fall einem Dendrogramm, visualisiert. Diese Grafik wird automatisch als PNG-Datei exportiert und im selben Verzeichnis wie die txt-Dateien gespeichert.

6.4.3 Analyse der Ergebnisse & Überprüfung der Hypothesen

Da Wordfish zunächst die Positionen der jeweiligen Dokumente schätzt, ergibt sich hieraus die folgende Tabelle (s. Abbildung 16).

	Partei	Position		SE
6	NSDAP	1.3474909	0.003744553	
5	GOP	1.0366607	0.005749220	
3	DEM	0.7210182	0.005397231	
2	CDU	-0.5879549	0.006222239	
7	SPD	-0.6443098	0.006469250	
4	FDP	-0.7427745	0.006627552	
1	AfD	-1.1301307	0.004762899	

Abbildung 16: Geschätzte Dokumentpositionen der jeweiligen Parteienkorpora in Wordfish, sortiert in absteigender Reihenfolge

²⁰⁸ Siehe hierzu <https://tutorials.quanteda.io/machine-learning/wordfish/>, letzter Zugriff am 08.12.2025.

²⁰⁹ Bei dem fertigen, finalen Code handelt es sich um das Skript „[wordfish.r](#)“ (s. GitHub), welches mit Hilfe von Microsoft Copilot (<https://m365.cloud.microsoft/chat/>) erstellt wurde.

Hierbei stehen die beiden Parameter *theta* und *SE* jeweils für die geschätzte Position jedes Dokuments (bzw. in diesem Fall Parteikorpus) auf der ideologischen Dimension sowie für die Standardfehler dieser Schätzung, wobei möglichst kleine Werte in der *SE* für eine sehr präzise Schätzung der Position stehen. Der Positionswert *theta* gibt die geschätzte ideologische Position auf einer kontinuierlichen Skala an, die Wordfish aus den Wortfrequenzen berechnet. Negative Werte bedeuten, dass die jeweiligen Parteien am linken Ende der Skala liegen, also in diesem Kontext beispielsweise eher „links“ oder progressiv sind, während positive Werte bedeuten, dass die Parteien am rechten Ende der Skala liegen, also eher „rechts“ oder konservativ sind. Auf die genaue Interpretation der jeweiligen Werte wird später noch genauer eingegangen werden.

Zusätzlich werden von Wordfish aber auch noch die *Estimated Feature Scores* aus dem Wordfish-Modell ausgegeben. In diesem Fall handelt es sich exemplarisch um die ersten 30 Begriffe des bereinigten AfD-Korpus (s. Abbildung 17).

Estimated Feature Scores:						
	programm deutschland	grundsatzprogramm	alternative	präambel	demokratie	
beta	0.2017	-0.7133	-2.649	-3.8440	-0.387287	0.08819
psi	2.2372	4.6255	-1.430	-0.9597	0.002483	3.61372
	grundwerte	volksabstimmungen	schweizer	vorbild	schlanker	staat
beta	-0.87868	-5.733	-5.049	-1.1679	-4.219	-0.3379
psi	0.03715	-4.740	-3.477	0.9084	-2.807	3.3465
	bürger	gewaltenteilung	gewährleisten	trennung	amt	mandat
beta	-0.8376	-3.613	0.1536	-1.7061	0.8428	-1.0822
psi	3.1253	-1.938	2.1408	-0.8329	1.9787	0.8449
	parteien	beschränken	parteienfinanzierung	verfassungsrecht	anpassen	
beta	-0.1952	-0.3026	-6.383	-0.8434	-1.6993	
psi	2.3552	0.5815	-5.064	-1.1496	0.2706	
	listenwahl	freies	verkleinerung	bundestages	wider	
beta	-4.737	-2.253	-3.841	-2.5908	-0.14553	
psi	-4.275	-1.581	-2.748	-0.5169	0.04605	

Abbildung 17: Estimated Feature Scores aus dem Wordfish-Modell für die ersten 30 Begriffe des bereinigten AfD-Korpus

Diese *Estimated Feature Scores* schätzen für jedes Wort (*Feature*), wie stark es zur ideologischen Dimension des Dokuments beiträgt. Für jedes Wort werden zwei Parameter angegeben, nämlich *beta* und *psi*. Der erste Wert gibt an, wie stark das jeweilige Wort die Position auf der ideologischen Skala beeinflusst. Ein hoher positiver *beta*-Wert bedeutet, dass das Wort häufiger bei Parteien am positiven Ende der Skala auftritt, zum Beispiel bei der NSDAP oder den Republikanern. Ein hoher negativer *beta*-Wert bedeutet hingegen, dass das jeweilige Wort häufiger bei Parteien am negativen Ende der Skala vorkommt, beispielsweise bei der AfD, FDP oder SPD. So besitzt das Wort „Programm“ beispielsweise einen *beta*-Wert von 0.2017. Da dieser Wert leicht positiv ist, ist dieses Wort eher bei Parteien rechts auf der Skala zu finden.

Der vergleichsweise hohe *psi*-Wert von 2.2372 gibt an, dass dieses Wort im gesamten Korpus eher häufig vorkommt. Das Wort „Listenwahl“ mit einem *beta*-Wert von -4.737 ist wiederum stark negativ und daher typisch für Parteien links auf der Skala. Der fast ebenso hohe negative *psi*-Wert deutet darauf hin, dass dieses Wort im gesamten Wordfish-Korpus nur sehr selten vorkommt. Wörter mit extremen *beta*-Werten, aber gleichzeitig auch möglichst hohen *psi*-Werten sind hierbei die interessantesten Indikatoren für die ideologische Dimension eines Korpus, da sie ideologisch besonders aussagekräftig sind (*beta*) und besonders häufig vorkommen (*psi*). Die negativsten *beta*-Werte bei gleichzeitig möglichst hohen *psi*-Werten besitzen hier Begriffe wie „Bürger“, „Staat“, „Parteien“ oder auch „Grundwerte“. Diese Begriffe sind somit typisch für Parteien am linken Ende der Skala, zu denen die AfD, FDP, SPD und CDU zählen. Diese Parteien scheinen also vorwiegend über rechtsstaatliche und demokratische Themen zu sprechen. Die positivsten *beta*-Werte bei gleichzeitig möglichst hohen *psi*-Werten besitzen die Wörter „Amt“, „Macht“ und „Programm“, welche daher besonders typisch für Parteien am rechten Ende der Skala (NSDAP, Republikaner und Demokraten) sind. Diese Parteien scheinen also vorwiegend über Themen wie institutionelle Macht und Autorität zu sprechen.

Visualisiert man wiederum die Werte aus Abb. 16, so entsteht das folgende zweidimensionale Koordinatensystem mit den von Wordfish geschätzten Dokumentpositionen der jeweiligen Parteienkorpora (s. Abbildung 18).

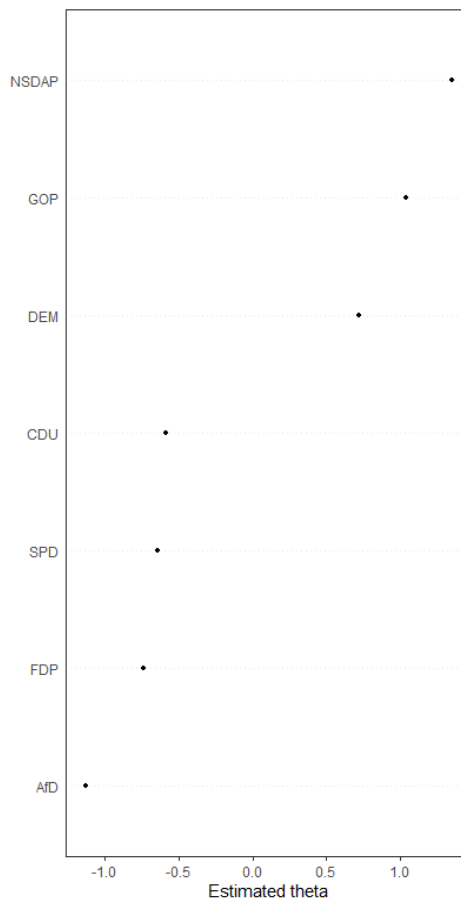


Abbildung 18: Visualisierung der von Wordfish geschätzten Dokumentpositionen der jeweiligen Parteienkorpora im zweidimensionalen Koordinatensystem

In dieser Visualisierung wird deutlich, dass die AfD mit einem *theta*-Wert von -1.1301 weit links auf der Wordfish-Skala in diesem Modell angeordnet ist. Darauf folgt die FDP mit einem Wert von -0.7428, dicht gefolgt von SPD (-0.6443) und CDU (-0.5880), welche ebenfalls im negativen Bereich angeordnet sind, aber etwas weniger extrem als die zuvor genannten Parteien. Demokraten (0.7210), Republikaner (1.0367) und NSDAP (1.3475) weisen hingegen deutlich positive Werte auf, weshalb sie am anderen Ende der Skala angeordnet werden. Dies erscheint auch durchaus schlüssig, da eine große Nähe zwischen der NSDAP und den Republikanern zuvor bereits vermutet und dieser Verdacht in den aufgestellten Hypothesen festgehalten wurde.

Auf den ersten Blick mögen diese Ergebnisse aber dennoch verwundern, da die AfD, entgegen allen zuvor aufgestellten Hypothesen, nicht nur mit den etablierten deutschen Parteien größere Ähnlichkeiten teilt als mit der NSDAP oder den Republikanern, sondern darüber hinaus auch noch die Partei am äußersten linken Rand der Wordfish-Skala ist. Das kann in diesem Kontext aber dadurch erklärt werden, dass die AfD sich bestimmter Begriffe bedient, welche im *beta*-Score des Wordfish-

Modells sehr negativ konnotiert, also auf der negativen „linken“ Seite der Skala verortet sind. Hierzu zählen Begriffe wie beispielsweise „Volksabstimmungen“, „Listenwahl“ oder auch „Gewaltenteilung“ (s. Abbildung 17). Dadurch wird der Eindruck erweckt, dass es sich bei der AfD aufgrund der Verwendung bestimmter Begriffe um eine „linke“ Partei handelt. Da diese Wordfish-Skala jedoch ausschließlich auf den Sprachmustern des ausgewählten Korpus basiert, ist sie lediglich relativ und nicht absolut. Außerdem ist sie nicht automatisch „links-rechts“ im politischen Sinne, sondern stellt vielmehr eine statistische Dimension basierend auf den Wortverteilungen innerhalb der jeweiligen Parteienkorpora dar.

Zusätzlich lassen sich diese Ergebnisse aber auch noch in Form eines hierarchischen Dendrogramms visualisieren (s. Abbildung 19).

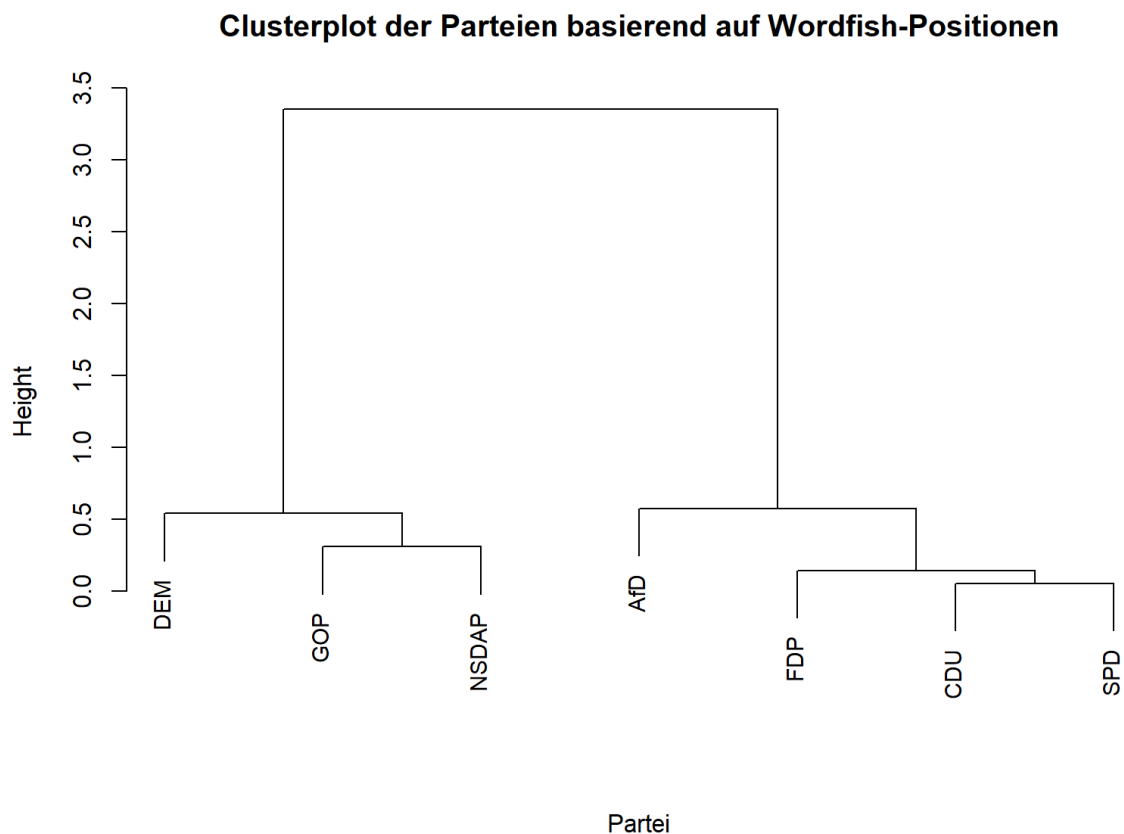


Abbildung 19: Clusterplot der verschiedenen Parteien, basierend auf deren jeweiligen Wordfish-Positionen

Abb. 19 zeigt ein hierarchisches Cluster-Dendrogramm, welches die Ähnlichkeit der verschiedenen Parteien, basierend auf ihren jeweiligen Wordfish-Positionswerten, darstellt. Die Parteien werden hierbei auf der x-Achse zu Clustern zusammengefasst, je nachdem, wie ähnlich ihre geschätzten ideologischen Positionen (*theta*-Werte) sind,

während die y-Achse die *Height*, also die Distanz zwischen den Clustern, welche während der hierarchischen Clusteranalyse berechnet wurde, repräsentiert. Die Werte auf der y-Achse geben an, wie weit zwei Cluster voneinander entfernt sind, bevor sie zusammengeführt werden. Eine geringe Höhe (z. B. < 1) bedeutet, dass die Parteien sich sehr ähnlich in ihren Wordfish-Positionswerten sind. So liegen beispielsweise CDU und SPD sehr nah beieinander. Eine große Höhe (z. B. > 3) bedeutet hingegen, dass sich bestimmte Parteien bzw. Cluster sehr unterschiedlich sind. So trennt beispielsweise der große „Ast“ bei *Height*-Wert ~ 3.5 das Cluster DEM, GOP und NSDAP vom zweiten Cluster AfD, FDP, CDU und SPD.

Konkret bedeutet dies, dass die drei Parteien aus Cluster 1 (links), also sprich DEM, GOP und NSDAP, auf der Wordfish-Skala aufgrund ihrer positiven *theta*-Werte nahe beieinander liegen. Dies bedeutet, dass sie ähnliche sprachliche Muster teilen, die vom Modell als „rechts“ bzw. konservativ geprägt interpretiert werden. Die Parteien AfD, FDP, CDU und SPD besitzen negative *theta*-Werte und bilden daher ein eigenes Cluster (Cluster 2, rechts). Ihre Sprache ist im Modell eher „links“ oder moderat positioniert.

Diese Visualisierung bestätigt demnach die Wordfish-Skala, da AfD und FDP näher an CDU und SPD als an GOP und NSDAP sind. DEM, GOP und NSDAP bilden ein deutlich getrenntes Cluster. Die y-Achse ist hierbei jedoch kein ideologischer Wert, sondern vielmehr ein Abstandswert aus der Clusteranalyse. Je höher der Verbindungsast ist, desto größer sind demnach auch die Unterschiede zwischen den zusammengeführten Gruppen. Dennoch ist es wichtig sich in Erinnerung zu rufen, dass diese Grafiken die ideologische Nähe auf Basis von Sprachmustern und nicht auf Basis von tatsächlicher politischer Selbstverortung visualisieren.

In Bezug auf die zuvor aufgestellten Hypothesen lässt sich somit sagen, dass das Wordfish-Modell die ersten beiden Hypothesen deutlich widerlegt, da weder eine hohe Übereinstimmung zwischen AfD & Republikanern (Wordfish-Scores von -1.13 zu +1.04), noch zwischen AfD & NSDAP (Wordfish-Scores von -1.13 zu +1.35) nachgewiesen werden kann. Die dritte Hypothese, welche von einer hohen Übereinstimmung von CDU, SPD & FDP mit den Demokraten ausgeht, trifft ebenso wenig zu, da die drei deutschen Parteien allesamt verhältnismäßig hohe negative Werte besitzen (-0.59 bis -0.74), die Demokraten aber einen vergleichsweise hohen positiven Wert (+0.72) aufweisen. Die vierte Hypothese, welche die niedrigste Übereinstimmung von Demokraten, SPD & FDP mit AfD & NSDAP prognostiziert, trifft nur teilweise zu, da SPD & FDP zwar einen großen Abstand zur NSDAP aufweisen (maximale Differenz beträgt 2.09), ebenso aber auch zu den Demokraten (maximale Differenz beträgt 1.46) und dafür näher an der AfD verortet werden (maximale Differenz beträgt 0.49). Die Demokraten wiederum liegen zwar weit entfernt von der

AfD (Differenz von 1.85), dafür aber auch näher an der NSDAP (Differenz von 0.63) als an SPD & FDP (maximale Differenz beträgt 1.46).

6.5 Word2Vec

6.5.1 Methodische Grundlagen & Einsatzgebiete

Vektorbasierte Methoden zählen die Wörter eines Korpus im Gegensatz zu frequenzbasierten Methoden nicht mehr, sondern stellen sie (oder auch ganze Texte) in Form von Vektoren in einem zwei- bzw. dreidimensionalen Raum dar. Dadurch ergibt sich eine semantische Ähnlichkeit, wobei jedes Wort bzw. jeder Text in Form eines Vektors abgebildet wird. Je näher die Vektoren beieinander liegen, desto ähnlicher sind sich entsprechend die beiden Wörter bzw. Texte. Aus diesem Grund sind verschiedene Maße notwendig, die numerisch ausdrücken, wie ähnlich sich zwei Vektoren sind.²¹⁰ Eine Möglichkeit hierfür bietet die sogenannte *Euclidian Distance* (deutsch „euklidische Distanz“), welche die Länge der direkten Verbindungslinie zwischen zwei Punkten mit Hilfe des Satzes von Pythagoras berechnet. Die Euclidian Distance bietet darüber hinaus den Vorteil, dass sie problemlos auch in mehr als zwei Dimensionen angewendet werden kann.

Eine weitere Möglichkeit zur mathematischen Berechnung der Ähnlichkeit zweier Vektoren wäre die sogenannte *Cosine Similarity* bzw. *Cosine Distance* (deutsch „Kosinusähnlichkeit“ bzw. „Kosinusdistanz“), die sich in der Computerlinguistik mittlerweile als Standardverfahren etabliert hat und auf dem Winkel, der zwischen den beiden Vektoren liegt, basiert. Da zwischen zwei Vektoren, die auf im Vektorraum eng benachbarte Punkte zeigen, ein spitzer Winkel liegt, und zwischen zwei Vektoren, die auf zwei Punkte mit großem Abstand zeigen, ein weiter Winkel liegt, besitzt dieses Maß gegenüber der *Euclidian Distance* den entscheidenden Vorteil, dass es von der Länge der Vektoren -und damit auch von der Text- bzw. Korpusgröße- unabhängig ist, da der Winkel zwischen den beiden Vektoren immer gleich bleibt.²¹¹ Dadurch bleibt auch die *Cosine Distance* stabil. Da die *Cosine Similarity* von der Korpusgröße unabhängig ist, ist sie als Ähnlichkeits- bzw. Distanzmaß für den Anwendungsfall dieser Arbeit am besten geeignet und soll daher -stellvertretend für Ähnlichkeits- bzw. Distanzmaße- im Folgenden für die politischen Texte des Korpus dieser Arbeit berechnet werden.

Bevor die *Cosine Similarity* jedoch berechnet werden kann, müssen die zu vergleichenden Texte bzw. Korpora zunächst in Vektoren umgewandelt werden. Dies

²¹⁰ Andresen, Melanie: Computerlinguistische Methoden für die Digital Humanities. Eine Einführung für Geisteswissenschaftler:innen. Tübingen 2024, S. 97. Gilt nachfolgend bis anders angegeben.

²¹¹ Ebd., S. 98. Gilt nachfolgend bis zum Ende des Absatzes.

geschieht beispielsweise mit Hilfe des *Word2Vec*-Modells von Mikolov et al., bei dem es sich um ein neuronales Modell zur Erzeugung von kontinuierlichen Wortvektoren (*word embeddings*) handelt.²¹² Das Ziel dieser Methode ist, sowohl semantische als auch syntaktische Ähnlichkeiten zwischen Wörtern durch ihre Verwendungskontexte mathematisch zu erfassen und nachzuweisen. Im Gegensatz zum weiter oben bereits beschriebenen *Bag-of-Words*-Ansatz, der Wörter als isolierte Einheiten einzelner Tokens betrachtet, stellt *Word2Vec* Wörter als Vektoren im mehrdimensionalen Raum dar, deren Ähnlichkeit im Sinne semantischer Nähe gemessen werden kann, da im multidimensionalen Vektorraum semantisch ähnliche Wörter nahe beieinander liegen. Die genaue Ähnlichkeit zwischen den Wörtern, die dann entsprechend als Wortvektoren dargestellt werden, kann anschließend wiederum über die zuvor bereits beschriebene *Cosine Similarity* gemessen werden.

In den *Digital Humanities* wird *Word2Vec* vorwiegend dazu eingesetzt, um große, historische Korpora zu analysieren und zu strukturieren. N. Tahmasebi führte beispielsweise eine Machbarkeitsstudie für diachrone *Word2Vec*-Analysen an historischen schwedischen Zeitungen von 1749–1925 durch.²¹³ Ihr Fokus liegt hierbei auf der Stabilität bzw. Qualität von Vektoren bei begrenzter historischer Datenlage. Ein ähnliches Beispiel bietet R. Heuser mit seiner exemplarischen Anwendung von *Word2Vec* auf literarische Korpora.²¹⁴ Dabei analysiert Heuser, wie durch den Einsatz von Vektoren *Close*²¹⁵- und *Distant-Reading*²¹⁶ Ansätze miteinander kombiniert

²¹² Vgl. hierzu Mikolov, Tomas; Chen, Kai; Corrado, Greg und Dean, Jeffrey: Efficient Estimation of Word Representations in Vector Space. In: *Workshop Proceedings of the 1st International Conference on Learning Representations (ICLR)*. Scottsdale 2013 (<https://arxiv.org/pdf/1301.3781>, letzter Zugriff am 12.10.2025), sowie Mikolov, Tomas; Yih, Wen-tau; und Zweig, Geoffrey: Linguistic Regularities in Continuous Space Word Representations. In: *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Atlanta 2013, S. 746-751 (<https://aclanthology.org/N13-1090.pdf>, letzter Zugriff am 12.10.2025).

²¹³ Vgl. hierzu Tahmasebi, Nina: A Study on Word2Vec on a Historical Swedish Newspaper Corpus. In: *CEUR Workshop Proceedings Vol. 2084* (DHN 2018), S. 25–37 (<https://ceur-ws.org/Vol-2084/paper2.pdf>, letzter Zugriff am 20.10.2025).

²¹⁴ Vgl. hierzu Heuser, Ryan James: Word Vectors in the Eighteenth Century. In: *DH2017* (<https://dh2017.adho.org/abstracts/582/582.pdf>, letzter Zugriff am 20.10.2025).

²¹⁵ In der Literaturwissenschaft bezeichnet *Close Reading* eine intensive Auseinandersetzung mit einem Text, bei der Wortwahl, Syntax und bestimmte Bilder untersucht werden, um deren Bedeutung zu erschließen. Diese Methode stammt aus dem *New Criticism* und stützt sich nicht auf historische oder biografische Forschungen, sondern zielt darauf ab, die Komplexität des jeweiligen Textes zu analysieren. Vgl. hierzu Wermer-Colan, Alex: Stylometry Methods and Practices. A review of various stylometry methods and programs for the digital humanities (<https://guides.temple.edu/stylometryfordh>, letzter Zugriff am 24.10.2025). Übersetzung des Autors.

²¹⁶ Beim sogenannten „Distant-Reading-Ansatz“ (englisch auch *Distant Reading Approach*) geht es darum, Texte nicht mehr einzeln zu lesen, sondern aus einer möglichst großen Menge an Texten mithilfe computergestützter Verfahren, basierend auf quantitativer Datenanalyse, neue Erkenntnisse zu gewinnen, ohne die jeweiligen Texte jedoch tatsächlich selbst zu lesen. Der italienische Literaturwissenschaftler Franco Moretti (*1950) führt diesen Ansatz in seiner für die *Digital Humanities* inzwischen einschlägig gewordenen Arbeit „Network Theory, Plot Analysis“ weiter aus. Vgl. hierzu Moretti, Franco: Network Theory, Plot Analysis. In: *New Left Review*, März/April 2011

werden können. Wevers und Koolen untersuchen ebenfalls die Anwendungsmöglichkeiten von *Word-Embeddings* (beispielsweise unter Verwendung von Methoden wie *Word2Vec*) in historischen Korpora, insbesondere im Hinblick auf Begriffsgeschichte und semantischen Wandel.²¹⁷ Ein zentrales Thema ihrer Studie ist die sogenannte *Bias*-Problematik. Die Autoren machen darauf aufmerksam, dass Embeddings zwar semantische Beziehungen erfassen können, aber auch sehr stark davon abhängen, welche Texte, welcher Zeitraum und welches Korpus verwendet werden. Je nachdem können durch diese unterschiedlichen Faktoren systematische Verzerrungen (*bias*) entstehen. Einen weiteren Anwendungsfall bieten Rheault und Cochrane, welche mit Hilfe von *Word2Vec* die semantische Nähe verschiedener Parteien über politische Reden analysierten.²¹⁸ Da diese Studie jedoch in Kapitel „2 Aktueller Forschungsstand“ bereits detailliert vorgestellt wurde, soll sie an dieser Stelle nur aus Gründen der Vollständigkeit kurz erwähnt werden.

6.5.2 Implementierung

In Python besitzt das Paket *gensim* bereits ein vordefiniertes *Word2Vec*-Modell, das über „pip install“ installiert werden kann. Der übrige Python-Code zur Erstellung eines lokalen *Word2Vec*-Modells wurde nach dem Vorbild eines YouTube-Tutorials übernommen.²¹⁹ Hierbei mussten die bereits vorverarbeiteten und lemmatisierten Dateien für alle Parteien zunächst wieder eingelesen werden. *Word2Vec* benötigt daraufhin eine Liste von Tokens pro Satz. Da die Korpora bereits lemmatisiert und bereinigt sind, reicht eine einfache Tokenisierung aus. Danach kann dann das eigentliche *Word2Vec*-Modell von *gensim* generiert werden. Dieses liest zunächst die Liste aller tokenisierten Sätze ein und kann beliebig angepasst werden. Für diesen konkreten Anwendungsfall bietet es sich an, in den Einstellungen zu definieren, dass alle Wörter (selbst diejenigen, die bloß ein einziges Mal im Korpus vorkommen), berücksichtigt werden, sowie ein Kontextfenster von 5 Wörtern vor und nach dem aktuellen Wort zu wählen. Die übrigen Werte für die Dimensionen der generierten Vektoren (=50) sowie die Anzahl der Iterationen über das Textkorpus (=10) werden

(<https://newleftreview.org/issues/ii68/articles/franco-moretti-network-theory-plot-analysis>, letzter Zugriff am 13.10.2025).

²¹⁷ Vgl. hierzu Wevers, Melvin und Koolen, Marijn: Digital begriffsgeschichte: Tracing semantic change using word embeddings. In: *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, Vol. 53(4), 2020, S. 226-243 (<https://pure.uva.nl/ws/files/66795875/01615440.2020.pdf>, letzter Zugriff am 20.10.2025).

²¹⁸ Vgl. hierzu Rheault, Ludovic und Cochrane, Christopher: Word Embeddings for the Analysis of Ideological Placement in Parliamentary Corpora. In: *Political Analysis*, Vol. 28(1), 2019/20, S. 112–133 (<https://www.cambridge.org/core/journals/political-analysis/article/word-embeddings-for-the-analysis-of-ideological-placement-in-parliamentary-corpora/017F0CEA9B3DB6E1B94AC36A509A8A7B>, letzter Zugriff am 20.10.2025).

²¹⁹ Vgl. hierzu “How to Build Your Own Word2Vec Model from Scratch in Python” (<https://www.youtube.com/watch?v=a3dxTcncUg4>, letzter Zugriff am 11.11.2025).

zunächst aus der Vorlage beibehalten. Daraufhin wird ein komplettes *Word2Vec*-Modell aus dem Korpus generiert und als .model-Datei gespeichert. Da diese Datei jedoch in Binärcode geschrieben und dadurch nicht lesbar ist, muss sie zunächst noch ins txt-Format umgewandelt werden. Dies geschieht ganz einfach mit Hilfe weniger Zeilen Code am Ende desselben Python-Skripts.²²⁰

Mit Hilfe eines zweiten Skripts²²¹ muss das zuvor generierte *Word2Vec*-Modell zunächst geladen werden, bevor es abgefragt werden kann. Dieses Modell besteht aus den gemischten Korpora aller Parteien und besitzt eine Vokabulargröße von 33.527 verschiedenen Wörtern, die im Modell enthalten sind, wobei jedes dieser Wörter in einem 50-dimensionalen Wortvektor repräsentiert wird. Hier lassen sich beispielsweise die 20 häufigsten Begriffe innerhalb des Modells (und damit auch der gemischten Korpora aller Parteien) sowie die Embeddings bestimmter Wörter, die in Form von Vektoren abgebildet werden, ganz gezielt abfragen. Darüber hinaus lassen sich zu beliebigen Begriffen die zehn ähnlichsten Wörter innerhalb des Modells abfragen.

Um nun jedoch die Ähnlichkeit zwischen zwei Parteien mithilfe des *Word2Vec*-Modells zu berechnen, muss die Cosine Similarity zwischen den dokumentbezogenen Vektoren bestimmt werden. Da *Word2Vec* zwar Wortvektoren, aber keine Dokumentvektoren erzeugt, muss aus den Wortvektoren eines Textes ein Textvektor berechnet werden. Dies geschieht typischerweise durch eine Mittelung der Wortvektoren (*Averaging*). Zu diesem Zweck wurde ein drittes Skript²²² erstellt, in welchem alle Wortvektoren gemittelt wurden, woraufhin die Cosine Similarity entweder für zwei einzelne, oder auch für alle Parteien gleichzeitig berechnet werden konnte. Die Ergebnisse dieser Berechnung werden nun im Folgenden diskutiert.

6.5.3 Analyse der Ergebnisse & Überprüfung der Hypothesen

Nach der Ausführung dieses Codes werden einmal die Ähnlichkeit zwischen CDU & SPD exemplarisch berechnet (0.9999823) sowie die der übrigen Parteien untereinander in einer Tabelle dargestellt (s. Abbildung 20).

²²⁰ Hierbei handelt es sich um das Skript „[Generierung w2v-Modell.py](https://m365.cloud.microsoft.com/chat/)“ (s. GitHub), welches mit Hilfe von Microsoft Copilot (<https://m365.cloud.microsoft.com/chat/>) erstellt wurde.

²²¹ Hierbei handelt es sich um das Skript „[Abfrage w2v-Modell.py](https://m365.cloud.microsoft.com/chat/)“ (s. GitHub), welches mit Hilfe von Microsoft Copilot (<https://m365.cloud.microsoft.com/chat/>) erstellt wurde.

²²² Hierbei handelt es sich um das Skript „[cosine similarity w2v.py](https://m365.cloud.microsoft.com/chat/)“ (s. GitHub), welches mit Hilfe von Microsoft Copilot (<https://m365.cloud.microsoft.com/chat/>) erstellt wurde.

	AfD	CDU	DEM	FDP	GOP	NSDAP	SPD
AfD	1.000000	0.999996	0.999892	0.999985	0.999875	0.999984	0.999978
CDU	0.999996	1.000000	0.999911	0.999988	0.999897	0.999983	0.999982
DEM	0.999892	0.999911	1.000000	0.999910	0.999991	0.999903	0.999855
FDP	0.999985	0.999988	0.999910	1.000000	0.999882	0.999962	0.999960
GOP	0.999875	0.999897	0.999991	0.999882	1.000000	0.999899	0.999847
NSDAP	0.999984	0.999983	0.999903	0.999962	0.999899	1.000000	0.999954
SPD	0.999978	0.999982	0.999855	0.999960	0.999847	0.999954	1.000000

Abbildung 20: Darstellung der Ähnlichkeitswerte aller Parteien mit Cosine Similarity in Matrizenform

Außerdem wird eine Liste mit den Top-Ähnlichkeitswerten zwischen allen Parteien in absteigender Reihenfolge ausgegeben (s. Abbildung 21).

```

AfD – CDU: 1.0000
DEM – GOP: 1.0000
CDU – FDP: 1.0000
AfD – FDP: 1.0000
AfD – NSDAP: 1.0000
CDU – NSDAP: 1.0000
CDU – SPD: 1.0000
AfD – SPD: 1.0000
FDP – NSDAP: 1.0000
FDP – SPD: 1.0000
NSDAP – SPD: 1.0000
CDU – DEM: 0.9999
DEM – FDP: 0.9999
DEM – NSDAP: 0.9999
GOP – NSDAP: 0.9999
CDU – GOP: 0.9999
AfD – DEM: 0.9999
FDP – GOP: 0.9999
AfD – GOP: 0.9999
DEM – SPD: 0.9999
GOP – SPD: 0.9998

```

Abbildung 21: Liste der Top-Ähnlichkeiten zwischen den Parteien in absteigender Reihenfolge

Hierbei fällt auf, dass die Werte zur Übereinstimmung zwischen den verschiedenen Parteien sehr hoch sind, was einem nahezu identischen Inhalt entsprechen würde. Diese derart hohen Werte könnten mitunter dadurch erklärt werden, dass die *Word2Vec*-Dokumentvektoren über eine einfache Mittelung aller Wortvektoren

gebildet wurden.²²³ Da das *Word2Vec*-Modell über alle Parteien gemeinsam trainiert wurde, liegen die Wortvektoren in einem gemeinsamen semantischen Raum. Hinzu kommt, dass viele Wörter in den Korpora aller Parteien vorkommen (s. Kapitel 6.1 Bag-of-Words), wodurch sich die Dokumentvektoren der verschiedenen Parteien strukturell sehr stark ähneln, obwohl die jeweiligen Parteienkorpora aufgrund der Auswahl der verschiedenen Parteien politisch sehr heterogen sind. Die Tatsache, dass im *Word2Vec*-Modell der Dokumentvektor aus dem Durchschnitt aller Wortvektoren gebildet (=„gemittelt“) wird, also unzähligen von teilweise sehr generischen Wörtern wie beispielsweise „Menschen“, „Europa“ oder „Deutschland“, könnte also dazu geführt haben, dass teils stark geglättete, fast identische Dokumentvektoren entstehen, wodurch die unterschiedlichen Parteien im Vektorraum sehr nahe zusammenrücken, was wiederum diese extrem hohen Ähnlichkeitswerte erklären würde.

Diese Problematik scheint in der Forschung bereits bekannt zu sein, sodass sich, um diese mangelnde Trennschärfe in den Ähnlichkeitswerten zu bereinigen, die Verwendung von *Doc2Vec*²²⁴ empfiehlt. Das neue Skript²²⁵ erzeugt an Stelle einer einfachen Mittelung aller Wortvektoren für jedes Parteienkorpus eigene *Doc2Vec*-Dokumentvektoren und stellt daraufhin erneut die Ähnlichkeitswerte aller Parteien mit Hilfe der Cosine Similarity in Matrizenform dar. Zusätzlich wird ein Ranking der ähnlichsten Parteien ausgegeben. Bei diesem neuen *Doc2Vec*-Modell wird zudem die Anzahl der Embeddings von bisher 50 auf 200 erhöht, um dadurch eine höhere Genauigkeit zu gewährleisten. Das Kontextfenster wurde ebenfalls von 5 auf 10 verdoppelt und die Anzahl der Iterationen von 10 auf 40 erhöht. Auch werden dieses Mal nur noch Wörter berücksichtigt, die mindestens zwei Mal im Korpus vorkommen, wodurch alle Wörter, die nur ein einziges Mal auftreten, automatisch wegfallen. Nach dem Training dieses Modells auf die Korpora der verschiedenen Parteien wurden die

²²³ Siehe hierzu das Skript „[cosine similarity w2v.py](#)“ auf GitHub.

²²⁴ *Doc2Vec* bzw. *Paragraph Vector* (Le & Mikolov) ist ein weiterentwickeltes *Word2Vec*-Modell, welches nicht nur Wortvektoren lernt, sondern darüber hinaus für jedes Dokument einen eigenen Vektor bildet und diesen trainiert, sodass dieser die kontextuellen Eigenschaften des Dokuments in dessen Repräsentation mit einbezieht. Die jeweiligen Dokumentvektoren werden also aktiv mittrainiert anstatt nachträglich berechnet, was für eine deutlich bessere Trennschärfe zwischen den verschiedenen Dokumenten sorgt. *Doc2Vec* wurde dabei bereits in vielen verschiedenen Bereichen des *NLP* erfolgreich eingesetzt, darunter bspw. *Text Classification*, *Sentiment Analysis* oder *Information Retrieval*, also genau dort, wo einfache Wortvektor-Mittelungen inhaltlich verschiedene Texte nicht ausreichend differenzieren können, wie es auch im konkreten Anwendungsfall dieser Arbeit der Fall ist. Vgl. hierzu Le, Quoc und Mikolov, Tomas: Distributed Representations of Sentences and Documents. In: *Proceedings of the 31st International Conference on Machine Learning*, PMLR 32(2), 2014, S. 1188-1196 (<https://proceedings.mlr.press/v32/le14.pdf>), sowie Lau, Jey Han und Baldwin, Timothy: An Empirical Evaluation of doc2vec with Practical Insights into Document Embedding Generation. In: *Proceedings of the 1st Workshop on Representation Learning for NLP*, Berlin 2016, S. 78-86 (<https://arxiv.org/pdf/1607.05368>).

²²⁵ Hierbei handelt es sich um das Skript „[doc2vec.py](#)“ (s. GitHub), welches mit Hilfe von Microsoft Copilot (<https://m365.cloud.microsoft/chat/>) erstellt wurde.

Dokumentvektoren für jede Partei generiert, deren Ähnlichkeit mit Hilfe der Cosine Similarity berechnet und in einer Ähnlichkeitsmatrix (s. Abbildung 22) visualisiert wurde. Auch wurden die verschiedenen Parteien paarweise wieder in einer Liste gemäß ihrer Ähnlichkeit in absteigender Reihenfolge sortiert (s. Abbildung 23).

	AfD	CDU	DEM	FDP	GOP	NSDAP	SPD
AfD	1.000000	0.080434	0.038087	0.132173	0.178738	0.215895	0.212094
CDU	0.080434	1.000000	0.043062	0.146436	0.112143	0.194325	0.280966
DEM	0.038087	0.043062	1.000000	0.064854	0.294161	0.182064	0.227297
FDP	0.132173	0.146436	0.064854	1.000000	0.062129	0.185373	0.261441
GOP	0.178738	0.112143	0.294161	0.062129	1.000000	0.144470	-0.049321
NSDAP	0.215895	0.194325	0.182064	0.185373	0.144470	1.000000	-0.001012
SPD	0.212094	0.280966	0.227297	0.261441	-0.049321	-0.001012	1.000000

Abbildung 22: Überarbeitete Cosine Similarity Matrix

Dadurch, dass keine Mittelwertbildung mehr erfolgt, sondern die Dokumentvektoren stattdessen für jede Partei einzeln erzeugt werden, ist nun eine deutlich bessere Trennschärfe zwischen den jeweiligen Parteien zu beobachten. Wie aus Abbildung 23 hervorgeht, besteht die höchste Ähnlichkeit zwischen der Demokratischen und der Republikanischen Partei der USA, was vermutlich an der ähnlichen Begriffsverwendung liegt (z.B. „dollar“, „amerika“, „amerikaner“, „biden“, „trump“ etc., siehe die beiden vorherigen Kapitel *BoW* und *Tf-Idf*). Vergleichbare Ähnlichkeitswerte teilt sich die SPD mit der CDU, der FDP und den Demokraten, dicht gefolgt von der Übereinstimmung zwischen AfD und NSDAP. Scheinbar besitzt die AfD auch eine vergleichsweise hohe Übereinstimmung mit der SPD, diese teilt aber wiederum einen der niedrigsten Werte mit der NSDAP. Ein Zirkelschluss scheint hier also wenig zielführend zu sein.

```

DEM - GOP: 0.2942
CDU - SPD: 0.2810
FDP - SPD: 0.2614
DEM - SPD: 0.2273
AfD - NSDAP: 0.2159
AfD - SPD: 0.2121
CDU - NSDAP: 0.1943
FDP - NSDAP: 0.1854
DEM - NSDAP: 0.1821
AfD - GOP: 0.1787
CDU - FDP: 0.1464
GOP - NSDAP: 0.1445
AfD - FDP: 0.1322
CDU - GOP: 0.1121
AfD - CDU: 0.0804
DEM - FDP: 0.0649
FDP - GOP: 0.0621
CDU - DEM: 0.0431
AfD - DEM: 0.0381
NSDAP - SPD: -0.0010
GOP - SPD: -0.0493

```

Abbildung 23: Überarbeitete Liste der Top-Ähnlichkeiten zwischen den Parteien in absteigender Reihenfolge

Bei Abbildung 23 fällt besonders auf, dass die SPD eine negative Cosine Similarity in ihrer Ähnlichkeit sowohl mit der NSDAP als auch der Republikanischen Partei der USA aufweist. Das liegt daran, dass der Winkel zwischen den beiden Parteien größer als 90 Grad ist, also im Multidimensionalen Vektorraum in entgegengesetzte Richtungen zeigt, woraus der negative Wert der Cosine Similarity resultiert. Besonders hohe Werte deuten hingegen darauf hin, dass die beiden Vektoren, welche die Parteien repräsentieren, in dieselben Richtung zeigen; bei einem Wert von >1 wäre die Richtung beider Vektoren also identisch. Wenn Vektoren aber in entgegengesetzte Richtungen zeigen, liegt das an den gegensätzlichen semantischen Mustern der beiden verglichenen Parteien.

Von den eingangs formulierten Hypothesen bezüglich der zu erwartenden Ähnlichkeitswerte der verschiedenen Parteien untereinander bestätigt sich nur eine einzige, nämlich die von der zu erwartenden besonders hohen Ähnlichkeit der AfD mit der NSDAP (Hypothese 2; fünfthöchster Wert von ~ 0.22). Eine besondere Nähe im „konservativen Spektrum“ der Parteien (also zwischen AfD und Republikanern, s. Hypothese 1) bleibt vollständig aus, ebenso wie eine hohe Übereinstimmung der Demokraten mit der CDU und FDP (Hypothese 3). Einzig und allein die Ähnlichkeit zwischen den Demokraten und der SPD ist eingetreten (kleiner Teil von Hypothese 3; vierthöchster Wert von ~ 0.23). Auch die letzte Hypothese (Hypothese 4) von den niedrigsten Übereinstimmungswerten zwischen Demokraten, SPD und FDP mit AfD

und NSDAP ist nur teilweise eingetreten. So teilt die SPD zwar, wie bereits erwähnt, einen negativen Wert mit der NSDAP (-0.001, der zweithöchste Wert überhaupt), jedoch teilt sie mit der AfD auch eine vergleichsweise eher hohe Ähnlichkeit (sechsthöchster Wert von ~0.21).

Diese teilweise sehr stark von den ursprünglichen Hypothesen abweichenden und damit doch sehr überraschenden Ergebnisse können unter anderem dadurch erklärt werden, dass sowohl *Word2Vec*, als auch dessen Erweiterung *Doc2Vec* keine kontextuellen Unterschiede in der Verwendung der Wörter berücksichtigen, da sie, im Gegensatz zu beispielsweise *Sentence Transformers*, nur ein globales Embedding pro Wort, aber keinen Kontext pro Verwendung berücksichtigen. So können Parteien, die grundsätzlich verschiedene Ansichten vertreten, aber über dieselben Themen sprechen, sehr ähnlich erscheinen. Darüber hinaus können auch dieselben Begriffe je nach Partei eine unterschiedliche Bedeutung haben. Beispielsweise nutzt die FDP den Begriff „Freiheit“ (im Sinne der Freiheit des Marktes) eventuell in einem anderen Sinn als die AfD („Freiheit“ des Volkes). *Word2Vec* erkennt diesen Unterschied jedoch nicht, da das Modell alle Begriffe kontextunabhängig voneinander betrachtet, was die potentiellen Ähnlichkeiten zwischen den Parteien deutlich erhöht.

Ein möglicher Lösungsansatz für dieses Problem wäre beispielsweise der Einsatz von *Sentence Transformers*. Diese beziehen auch den Kontext der jeweiligen Wörter in ihre Analyse mit ein, um daraus bedingte Ähnlichkeiten zu berechnen. Genau das soll im nachfolgenden Kapitel anhand dieses konkreten Anwendungsfalls genauer untersucht werden.

6.6 Sentence Transformers

6.6.1 Methodische Grundlagen & Einsatzgebiete

Sentence Transformers sind innovative Algorithmen, die umfassende semantische Darstellungen von Textdokumenten generieren und dadurch semantische Abfragen ermöglichen.²²⁶ Dabei sind Transformer-Modelle derzeit bei fast allen computerlinguistischen Aufgaben die leistungsfähigsten Modelle und stellen einen wesentlichen Durchbruch der letzten Jahre dar.²²⁷ Hierbei handelt es sich um einen neuen Typ neuronaler Netze, der von Vaswani et al.²²⁸ 2017 entwickelt wurde. Ihr

²²⁶ Galli, Carlo; Donos, Nikolaos und Calciolari, Elena: Performance of 4 Pre-Trained Sentence Transformer Models in the Semantic Query of a Systematic Review Dataset on Peri-Implantitis (<https://doi.org/10.3390/info15020068>, letzter Zugriff am 28.10.2025). Übersetzung des Autors.

²²⁷ Andresen: Computerlinguistische Methoden, S. 188-89. Gilt nachfolgend bis zum Ende des Absatzes.

²²⁸ Vgl. hierzu Vaswani, Ashish; Shazeer, Noam; Parmar, Niki; Uszkoreit, Jakob; Jones, Llion; Gomez, Aidan N.; Kaiser, Lukasz und Polosukhin, Illia: Attention is All you Need. In: *Advances in Neural*

Erfolg beruht dabei unter anderem auf einer deutlich erhöhten Effizienz in der Datenverarbeitung, da sie nicht mehr –wie bisher üblich– Sätze und Texte sequentiell, sondern den gesamten Input parallel verarbeiten können. Dadurch kann eine höhere Menga an Trainingsdaten in kürzerer Zeit verarbeitet werden, da sehr viele Berechnungen parallel durchgeführt werden. Dies bietet auch einen Vorteil für die Erkennung von Zusammenhängen zwischen weiter entfernten Elementen des Inputsatzes bzw. -textes. Obwohl die Größe des möglichen Inputs bei den Modellen aktuell noch begrenzt ist²²⁹, können neuere Modelle inzwischen auch immer längere Tokensequenzen parallel verarbeiten.

Zentral für diese Verarbeitung der Inputsequenz ist das Prinzip der *Attention*, welches davon ausgeht, dass für die Interpretation eines Wortes manche Wörter im Kontext wichtiger sind als andere.²³⁰ Diese Wörter werden entsprechend unterschiedlich gewichtet, sodass Wörter, die für die Interpretation wichtig sind, mathematisch höher gewichtet und weniger wichtige Wörter entsprechend niedriger gewichtet werden. Die optimale Gewichtung wird in einem neuronalen Netz durch das Training auf großen Datenmengen ermittelt. Durch dieses *Pretraining* ist das Training eines ganz neuen Sprachmodells in der Regel nicht mehr erforderlich, da zahlreiche dieser Modelle mittlerweile online für die kostenlose Nachnutzung zur Verfügung stehen.²³¹ Aus diesem Grund empfiehlt sich der Einsatz von Transformer-Modellen insbesondere auch in geisteswissenschaftlichen Anwendungsbereichen, die oftmals nicht auf besonders große Datenmengen zurückgreifen können.

In den *Digital Humanities* werden *Sentence Transformers* beispielsweise eingesetzt, um semantische Suchen über historische Korpora durchzuführen. Lucian Li präsentiert eine Methode, die es ermöglicht, intellektuelle Einflüsse in umfangreichen historischen Textkorpora durch den Einsatz von *Sentence-Embeddings* nachzuweisen.²³² Hierbei ermöglicht der Aufbau eines *Sentence-Embedding-Index* die effiziente Identifizierung semantisch ähnlicher Ideen, Paraphrasen und indirekter Einflüsse, selbst in Texten mit

Information Processing Systems, vol. 30 (2017), S. 1-11 (https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf, letzter Zugriff am 12.10.2025).

²²⁹ Bei den BERT-Modellen ist die Maximalgröße beispielsweise aktuell noch auf 512 Token beschränkt, sodass vor allem längere Texte noch nicht am Stück verarbeitet werden können. Vgl. hierzu Andresen: *Computerlinguistische Methoden*, S. 188-89.

²³⁰ Andresen: *Computerlinguistische Methoden*, S. 189-90. Gilt nachfolgend bis zum Ende des Absatzes.

²³¹ Dies gilt vor allem für die Plattform *Hugging Face*, der auch das Modell des Sentence-Transformers für die weitere Analyse in dieser Arbeit entstammt. Vgl. hierzu <https://huggingface.co/sentence-transformers>, letzter Zugriff am 11.10.2025.

²³² Vgl. hierzu Li, Lucian: *Tracing the Genealogies of Ideas with Sentence Embeddings*. In: *Proceedings of the 4th International Conference on Natural Language Processing for Digital Humanities*, 2024, S. 9–16 (<https://aclanthology.org/2024.nlp4dh-1.2.pdf>, letzter Zugriff am 28.10.2025). Übersetzung des Autors; gilt nachfolgend bis zum Ende des Absatzes.

hohen OCR-Fehlerraten. Die Methode wurde auf einen Korpus von 250.000 Sachtexten des 19. Jahrhunderts angewendet und die Ergebnisse zeigen eine hohe Übereinstimmung mit etablierten historischen Erkenntnissen. Dieser Ansatz ermöglicht eine erweiterte und inklusivere Analyse der Verbreitung von Ideen, die über traditionelle Kanons und bekannte Persönlichkeiten hinausgeht und auch marginalisierte Gruppen berücksichtigt.

Einen weiteren, sehr praxisnahen Anwendungsfall bietet die Studie von Fischer et al., welche mit der sogenannten „Concept Over Time Analysis“ ein Tool entwickelt haben, das in der *Discourse Analysis Tool Suite* integriert ist.²³³ Es ermöglicht Forschenden, Konzepte interaktiv zu definieren, zu klassifizieren und zeitlich zu visualisieren. Der Prozess der Berechnung semantischer Ähnlichkeiten erfolgt dabei unter Zuhilfenahme von *Sentence Transformers* und *Few-Shot Sentence Classification*. Die Optimierung dieser Ähnlichkeiten wird durch Benutzerfeedback und iterative Feinabstimmung gewährleistet. Das Endergebnis ist eine Zeitleistenvisualisierung, die demonstriert, wie spezifische Konzepte im Diskurs über die Zeit hinweg in Erscheinung treten und eine Transformation durchlaufen, was sich insbesondere für qualitative Analysen im Rahmen der *Digital Humanities* als vorteilhaft erweist.

Ein weiteres Anwendungsbeispiel bieten Baer und Purves, die demonstrieren, wie *Sentence Transformers* dazu genutzt werden können, um relevante Dokumente aus umfangreichen Sammlungen, wie beispielsweise „Geograph“ oder „WikiHow“, zu extrahieren.²³⁴ Sie präsentieren hierbei einen Workflow zur Identifikation landschaftsbeschreibender Texte in großen, unstrukturierten Korpora unter Verwendung von *Sentence Transformers* in Kombination mit der bereits beschriebenen *Cosine Similarity*. Den Ausgangspunkt der Untersuchung bildet eine kleine, qualitativ hochwertige Sammlung aktiv crowd-gesourcter Landschaftsbeschreibungen. Anhand dieser Sammlung werden dann anschließend ähnliche Texte in „Geograph“- und „WikiHow“-Korpora gesucht. Baer und Purves gelangen zu dem Schluss, dass der hier beschriebene Ansatz herkömmliche *TF-IDF*-Methoden übertrifft, da er thematisch zusammenhängende Dokumente identifiziert und somit die Erstellung neuer, domänenspezifischer Korpora erleichtert. Darüber hinaus kann diese Methode ebenso auch in anderen wissenschaftlichen Disziplinen

²³³ Vgl. hierzu Fischer, Tim; Schneider, Florian; Geislinger, Robert; Helfer, Florian; Koch, Gertraud und Biemann, Chris: Concept Over Time Analysis: Unveiling Temporal Patterns for Qualitative Data Analysis. In: *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Volume 3: System Demonstrations (2024), S. 148–157 (<https://aclanthology.org/2024.naacl-demo.15/>, letzter Zugriff am 28.10.2025). Übersetzung des Autors; gilt nachfolgend bis zum Ende des Absatzes.

²³⁴ Vgl. hierzu Baer, Manuel F. und Purves, Ross S.: Identifying Landscape Relevant Natural Language using Actively Crowdsourced Landscape Descriptions and Sentence-Transformers. In: *KI – Künstliche Intelligenz*, Volume 37 (2023), S. 55–67 (<https://link.springer.com/article/10.1007/s13218-022-00793-3>, letzter Zugriff am 28.10.2025). Übersetzung des Autors; gilt nachfolgend bis zum Ende des Absatzes.

eingesetzt werden, die mit Korpora in natürlicher Sprache als zugrunde liegenden Daten arbeiten.

6.6.2 Implementierung

Für die Implementierung in Python²³⁵ wird ein vortrainiertes *Sentence Transformers* Modell²³⁶ der Open-Source Plattform *Hugging Face* verwendet, welches in Python mit der *Sentence Transformers Library* genutzt werden kann. Anschließend werden die nicht-vorverarbeiteten, kombinierten Korpora für alle Parteien eingelesen. Dies ist insofern entscheidend, da die *Sentence Transformers* auch ganze Sätze und deren Kontext analysieren und nicht bloß einzelne Wörter, wie dies bei den meisten der vorangegangenen Methoden der Fall war. Als nächster Schritt werden daraufhin die Embeddings, also die kodierten Sätze, für jedes Parteikorpus generiert. Für diese Embeddings wird dann wiederum die *Cosine Similarity* berechnet und das Ergebnis in Form einer Matrix mit den Ähnlichkeitswerten für jede Partei ausgegeben. Zusätzlich werden diese Ergebnisse auch noch in Form einer Heatmap visualisiert, um mit Hilfe verschiedener Farben die Ähnlichkeitswerte dieser Matrix visuell ansprechender zu gestalten. Darüber hinaus wird auch noch eine *PCA* aller Parteien durchgeführt, wobei die insgesamt 768 Dimensionen der Embeddings auf zwei Dimensionen reduziert werden, um diese auf einer entsprechenden Grafik ebenfalls darstellen zu können. Diese Ergebnisse sollen nun im folgenden Kapitel dargestellt und weiter ausgeführt werden.

6.6.3 Analyse der Ergebnisse & Überprüfung der Hypothesen

Die Ergebnisse der auf den *Sentence Transformers* basierenden *Cosine Similarity* zwischen den verschiedenen Parteien werden in der folgenden Ähnlichkeitsmatrix dargestellt (s. Abbildung 24).

²³⁵ Hierbei handelt es sich um das Skript „[sentence transformers.py](#)“ (s. GitHub), welches mit Hilfe von ChatGPT (<https://chatgpt.com/>) erstellt wurde.

²³⁶ Hierbei handelt es sich um das Modell „[jina-embeddings-v2-base-de](#)“. Siehe hierzu auch Hugging Face: Sentence Transformers (<https://huggingface.co/sentence-transformers>, letzter Zugriff am 11.10.2025).

Cosine Similarities zwischen allen Parteien

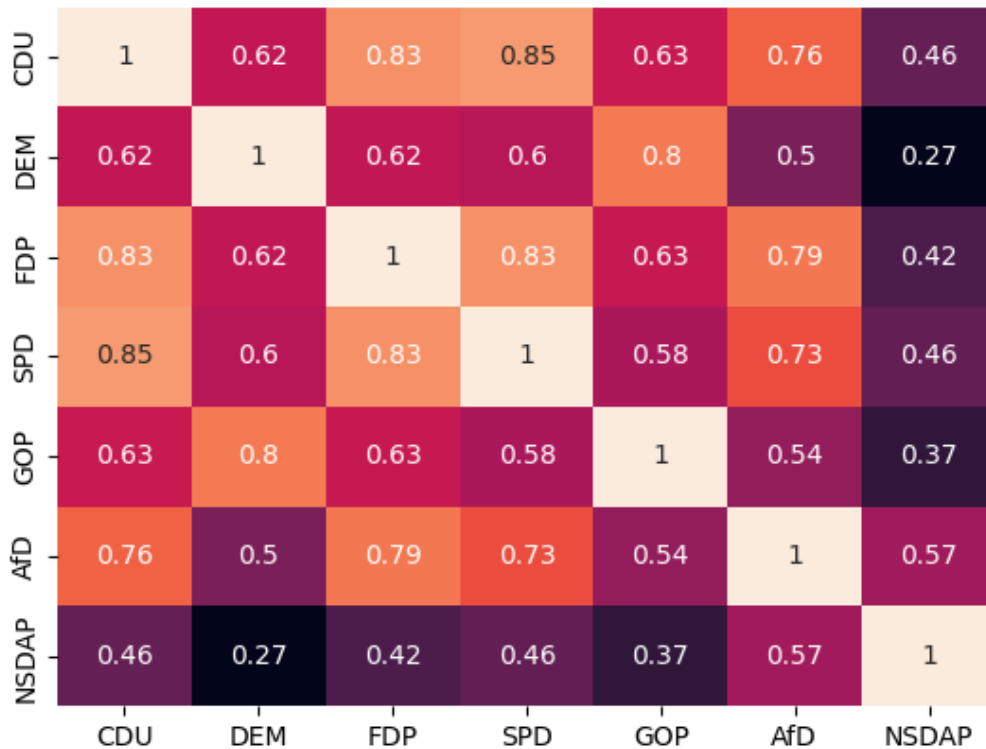


Abbildung 24: Heatmap der Cosine Similarities zwischen allen Parteien, basierend auf Sentence Transformers

Abbildung 24 zeigt eine Heatmap der Ergebnisse der *Cosine Similarities* zwischen den verschiedenen Parteien. Hierbei wird eine sehr hohe Ähnlichkeit mit sehr hellen Farbtönen widergespiegelt, welche mit abnehmenden Ähnlichkeitswerten immer dunkler werden, sodass die maximale Übereinstimmung von $\langle 1 \rangle$ = weiß und $\langle 0 \rangle$ = schwarz entsteht. Die diagonale Reihe mit dem maximalen Wert von $\langle 1 \rangle$ entsteht durch den Vergleich jeder Partei mit sich selbst.

Besonders auffällig hierbei sind die hohen Ähnlichkeitswerte der AfD mit FDP (0.79) und CDU (0.76), die –entgegen der zuvor aufgestellten Hypothese– höher als diejenigen der AfD zur NSDAP (0.57) oder gar den Republikanern (0.54) sind. Die CDU besitzt erwartungsgemäß eine hohe Ähnlichkeit mit der SPD (0.85) und FDP (0.83), teilt sich aber einen vergleichsweise eher niedrigen Wert mit den Demokraten (0.62). Wie auch schon zuvor in den anderen Methoden, so teilen sich die Demokraten auch hier ihren höchsten Wert (0.8) mit den Republikanern. Die möglichen Gründe hierfür sollen jedoch erst im nachfolgenden Kapitel ausführlich erörtert und diskutiert werden. Die niedrigsten Werte teilen sich die Demokraten mit der NSDAP (0.27) und

AfD (0.5). CDU, FDP und SPD besitzen eine ähnliche, etwas höhere Übereinstimmung (~0.6).

Die FDP besitzt die höchste Ähnlichkeit mit CDU und SPD (jeweils 0.83), dicht gefolgt von der AfD (0.79), und die niedrigste mit der NSDAP (0.42). Die Republikaner besitzen, wie bereits erwähnt, die höchste Ähnlichkeit mit den Demokraten (0.8), während sie von den deutschen Parteien der CDU und FDP am ähnlichsten sind (jeweils 0.63). Mit der AfD lässt sich also, anders als erwartet, keine besonders hohe Übereinstimmung feststellen (0.54). Auffällig ist auch, dass die Republikaner die geringste Ähnlichkeit zur NSDAP besitzen (0.37), was sich vermutlich mit der „altdeutschen“ Sprache des NSDAP-Korpus erklären lässt, welches dem maschinell übersetzten Englisch der beiden US-amerikanischen Parteien besonders unähnlich sein muss. Denn auch die NSDAP teilt sich im Umkehrschluss ihre niedrigsten Übereinstimmungswerte mit den beiden US-amerikanischen Parteien (0.27 & 0.37), während sie der AfD (0.57) am ähnlichsten ist. Die SPD wiederum besitzt die größte Ähnlichkeit zu CDU (0.85) und FDP (0.83), gefolgt von der AfD (0.73). Die niedrigste Übereinstimmung teilt die SPD mit der NSDAP (0.46).

Wählt man nun als weiteres Distanzmaß die *PCA* und reduziert die 768 Dimensionen der Embeddings auf einen zweidimensionalen Raum, so entsteht das folgende Schaubild (s. Abbildung 25).

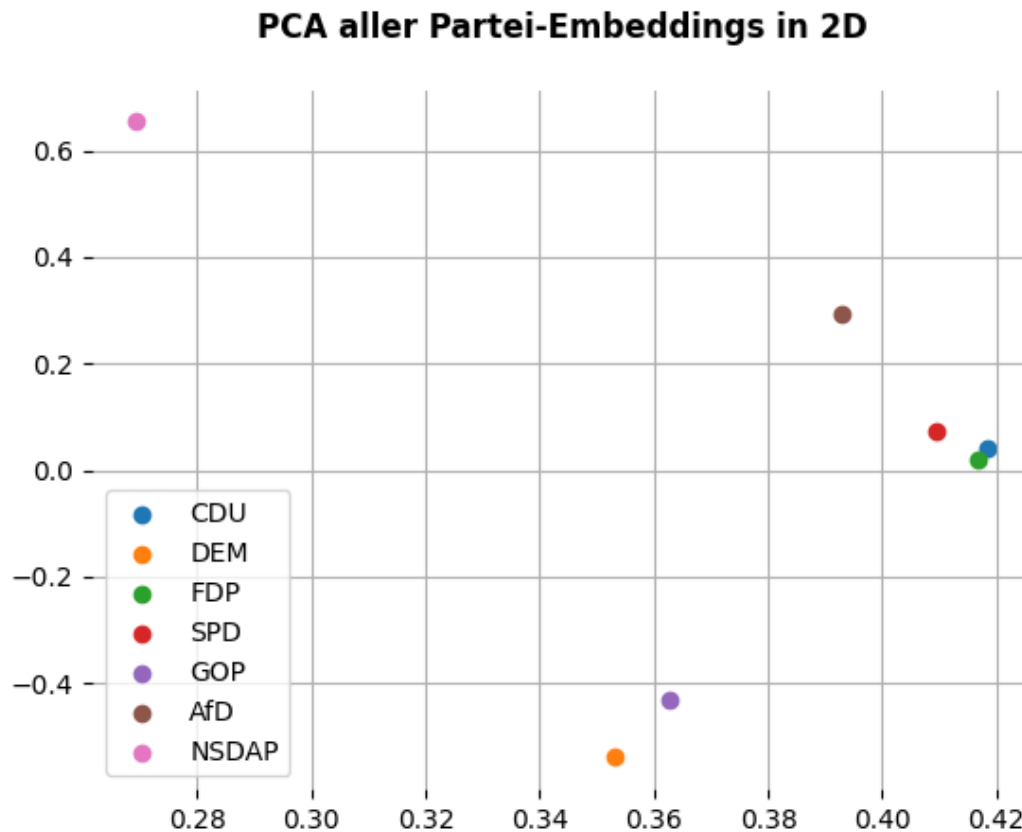


Abbildung 25: Ergebnisse der PCA aller Partei-Embeddings in 2D, basierend auf Sentence Transformers

Aus dieser Visualisierung geht hervor, dass sich die deutschen Parteien, mit Ausnahme der NSDAP, alle auf einer ähnlichen semantischen Ebene bewegen, während die beiden US-amerikanischen Parteien ein weit abgetrenntes, eigenes semantisches Cluster bilden. Auch dies deckt sich größtenteils mit den bisherigen Beobachtungen aus den vergangenen methodischen Analysen dieser Arbeit. Doch auch die jeweiligen Unterschiede zwischen den einzelnen deutschen Parteien werden hier nochmals etwas deutlicher sichtbar. Ähnlich wie bei der Visualisierung der *Cosine Similarities* in Abbildung 24 bildet sich das engste Cluster aus CDU, FDP und SPD. Die AfD ist den etablierten deutschen Parteien am ähnlichsten, gleichzeitig aber auch die Partei mit der größten Nähe zur NSDAP. Insofern entsprechen diese Ergebnisse der *PCA* also größtenteils der zuvor erfolgten Visualisierung der *Cosine Similarities* und spiegeln diese wider.

Zu den verschiedenen Hypothesen lässt sich in Bezug auf die *Sentence Transformers* zunächst festhalten, dass fast alle der insgesamt vier Hypothesen erfüllt wurden. Es liegt zwar keine hohe Übereinstimmung zwischen der AfD und den Republikanern vor (Hypothese 1), da sie den zweitniedrigsten Wert von 0.54 mit der AfD teilen und nur

die Demokraten eine noch geringere Ähnlichkeit von 0.5 zur AfD aufweisen. Und auch Hypothese 2 mag auf den ersten Blick nicht zutreffen, da keine besonders hohe Übereinstimmung zwischen AfD & NSDAP vorliegt (drittniedrigster Wert der AfD von 0.57). Betrachtet man jedoch die andere Seite, so ist die AfD von allen Parteien diejenige mit der höchsten Ähnlichkeit zur NSDAP, weshalb Hypothese 2 zumindest teilweise als zutreffend gewertet werden kann. Hypothese 3, welche von einer hohen Übereinstimmung von CDU, SPD & FDP mit der Demokratischen Partei der USA ausgeht, trifft ebenfalls zu, da die Ähnlichkeitswerte in diesem Fall von 0.6 bis 0.62 reichen. Die vierte Hypothese, welche von der niedrigsten Übereinstimmung von Demokraten, SPD & FDP mit AfD & NSDAP ausgeht, trifft nur teilweise zu, da die Demokraten zwar zu AfD & NSDAP die niedrigste Ähnlichkeit aufweisen, SPD & FDP jedoch nur mit der NSDAP die niedrigsten Übereinstimmungswerte teilen. Beide Parteien sind der AfD deutlich ähnlicher als den Demokraten, was aber höchstwahrscheinlich der gemeinsamen Sprache und dem politischen Fokus auf Deutschland geschuldet ist.

Im nachfolgenden Kapitel sollen nun die gewonnenen Erkenntnisse noch einmal zusammenfassend dargestellt sowie kritisch über die verwendeten digitalen Methoden reflektiert werden, um auf diese Weise einen übergeordneten Methodenvergleich zu ermöglichen.

7 Fazit: Vergleich der Ergebnisse & Reflexion der verwendeten Methoden

In diesem Kapitel sollen nun die Ergebnisse der verschiedenen, in dieser Arbeit vorgestellten und implementierten digitalen Methoden miteinander verglichen werden. Dies geschieht insbesondere auf der Grundlage der zuvor aufgestellten vier Hypothesen bezüglich der Ähnlichkeit der verschiedenen Parteien untereinander. Hypothese 1 ging von einer hohen Übereinstimmung zwischen AfD & Republikanern, basierend auf den ähnlichen innen- und außenpolitischen Zielsetzungen der letzten Jahre, aus, wohingegen Hypothese 2 eine besonders hohe Übereinstimmung zwischen AfD & NSDAP prognostiziert. Hypothese 3 geht wiederum von einer hohen Übereinstimmung von CDU, SPD & FDP mit der Demokratischen Partei der USA aus, während Hypothese 4 die niedrigste Übereinstimmung von Demokraten, SPD & FDP mit AfD & NSDAP vorhersagt. In der folgenden Tabelle soll daher noch einmal festgehalten und dokumentiert werden, welche der vier genannten Hypothesen mit welchen der verschiedenen Methoden verifiziert werden konnten (s. Tabelle 4).

Tabelle 4: Übersicht der überprüften Hypothesen anhand der verwendeten Methoden

Methoden	Hypothese 1	Hypothese 2	Hypothese 3	Hypothese 4
Bag-of-Words	X	X	X	X
Term Frequency– Inverse Document Frequency	X	X	X	X
Stilometrische Analyse	X	✓	✓ / X	X
Wordfish for Ideological Scaling	X	X	X	✓ / X
Word2Vec (Doc2Vec)	X	✓	X	✓ / X
Sentence Transformers	X	✓ / X	✓	✓ / X

Aus dieser tabellarischen Übersicht wird klar ersichtlich, dass die aufgestellten Hypothesen in lediglich 3 von insgesamt 24 Fällen uneingeschränkt zutreffen, was einem Gesamtanteil von 12,5% entspricht. In 5 von 24 Fällen treffen die Hypothesen teilweise zu, was fast 21% der Fälle ausmacht. In zwei Drittel der Fälle aber treten die aufgestellten Hypothesen nicht ein (16/24). Diese Ergebnisse wurden in der folgenden Grafik für eine bessere Anschaulichkeit nochmals visualisiert (s. Abbildung 26).

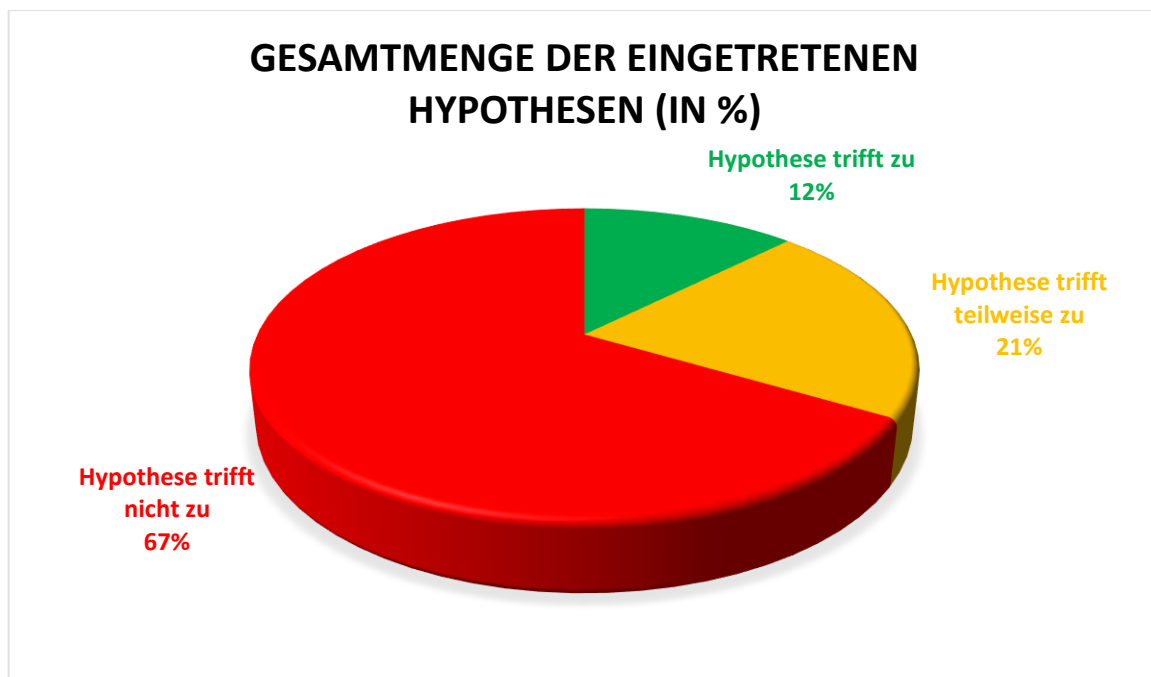


Abbildung 26: Übersicht über die Gesamtmenge der eingetretenen Hypothesen (in %)

Um besser nachvollziehen zu können, warum die aufgestellten Hypothesen lediglich in so wenigen Fällen, gemessen an der Gesamtanzahl, zutreffen, erscheint es sinnvoll, zunächst eine Übersicht der größten und geringsten Ähnlichkeiten zwischen den jeweiligen Parteien pro Methode zu erstellen. Diese Top-Ähnlichkeiten der verschiedenen Parteien sind methodenweise in der folgenden Tabelle noch einmal zusammengefasst (s. Tabelle 5).

Tabelle 5: Übersicht der größten und geringsten Ähnlichkeiten zwischen den Parteien pro Methode, (falls vorhanden) basierend auf *Cosine Similarity*

Methoden	1.	2.	3.	Geringste Ähnlichkeit
Bag-of-Words	CDU – SPD	CDU – FDP	AfD – NSDAP bzw. FDP – SPD	DEM – CDU/SPD bzw. NSDAP – GOP
Term Frequency–Inverse Document Frequency	DEM – GOP	CDU – FDP	AfD – CDU/FDP	SPD – AfD/DEM/GOP
Stilometrische Analyse	SPD – FDP	CDU – SPD bzw.	AfD – DEM	GOP – SPD/FDP

		CDU – FDP		
Wordfish for Ideological Scaling	CDU – SPD	SPD – FDP	CDU – FDP	NSDAP – AfD
Word2Vec (Doc2Vec)	DEM – GOP	CDU – SPD	FDP – SPD	GOP – SPD
Sentence Transformers	CDU – SPD	CDU – FDP	DEM – GOP	DEM – NSDAP

Wie aus dieser Übersicht ersichtlich wird, lässt sich also erst einmal festhalten, dass die Ergebnisse bezüglich der Ähnlichkeit der verschiedenen Parteien untereinander, abhängig von der verwendeten Methode, teilweise sehr stark variieren können. Nichtsdestotrotz ist aber dennoch ein gewisser „roter Faden“ vorhanden, der alle Methoden miteinander verbindet. Dieser rote Faden besteht vorwiegend aus der überaus großen Ähnlichkeit der drei etablierten deutschen Parteien CDU, SPD und FDP untereinander, welche durchweg bei allen Methoden nachgewiesen werden kann. Darüber hinaus ist auch die große Ähnlichkeit der Demokraten und Republikaner zueinander evident, da sie immerhin bei der Hälfte aller verwendeten Methoden die Top 3 der höchsten Ähnlichkeitswerte erreicht haben. Die geringste Übereinstimmung erzielte am häufigsten die SPD mit der Republikanischen Partei der USA (in der Hälfte aller Fälle), dicht gefolgt von der Demokratischen Partei (2/6).

Zusätzlich soll in diesem Kapitel aber ebenso auch über die Eignung der verschiedenen, in dieser Arbeit thematisierten und vorgestellten digitalen Methoden kritisch reflektiert werden. Dies geschieht auf der Grundlage des gemeinsamen Anwendungsfalles, nämlich dem Vergleich dieser digitalen Methoden der Textanalyse zur Identifizierung rechtspopulistischer Sprache in politischen Texten. Wie in der Einleitung bereits ausgeführt, wurde im speziellen Anwendungsfall dieser Arbeit rechtspopulistische Sprache durch eine möglichst große semantische Nähe einer Partei zum politischen Korpus der NSDAP definiert. Im Folgenden wird daher in einer tabellarischen Übersicht dargestellt, bei welcher Methode welche Partei die größte Nähe zur NSDAP aufweist (s. Tabelle 6).

Tabelle 6: Übersicht der Parteien mit den höchsten Ähnlichkeitswerten zur NSDAP pro Methode, (falls vorhanden) basierend auf *Cosine Similarity*

Methoden	1.	2.	3.
Bag-of-Words	AfD	CDU/SPD	GOP
Term Frequency–Inverse Document Frequency	AfD/CDU/FDP	–	–

Stilometrische Analyse	GOP	AfD/DEM	CDU
Wordfish for Ideological Scaling	GOP	DEM	CDU
Word2Vec (Doc2Vec)	AfD	CDU	FDP
Sentence Transformers	AfD	CDU/SPD	FDP

Aus dieser Tabelle geht hervor, dass die AfD die Partei mit der höchsten Ähnlichkeit zur NSDAP ist, da sie bei 4 von 6 Methoden die höchsten Übereinstimmungswerte zur NSDAP aufweist. Darüber hinaus wird aus dieser Übersicht deutlich, dass die CDU bei allen Methoden durchschnittlich mindestens den zweiten bis dritten Platz belegt. Auch die Republikaner liegen, zumindest bei der Stilometrischen Analyse und Wordfish, jeweils auf dem ersten Platz.

BoW und *Tf-Idf* scheinen vor allem dazu geeignet zu sein, einen ersten, groben Überblick über die zentralen Themen zu erlangen, welche in einem bestimmten Korpus vertreten sind. Zur Identifizierung rechtspopulistischer Sprache sind sie daher nur geeignet, falls ein wörterbuchbasierter Ansatz mit einer zuvor festgelegten Stoppwortliste rechtspopulistischer Begriffe durchgeführt werden soll. Dies liegt auch daran, dass sie mit reinen Wortfrequenzen arbeiten und diese im Sinne der *mfw* zählen, ohne dabei jedoch mit Ähnlichkeits- bzw. Distanzmaßen, wie beispielsweise der *Cosine Similarity* oder *PCA* zu arbeiten.

Die Stilometrische Analyse hat sich in diesem konkreten Anwendungsfall als sehr zuverlässige Methode erwiesen, um semantische Ähnlichkeiten in politischen Texten nachzuweisen. Dies gilt besonders für die Analyse der unverarbeiteten Parteienkorpora, wo eine große Nähe zwischen AfD, NSDAP und der Republikanischen Partei der USA nachgewiesen werden konnte. Führt man dieselbe Analyse jedoch mit den bereinigten Parteienkorpora durch, fällt eine starke Trennung zwischen den deutschen und amerikanischen Parteien auf, welche vorwiegend auf die unterschiedlichen, landesspezifischen behandelten Themen zurückzuführen ist. Jedoch ist auch hier die NSDAP der AfD am nächsten.

Das von Proksch et al. entwickelte Modell *Wordfish for Ideological Scaling* weist jedem Begriff eines Korpus einen konkreten Wert zu (*Estimated Feature Scores*), wodurch es sich besonders zur Analyse und Klassifizierung politischer Texte anbietet. Im Fall von Wordfish ist jedoch besonders aufgefallen, dass sich das Modell aufgrund der Gewichtung der verwendeten Begriffe auch sehr leicht in die Irre führen lässt. So war es beispielsweise bei der AfD der Fall, dass diese am „linken“ Rand der Skala verortet wurde, sogar deutlich linker als die SPD und am entgegengesetzten Ende der Skala zu NSDAP und Republikanern (s. Abbildung 18). Dies stellt eine Einordnung im Kontext der Identifizierung rechtspopulistischer Sprache insofern vor Herausforderungen, als dass die Verwendung bestimmter Begriffe die Gesamtwertung

eines Textes sehr stark beeinflusst, es aber nicht ersichtlich ist, wie sich die bestimmten Werte für die verwendeten Begriffe eines Textes explizit zusammensetzen oder wie genau diese berechnet werden, was den Anschein einer gewissen Willkür erweckt.

Word2Vec wandelt jedes Wort eines Textes bzw. Korpus in einen Vektor um und stellt diese Vektoren im multidimensionalen Vektorraum dar. Hierbei stellte sich zunächst das Problem, dass die Werte zur Übereinstimmung zwischen den verschiedenen Parteien extrem hoch waren, was einem fast identischen Inhalt entsprechen würde. Dies lag mitunter daran, dass die *Word2Vec*-Dokumentvektoren zunächst über eine einfache Mittelung aller Wortvektoren gebildet wurden und das *Word2Vec*-Modell über alle Parteien gemeinsam trainiert wurde, was dazu führte, dass die Wortvektoren in einem gemeinsamen semantischen Raum liegen. Dieses Problem konnte in diesem Fall dadurch behoben werden, ein weiteres Skript anzufertigen, welches an Stelle einer einfachen Mittelung aller Wortvektoren für jedes Parteikorpus eigene *Doc2Vec*-Dokumentvektoren erzeugt, was dazu führte, dass nun eine deutlich bessere Trennschärfe zwischen den jeweiligen Parteien zu beobachten war. Dadurch stellte sich *Doc2Vec*, als Weiterentwicklung zu *Word2Vec*, als hervorragend geeignete Methode für den Anwendungsfall dieser Arbeit heraus.

Die in dieser Arbeit verwendeten *Sentence Transformers* führten zu dem Ergebnis, dass unter anderem eine besonders hohe Ähnlichkeit zwischen Demokraten und Republikanern, aber auch den etablierten deutschen Parteien untereinander nachgewiesen werden konnte. Die NSDAP liegt deutlich abseits dieser beiden großen Hauptcluster (s. Abbildung 25), teilt sich jedoch ihre höchste Übereinstimmung mit der AfD (0.57; s. Abbildung 24). Hieraus wird klar ersichtlich, dass sich *Sentence Transformers* nicht nur für die Bestimmung semantischer Ähnlichkeiten verschiedener Parteien, sondern darüber hinaus auch sehr gut und zuverlässig für den Anwendungsfall dieser Arbeit, nämlich die Identifizierung rechtspopulistischer Sprache in politischen Texten, eignen.

Die vorliegenden Ergebnisse zeigen demnach, dass jede der verwendeten Methoden unterschiedliche, ganz eigene Stärken und Schwächen besitzt. Gemessen an der Anzahl der zutreffenden Hypothesen haben sich jedoch hauptsächlich die *Sentence Transformers*, die *Stilometrische Analyse* und *Word2Vec*, sowie in Teilen auch *Wordfish for Ideological Scaling*, als die erfolgreichsten der verwendeten Methoden herausgestellt. Auf den konkreten Anwendungsfall dieser Arbeit bezogen, nämlich die Identifizierung rechtspopulistischer Sprache in politischen Texten, bilden sich bei diesen vier Methoden ganz deutlich zwei verschiedene Lager heraus, welche untereinander nahezu identische Ergebnisse in Bezug auf die semantische Nähe zur NSDAP vorweisen. Nämlich einerseits die beiden in *R* implementierten Methoden *Stilometrische Analyse* und *Wordfish*, welche beide jeweils Republikaner, Demokraten

und CDU am ähnlichsten zur NSDAP verorten, und andererseits die *Sentence Transformers* und *Word2Vec*, welche übereinstimmend AfD, CDU und FDP als die drei Parteien mit den höchsten Ähnlichkeitswerten zur NSDAP ausweisen.

Dementsprechend bedeutet dies, dass –abhängig von der verwendeten Methode– sowohl der AfD als auch der Republikanischen Partei der USA gemäß der These dieser Arbeit die Verwendung rechtspopulistischer Sprache nachgewiesen werden konnte. Dies bedeutet jedoch nicht, dass es sich bei diesen beiden Parteien damit automatisch auch um rechtspopulistische Parteien handelt, da diese Ergebnisse lediglich auf einer stichprobenartig ausgewählten Sammlung politischer Texte basieren, welche nicht als repräsentativ für die gesamte Partei und all ihre Mitglieder angesehen werden können. Dennoch untermauert dieses Ergebnis sehr eindrücklich die zuvor bereits in der Einleitung formulierten Beobachtungen bezüglich der aktuellen politischen Entwicklungen und unterstreicht zugleich die Wichtigkeit einer wehrhaften Demokratie als solidem Gegengewicht zu erstarkenden rechtspopulistischen Strömungen in Deutschland und Amerika.

8 Ausblick

In Zukunft könnte eine ähnliche Analyse dazu eingesetzt werden, um Prognosen für die Ergebnisse von möglichen Koalitionsverhandlungen nach der nächsten Bundestagswahl aufzustellen. Denn um erfolgreiche (Regierungs-) Koalitionen zu bilden, ist zunächst eine möglichst große Schnittmenge am Inhalt der politischen Themen und Zielsetzungen notwendig. Genau dies wurde in der vorliegenden Arbeit untersucht, weshalb sich ein solches Vorgehen auf dem Gebiet der Politikwissenschaft und -forschung durchaus anbieten würde.

Die semantische Plattform „Semantha“²³⁷ hat das bereits in die Tat umgesetzt. Bei dieser Internetplattform handelt es sich um eine Kooperation des *KI Bundesverbandes e.V.* mit den Unternehmen *Aleph Alpha* und *thingsTHINKING*. Ziel ist hierbei, die Parteiprogramme verschiedener, im Deutschen Bundestag verteilter Parteien semantisch zu analysieren und dabei die jeweiligen Überschneidungen und Differenzen aufzuzeigen. Darüber hinaus wurden dabei beispielhaft aus den Ergebnissen bereits mögliche Formulierungen für einen potenziellen Koalitionsvertrag generiert. *Semantha* beziehungsweise der KI Bundesverband verfolgen also kommerzielle Ansätze zur semantischen Analyse von Parteiprogrammen, inklusive der Analyse von Überschneidungen. Jedoch betreibt *Semantha* diese Ansätze nur in reiner Eigenentwicklung und es sind weder wissenschaftliche Publikationen vorhanden, welche auf die technischen und methodischen Vorgänge hindeuten könnten, welche im Hintergrund ablaufen, noch wird auf solche an irgendeiner Stelle verwiesen. Dadurch werden diese Hintergrundprozesse sehr undurchsichtig und nur schwer nachvollziehbar, was wiederum der Glaubwürdigkeit der Qualität der Plattform schadet.

Dennoch könnte eventuell gerade eine Kombination dieser beiden Ansätze, nämlich einerseits dem vollkommen kommerzialisierten Ansatz von *Semantha*, kombiniert mit ausreichend Transparenz in der wissenschaftlichen Forschung und Offenlegung der, vor allem in den *Digital Humanities* sehr weit verbreiteten und gern gesehenen *FAIR*-Prinzipien, also der Auffindbarkeit (*Findability*), Zugänglichkeit (*Accessibility*), Austauschbarkeit (*Interoperability*) und Wiederverwendbarkeit (*Reusability*) von Informationen, in diesem Fall für zukünftige Projekte den entscheidenden Ausschlag geben, um das Forschungsfeld der „klassischen“ Politikwissenschaft durch den Einsatz und die Verwendung digitaler Methoden weiter auszubauen und gewinnbringend zu ergänzen, indem die dadurch gewonnen Erkenntnisse nicht nur einzelnen Unternehmen in Form von kommerziellem Profit, sondern auch einer offen

²³⁷ Vgl. hierzu Semantha (<https://www.semantha.de/de/parteiprogramme/>, letzter Zugriff am 07.08.2025).

geteilten Forschung und letzten Endes damit auch der interessierten, breiten Öffentlichkeit zugänglich gemacht werden können.

Quellen- & Literaturverzeichnis

Internetquellen

- [1] Allison, Sarah; Heuser, Ryan; Jockers, Matthew; Moretti, Franco; Witmore, Michael: Quantitative Formalism: An Experiment. In: *Pamphlets of the Stanford Literary Lab* (Pamphlet 1). Stanford Literary Lab 2011 (<https://d-nb.info/1164077457/34>, letzter Zugriff am 13.10.2025).

- [2] Analyse der Wahlprogramme: Was sind die häufigsten Substantive pro Partei? Ein Vergleich der relativen Worthäufigkeiten in den Wahlprogrammen zur BTW21 (https://www.reddit.com/r/de/comments/pujwxr/analyse_der_wahlprogramme_was_sind_die_h%C3%A4ufigsten/#lightbox, letzter Zugriff am 27.09.2025).

- [3] Aydogan, Abdullah; Tuna, Tayfun und Yildirim, A. Kadir: Ideological Congruence and Social Media Text as Data. In: *Journal of Representative Democracy*, Volume 55, Issue 2/2019, S. 159-178 (<https://doi.org/10.1080/00344893.2019.1592012>, letzter Zugriff am 18.10.2025).

- [4] Baer, Manuel F. und Purves, Ross S.: Identifying Landscape Relevant Natural Language using Actively Crowdsourced Landscape Descriptions and Sentence-Transformers. In: *KI – Künstliche Intelligenz*, Volume 37 (2023), S. 55–67 (<https://link.springer.com/article/10.1007/s13218-022-00793-3>, letzter Zugriff am 28.10.2025).

- [5] Brettschneider, Frank und Thoms, Claudia: Die Wahlprogramme zur Bundestagswahl 2021 im Vergleich aller Bundestagswahlen seit 1949. Eine Studie der Universität Hohenheim (https://www.uni-hohenheim.de/uploads/media/Wahlprogramm-Check_2021_Bundestagswahl.pdf, letzter Zugriff am 10.08.2025).

- [6] Bundeszentrale für Politische Bildung (bpb, <https://www.bpb.de/shop/zeitschriften/apuz/279819/jenseits-von-links-und-rechts/>, letzter Zugriff am 17.09.2025).
- [7] Byszuk, Joanna: The Voices of Doctor Who – How Stylometry Can be Useful in Revealing New Information about TV Series. In: *Digital Humanities Quarterly*, Vol. 14, Nr. 4 (2020), Absatz 28 (<https://dhq.digitalhumanities.org/vol/14/4/000499/000499.html>, letzter Zugriff am 24.10.2025).
- [8] Ceron, Andrea: Changing politics, changing language: The effect of institutional and communicative changes on political language measured through content analysis of Italian intra-party debates. In: *Journal of Language and Politics*, Band 14, Ausgabe 4 (2015), S. 528-551 (<https://air.unimi.it/bitstream/2434/365263/2/changinglanguage.pdf>, letzter Zugriff am 18.10.2025).
- [9] Ceron, Tanise; Blokker, Nico und Padó, Sebastian: Optimizing text representations to capture (dis)similarity between political parties (<https://arxiv.org/abs/2210.11989>, letzter Zugriff am 10.08.2025).
- [10] Ceron, Tanise; Nikolaev, Dmitry und Padó, Sebastian: Additive manifesto decomposition: A policy domain aware method for understanding party positioning (<https://arxiv.org/abs/2305.10136>, letzter Zugriff am 13.08.2025).
- [11] Decker, Frank: Jenseits von links und rechts. Lassen sich Parteien noch klassifizieren? (<https://www.bpb.de/shop/zeitschriften/apuz/279819/jenseits-von-links-und-rechts/>, letzter Zugriff am 17.09.2025).
- [12] Digital Humanities Workbench (https://www2.fgw.vu.nl/werkbanken/dighum/data_analysis/text_analysis/stylometry.php, letzter Zugriff am 24.10.2025).
- [13] Fischer, Tim; Schneider, Florian; Geislinger, Robert; Helfer, Florian; Koch, Gertraud und Biemann, Chris: Concept Over Time Analysis: Unveiling

Temporal Patterns for Qualitative Data Analysis. In: *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Volume 3: System Demonstrations (2024), S. 148–157 (<https://aclanthology.org/2024.naacl-demo.15/>, letzter Zugriff am 28.10.2025).

- [14] Galli, Carlo; Donos, Nikolaos und Calciolari, Elena: Performance of 4 Pre-Trained Sentence Transformer Models in the Semantic Query of a Systematic Review Dataset on Peri-Implantitis (<https://doi.org/10.3390/info15020068>, letzter Zugriff am 28.10.2025).
- [15] Grasnick, Belinda und Kumpfmüller, Konstantin: Einstufung des Verfassungsschutzes: Warum die AfD "gesichert rechtsextremistisch" ist. In: Tagesschau.de (<https://www.tagesschau.de/inland/innenpolitik/afd-verfassungsschutz-rechtsextremismus-100.html>, letzter Zugriff am 03.10.2025).
- [16] Grootendorst, Maarten: BERTopic: Neural topic modeling with a class-based TF-IDF procedure (<https://arxiv.org/abs/2203.05794>, letzter Zugriff am 14.10.2025).
- [17] Gschliesser, Nathalie: Der Gebrauch von Sprache in Deutschen Wahlkampagnen : Langage utilisé pendant les campagnes électorales en Allemagne. Masterarbeit an der Faculté de philosophie, arts et lettres, Université catholique de Louvain, 2018 (<https://thesis.dial.uclouvain.be/entities/masterthesis/38a91b1e-5334-4408-a337-45ee560b4241>, letzter Zugriff am 27.09.2025).
- [18] Hartmann, Stefan: Quantitative Analyse von Wahlprogrammen (<https://empirical-linguistics.github.io/wahlprogramme2021/#Pakete>, letzter Zugriff am 13.09.2025).
- [19] He, Zihao; Mokherian, Negar; Câmara, António; Abeliuk, Andrés und Lerman, Kristina: Detecting Polarized Topics Using Partisanship-aware Contextualized Topic Embeddings (<https://arxiv.org/abs/2104.07814>, letzter Zugriff am 12.08.2025).

- [20] Heuser, Ryan James: Word Vectors in the Eighteenth Century. In: *DH2017* (<https://dh2017.adho.org/abstracts/582/582.pdf>, letzter Zugriff am 20.10.2025).
- [21] Jacobs, Arthur M. und Kinder, Anette: Electoral Programs of German Parties 2021: A Computational Analysis Of Their Comprehensibility and Likeability Based On ‘SentiArt’ (<https://arxiv.org/abs/2109.12500?utm>, letzter Zugriff am 09.08.2025).
- [22] Juola, Patrick: The Rowling Case: A Proposed Standard Analytic Protocol for Authorship Questions. In: *Digital Scholarship in the Humanities*, Volume 30, Issue 1 (2015), S. 100–113 (<https://doi.org/10.1093/lc/fqv040>, letzter Zugriff am 27.10.2025).
- [23] Kleiser, Thomas: Wer und was ist Links oder Rechts? Teil I: Bedeutung von Links und Rechts (<https://socialsciencemeetsdatascience.de/analysis/article1.1.html>, letzter Zugriff am 24.09.2025).
- [24] Lau, Jey Han und Baldwin, Timothy: An Empirical Evaluation of doc2vec with Practical Insights into Document Embedding Generation. In: *Proceedings of the 1st Workshop on Representation Learning for NLP*, Berlin 2016, S. 78-86 (<https://arxiv.org/pdf/1607.05368>).
- [25] Le, Quoc und Mikolov, Tomas: Distributed Representations of Sentences and Documents. In: *Proceedings of the 31st International Conference on Machine Learning*, PMLR 32(2), 2014, S. 1188-1196 (<https://proceedings.mlr.press/v32/le14.pdf>).
- [26] Li, Lucian: Tracing the Genealogies of Ideas with Sentence Embeddings. In: *Proceedings of the 4th International Conference on Natural Language Processing for Digital Humanities*, 2024, S. 9–16 (<https://aclanthology.org/2024.nlp4dh-1.2.pdf>, letzter Zugriff am 28.10.2025).
- [27] Manifesto Corpus: <https://manifesto-project.wzb.eu/information/documents/corpus>, letzter Zugriff am 20.08.2025.
- [28] Mikolov, Tomas; Chen, Kai; Corrado, Greg und Dean, Jeffrey: Efficient Estimation of Word Representations in Vector Space. In: *Workshop*

Proceedings of the 1st International Conference on Learning Representations (ICLR). Scottsdale 2013 (<https://arxiv.org/pdf/1301.3781>, letzter Zugriff am 12.10.2025).

- [29] Mikolov, Tomas; Yih, Wen-tau; und Zweig, Geoffrey: Linguistic Regularities in Continuous Space Word Representations. In: *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Atlanta 2013, S. 746-751 (<https://aclanthology.org/N13-1090.pdf>, letzter Zugriff am 12.10.2025).
- [30] Moretti, Franco: Network Theory, Plot Analysis. In: *New Left Review*, März/April 2011 (<https://newleftreview.org/issues/ii68/articles/franco-moretti-network-theory-plot-analysis>, letzter Zugriff am 13.10.2025).
- [31] Mudde, Cas: The Populist Zeitgeist. In: Hertner, Isabelle und Jones, Erik (Hgg.): *Government and Opposition. An International Journal of Comparative Politics*, Band 39, Ausgabe 4. Oxford/Malden 2004, S. 541–563 (<https://www.cambridge.org/core/journals/government-and-opposition/article/populist-zeitgeist/2CD34F8B25C4FFF4F322316833DB94B7>, letzter Zugriff am 21.08.2025).
- [32] Murel, Jacob und Kavlakoglu, Eda: What is bag of words? (<https://www.ibm.com/think/topics/bag-of-words>, letzter Zugriff am 13.10.2025).
- [33] Pandey, Yogesh; Sharma, Monika; Kashaf Siddiqui, Mohammad und Singh Yadav, Sudeept: "Hate Speech Detection Model Using Bag of Words and Naïve Bayes". In: *Advances in Data and Information Sciences*, S. 457–470, 2020 (https://link.springer.com/chapter/10.1007/978-981-16-5689-7_40, letzter Zugriff am 13.10.2025).
- [34] Plenarprotokolle des Deutschen Bundestages (<https://www.bundestag.de/dokumente/protokolle/plenarprotokolle>, letzter Zugriff am 27.08.2025).
- [35] Rauh, Christian: Clear messages to the European public? The language of European Commission press releases 1985–2020. In: *Journal of European Integration*, Vol. 45, Nr. 4, S. 683–701 (<https://doi.org/10.1080/07036337.2022.2134860>, letzter Zugriff am 14.08.2025).

- [36] Rheault, Ludovic und Cochrane, Christopher: Word Embeddings for the Analysis of Ideological Placement in Parliamentary Corpora. In: *Political Analysis*, Vol. 28(1), 2019/20, S. 112–133 (<https://www.cambridge.org/core/journals/political-analysis/article/word-embeddings-for-the-analysis-of-ideological-placement-in-parliamentary-corpora/017F0CEA9B3DB6E1B94AC36A509A8A7B>, letzter Zugriff am 20.10.2025).
- [37] Riesch, Benjamin: Wordkrill: Extending Wordfish into the multidimensional political space (<https://www.arxiv.org/pdf/2506.20275>, letzter Zugriff am 18.10.2025).
- [38] Rooduijn, Matthijs und Akkerman, Tjitske: Flank attacks: Populism and left-right radicalism in Western Europe. In: *Party Politics*, Band 23, Ausgabe 3, S. 193–204 (<https://journals.sagepub.com/doi/full/10.1177/1354068815596514>, letzter Zugriff am 21.08.2025).
- [39] Salminen, Joni; Hopf, Maximilian; Chowdhury, Shammur A.; Jung, Soon-gyo; Almerexhi, Hind, und Jansen, Bernard J.: “Developing an online hate classifier for multiple social media platforms”. In: *Human-centric Computing and Information Sciences*, Vol. 10, 2020 (<https://hcis-journal.springeropen.com/articles/10.1186/s13673-019-0205-6>, letzter Zugriff am 13.10.2025).
- [40] Sammlung autorisierter Reden der Bundespräsidenten, der Bundeskanzler/in und von Mitgliedern der Bundesregierung (<https://www.bundesregierung.de/breg-de/service/newsletter-und-abos/bulletin>, letzter Zugriff am 23.08.2025).
- [41] Sauter, Anke: Neue Grundlage für die Hitler-Forschung. In: Philosophie und Geschichtswissenschaften, UniReport 2.24 (<https://aktuelles.uni-frankfurt.de/unireport/neue-grundlage-fuer-die-hitler-forschung/>, letzter Zugriff am 13.09.2025).
- [42] Scheufele, Bertram: Warum Frames und Ideologien keine Synonyme sind und Populismus keines von beiden ist – Konzeptionelle Brückenschläge zwischen Frames und Ideologien und Einwände zur aktuellen Populismus-Forschung. In: Tepe, Peter und Oehm, Stefan (Hgg.): *Mythos-Magazin: Politisches Framing 1* (2019), S. 1–28 (https://mythos-magazin.de/politisches-framing/bs_warum_frames_und_ideologien.pdf, letzter Zugriff am 21.08.2025).

- [43] Semantha (<https://www.semantha.de/de/parteiprogramme/>, letzter Zugriff am 07.08.2025).
- [44] Silge, Julia und Robinson, David: Text Mining with R. A Tidy Approach. 1. Auflage, Beijing et al. 2017 (<https://www.tidytextmining.com/tfidf#tfidf>, letzter Zugriff am 04.10.2025).
- [45] Slapin, Jonathan B. und Proksch, Sven-Oliver: A Scaling Model for Estimating Time-Series Party Positions from Texts. In: American Journal of Political Science, Volume 52, Issue 3 (2008), S. 705-722 (<https://doi.org/10.1111/j.1540-5907.2008.00338.x>, letzter Zugriff am 17.10.2025).
- [46] Stein Rokkan: Eine Familie von Modellen für die vergleichende Geschichte Europas. In: *Zeitschrift für Soziologie* 2/1980, S. 121, zitiert nach Decker, Frank: Jenseits von links und rechts. Lassen sich Parteien noch klassifizieren? (<https://www.bpb.de/shop/zeitschriften/apuz/279819/jenseits-von-links-und-rechts/>, letzter Zugriff am 17.09.2025).
- [47] Tagesschau.de:
<https://www.tagesschau.de/inland/bundestagswahl/parteien/musk-afd-wahlkampfshow-100.html>, letzter Zugriff am 03.10.2025.
- [48] Tahmasebi, Nina: A Study on Word2Vec on a Historical Swedish Newspaper Corpus. In: *CEUR Workshop Proceedings Vol. 2084* (DHN 2018), S. 25–37 (<https://ceur-ws.org/Vol-2084/paper2.pdf>, letzter Zugriff am 20.10.2025).
- [49] Tang, Muh-Chyun; Cheng, Yun Jen und Chen, Kuang Hua: A longitudinal study of intellectual cohesion in digital humanities using bibliometric analyses. In: *Scientometrics*, Volume 113, S. 985–1008, 2017 (<https://link.springer.com/article/10.1007/s11192-017-2496-6>, letzter Zugriff am 14.10.2025).
- [50] Trilcke, Peer: Wortwelten der Wahlprogramme: Digitale Analysen der Programme der Parteien zur brandenburgischen Landtagswahl 2019. In:

Netzwerk Digitale Geisteswissenschaften (<https://www.uni-potsdam.de/de/digital-humanities/blog/wahlanalyse2019>, letzter Zugriff am 07.08.2025).

- [51] Wermer-Colan, Alex: Stylometry Methods and Practices. A review of various stylometry methods and programs for the digital humanities (<https://guides.temple.edu/stylometryfordh>, letzter Zugriff am 24.10.2025).
- [52] Wermer-Colan, Alex; Lemire-Garlic, Nicole und Antsen, Jeff: Text Mining YouTube Comment Data with Wordfish in R (<https://doi.org/10.46430/phen0120>, letzter Zugriff am 08.12.2025).
- [53] Wevers, Melvin und Koolen, Marijn: Digital begriffsgeschichte: Tracing semantic change using word embeddings. In: *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, Vol. 53(4), 2020, S. 226-243 (<https://pure.uva.nl/ws/files/66795875/01615440.2020.pdf>, letzter Zugriff am 20.10.2025).

Sekundärliteratur

- [1] Ahlheim, Hannah: Der Nationalsozialismus 1933-1939. Paderborn 2025.
- [2] Andresen, Melanie: Computerlinguistische Methoden für die Digital Humanities. Eine Einführung für Geisteswissenschaftler:innen. Tübingen 2024.
- [3] Angeli, Oliviero und Otteni, Cyrill: Migration und Wahlverhalten in Deutschland. In: Brinkmann, Heinz Ulrich und Reuband, Karl-Heinz (Hgg.): *Rechtspopulismus in Deutschland. Wahlverhalten in Zeiten politischer Polarisierung*. Wiesbaden 2022, S. 371-393.
- [4] Baum, Gerhart: Besinnt Euch! Das Vermächtnis eines großen Liberalen. Berlin 2025.
- [5] Becher, Phillip: Rechtspopulismus. Köln 2013.
- [6] Bierling, Stephan: Die Unvereinigten Staaten. Das politische System der USA und die Zukunft der Demokratie. 2. Auflage, München 2024.
- [7] Brechenmacher, Thomas: Die CDU unter Angela Merkel (2000-2018). In: Lammert, Norbert: *Christlich Demokratische Union. Beiträge und Positionen zur Geschichte der CDU*. 1. Auflage, München 2020, S. 81-135.
- [8] Budge, Ian; Klingemann, Hans-Dieter; Volkens, Andrea; Bara, Judith und Tanenbaum, Eric: Mapping Policy Preferences: Estimates for Parties, Electors, and Governments 1945-1998. Oxford 2001.
- [9] Budge, Ian; Robertson, David und Hearl, Derek: Ideology, Strategy, and Party Change: Spatial Analyses of Post-War Election Programmes in 19 Democracies. Cambridge 1987.
- [10] Bücken, Sinah; Schade, Sarah Maria und Wiegerling, Ulrike: Die AfD: Woher sie kommt, wie sie funktioniert, wer sie unterstützt. In: Walther, Eva und

Isemann, Simon D. (Hgg.): *Die AfD – psychologisch betrachtet*. Wiesbaden 2019, S. 27-56.

- [11] Burrows, John F.: *Computation into criticism: a study of Jane Austen's novels and an experiment in method*. Oxford/New York: 1987.
- [12] Burrows, John F.: 'Delta': a Measure of Stylistic Difference and a Guide to Likely Authorship. In: *Literary and Linguistic Computing*, 17(3), 2002, S. 267–287.
- [13] Eve, Martin Paul: *The Digital Humanities and Literary Studies*. Oxford 2022, S. 28-29.
- [14] Forschungsgruppe Wahlen, Politbarometer Februar 2024, S. 3, zitiert nach Patzelt, Werner J.: *Deutschlands blaues Wunder. Die AfD und der Populismus*. München 2025, S. 41.
- [15] Friedrich, Sebastian: *Die AfD. Analysen – Hintergründe – Kontroversen*. Berlin 2017.
- [16] Gellner Winand und Kleiber, Martin: *Das Regierungssystem der USA. Eine Einführung*. 1. Auflage, Baden-Baden 2007.
- [17] Greven, Thomas: *Die Republikaner. Anatomie einer amerikanischen Partei*. München 2004.
- [18] Hemmelmann, Petra: *Der Kompass der CDU. Analyse der Grundsatz- und Wahlprogramme von Adenauer bis Merkel*. Wiesbaden 2017.
- [19] Hillje, Johannes: *Das „Wir“ der AfD. Kommunikation und kollektive Identität im Rechtspopulismus*. Frankfurt am Main 2022.
- [20] Jockers, Matthew L.: *Macroanalysis: Digital Methods and Literary History*. Urbana/Chicago 2013.
- [21] Kühnel, Steffen und Leibold, Jürgen: Die Auswirkungen von Migration und Integration auf das Wahlverhalten in Deutschland. In: Brinkmann, Heinz Ulrich und Reuband, Karl-Heinz (Hgg.): *Rechtspopulismus in Deutschland. Wahlverhalten in Zeiten politischer Polarisierung*. Wiesbaden 2022, S. 395-421.

- [22] Laver, Michael; Benoit, Kenneth und Garry, John: Extracting Policy Positions from Political Texts Using Words as Data. In: *American Political Science Review* 97/2 (2003), S. 311–332.
- [23] Lempp, Jakob; Serfling, Oliver und Rolf, Jan Niklas: Parteianhängerschaft in Deutschland. Eine Analyse der Parteien und ihrer Anhängerschaften in Bund und Ländern. Wiesbaden 2023.
- [24] Lütjen, Torben: Partei der Extreme: Die Republikaner. Über die Implosion des amerikanischen Konservatismus. Bielefeld 2016.
- [25] Möller, Kolja: Volk und Elite. Eine Gesellschaftstheorie des Populismus. 2. Auflage, Berlin 2024.
- [26] Mosteller, Frederick und Wallace, David L.: Inference and disputed authorship: The Federalist. Reading 1964.
- [27] Patzelt, Werner J.: Deutschlands blaues Wunder. Die AfD und der Populismus. München 2025.
- [28] Pfahl-Traugber, Armin: Die AfD und der Rechtsextremismus. Eine Analyse aus politikwissenschaftlicher Perspektive. Wiesbaden 2019.
- [29] Tarrow, Sidney: Movements and Parties. Critical Connections in American Political Development. Cambridge et al. 2021.
- [30] Treibel, Jan: Die FDP. Prozesse innerparteilicher Führung 2000-2012. 1. Auflage, Baden-Baden 2014.
- [31] Walter, Franz; Werwath, Christian und D'Antonio, Oliver: Die CDU. Entstehung und Verfall christdemokratischer Geschlossenheit (=Die politischen Parteien der Bundesrepublik Deutschland). 1. Auflage, Baden-Baden 2011.

Tools & Code Sources

- ChatGPT: <https://chatgpt.com/>.
- DeepL Übersetzer: <https://www.deepl.com/de/translator>.
- GitHub-Repository: https://github.com/Stevens29081997/Masterarbeit_Kevin_Stevens.
- How to Build Your Own Word2Vec Model from Scratch in Python (<https://www.youtube.com/watch?v=a3dxTcncUg4>, letzter Zugriff am 11.11.2025).
- Hugging Face: Sentence Transformers (<https://huggingface.co/sentence-transformers>, letzter Zugriff am 11.10.2025).
- Latent Semantic Scaling: <https://github.com/umanlp/SemScale>, letzter Zugriff am 08.12.2025.
- Microsoft Copilot (<https://m365.cloud.microsoft/chat/>).
- Python Bag of Words Model: A Complete Guide (https://www.datacamp.com/tutorial/python-bag-of-words-model?dc_referrer=https%3A%2F%2Fwww.google.com%2F, letzter Zugriff am 11.11.2025).
- spaCy-Modell für die deutsche Sprache: <https://spacy.io/models/de>, letzter Zugriff am 10.11.2025.
- Sentence Transformers Modell für die deutsche Sprache: [jinaai/jina-embeddings-v2-base-de · Hugging Face](https://huggingface.co/jinaai/jina-embeddings-v2-base-de), letzter Zugriff am 25.01.2026.
- Wordclouds: <https://www.wordclouds.com>, letzter Zugriff am 18.11.2025.
- Wordfish-Code für R: <http://www.wordfish.org/software.html>, letzter Zugriff am 17.10.2025, sowie <https://tutorials.quanteda.io/machine-learning/wordfish/>, letzter Zugriff am 08.12.2025.
- Wordscores: <https://tutorials.quanteda.io/machine-learning/wordscores/>, letzter Zugriff am 08.12.2025.

Anhang

Tabelle 7: Ungefilterte Ergebnisse der Tf-Idf-Implementierung

Partei	Top 10 Tf-Idf Begriffe (inkl. Wert)
AfD	afd (0.5596), euro (0.2186), p (0.1622) , deutschen (0.1472), lehnt (0.1435), ezb (0.1305), eeg (0.1109), bundestag (0.1020), pepp (0.1013), i (0.0913) .
CDU	schäuble (0.3273), wolfgang (0.3092) , bundeskanzlerin (0.1546), ddr (0.1171), worauf (0.0995) , deutschen (0.0991), bürgerinnen (0.0871), zudem (0.0837) , ampel (0.0831), politikwechsel (0.0820).
DEM	dollar (0.6342), biden (0.4027), amerikaner (0.2458), amerika (0.2388), amerikas (0.1787) , lgbtqi (0.1787), amerikanern (0.1648) , saubere (0.1593) , medicare (0.1492), harris (0.1306).
FDP	bürgerinnen (0.1963), euro (0.1771), lindner (0.1429), zudem (0.1233) , demokraten (0.1176), june (0.1099) , wachstumsinitiative (0.0880), deutschen (0.0836), subventionen (0.0790), ängste (0.0779).
GOP	republikaner (0.3980), joe (0.3733), trump (0.3365), stimmzettel (0.3243), dollar (0.3164), donald (0.3081) , amerika (0.2855), sir (0.2654), heritage (0.2359), biden (0.2205).
NSDAP	muß (0.7178) , jude (0.4147), deutschen (0.2689), führer (0.2431), volke (0.2413), bolschewismus (0.2393), händeklatschen (0.2153) , heiterkeit (0.1914) , stürmischer (0.1914) , mußte (0.1914) .
SPD	regierungsprogramm (0.5733), spd (0.2448), bundestagswahl (0.1674), ebert (0.1651), brandt (0.1230), sozialdemokratie (0.1216), bürgerinnen (0.1173), süd (0.1129) , willy (0.1007) , zukunftsinvestitionen (0.0956).

Anhand dieser Tabelle wird sehr gut ersichtlich, aus welchem Grund eine zusätzliche, manuelle Nachfilterung der Ergebnisse notwendig war. Alle Änderungen wurden in der obigen Tabelle markiert. Im Korpus der AfD steht an dritter Stelle beispielsweise ein „p“ und an zehnter Stelle ein „i“. Hierbei kann es sich nur um Artefakte aus dem Preprocessing handeln, die bei der Entfernung von Bindestrichen entstanden sind.

Bei der CDU schien es sinnvoll, die Begriffe „Wolfgang“ und „Schäuble“ zusammenzufassen, da sie dieselbe Person repräsentieren. Auch die Stoppwörter „worauf“ und „zudem“ scheinen im Preprocessing wohl übersehen worden zu sein. Da sie keinerlei Aussagekraft besitzen, mussten sie nachträglich entfernt werden.

Im Korpus der Demokraten bot es sich an, die Begriffe „amerikas“ und „amerikanern“ rauszunehmen, da sie Varianten von „amerika“ und „amerikaner“ sind, die nicht korrekt lemmatisiert wurden. Auch das Adjektiv „saubere“ ist ohne weiteren Kontext nur wenig aussagekräftig und wurde daher ebenfalls entfernt.

Im FDP-Korpus wurde das Stoppwort „zudem“ entfernt, ebenso wie das Wort „june“. Dieses Wort stammt aus dem Datum des letzten Zugriffs und ist in dem Sinne „FDP-spezifisch“, da die anderen Korpora bereits im Mai erstellt wurden, und „May“ als Gemeinsamkeit aller anderen daher gar nicht erst erwähnt wurde.

Im Korpus der Republikanischen Partei wurde lediglich „donald“ entfernt, da dieses Wort bereits mit dem Wort „trump“ repräsentiert wird, welches einen weitaus höheren *Tf-Idf*-Wert besitzt. Im Fall von Joe Biden wurde der Vorname nur in der Trefferliste beibehalten, weil er deutlich häufiger genannt wurde als der Nachname.

Im NSDAP-Korpus wurden „muß“ und „mußte“ entfernt, da beide Wörter Varianten von „müssen“ sind, aufgrund ihrer altdeutschen Schreibweise aber nicht vom Lemmatizer erkannt wurden. Die Wörter „Händeklatschen“, „Heiterkeit“ und „stürmischer“ wurden entfernt, da es sich bei diesen Begriffen um Anmerkungen der Transkriptionen handelt, welche diese öffentlichen festgehalten hatten. „Stürmischer“ bezieht sich zudem auf „Beifall“, welcher aber aufgrund seines Vorkommens in allen anderen Korpora einen viel zu geringen *Tf-Idf*-Wert besitzt, um in dieser Trefferliste überhaupt angezeigt zu werden.

Im Korpus der SPD wurden die Begriffe „süd“ und „willy“ aus der Trefferliste entfernt, da der erste eine geografische Richtung angibt und der zweite sich auf den bereits repräsentierten „brandt“ bezieht und daher redundant ist.

Eigenständigkeitserklärung

Ich erkläre hiermit, die Arbeit selbständig verfasst und bei der Erstellung dieser Arbeit die einschlägigen Bestimmungen, insbesondere zum Urheberrechtsschutz fremder Beiträge, eingehalten zu haben. Soweit meine Arbeit fremde Beiträge (z. B. Bilder, Zeichnungen, Textpassagen) enthält, erkläre ich, dass diese Beiträge als solche gekennzeichnet sind (z. B. Zitat, Quellenangabe) und ich eventuell erforderlich gewordene Zustimmungen der Urheber zur Nutzung dieser Beiträge in meiner Arbeit eingeholt habe.

Unterschrift:

K. Sturm, Stuttgart, 01.02.2026