

Universität Potsdam
Automatische Textanalyse in den Politikwissenschaften
Dozent:
Prof. Dr. Manfred Stede
Wintersemester 2020/21

Bericht

Untersuchungen der Wahlprogramme zur Bundestagswahl

Gruppe 4: Colorless Green Ideas minus L
Katja Konermann, 802658 (katja.konermann@uni-potsdam.de)
Anina Klaus, 802682 (aklaus@uni-potsdam.de)
Niklas Stepczynski, 797542 (stepczynski@uni-potsdam.de)

Inhaltsverzeichnis

1	Fragestellung	1
2	Korpus und Vorverarbeitung	1
3	Vokabular der Parteien	1
4	Klimabegriffe in Wahlprogrammen	8
5	Sentiment in Wahlprogrammen	10
6	Selbstreferentielle Sätze	14
7	Weitere Kategorien im Kontext Regierung - Opposition	17
7.1	Substantive	17
7.2	Adjektive	18
8	Fazit und Ausblick	20
9	Literatur	21

1 Fragestellung

Im Wahlkampf versuchen die verschiedenen Parteien Wähler*innen von ihrem Standpunkt zu überzeugen und für sich einzunehmen. In den Wahlprogrammen legen Parteien zu vielfältigen Themen ihre Meinung dar, stellen Forderungen und machen Versprechungen. Gerade deshalb eignen sich Wahlprogramme besonders, um verschiedene Untersuchungen vorzunehmen: Wie unterscheidet sich die Wortwahl der Parteien? Wer benutzt Begriffe wie *Klimakrise*, um die Dringlichkeit des Klimawandels und die Bedrohung bei fehlendem Handeln zu verdeutlichen, und wer nicht? Wer benutzt eher negativ konnotierte Begriffe, welche Parteien *fordern*, welche schauen eher auf die Vergangenheit? Für dieses Projekt haben wir Wahlprogramme zur Bundestagswahl mithilfe automatischer Analysetools betrachtet. Wir untersuchen hier so, ob Parteien sich in ihrer Wortwahl unterscheiden und wie Begriffe zum Thema Klimawandel verwendet werden. Der Hauptaugenmerk wird jedoch auf die Frage gerichtet, ob zwischen den Wahlprogrammen von Regierungs- und Oppositionsparteien der jeweiligen Jahren Unterschiede zu erkennen sind.

2 Korpus und Vorverarbeitung

Als Grundlage für diese Thesen dient ein Korpus der Wahlprogramme der sechs größten deutschen Parteien zu den Bundestagswahlen 2002 bis 2017. Hierzu haben wir uns für die Parteien CDU, SPD, AfD (seit 2013), FDP, DIE LINKE (vor 2007 die PDS) und Bündnis 90/die Grünen entschieden. Deren Wahlprogramme wurden in ein Plaintext-Format umgewandelt und dabei Elemente wie Seitenzahlen, Zeilenumbrüche, Codierungsfehler etc. herausgenommen. Zusätzlich haben wir uns dazu entschieden, den Plaintext zu lemmatisieren und die Liste der Stopppwörter so zu erweitern, sodass auch Wörter wie *dass*, *dabei* oder *dafür* herausgefiltert werden.

Die Untersuchungen des Korpus wurden größtenteils mit der R-Bibliothek *quanteda* durchgeführt. Ein ausführliche Besprechung des Codes und zusätzliche Daten können unter katjakon.github.io/Text-Mining-Wahlprogramme/ abgerufen werden.

	2002	2005	2009	2013	2017
Regierung	SPD, Bündnis90/die Grünen	SPD, Bündnis90/die Grünen	CDU, SPD	CDU, FDP	CDU, SPD
Opposition	CDU, FDP, PDS	CDU, FDP, PDS	FDP, Die Linke, Bündnis90/ die Grünen	SPD, Die Linke, Bündnis90/die Grünen, AfD	Die Linke, FDP, Bündnis90/die Grünen, AfD

Tabelle 1: Regierungs- und Oppositionsparteien in den Wahljahren

3 Vokabular der Parteien

Code mit Erklärungen zu diesem Abschnitt unter:

katjakon.github.io/Text-Mining-Wahlprogramme/vocabulary.html

Schon im ersten Teil des Projektes wurden die häufigsten Terme in Wahlprogrammen betrachtet. Wir haben diese Untersuchungen wiederholt, aber diesmal die Korpustexte lemmatisiert.

Die häufigsten Terme für jede Partei sind im Verzeichnis *data* zu finden. Dabei sind die 100 frequentesten Begriffe für die CDU etwa in der Datei *data/top_cdu.csv* gespeichert. In Tabelle 2 werden die 10 häufigsten Begriffe jeder Partei gegenübergestellt.

Wenig überraschend ist dabei, dass der eigene Parteiname jeweils relativ häufig auftritt.

Rang	AfD	CDU	SPD	DIE LINKE	PDS	FDP	B90/Die Grünen
1	müssen	deutschland	müssen	müssen	sozial	müssen	müssen
2	deutschland	müssen	gut	sozial	müssen	sollen	mensch
3	afd	land	sozial	öffentlich	sollen	fdp	mehr
4	deutsch	gut	deutschland	mensch	öffentlich	deutschland	gut
5	fordern	mensch	mensch	linke	mensch	mehr	neu
6	sollen	neu	mehr	sollen	pds	mensch	sollen
7	staat	mehr	neu	gut	mehr	neu	brauchen
8	land	deutsch	land	mehr	neu	frei	sozial
9	kind	stärken	sollen	recht	demokratisch	setzen	setzen
10	bürger	deshalb	europäisch	arbeit	deutschland	gut	deshalb

Tabelle 2: Häufigste Terme nach Partei

Terme wie *deutschland* oder *mensch* sind parteiübergreifend frequent. Auch die Verben *müssen* und *sollen* kommen bei beinahe allen Parteien in den oberen Rängen vor.

Terme wie *neu* und *gut* weisen - wie schon im ersten Teil festgestellt - daraufhin, dass Wahlprogramme eher positiv besetzt sind (Näheres dazu im Abschnitt *Sentiment in Wahlprogrammen*).

In Abbildung 1 sind die Terme dargestellt, die insgesamt am häufigsten vorkommen. Dabei wird für jede Partei die relative Häufigkeit auf einer Skala dargestellt. Um so höher die relative Häufigkeit ist, um so mehr verschiebt sich die Farbe ins Rote. Interessant ist hier zu betrachten, bei welchen Termen sich Unterschiede zwischen den Parteien zeigen. So fällt etwa auf, dass DIE LINKE und die PDS den Begriff *sozial* im Vergleich öfter gebrauchen. Die AfD und die CDU benutzen den Term *deutschland* mehr als andere Parteien. Das Wort *gut* wird besonders von CDU und SPD benutzt. Vor allem bei dem Verb *müssen* zeigen sich Unterschiede: DIE LINKE verwendet es sehr häufig, während es bei der CDU eher unterrepräsentiert ist.

Durch *Term Frequency - Inverted Document Frequency* (TF-IDF) werden für jede Partei zudem relevante Terme identifiziert. In den Abbildungen 2 bis 8 werden für jede Partei die Terme, mit dem höchsten TF-IDF-Wert dargestellt. Dabei ist zu beachten, dass der eigene Parteiname jeweils herausgefiltert wurde, da dieser nicht überraschend und kaum aussagekräftig ist. Im Gegensatz zu den einfachen *bag of words* Untersuchungen zeigen sich hier deutlicher Unterschiede zwischen den Parteien. Es treten hier Terme auf, die auf Themen hindeuten könnten, die besonders bei einzelnen Parteien im Vordergrund stehen. Begriffe wie *zuwanderung*, *massenimmigration*, *koranunterricht*, *islamunterricht* und *multikulturalismus*, die bei der AfD einen hohen TF-IDF-Wert erreicht haben, sind bezeichnend für den Fokus der Partei auf Immigration.

Während die Terme der AfD eher kritisch zu sein scheinen, sind die Begriffe, die bei der CDU einen hohen TF-IDF-Wert erreicht haben, weitaus positiver: Worte wie *zukunft*, *erfolg*, *wettbewerbsfähig* oder *wachstum* könnten zudem darauf verweisen, dass die CDU häufig (siehe Tabelle) an der Regierung beteiligt war. Ihre christliche Ausrichtung zeigt sich an Termen wie *schöpfung* oder *christlich*.

Mit der CDU gemeinsam hat die FDP Terme wie *mittelstand* und *marktwirtschaft*. Relevant für

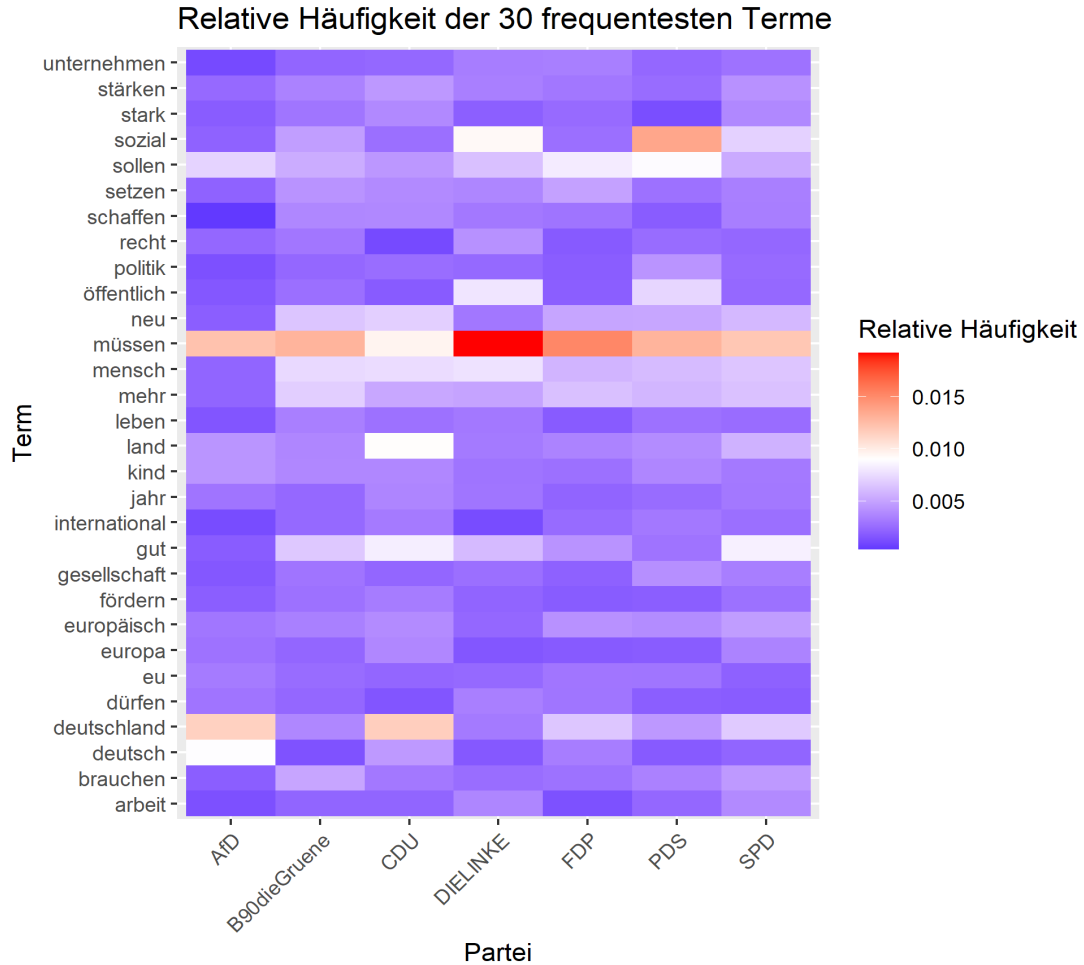


Abbildung 1: Vergleich der 30 häufigsten Terme zwischen den Parteien

die FDP sind zudem *liberal*¹, *eigenverantwortung* und *progression*. Auch diese Termini deuten auf die Ausrichtung der FDP hin.

Die Themen der Partei B90/Die Grünen zeigen sich in Begriffen wie *massentierhaltung*, *klimakrise*, *atomausstieg* oder *agrarwende*. In den Termen der SPD mit den höchstem TF-IDF Wert zeigen sich nicht so deutlich Themen auf. Zwar treten Begriffe auf, die mit *sozialdemokratisch* verwandt sind, aber es ist hier anzunehmen, dass diese durch die häufige Verwendung des eigenen Parteinamen als relevant eingestuft werden. Aussagekräftiger sind dagegen Begriffe wie *arbeitnehmerrechte* und *arbeitnehmerinnen*.

Ähnlich scheint es bei den Begriffen von der Partei DIE LINKE zu sein: Zwar gibt es Begriffe *kapitalismus*, *massenerwerbslosigkeit* oder *mindestsicherung*, die als durchaus relevant eingestuft werden können, aber es treten auch Token wie *xiv*, *vgl* oder *000* auf, die keine Bedeutung zu besitzen scheinen und eher auf Formatierungen im Wahlprogramm zurückzuführen sind.

Für die Wahlprogramme der PDS scheint zum Beispiel der Begriff *ostdeutschland* relevant.

¹Die Lemmatisierung scheint hier nicht gut zu funktionieren: Von *liberalen*, *liberale* und *liberaler* ist kein Lemma erkannt worden.

Das könnte darauf zurückzuführen sein, dass die Partei nur in den frühen 2000ern existierte, wo die Wiedervereinigung Deutschlands ein wichtiges Thema war. Es zeigen sich hier verschiedene Terme, die für die Vorläuferpartei der DIE LINKE durchaus plausibel erscheinen, wie etwa *sozialdarprinzip*, *arbeitslosenunterstützung* oder *integrationshilfe*.

Insgesamt können schon einfache Methoden wie das *bag of words* Modell und TF-IDF Berechnungen aufschlussreiche Einblicke in die Themen in Wahlprogrammen geben. Sie dienen so als ein erster Überblick und zeigen, dass Parteien sich unterschiedlich ausdrücken und unterschiedlich Schwerpunkte setzen. Es können außerdem Gemeinsamkeiten herausgestellt werden, auch wenn diese wenig überraschend sind: In Wahlprogrammen zur Bundestagswahl geht es nunmal um *Deutschland* und der Fokus liegt auf *Menschen*. Verben wie *müssen* und *sollen* deuten auf den deontischen Charakter in Wahlprogrammen hin: Es werden Verpflichtungen und Notwendigkeiten besprochen.

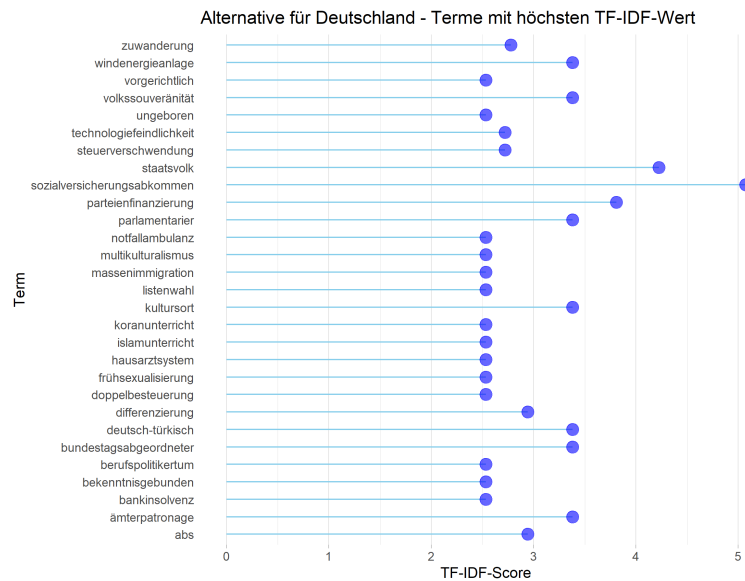


Abbildung 2: TF-IDF Scores für die AfD

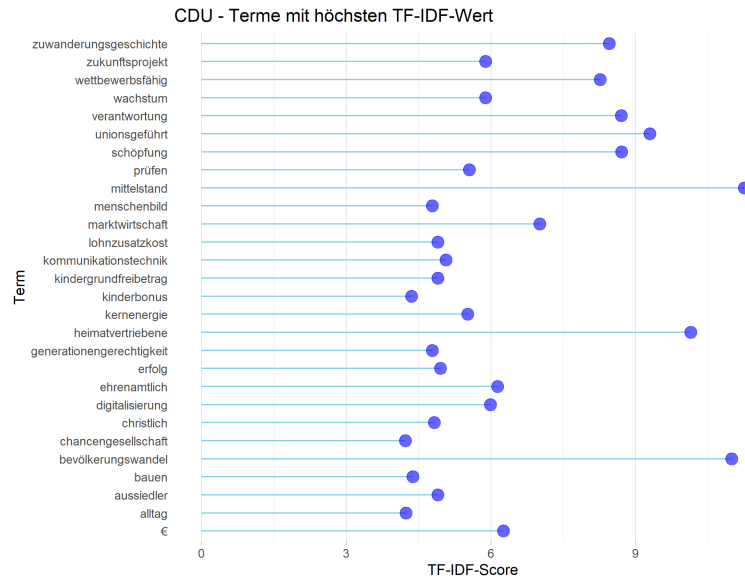


Abbildung 3: TF-IDF Scores für die CDU

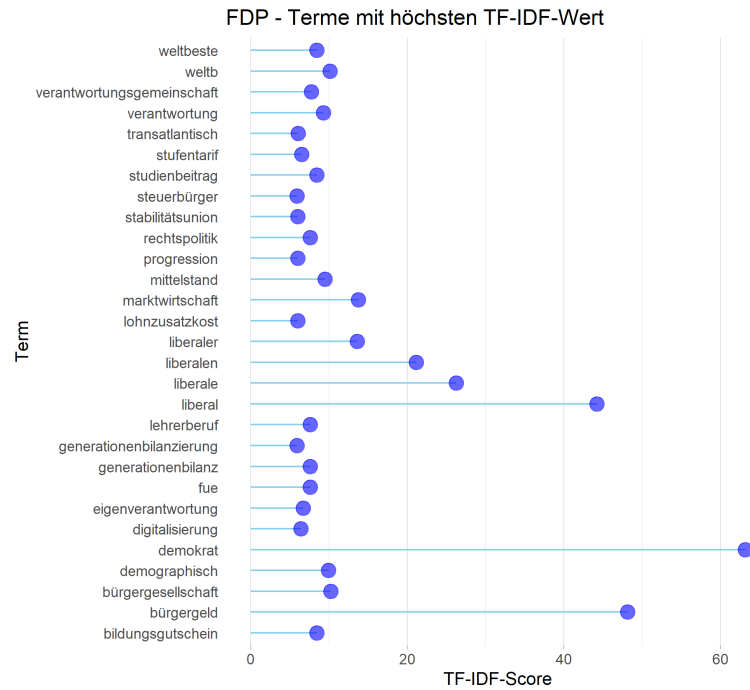


Abbildung 4: TF-IDF Scores für die FDP

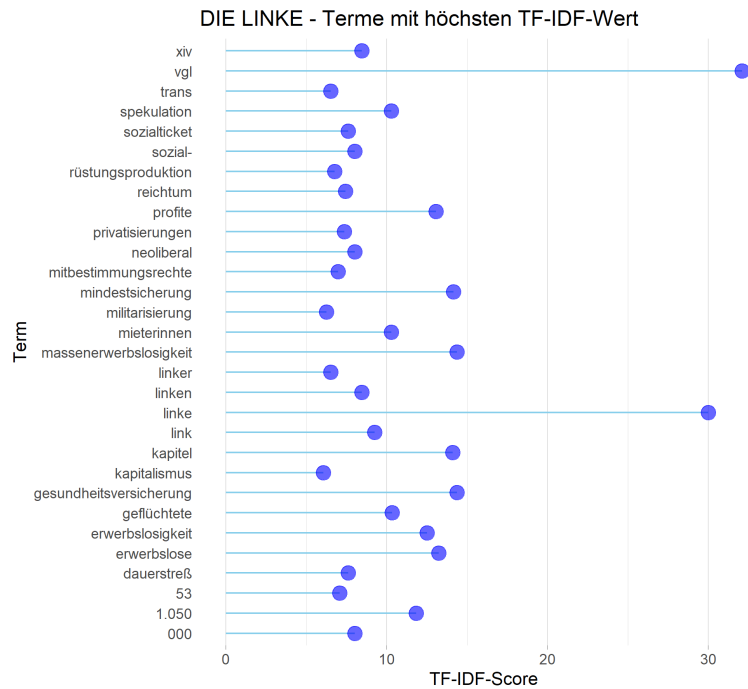


Abbildung 5: TF-IDF Scores für DIE LINKE

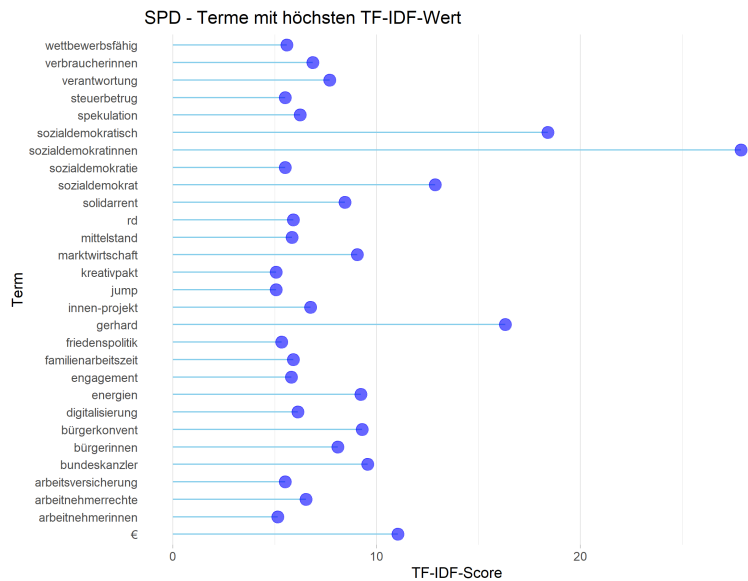


Abbildung 6: TF-IDF Scores für die SPD

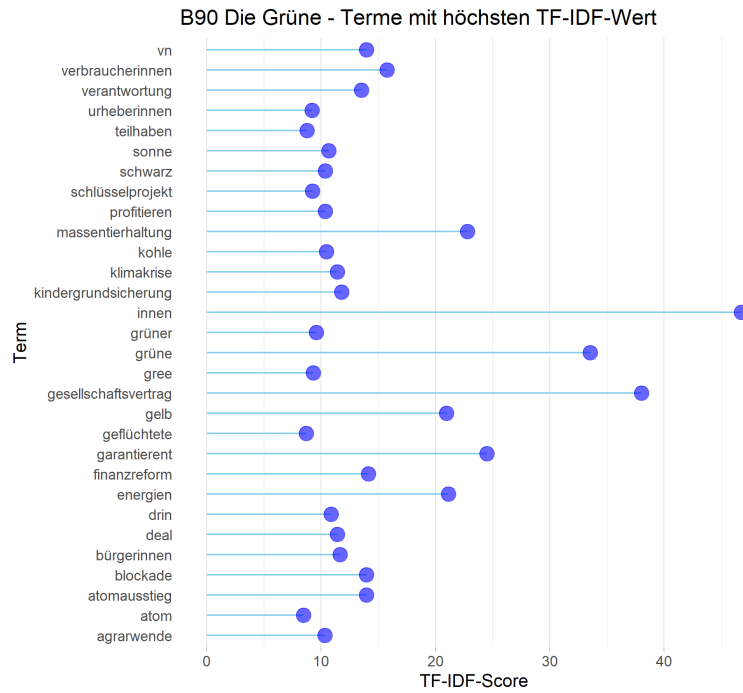


Abbildung 7: TF-IDF Scores für die B90/Die Grünen

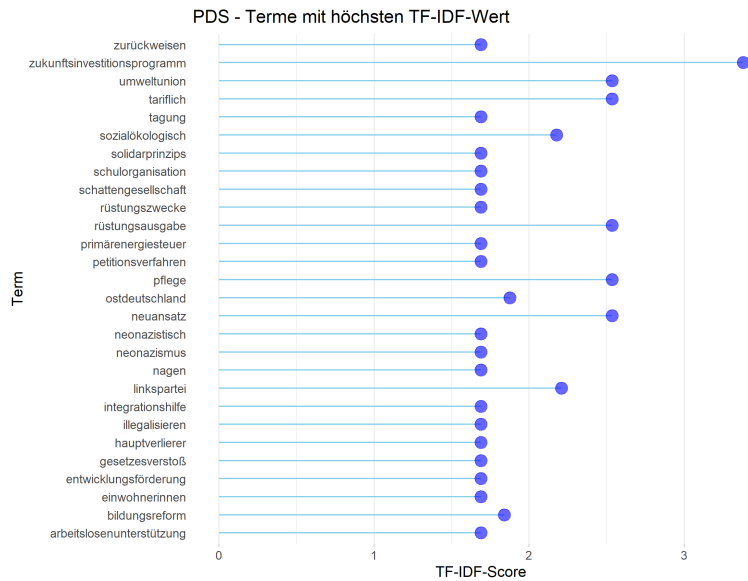
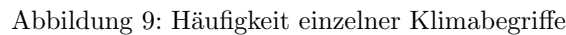


Abbildung 8: TF-IDF Scores für die PDS

Code mit Erklärungen zu diesem Abschnitt unter:
katjakon.github.io/Text-Mining-Wahlprogramme/climate.html

klimawandel, treibhaus*, CO2, erderwärmung, erneuerbare energien, 2-Grad-Ziel,
 zwei-grad-ziel, klimakrise, klimakatastrophe, klimaschutz, abholzung, fossile ener-
 gie*, atmosphäre, kohlenstoffdioxid, emission*



8

Parteiübergreifend werden eher Begriffe wie *co2*, *klimaschutz* und *klimawandel* benutzt, die neutralen scheinen.

In Abbildung 10 wird die Häufigkeit der Klimabegriffe über die Wahljahre hinweg dargestellt. Hier ist deutlich zu erkennen, dass die Verwendung von Klimatermen bei allen Parteien ab 2009 zunimmt. Dies deutet auf eine zunehmende Wichtigkeit dieses Themas für Politik und Gesellschaft hin. Im Wahljahr 2017 sinkt die Häufigkeit der Klimaterme dagegen wieder. Die Gründe dafür sind nicht offensichtlich. Eventuell rückten hier andere Themen in den Vordergrund.

Auch in dieser Grafik zeigt sich, dass Klimabegriffe am häufigsten bei der Partei B90/Die Grünen auftreten. Am seltensten scheinen sie von der AfD und der PDS verwendet zu werden. Die Nachfolgepartei der PDS - DIE LINKE - gebraucht dagegen Begriffe zum Klima im Vergleich häufiger, wobei sie im Jahr 2017 ihr bisheriges Maximum erreicht. Die Häufigkeit von Klimabegriffen in den Wahlprogrammen der FDP scheint sich nur wenig zu verändern. Zwar steigt auch die Häufigkeit im Vergleich von 2005 und 2009 an, allerdings nicht so stark wie etwa bei der CDU.

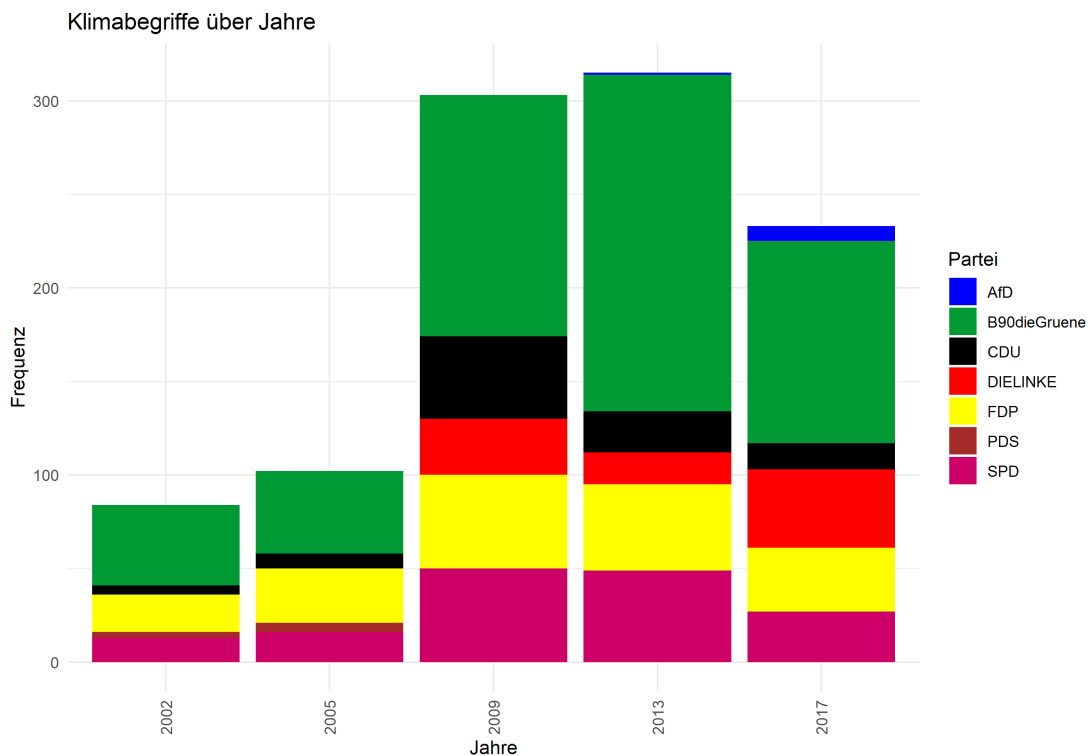


Abbildung 10: Häufigkeit von Klimatermen über die Jahre

5 Sentiment in Wahlprogrammen

Code mit Erklärungen zu diesem Abschnitt unter:

katjakon.github.io/Text-Mining-Wahlprogramme/sentiment.html

Für die Untersuchung des Sentiments in Wahlprogrammen haben wir das Wörterbuch *SentiWS* der Universität Leipzig genutzt. In diesem Projekt sind es im Verzeichnis *data* unter *SentiWS_v1.8c_Negative.txt* und *SentiWS_v1.8c_Positive.txt* abgespeichert. Das Sentimentwörterbuch enthält etwa 1800 negativ und 1600 positiv besetzte Terme.

Zunächst wird in Abbildung 11 die relative Häufigkeit von negativen und positiven Termen über die Jahre dargestellt. Es ist hier sinnvoll, nicht die absolute Häufigkeit zu nutzen, da sich die Länge der Wahlprogramme zwischen den Wahljahren mitunter sehr unterscheidet. Auffallend ist hier, dass positive Begriffe weitaus häufiger als negative auftreten. Das deckt sich mit den Erkenntnissen aus den *bag of words* Betrachtungen, wo positiv besetzte Begriffe oft in den oberen Rängen aufgetreten sind. Über die Jahre scheint es keine großen Schwankungen im Sentiment zu geben. Sowohl die relative Häufigkeit von positiven als auch negativen Termen ist relativ stabil. Die wenigsten positiven und die meisten negativen Terme treten im Wahljahr 2005 auf. Hier scheint das Sentiment also am negativsten zu sein. Dass mehr negative Sentimentterme auftreten, wenn weniger positive Begriffe vorhanden sein, kann hier ebenfalls als leichter Trend ausgemacht werden.

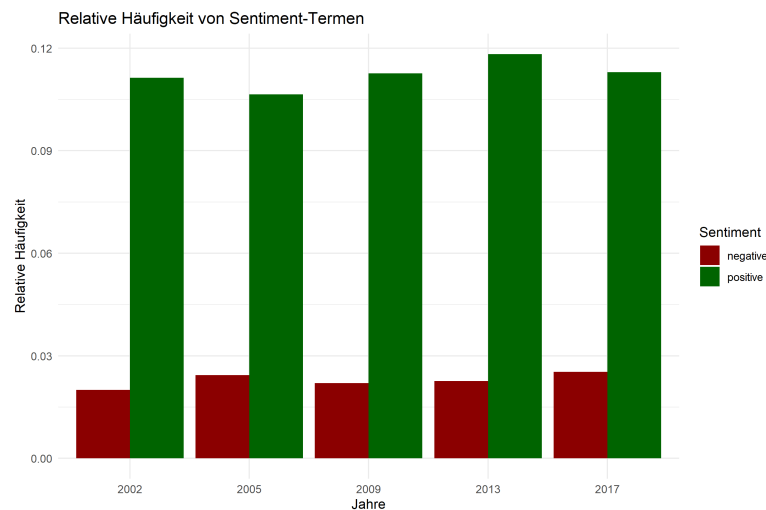


Abbildung 11: Sentiment über Jahre

In Abbildung 12 wird die relative Häufigkeit von Sentimenttermen über die Parteien verglichen. Auch hier ist zu beobachten, dass positive Terme seltener auftreten, wenn die relative Häufigkeit von negativen Termen höher ist. Diese Korrelation scheint durchaus plausibel: Wenn eine Partei eher positive Aspekte betont, werden negative Entwicklungen eher außer Acht gelassen.

Die CDU und die SPD - Parteien, die am häufigsten an Regierungen beteiligt waren (siehe Tabelle 1) - weisen die höchste Häufigkeit von positiven Termen und gleichzeitig die niedrigste

Häufigkeit von negativen Termen auf. In den Wahlprogrammen der AfD sind Terme mit positiven Sentiment im Vergleich zu anderen Parteien am seltensten. Die Parteien AfD, PDS und DIE LINKE, die im hier betrachteten Zeitraum an keiner Regierung beteiligt waren, benutzen vergleichsweise wenig positiv besetzte Terme. Die Häufigkeit der negativ besetzten Terme unterscheidet sich dagegen nicht so stark. Eine mögliche Erklärung hierfür wäre, dass Oppositionsparteien die Lage Deutschlands nicht unbedingt schlechter darstellen, dafür aber eher auf eine übermäßige positive Darstellung verzichten.

In Abbildung 13 und 14 wird die Entwicklung der relativen Häufigkeit von Sentimenttermen

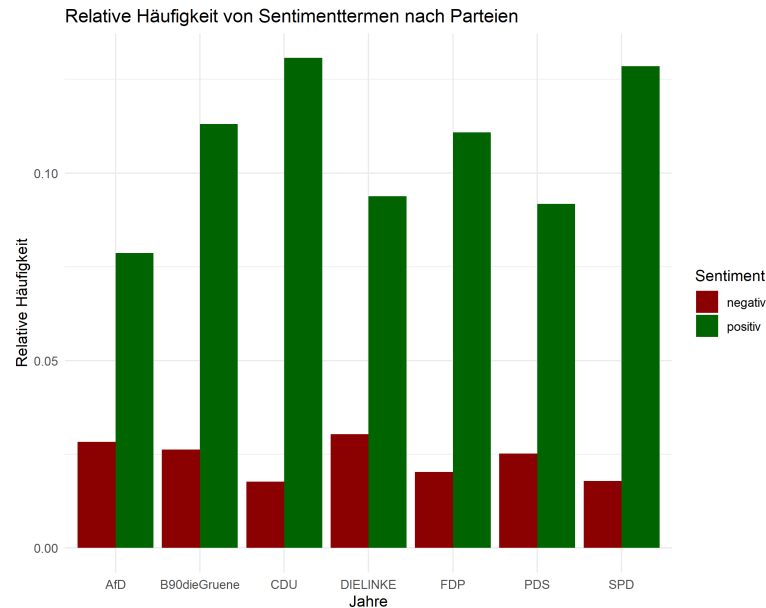


Abbildung 12: Sentiment der Parteien

über Parteien und Jahre hinweg betrachtet. Hier kann nun genau untersucht werden, inwieweit sich die Wahlprogramme von Regierungs- und Oppositionsparteien in ihrem Sentiment unterscheiden. So kann einmal für jede Partei betrachtet werden, ob sich ihr Sentiment ändert, wenn sie an der Regierung beteiligt ist. Für die Wahlprogramme der CDU zeigt sich beispielsweise, dass die Verwendung von positiv konnotierten Begriffen in den Jahren zunimmt, in denen sie zur Zeit der Wahl an Regierungen beteiligt waren (2009, 2013, 2017). Für andere Parteien lässt sich dies nicht so eindeutig feststellen. Die relative Häufigkeit von positiven Sentimenttermen in den Wahlprogrammen der Partei B90/ Die Grünen ist zum Beispiel 2017 höher als 2005, obwohl die Partei in diesem Jahr zum Zeitpunkt der Wahl nicht an der Regierung beteiligt war. Für die Wahlprogramme der FDP lässt sich ähnliches beobachten. Die relative Häufigkeit von positiven Begriffen in Wahlprogrammen der SPD bleibt relativ konstant.

Auch bei Parteien, die im betrachteten Zeitraum nicht an Regierungen beteiligt waren, zeigt sich kein eindeutiges Bild: Während die Verwendung positiver Terme bei der PDS und DIE LINKE über die Jahre relativ konstant bleibt, sinkt die relative Häufigkeit in den Wahlprogrammen der AfD von 2013 auf 2017.

Wie auch bei der relativen Häufigkeit von positiven Termen lässt sich auch der Verwendung

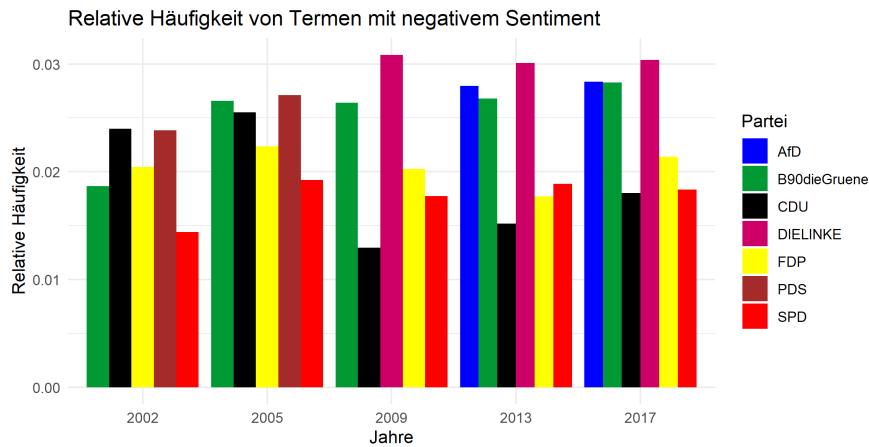


Abbildung 13: Negatives Sentiment über Parteien und Jahre

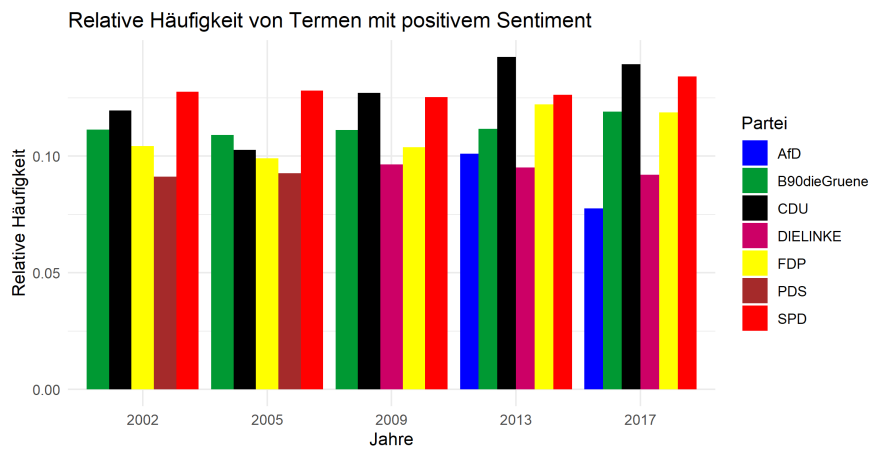


Abbildung 14: Positives Sentiment über Parteien und Jahre

von negativen Begriffen keine klare Korrelation für einzelne Parteien feststellen. Während die Häufigkeit von negativen Sentimenttermen für die CDU in den Jahren höher ist, in denen sie nicht an Regierungen beteiligt war (2002 und 2005), lässt sich etwa bei der SPD beobachten, dass die Verwendung von negativen Sentimenttermen im Wahljahr 2005 am höchsten war, obwohl die Partei zu diesem Zeitpunkt an der Regierung beteiligt war. Für die FDP kommen negative Terme wiederum am seltensten vor, als die Partei an der Regierung beteiligt war, also vor der Wahl 2013.

Betrachtet man die Wahljahre für sich und vergleicht Oppositions- und Regierungsparteien, lassen sich durchaus interessante Beobachtungen machen: 2002 ist die Verwendung von negativen Sentimenttermen bei den Regierungsparteien SPD und B90/Die Grünen am niedrigsten. Während dies für das Wahljahr 2005 nicht gilt, ist es für die darauffolgenden Jahren wieder zu beobachten: Die Regierungsparteien der jeweiligen Jahre sind gleichzeitig die Parteien, bei denen die relative Häufigkeit von negativ konnotierten Begriffe am niedrigsten ist. Dies scheint die These zu unterstützen, dass Regierungsparteien eher auf eine negative Darstellung verzichten. Aller-

dings muss hier angemerkt werden, dass die Unterschiede zwischen den relativen Häufigkeiten sehr gering sind und damit nicht unbedingt eine starke Aussagekraft besitzen.

Die Terme, die in den Wörterbüchern *SentiWS* enthalten sind, werden nicht nur in positiv und negativ unterteilt, sondern ihnen werden auch Werte zugeordnet: Um so positiver das Sentiment eines Terms ist, um so näher liegt sein Wert an eins. Je negativer ein Term konnotiert wird, desto näher liegt sein Wert an minus eins. So kann nicht nur die Häufigkeit positiver und negativer Terme Auskunft über das Sentiment in Wahlprogrammen geben, sondern auch die Werte der einzelnen Terme. In Abbildung 15 werden die aufsummierten Werte der positiven und negativen Sentimentterme gegenübergestellt. Dafür wurde der Sentimentwert jedes Terms mit seiner relativen Frequenz der jeweiligen Partei multipliziert. Diese Werte wurden dann pro Partei aufsummiert. Die Erkenntnisse der vorherigen Untersuchungen werden hier bestätigt. Große Verschiebungen scheint es hier nicht zu geben. Die CDU besitzt den höchsten Wert für positive Terme, während der Wert für negative Terme nicht besonders ausschlägt. Ähnlich erscheint die Werte der SPD. Die AfD weist den niedrigsten positiven Wert auf. Den kleinsten Wert für negative Terme ist bei der Partei DIE LINKE zu finden.

Sentimentanalyse hat sich hier insgesamt als durchaus gute Methode zur Untersuchung der

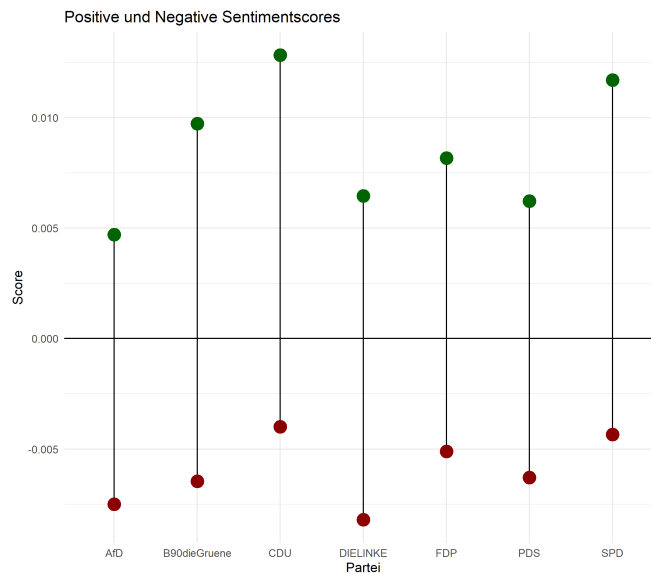


Abbildung 15: Sentimentwerte der Parteien

Wahlprogramme erwiesen. Es lassen sich Ergebnisse erzielen, die zumindest plausibel und interessant erscheinen. Schon durch eine simple Betrachtung der Häufigkeiten von negativen und positiven Termen können Unterschiede zwischen den Parteien festgestellt werden, die zumindest teilweise mit Regierungsbeteiligung korreliert werden können, auch wenn der Datensatz hier noch vergleichsweise klein ist und natürlich nicht ausgeschlossen werden kann, dass die Unterschiede zufällig entstanden sind.

6 Selbstreferentielle Sätze

Code mit Erklärungen zu diesem Abschnitt unter:

katjakon.github.io/Text-Mining-Wahlprogramme/annotated_data.html

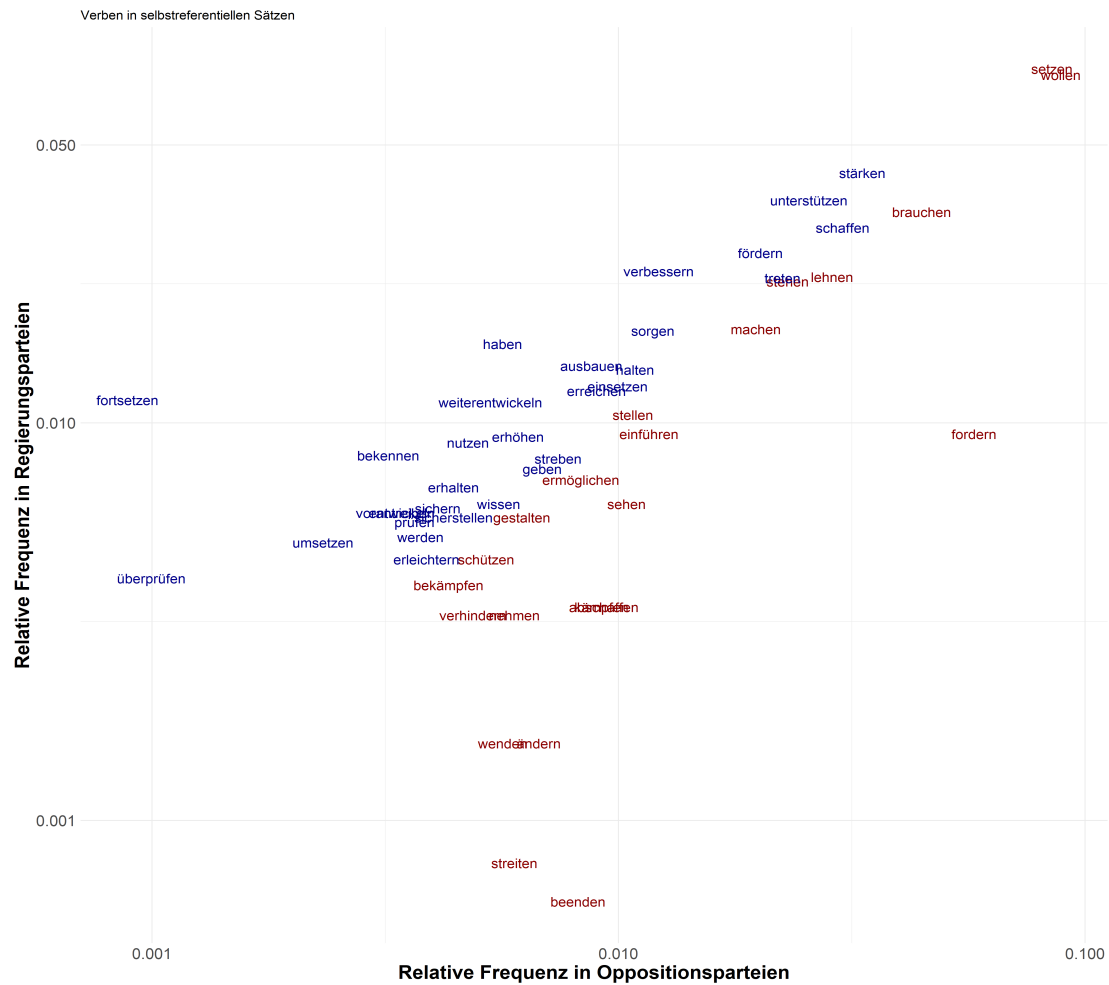


Abbildung 16: Selbstreferentielle Prädikate nach Regierungs- und Oppositionswahlprogrammen

Regierungsparteien preisen in ihren Wahlprogrammen die Erfolge ihrer Arbeit der letzten Legislaturperiode an, während Oppositionsparteien die Versäumnisse anprangern oder Forderungen aufstellen. Um diese These zu überprüfen, haben wir *selbst-referenzielle Prädikate* untersucht. Als solche verstehen wir Prädikate von Sätzen der Form „Wir fordern..“ oder „Die FDP befürwortet...“, also Sätze, in denen Parteien ihr eigenes Tun und Handeln beschreiben. Um solche Prädikate zu extrahieren, haben wir zunächst die Wahlprogramme mithilfe der R-Library *udpipe* annotiert. Hierdurch haben wir Zugriff auf allerlei Daten über jedes Token, etwa POS-tag, Features und Dependenz-Relation. Im nächsten Schritt kann dann das annotierte Modell in Oppositions- und Regierungswahlprogramme aufgeteilt werden, um für jede der beiden Gruppen jene Prädikate zu extrahieren, welche eine Eigenbezeichnung der Partei (also *wir* oder den jeweiligen Parteinamen) zum Subjekt haben. Diese werden dann mit den dazugehörigen Informationen in einem *data.frame* abgespeichert. Dies ermöglicht eine Visualisierung und Analyse der Daten.

Jahr	Partei	Kontext
2013	SPD	Gerade in den ostdeutschen Ländern müssen wir weiterhin dafür streiten , dass möglichst viele Menschen die Demokratie und unser Gemeinwesen mitgestalten.
2013	Die Linke	Wir fordern verbindliche Regeln für alle öffentlichen Unternehmen – die Begrenzung von Managergehältern bei den Landesbanken kann hier ein Vorbild sein – und streiten dafür, dass solche Regelungen in allen Unternehmen gelten.

Tabelle 3: Kontexte für das Wort *streiten* in Oppositionsparteien

Jahr	Partei	Kontext
2017	AfD	Wir fordern außerdem eine stärkere Kontrolle und Berücksichtigung gesamtwirtschaftlicher deutscher Interessen, wenn ausländische Konzerne eine deutsche Firma übernehmen wollen.
2013	Die Linke	Wir wollen, dass niemand im Alter und in der Arbeit arm ist. Wir fordern einen gesetzlichen Mindestlohn von zehn Euro pro Stunde.

Tabelle 4: Kontexte für das Wort *fordern* in Oppositionsparteien

Jahr	Partei	Kontext
2009	Bündnis90/die Grünen	Und wir setzen uns weiterhin dafür ein, den gesellschaftlichen Skandal zu beenden , dass Frauen für gleichwertige Arbeit schlechter bezahlt werden.
2017	AfD	Nur das Staatsvolk der Bundesrepublik Deutschland kann durch das Mittel der unmittelbaren Demokratie diesen illegalen Zustand beenden .

Tabelle 5: Kontexte für das Wort *beenden* in Oppositionsparteien

Jahr	Partei	Kontext
2009	CDU	[...] wenn wir unsere Wachstumspolitik fortsetzen .
2002	SPD	Wir haben mit unserer Politik der Mitte die Erneuerung in Deutschland begonnen. Und wir wollen sie fortsetzen und den Zusammenhalt sichern.

Tabelle 6: Kontexte für das Wort *fortsetzen* in Regierungsparteien

Jahr	Partei	Kontext
2009	CDU	Wir werden daher die Besteuerung von Unternehmen auf krisenverschärfende Wirkungen überprüfen und die notwendigen Anpassungen vornehmen.
2002	Bündnis90/die Grünen	Wir wollen die Geheimdienste durch eine unabhängige Geheimdienst-Struktur-Kommission – ihre Kosten und ihre Schaden und Nutzen für die Politik – überprüfen .

Tabelle 7: Kontexte für das Wort *überprüfen* in Regierung

In Abbildung 16 sind die Frequenzen der häufigsten selbstreferentiellen Prädikate für Oppositions und Regierungswahlprogramme gegenübergestellt. Je weiter oben ein Term steht, desto stärker ist er in Programmen von Regierungsparteien vertreten. Je weiter rechts er steht, desto häufiger findet er sich in Wahlprogrammen der Opposition. Terme, die entlang der „gedachten $x = y$ – Linie gruppiert sind, treten also in etwa gleich häufig auf. Die Farbe gibt zusätzlich an, in welcher Gruppe ein Prädikat schlussendlich frequenter ist. Regierungsparteien nutzen also häufiger Prädikate wie *fortsetzen*, *überprüfen* und *verbessern*, während Oppositionsparteien eher *fordern*, *beenden* oder *ändern* verwenden. Viele der Begriffe entsprechen den Erwartungen: Ein Begriff wie *weiterentwickeln* fügt sich sehr gut in den Kontext einer Regierungspartei ein, welche weiterregieren möchte, während ein Begriff wie *fordern* impliziert, dass der Fordernde selbst (noch) nicht in der Position ist, seine Forderung umzusetzen. Es scheint also durchaus eine Klasse von Verben zu geben, welche auf bestehende Regierungsarbeit einer Partei hindeutet, sowie eine weitere, welche die Verfasserin in der eher Opposition verortet. Selbst-referenzielle Prädikate können somit als viel-versprechender Indikator für die Position einer Partei im politischen System zum Zeitpunkt der Äußerung angesehen werden.

Auch Zeit- und Verbform der Prädikate lassen sich aus dem Modell erschließen. Hierbei wird deutlich, dass die Regierung häufiger Vergangenheitsformen verwendet. Darauf deuten sowohl der höhere Anteil an Prädikaten der Kategorie *Past* hin als auch der höhere Anteil der Verbform *Part*, welche auf Partizipien im Perfekt hinweist.

Position	Gegenwart	Vergangenheit	Infinitiv	Finit	Partizip
Regierung	0.4204848	0.03424550	0.4448788	0.4547303	0.06176701
Opposition	0.5357617	0.02325066	0.3799823	0.5590124	0.01494686

Tabelle 8: Zeitform und Verbform der Prädikate

7 Weitere Kategorien im Kontext Regierung - Opposition

Neben den Prädikaten haben wir auch die Verwendung von Substantiven und Adjektiven in Wahlprogrammen betrachtet.

7.1 Substantive

Das Vorgehen hierbei war sehr ähnlich, diesmal wurde das annotierte Modell danach gefiltert, ob das Token in der Kategorie *upos* den Wert *NOUN* hat. Aus den Frequenzen ließ sich die folgende Grafik berechnen:

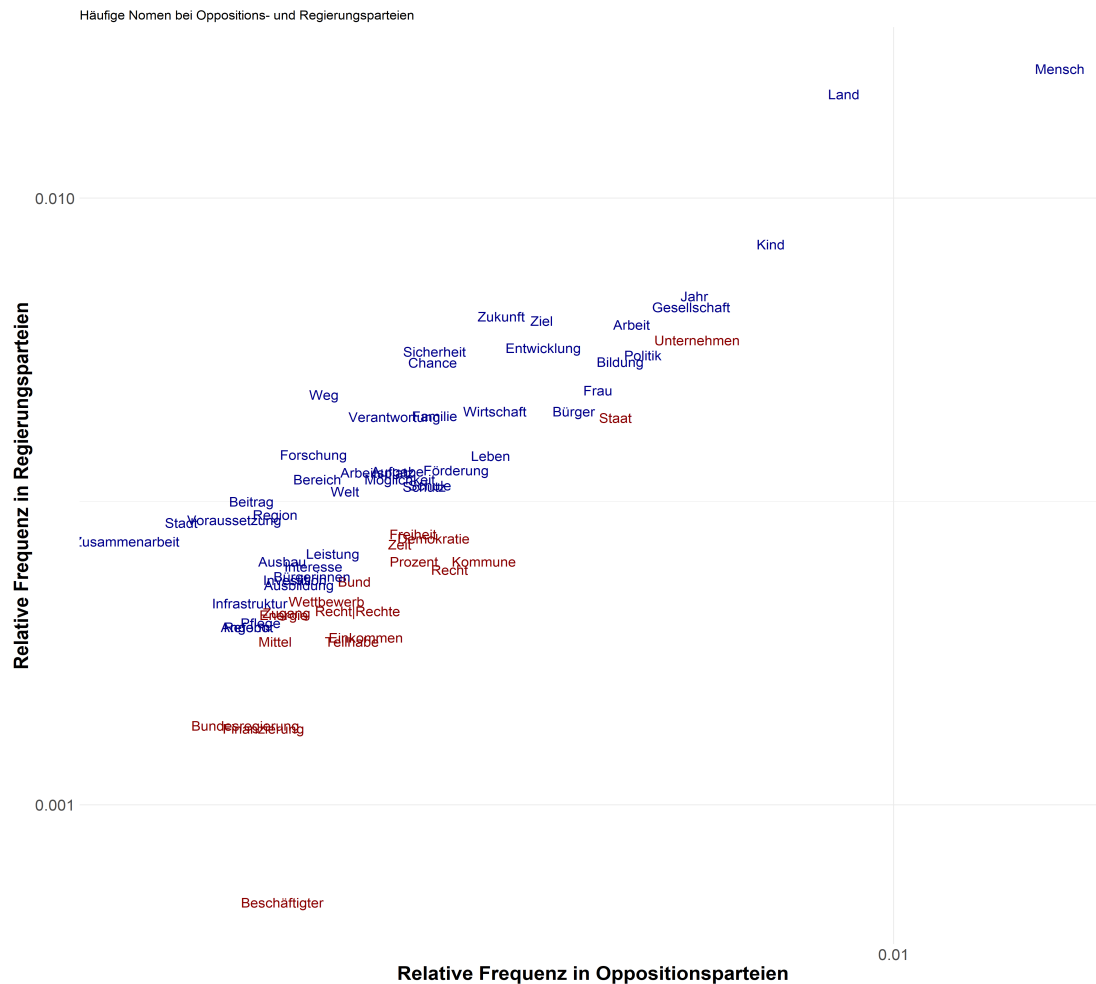


Abbildung 17: Frequentive Nomen bei Oppositions- und Regierungsparteien

Auch hier kann man versuchen, den Begriffen eine Aussagekraft bezüglich der Regierungsbe-
teiligung der Verfasser beizumessen. So scheinen Regierungsparteien häufiger über *Zusammen-
arbeit* oder *Zukunft* zu reden. Begriffe wie *Chance* können darauf hindeuten, dass Regierungs-
parteien eher positiv konnotierte Wörter verwenden. Der am deutlichsten in der Opposition
vertretende Begriff ist wiederum aber *Beschäftigter*, ein Begriff, der wiederum lediglich ein The-
ma andeutet. Es ist also zu vermuten, dass Substantive eher Aufschluss über die Hauptthemen

der Parteien gibt. Die Vermutung ist hier, dass die Verteilung der Begriffe anders wäre, wenn andere Parteien die Regierungsarbeit dominieren würden. So wären zum Beispiel Begriffe im Umweltkontext in der Gruppe zu vermuten, in der die Partei Bündnis 90/die Grünen in der jeweiligen Legislatur-Periode anzutreffen sind, welche wie im Abschnitt 4 dieses Thema dominieren. Auch Sentimentwerte von Substantiven können unzuverlässig sein: Zwar deuten Begriffe wie *Chance* und *Leistung* auf ein positiveres Sentiment in Texten von Regierungsparteien hin, doch auch diese Parteien müssen sich mit Themen wie *Krieg* und *Umweltverschmutzung* befassen.

7.2 Adjektive

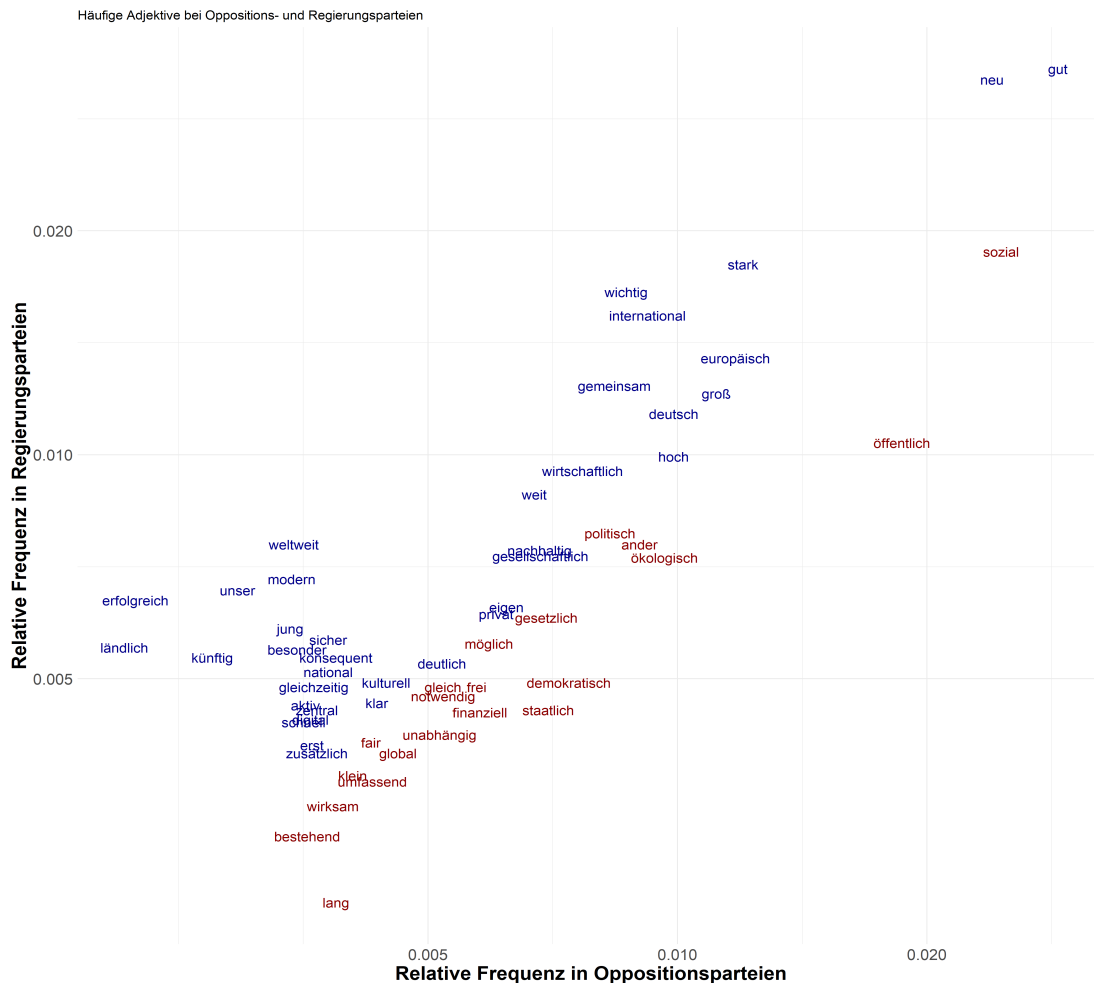


Abbildung 18: Frequente Adjektive bei Oppositions- und Regierungsparteien

Auch bei den Adjektiven zeigt sich ein gespaltenes Bild. Zwar gibt es auch hier Begriffe, die eher auf Themenbereiche einer Partei hindeuten, etwa *finanziell*, *ökologisch* oder *europäisch*. Jedoch gibt es hier besonders in der Gruppe der eher auf Regierungsseite verwendeten Adjektive Begriffe, die sich sehr gut in den Kontext eines Wahlprogrammes, welches bisherige Regierungsarbeit anpreist, einpassen lassen. Beispiele hierfür sind *erfolgreich*, *künftig* und *stark*. Auch eine

Sentimentanalyse der Adjektive bestätigt die Vermutung, dass Regierungsparteien eher ein positives Bild zeichnen wollen als Oppositionsparteien (siehe Abbildung 19). Es ist also zu vermuten, dass Adjektive bessere Indikatoren für Regierungsbeteiligung sind als Substantive, da sie einen stärkeren Beitrag zum Sentiment des Textes leisten.

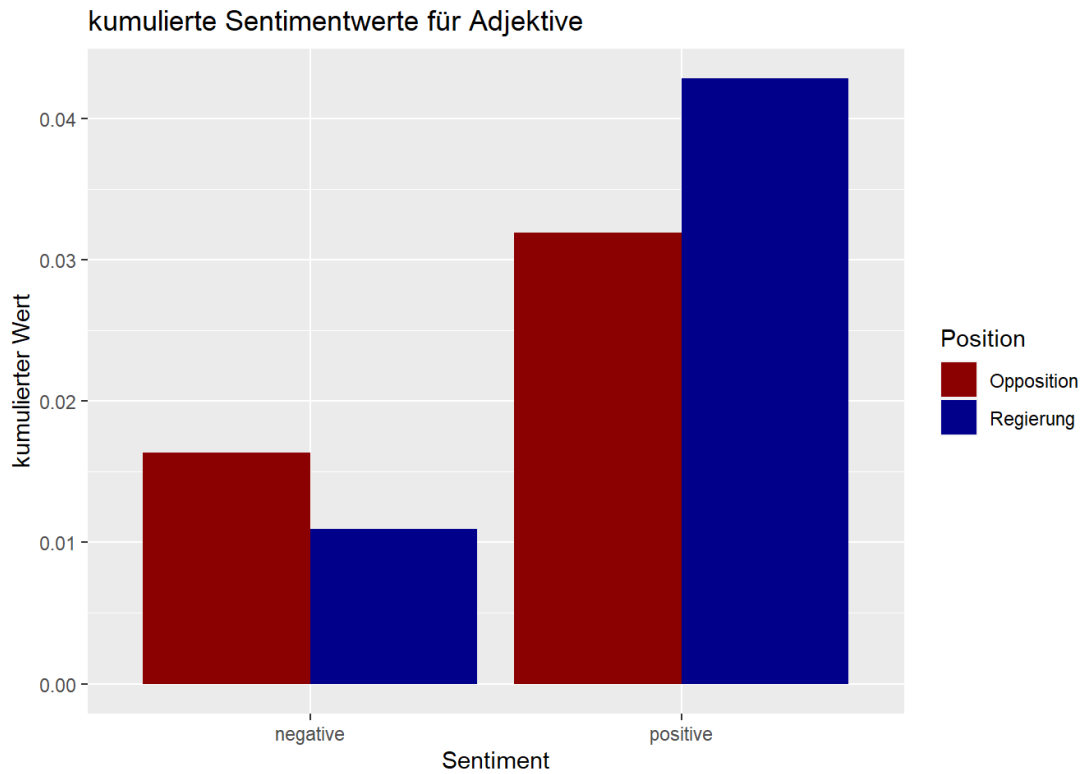


Abbildung 19: Frequentive Nomen bei Oppositions- und Regierungsparteien

8 Fazit und Ausblick

Durch unsere Untersuchungen konnten wir feststellen, dass durchaus Unterschiede zwischen Regierungs- und Oppositionsparteien festzustellen sind. Durch Sentiment und die Art der Prädikate, die in selbstreferentiellen Sätzen auftreten, zeigt sich, dass Parteien, die zum Zeitpunkt der Wahl an der Regierung beteiligt waren, eher auf eine negative Darstellung verzichten, sich mehr auf die Vergangenheit beziehen und ihre bisherigen Erfolge besprechen, was sich vor allem an den Vergangenheitsformen der Verben ablesen lässt. Oppositionsparteien verwenden dagegen vergleichsweise oft weniger positive Begriffe und konzentrieren sich mit Verben wie *fordern* eher auf die Zukunft.

Ein Problem der Betrachtungen hier ist, dass nicht endgültig festgestellt werden kann, ob die Unterschiede wirklich durch die Gegenüberstellung von Regierung und Opposition entstehen, oder ob sie einfach durch Unterschiede in den Parteien zustande kommen. Schließlich war die CDU und die SPD überwiegend an Regierungen beteiligt, während etwa DIE LINKE nur in der Opposition vertreten war. Die TF-IDF-Terme können dabei einen Hinweis auf das Vokabular von einzelnen Parteien geben. Auch mit *bag of words* Untersuchungen kann ein erster Überblick über die Wortwahl der Parteien gegeben werden. Zukünftig könnten diese Untersuchungen noch ausgebaut werden. Es wäre beispielsweise interessant, einen weiteren Zeitraum zu betrachten und Programme anderer Wahlen wie etwa der Landtags- oder Europawahl hinzuzuziehen.

Die hier ausgewählten Kriterien Sentiment, selbstreferentielle Sätze, Adjektive und Nomen könnten dann etwa dazu genutzt werden, einen Klassifizierer zu entwickeln, der ein unbekanntes Wahlprogramm als Regierungs- oder Oppositionsprogramm identifiziert. Damit könnten diese Kriterien weiter evaluiert werden und eventuell neue Parameter aufgedeckt werden, die Regierungs- von Oppositionsprogrammen unterscheiden.

9 Literatur

- R. Remus, U. Quasthoff G. Heyer: *SentiWS - a Publicly Available German-language Resource for Sentiment Analysis*. In: Proceedings of the 7th International Language Resources and Evaluation (LREC'10), pp. 1168-1171, 2010
- Bundeszentrale für politische Bildung. *Bundestagswahlen 1949 - 2009*. Abgerufen von <https://www.bpb.de/politik/wahlen/bundestagswahlen/62559/bundestagswahlen-1949-2009>
- Bundeszentrale für politische Bildung. *Die Bundestagswahlen seit 1949: Parteiensystem und Koalitionen*. Abgerufen von <https://www.bpb.de/izpb/250399/bundestagswahlen-seit-1949>.

—