

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

Bericht Korpuserstellung und erste Untersuchungen

Framing in Wahlprogrammen

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	Ansatz	1
2	Aufbau Korpus	1
3	Erste Untersuchungen	2
3.1	Vorverarbeitung	2
3.2	Bag of Words	2
3.3	Keywords in Context	3
3.4	Kollokationen	6
3.5	TF-IDF	6
3.6	Topic-Modeling	7
4	Ausblick	8
5	Anhang	9

1 **Ansatz**

Welche Macht eine geschickte Rhetorik im Prozess der öffentlichen Meinungsbildung haben kann, verdeutlichen bereits die "großen Redner der Antike". Dabei ist das sogenannte Framing im politischen Tagesgeschäft allgegenwärtig. Um die Allgemeinheit vom eigenen Standpunkt zu überzeugen, werden möglichst eindrücklicher Bilder und Assoziationen verwendet. Durch bestimmte Formulierungen soll die Denkweise der Adressat:innen beeinflusst werden. Mit dem Begriff *Klimakrise* soll die Dringlichkeit eines Entgegensteuerns verdeutlicht werden. Wer von *Asyltourismus* spricht, versucht, Fluchtursachen in einem bestimmten Licht erscheinen zu lassen.

Alle vier Jahre steigen die Bemühungen der deutschen Politiker:innen, ihre Ansichten unter Volk zu bringen, schlagartig an: Der Bundestag wird gewählt. Hierbei entscheidet sich, welche Parteien die Wähler am meisten ansprechen und folglich die politischen Entscheidungen der nächsten Legislaturperiode prägen werden.

Ein wichtiges Instrument des Wahlkampfes sind Wahlprogramme. In diesen Schriftstücken stellen die jeweiligen Parteien ihre unterschiedlichen Ziele und Ansichten dar. Wir wollen untersuchen, wie sich Framing in Wahlprogrammen gestaltet. In diesem Bericht beschreiben wir den Prozess der Korpus-Erstellung sowie wie die Auswertung einiger erster Untersuchungen

2 **Aufbau Korpus**

Als Grundlage für unseren Korpus dienen die Wahlprogramme der größeren deutschen Parteien zur Bundestagswahl von 2002 bis 2017. Wir haben uns dabei für die Parteien CDU, SPD, AfD, FDP, DIE LINKE (vor 2007 die PDS) und Bündnis 90/die Grünen entschieden. Der Zeitraum wurde ebenfalls bewusst gewählt, um eine größtmögliche Ausgeglichenheit des Korpus zu gewährleisten und den Einfluss späterer Gründungsjahre zu minimieren. Ziel war außerdem, die Entwicklung der politisch wichtigen Themen der letzten 20 Jahre untersuchen zu können.

Zugänglich sind die Parteiprogramme im Internet, beispielsweise auf den offiziellen Webseiten der Parteien selbst oder durch parteinahe Stiftungen. Um automatische Textanalysen durchführen zu können, wurden die Parteiprogramme automatisch mittels PDF-zu-txt-Konvertierung in Plaintext-Format umgewandelt. Hierbei haben wir uns an folgende Herangehensweise gehalten:

- Seitenzahlen, Impressum und Stichwortverhältnis entfernen
- Zeilenumbrüche auflösen
- beim Übertragen des Parteiprogrammes in Plaintext-Format entstandene Codierungsfehler beheben.
- Teils Korrektur aufgrund zweispaltigen Layouts vertauschter Textabschnitte.
- teils Korrektur falsch erkannter Zeichen (entstanden bei Konvertierung bildbasierter PDFs)

3 Erste Untersuchungen

Für genauere Erklärungen zum Code siehe *exploration.Rmd*

3.1 Vorverarbeitung

Alle Untersuchung des Korpus wurde mithilfe der R-Bibliothek *quanteda* durchgeführt. Um aussagekräftige Ergebnisse zu erzielen, wurden zunächst Stoppwörter und Satzzeichen aus dem Korpus entfernt. Dabei wurde die Liste von Token genutzt, die in *quanteda* enthalten ist.

3.2 Bag of Words

In Abbildung 1 sind die 100 häufigsten Token im gesamten Korpus dargestellt. Die Größe eines Token korrespondiert dabei in diesem Plot mit der Häufigkeit seines Auftretens. Die absolute Häufigkeit der Token in dieser Abbildung reicht von etwa 3500 (*dass*) bis zu ca. 400 (*teilhabe*).



Abbildung 1: Die 100 häufigsten Token nach Entfernung von Stoppwörtern

Obwohl kein Token dominiert, zeigen sich schon hier Themen, die in Wahlprogrammen behandelt werden. So tauchen etwa Begriffe wie *eu*, *europäisch* und *euro* auf, die im Zusammenhang mit der Europäischen Union und Europa stehen. Hervorzuheben sind zudem die auftretenden Verben wie *stärken*, *ermöglichen*, *verbessern* und *fördern*. Sie könnten darauf hindeuten, dass Wahlprogramme einen zukunftsorientierten, positiven Charakter besitzen.

Viele der 100 häufigsten Token - wie etwa *chancen*, *besser*, *freiheit* - scheinen positiv konnotiert zu sein. Unter ihnen befindet sich kein Begriff, der offensichtlich negative Assoziationen weckt.

In Abbildung 2 werden jeweils die häufigsten Token der verschiedenen Parteien gegenübergestellt. Laut der Dokumentation¹ wird dabei die relative Häufigkeit eines Terms in mehreren Dokumenten² miteinander verglichen, wobei ein Term einem Dokument zugeordnet wird, wenn er am häufigsten in diesem Dokument verwendet wurde. Die Größe eines Terms ist abhängig von der Differenz der Häufigkeit in der zugeordneten Gruppe und der durchschnittlichen Häufigkeit in den übrigen Gruppen.

Wenig überraschend zeigt sich, dass der Term, der am häufigsten gegenüber dem Durchschnitt auftritt, bei vielen Parteien der eigene Parteiname ist. Auffällig ist aber, dass die Terme nicht gleichmäßig zwischen den Parteien aufgeteilt sind. Vor allem der Anteil der Terme der SPD und von Die Grünen ist geringer.

¹www.rdocumentation.org/packages/wordcloud/versions/2.6/topics/comparison.cloud

²Dokumente bezeichnen in diesem Fall Parteien

Das deutet darauf hin, dass in den Wahlprogrammen dieser Parteien die Häufigkeit vieler Terme nicht sehr vom Durchschnitt der Wahlprogramme der anderen Parteien abweicht. Dagegen scheint die AfD viele Begriffe häufiger in ihren Wahlprogrammen zu verwenden als andere Parteien: Terme wie *zuwanderung*, *asyl* und *islam* könnten auf wichtige Themen der AfD hinweisen. Zumindest zeigt sich aber, dass die AfD diese Begriffe eher verwendet als andere Parteien.

Auch in den Termen anderer Parteien zeigen sich Themen und Einstellungen, die plausibel erscheinen. Die FDP verwendet etwa häufiger Begriffe wie *liberal*, *wettbewerb* und *freiheit*, die im Zusammenhang mit einer liberalen Marktwirtschaft zu stehen scheinen. Auch die Terme der PDS und von DIE LINKE - beispielsweise *arbeitslosigkeit*, *armut* und *grundsicherung* - deuten auf Themen wie soziale Gerechtigkeit hin. Weniger aussagekräftig scheinen dagegen die Begriffe der SPD.

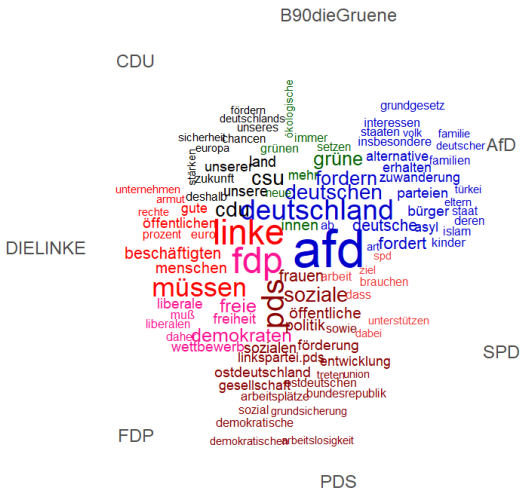


Abbildung 2: Vergleich der Parteien

3.3 Keywords in Context

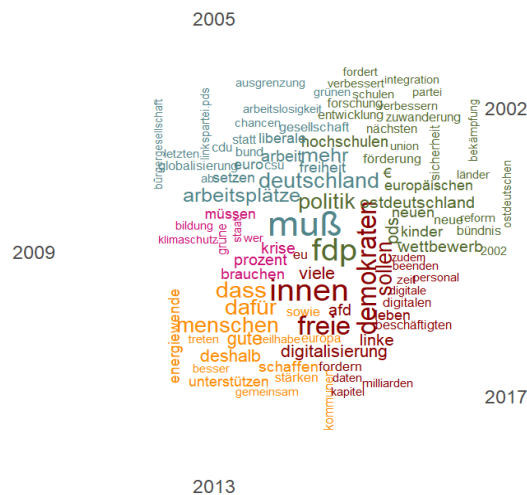


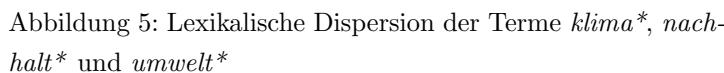
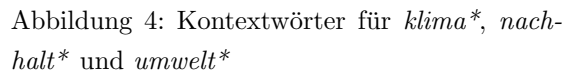
Abbildung 3: Vergleich der Jahre

Mit dem gleichen Verfahren werden die verschiedenen Jahre, aus denen die Wahlprogramme stammen, in Abbildung 3 verglichen. Auch hier lassen sich Themen erahnen, die in den jeweiligen Jahren im Vordergrund standen. So werden beispielsweise im Jahr 2017 häufiger Begriffe wie *digitalisierung*, *digital* und *daten* verwendet. 2002 werden dagegen eher Terme wie *ostdeutschland* und *ostdeutschen* gebraucht.

Terme wie *arbeitslosigkeit*, *arbeitsplätze* und *arbeit* im Jahr 2005 deuten auf die hohe Arbeitslosigkeit³ in den Jahren zuvor hin. Dieses

³www.bpb.de/nachschlagen/zahlen-und-fakten/soziale-situation-in-deutschland/61718/arbeitslose-und-arbeitslosenquote

Außerdem treten verschiedene Bereiche wie *landwirtschaft*, *wirtschaft* und *mobilität* auf, die mit diesen Themen in Verbindung stehen. Ein weiterer Aspekt dieser Themen zeigt sich in den Termen *zukunft*, *gerechtigkeit* und *verantwortung*, die eher moralische Konnotationen zu betrachten, wie und wie oft diese B



4

erweitert werden. Außerdem sollten die Kontextwörter der Begriffe nach Parteien unterschieden werden, um zu untersuchen, welche Partei welche Begriffe auf welche Art und Weise verwendet. Zusätzlich könnte die Häufigkeit der Begriffe über die Wahljahre hinweg betrachtet werden.

3.4 Kollokationen

Die Betrachtung häufiger Bigramme erfolgt zunächst ohne Stoppwörter. Das häufigste Bigramme ist dabei *europäische union*⁵, was die Wichtigkeit des Themas Europa auch in Wahlprogrammen zur Bundestagswahl hervorhebt.

Kollokationen wie *unser land* und *unsere gesellschaft* zeigen außerdem, dass Parteien in Wahlprogrammen oft eine gemeinschaftliche Perspektive einnehmen. Das Thema Klima klingt durch das Bigramm *erneuerbare energien* an.

Werden Stoppwörter in die Berechnung der häufigsten Bigramme einbezogen, so zeigen sich zwar größtenteils inhaltslose Kollokationen (*für den, in der*). Auffallend ist dagegen die häufige Verwendung von Wortfolgen wie *wir werden*, *wir setzen* und *wir wollen*, die alle auf Absichten und Versprechungen der Parteien hindeuten. Auch auf Deutschland wird sich häufig durch die Kollokation *in deutschland* bezogen.

3.5 TF-IDF

Nach Parteien gruppiert können durch die Berechnung des TF-IDF Scores relevante Terme für jede Partei bestimmt werden. Für die AfD zeigen sich dabei Terme wie *gender-forschung*, *gender-ideologie*, also Bereiche, denen die AfD eher kritisch gegenübersteht. Begriffe wie *deutsch-türkisch* und *kultursorten* könnten sich auf Themen wie Immigration und Integration beziehen, auf die die AfD einen großen Fokus legt.

Für die SPD zeigen sich zunächst wenig interessante Terme wie *sozialdemokratisch* und *sozialdemokraten*. Charakteristischer scheinen dagegen Begriffe wie *solidarrente* und *familienarbeitszeit*: Konzepte, die von der SPD vorgeschlagen und vertreten werden. Auch *verantwortung* scheint in den Wahlprogrammen der SPD ein relevanter Term zu sein. Interessanterweise taucht *verantwortung* auch bei der CDU als relevanter Term auf. Der Begriff *schöpfung* bezieht sich auf den christlichen Hintergrund der Partei. Zudem zeigen sich in den Programmen der CDU Wörter wie *heimatvertrieben* und *zuwanderungsgeschichte*, die sich auf Immigration und Flucht beziehen.

Für DIE LINKE ergeben sich Terme wie *erwerbslosigkeit*, *massenerwerbslosigkeit* und *mindestsicherung*. Bezeichnende Terme für die Wahlprogramme von Die Grünen sind unter anderem *massentierhaltung*, *gesellschaftsvertrag* und *einmischen*. Für die Programme der FDP zeigen sich Begriffe wie *liberal*, *marktwirtschaft* und *weltbeste*. In den Wahlprogrammen der PDS sind Terme wie *bedarfsorientiert*, *umweltunion* und *beschäftigungssektoren* besonders relevant.

⁵Dabei werden hier die Bigramme *europäischen union* und *europäische union* zusammengezählt. Diese kommen zusammen 290 mal vor.

3.6 Topic-Modeling

Über das package „seededlda“ lassen sich Topic Models erstellen, dabei kann man über den Parameter k die Anzahl der erstellten Topics festlegen. Gerade unter Verwendung des Gesamtkorpus zeigen sich hier aussagekräftige Ergebnisse. Erstellt man etwa ein Modell mit 10 Topics, wie in Abbildung 8 im Anhang, so lassen sich bereits die Standpunkte und Hauptaugenmerke einzelner Parteien erkennen. B90/Die Grünen scheint in ihren Wahlprogrammen Wert auf *ökologie*, *verbraucherinnen* und *teilhabe* zu legen, außerdem *geflüchtete*, *digitalisierung* sowie *jobs*. In den zwei Topics, die der FDP zugeordnet werden können, finden sich die Begriffe *wettbewerb*, *freiheit*, *digitalisierung*, *altersvorsorge*, *weltbeste* und *bürokratie*. Auch DIE LINKE und die PDS haben in diesem Beispiel je ein Topic. Hier stechen Begriffe wie *ostdeutschland*, *euro*, *beschäftigten*, *globalisierung* und *arbeitslosigkeit* hervor. Mit der Situation der Beschäftigten scheint sich auch die SPD ganz besonders zu befassen, sie wird mit Begriffen wie *arbeitnehmerinnen*, *verbraucherinnen*, *industriepolitik* und *gerechtigkeit* in einem Topic gruppiert. Weniger konkret sind dafür die Begriffe im Topic der CDU/CSU. Wichtig scheint hier das allgemeine Wohlergehen Deutschlands und weniger spezifische Kernthemen. Die Begriffe in diesem Topic sind *deutschland*, *land* und *unseres*. Schon etwas konkreter sind *wachstum* und *chancen*. Zentrale Begriffe im Topic der AfD sind *zuwanderung*, *lehnt*, *parteien* und *deutsche*. Ein letztes Topic ist keiner Partei zuzuordnen. Es scheint mit Begriffen wie *menschen* und *deutschland* und *stärken* auch kein klar umgrenztes Thema abzubilden, vielmehr finden sich hier auch einige Begriffe, die man eventuell der Liste der Stopwörter hinzufügen könnte, etwa *dass*.

Insgesamt bestätigen die Topics im erstellten Modell die Erwartungen bezüglich der Kernthemen der Parteien. Mithilfe von Topic Modeling lassen sich also diejenigen Aspekte herausarbeiten, welche die betrachteten Parteien voneinander abgrenzen und definieren. Wichtig ist dabei, zu beachten, dass keine Vollständigkeit gewährleistet werden kann. So geht aus dem Beispiel-Modell hervor, dass die AfD sich stärker als die anderen Parteien mit dem Thema „Zuwanderung“ beschäftigt, auf ihr ursprüngliches Kernthema, die Ablehnung des Euro, deutet jedoch kein Begriff mehr hin.

Der Versuch, Topic Models⁶ auf Basis der Programme einzelner Parteien zu generieren, verlief weniger erfolgreich als erwartet. Bei einem k -Wert von 20 zeigen sich noch keine klar umgrenzten Bereiche. Die Erwartung, so eine Übersicht über die verschiedenen Themen in den Wahlprogrammen zu erhalten, hat sich nur in Teilen bestätigt. Zwar gibt es Topics, die ein Thema aus dem Wahlprogramm darzustellen scheinen, so scheint es in Topic 6 aus dem Beispiel eines Topic Models auf Grundlage der grünen Wahlprogramme um die Rechtfertigung von Auslandseinsätzen zu gehen. Teilweise scheinen jedoch mehrere Themenbereiche in einem Topic vereint, etwa „geflüchtete“, „verbraucher“ und „digitalisierung“ in Topic 20. Vielleicht würde hier in Anbetracht der Vielfalt von Themen, die in einem Wahlprogramm abgedeckt werden ein höherer k -Wert helfen, jedoch zeigen sich auch bei k -Werten von 30 oder 40 keine eindeutigen Ergebnisse, ebenso wenig bei einem Niedrigerem.

Erstellt man Topic Models für einzelne Jahre, so kann man auch den Wandel der bestimmenden Themen im zeitlichen Kontext verfolgen. Zwar erhält sich die deutliche Struktur von klar einer Partei zuzuordnenden Topics nur in Teilen, vermutlich aufgrund der verringerten

⁶Alle weiteren Models siehe `exploration.Rmd`

Korpus-Größe, jedoch lassen sich durchaus einige Erkenntnisse gewinnen: So war *rot-grün* für die CDU/CSU im Jahr 2002 ein prägendes Thema. Dies ist sicherlich darauf zurückzuführen, dass die Union zu dieser Zeit in der Opposition war, während eine rot-grüne Koalition regierte. Auch Ablehnung scheint eine vorherrschende Einstellung zu sein, darauf deutet der Begriff *lehnen* hin. 2017 dominieren hingegen die Begriffe *erfolg*, *deutschland* und Arbeitsplätze das Topic der CDU/CSU. Für die Grünen wiederum ist (den Erwartungen entsprechend) sowohl 2002 als auch 2017 das Thema Ökologie wichtig.

4 Ausblick

Mit den verschiedenen Werkzeugen, welche Quanteda bereitstellt, lassen sich bereits zahlreiche Aussagen über die Wahlprogramme treffen. Besonders die Kernthemen der einzelnen Parteien lassen sich auch im Vergleich der Wahlprogramme darstellen. Immer wieder begegnen einem auch bei solch allgemeinen Untersuchungen Begriffe, die ein bestimmtes Sichtweise darstellen. So ist in den Wahlprogrammen der LINKEn von *massenarbeitslosigkeit* die Rede, für die FDP scheint der Begriff *freiheit* zentral und die AfD beschäftigt sich gar mit *kultursorten*.

Mit der Verwendung derartiger Begriffe wollen wir uns in der Projektleistung tiefergehend befassen. Nicht nur wollen wir herausarbeiten, welche Partei welche Frames und konkret welche wertenden Begriffe verwendet. Ziel ist außerdem, zu untersuchen, ob die verwendeten Frames andere sind, wenn eine Partei an der Regierung beteiligt ist. Schlussendlich wollen wir also fragen: „Welche Partei verwendet welches Vokabular - und wie wirkt sich eine Regierungsbeteiligung darauf aus?“

5 Anhang

Abbildung 8: Topic Model mit k=10 auf Grundlage des Gesamtkorpus

```
##      topic1      topic2      topic3      topic4      topic5
## [1,] "deutschland" "fdp"      "spd"      "afd"      "ostdeutschland"
## [2,] "cdu"        "liberale" "arbeitnehmerinnen" "deutschen" "politik"
## [3,] "csu"        "muß"      "verbraucherinnen" "fordert"   "pds"
## [4,] "land"       "wettbewerb" "marktwirtschaft" "zuwanderung" "bündnis"
## [5,] "unserer"    "liberalen" "sozialdemokraten" "deutsche"  "globalisierung"
## [6,] "chancen"   "setzt"    "mitbestimmung"  "lehnt"    "ostdeutschen"
## [7,] "unsere"    "freiheit" "sozialdemokratinnen" "parteien" "\200"
## [8,] "unseres"   "fordert"  "fortsetzen"     "fordern"  "arbeitslosigkeit"
## [9,] "wachstum"  "staat"    "industriepolitik" "deutscher" "reform"
## [10,] "deutschlands" "hochschulen" "idee"          "alternative" "erneuerung"
##      topic6      topic7      topic8      topic9      topic10
## [1,] "linke"      "freie"      ""          "innen"     "grüne"
## [2,] "müssen"     "demokraten" "dass"      "grüne"     "grünen"
## [3,] "öffentliche" "fordern"    "menschen" "digitalisierung" "ökologischen"
## [4,] "beschäftigten" "digitalisierung" "mehr"     "gutes"     "verbraucherinnen"
## [5,] "soziale"    "daher"      "müssen"   "geflüchtete" "grün"
## [6,] "öffentlichen" "daten"      "deutschland" "sorgen"    "ökologische"
## [7,] "euro"       "beispiel"   "dafür"    "kämpfen"   "teilhabe"
## [8,] "statt"      "altersvorsorge" "deshalb"  "drei"      "endlich"
## [9,] "sozial"     "weltbeste"  "setzen"   "oft"       "wählt"
## [10,] "fordert"   "digitalen"  "stärken"  "viele"     "jobs"
```