# Erste Untersuchungen

#### Katja Konermann

Für genauere Erklärungen zum Code siehe exploration. Rmd

### 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.

#### 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).

internationalen

bessere öffentliche
prozent förderung möglich
öffentlicherwirtschaftforschung
dürfen freiheit unterstützen europäische
insbesondere kinderlymmer stehen
beim andern schaffen frauen stehen
besser spziale
leben habeit neuen
länder stärken dass Gopolitikozialen j
stärken stehen schaffen scha

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. könnten darauf hindeuten, Wahlprogramme einen zukunftsorientierpositiven Charakter 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.

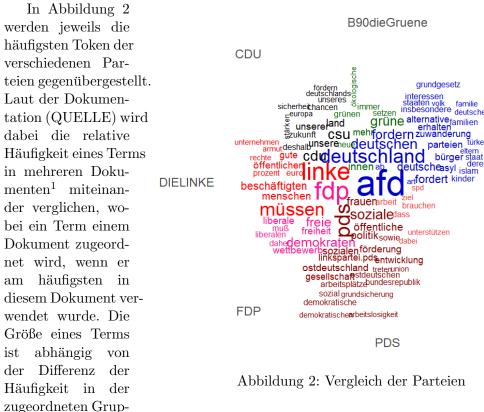


Abbildung 2: Vergleich der Parteien

**PDS** 

grundgesetz

bürger

SPD

artfordert kinde

grüne alternative familien

pe und der durchschnittlichen Häufigkeit in den anderen 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 Grüne ist geringer. 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 Begriff der SPD.

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

<sup>&</sup>lt;sup>1</sup>Dokumente bezeichnen in diesem Fall Parteien

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 Arbeitslosigkeit (QUELLE) in den Jahren zuvor hin. Dieses Thema scheint sich so auch in den Wahlprogrammen wiederzufinden. In den Jahren 2009 und 2013 treten die Terme klimaschutz, grüne und energiewende häufiger und könnten anzeigen, dass Themen wie Klimawandel und Umweltschutz an Wichtigkeit im Wahlkampf gewinnen.

### 3 Keywords in Context

Die Betrachtung der Kontexte von bestimmten Schlüsselwörtern kann einen ersten Einblick in die Darstellung von Themen wie Klimawandel und Europa geben. Häufige Begriffe, die in einem Fenster von 10 Token um die Terme umwelt\*, klima\* und nachhalt\* auftreten, sind in Abbildung 4 dargestellt. Das dominierende Verb müssen könnte darauf hinweisen, dass den Themen Klimaschutz und -wandel in Wahlprogrammen oftmals eine große Dringlichkeit zugemessen wird. Außerdem treten verschiedene Berei-

2009

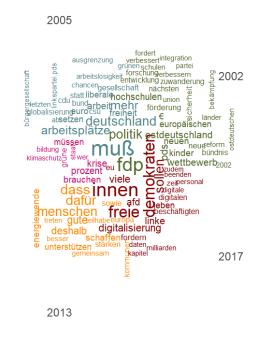


Abbildung 3: Vergleich der Jahre

che 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 besitzen scheinen. Es ist weiterhin interessant zu betrachten, wie und wie oft diese Begriffe in den Wahlprogrammen verteilt sind. In Abbildung 5 ist die lexikalische Dispersion der Begriffe rund um Klima- und Umweltschutz dargestellt. Hier zeigt sich, dass die Grüne die Terme am häufigsten und am großflächigsten verwendet. Die AfD dagegen verwendet die Begriffe weitaus seltener: Im Wahlprogramm von 2013 tauchen die Terme umwelt\* und klima\* sogar gar nicht auf. Auch die FDP und die PDS gebrauchen diese Terme nur selten. In manchen Programmen wie CDU 2009 und 2013 scheinen die Begriffe über das gesamte Programm verteilt zu sein, während es in anderen Programmen wie etwa

DIE LINKE von 2013 und 2017 bestimmte Stellen im Text gibt, wo sich das Auftreten der Terme ballt.

Auf die gleiche Weise werden in Abbildung 6 die Kontextwörter für  $eu^*$  und  $europ^*$  dargestellt. Viele der dominierenden Terme sind nicht besonders interessant (europa, europäische, union). Aber auch hier tritt als dominierendes Verb  $m\ddot{u}ssen$  auf. Weitere Verben wie  $st\ddot{u}rken$ , schaffen und (ein-)setzen deuten darauf hin, dass das Thema Europa in Wahlprogrammen meist positiv besetzt ist. Nomen wie frieden und sicherheit zeigen wichtige Ziele der EU. Dass Europa in Wahlprogrammen als ein gemeinschaftliches Unterfangen gesehen wird, zeigt sich an Termen wie gemeinsam, zusammenarbeit und  $unterst\ddot{u}tzen$ . Abbildung 7 zeigt die lexikalische Dispersion für die Terme  $eu^*$  und  $europ^*$ . Dabei ist auffallend, dass die beiden Terme häufiger gebraucht werden als die Terme zum Thema Klima. Fast in allen Programmen finden sich Stellen wieder, an denen die Terme sehr verdichtet auftreten. Häufig tritt diese Ballung eher zum Ende der Wahlprogramme auf. Besonders oft scheint die FDP und Die Grüne diese Terme zu verwenden. Selten dagegen treten sie wiederum in dem Programm der AfD von 2013 auf. Auch die PDS scheint die Begriffe nicht so häufig zu gebrauchen.

Im weiteren Verlauf des Projektes wäre es interessant, die Darstellung die Themen Klima und Europa noch weiter aufzuschlüsseln. Dafür könnte etwa das Wörterbuch noch 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.

#### 4 Kollokationen

Die Betrachtung häufiger Bigramme erfolgt zunächst ohne Stoppwörter. Das häufigste Bigramme ist dabei europäische union<sup>3</sup>, 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.

 $<sup>^2 {\</sup>rm Für}$  das Schlüsselwort  $eu^*$ ist dies mit<br/>unter nicht so eindeutig, da hier auch Geldangaben wi<br/>e25 Milliarden Euro (CDU, 2013) miteinbezogen werden

<sup>&</sup>lt;sup>3</sup>Dabei werden hier die Bigramme europäischen union und europäische union zusammengezählt. Diese kommen zusammen 290 mal vor.

#### 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 genderforschung, 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 Grüne 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.

## 6 Topic Modelling



Abbildung 4: Kontextwörter für  $klima^*$ ,  $nachhalt^*$  und  $umwelt^*$ 

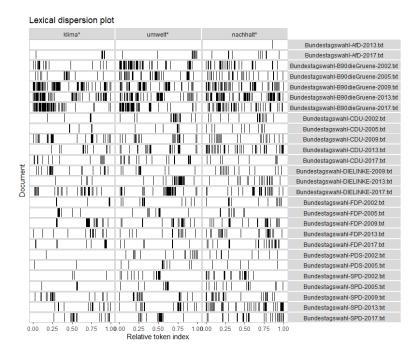


Abbildung 5: Lexikalische Dispersion der Terme  $klima^*$ ,  $nach-halt^*$  und  $umwelt^*$ 



Abbildung 6: Kontextwörter für  $eu^*$  und  $europ^*$ 

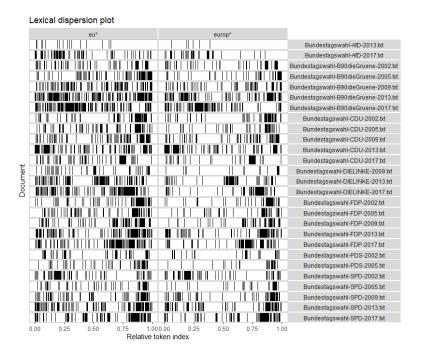


Abbildung 7: Lexikalische Dispersion der Terme  $eu^*$  und  $eu-rop^*$