

Schlagwörter im Schweizer Energiediskurs. Korpuslinguistische Methoden zur Identifikation von salienter Lexik im Diskurs

Maren Runte & Julia Krasselt

1. Einleitung: Schlagwörter im Schweizer Energiediskurs¹

Welche sprachlichen Merkmale charakterisieren den öffentlichen Energiediskurs in der Schweiz? Welche datengeleiteten, korpuslinguistischen Methoden eignen sich, Schlagwörter im Energiediskurs zu identifizieren? Der vorliegende Beitrag zeigt Zugänge der digitalen Linguistik zur Beantwortung dieser Forschungsfragen.

Voraussetzung für die Bearbeitung dieser Forschungsfragen ist ein themenspezifisches Korpus, welches möglichst umfassend den Energiediskurs in der Schweiz repräsentiert. Ausgangsbasis für die Erstellung eines solchen Korpus ist das von der ‚Zürcher Hochschule für Angewandte Wissenschaften‘ (ZHAW) erstellte, linguistisch annotierte Korpus zum Schweizer Sprachgebrauch („Swiss-AL“), in dem zahlreiche, öffentlich zugängliche Schweizer Onlinequellen vertreten sind.² Das mehrsprachige Korpus umfasst aktuell insgesamt ca. 1,19 Mia. Token in den Sprachen Deutsch, Französisch und Italienisch, das Subkorpus mit Texten von Akteuren aus der Deutschschweiz („Swiss-AL-DE“) zählt ca. 700 Mio. Token. Mittels eines Topic-Modeling-Verfahrens (vgl. Blei/Ng/Jordan 2003) konnte aus dem deutschsprachigen Gesamtkorpus ein themen-, das heißt energiespezifisches, deutschsprachiges Subkorpus ‚Swiss-AL-ED-DE‘ (ED = Energiediskurs) mit einem Umfang von ca. 15,7 Mio. Token gebildet werden, welches die Basis für die vorliegende Untersuchung bildet. Das Ener-

1 Die hier vorgestellten Daten und Analysen stammen aus dem Forschungsprojekt ‚Energiediskurse in der Schweiz‘ (vgl. Stücheli-Herlach/Ehrensberger-Dow/Dreesen 2018). Das Projekt wird vom Schweizerischen Bundesamt für Energie (BFE) gefördert und an der Zürcher Hochschule für Angewandte Wissenschaften (ZHAW) durchgeführt (Projektzeitraum: 2016–2019).

2 Weitere Informationen unter <https://www.zhaw.ch/de/linguistik/forschung/swiss-al/> (18.07.2019).

giediskurskorpus enthält öffentlich und online publizierte Aussagen von Akteuren aus den Bereichen Medien, Politik, Wirtschaft und Wissenschaft (vgl. Stücheli-Herlach/Ehrensberger-Dow/Dreesen 2018, 28). Der Akteursliste liegt ein sorgfältiges Sampling der Quellen zugrunde: Wichtig ist die Bedeutung des Akteurs in der gesamten Schweiz, die Bedeutung für die jeweilige Sprachregion und die Bedeutung für den Schweizer Energiediskurs.

Das Korpus ‚Swiss-AL-ED-DE‘ ist, gemeinsam mit seinen französisch- und italienischsprachigen Pendanten, die Datengrundlage für das Projekt ‚Energiediskurse in der Schweiz‘ (vgl. Fußnote 1). Neben zahlreichen weiteren Fragestellungen, etwa der unterschiedlichen Rahmung des Energiediskurses in den Landessprachen, der unterschiedlichen Themengewichtung, der Rekonstruktion von Akteursnetzwerken etc., zentriert sich das Forschungsinteresse des Projektes auch auf die Untersuchung von salienter Lexik im Diskurs. Eine wichtige Frage hierzu ist die nach dem Verhältnis zwischen korpuslinguistisch ermittelten Keywords und salienter Lexik im Sinne von Schlagwörtern (zu den Termini ‚saliente Lexik‘ und ‚Schlagwort‘ vgl. Schröter 2015, 395-396). Schlagwörter sind deshalb so wichtig, weil sie das „gemeinsame Bewußtsein oder Wollen“ (Dieckmann, zit. nach Wengeler 2017, 22) in einem Wort zusammenfassen und daher ein guter Indikator für diskursive Auseinandersetzungen sind. Dass sich Schlagwörter – da auf lexikalischer Ebene – zudem mit Mitteln der quantitativen Linguistik analysieren lassen, verstärkt noch das Interesse diese auf der Grundlage eines linguistisch annotierten Textkorpus zu untersuchen, da sich bei der Ermittlung von Schlagwörtern quantitative und qualitative Verfahren experimentell verbinden lassen. In diesem Aufsatz wollen wir daher einen methodischen Vorschlag vorstellen, wie Schlagwörter mittels Verfahren der Korpuslinguistik und des maschinellen Lernens ermittelt werden können.

2. Schlagwörter korpuslinguistisch bestimmen

Bisher werden korpuslinguistische Methoden zur Analyse von zuvor qualitativ bestimmten Schlagwörtern eingesetzt, jedoch nicht zu deren Identifikation:

„Wie oben angedeutet, dürfte es auch und gerade im Zusammenhang mit der Verbreitung korpuslinguistischer Analysemethoden für die Bestimmung und Analyse salienter Lexik in öffentlich-politischen Diskursen von Bedeutung werden zu spezifizieren, woraus sich diese Salienz begründet und worin sie sich zeigt.“ (Schröter 2015, 408)

Storjohann und Schröter sehen es als Desiderat, Schlagwörter nicht nur durch Korpusanalysen semantisch zu fassen, sondern quantitative Analyseverfahren

auch zu deren Identifikation und weiteren Analyse einzusetzen (vgl. Storjohann/Schröter 2013, 206). In den folgenden Abschnitten möchten wir zunächst die Möglichkeiten und Grenzen klassischer korpuslinguistischer Analyseverfahren wie Keyword- und Kollokationsanalysen zur Identifikation parteilicher (Kap. 2.1) und überparteilicher Schlagwörter (Kap. 2.2) aufzeigen. Kapitel 3 und 4 zeigen, wie diese Grenzen mit Hilfe von Methoden der distributionellen Semantik und des maschinellen Lernens überwunden werden können.

2.1. Keywordanalysen zur Identifikation parteilicher Schlagwörter

Die „Schlagwort“-Definition von Fritz Hermanns hat die politolinguistische Schlagwortforschung stark geprägt: „*Prototypisch* ist ein Schlagwort ein Lexem mit den folgenden Eigenschaften: Es ist *aktuell*, *parteiisch* und daher auch *appellativ*.“ (Hermanns 2007, 472; Herv. i. O.). Ausgehend von dieser Definition des Begriffs kann man Standard-Verfahren der Korpuslinguistik anwenden, um Schlagwörter zu ermitteln.

Typischerweise ermittelt man mit Hilfe von zwei Subkorpora die Lexik, durch welche sich das eine Subkorpus von dem anderen unterscheidet; das Ergebnis eines solchen Vergleichs sind Keyword-Listen für beide Subkorpora. Als Beispiele seien hier Schlagwörter von den Webseiten zweier Schweizer Parteien angeführt: Nachdem bei der Keywordbestimmung mit Loglikelihood zunächst die Keywords im Vergleich zu ‚Swiss-AL‘ als Referenzkorpus ermittelt und hochfrequente Funktionswörter herausgefiltert worden sind, zeigt Abbildung 1 die Visualisierung der zentralen 50 Keywords der Schweizerischen Volkspartei (SVP), basierend auf den Texten der Parteiwebseite.

Einige lexikalische Cluster sind hierbei auffällig:

1. Eigennennung der Parteien (*SVP, Volkspartei, SVP-Fraktion*) bzw. parteispezifische Angaben wie (*Zentralsekretariat*).
2. Politischer Allgemeinwortschatz bzw. „Institutionsvokabular“ (Niehr 2014, 65–66) der Schweiz (*Volksentscheid, Parlament, Bundesrat, Bund, Bundesverfassung*) und
3. Parteispezifische Lexik, welche wichtige Punkte der Parteiprogramme aufnimmt, wie *Masseneinwanderungsinitiative, Ausschaffungsinitiative, Zuwanderung, kriminell*. Dieses lexikalische Cluster wird in der Schlagwortforschung auch als „Ideologievokabular“ (ebd.) bezeichnet, wobei auch die Eigennennungen diesem Vokabular zuzurechnen wären.



Abb. 1: Keywords der gesamten SVP-Webseite



Abb. 2: Keywords der SVP-Webseite zu Energiethemen



Abb. 3: Keywords der SP-Webseite zu Energiethemen

Abb. 1–3: Die Größe der Wörter basiert auf dem LLR-Wert. Vergleichskorpus ist jeweils ‚Swiss-AL‘.

Ermittelt man die Keywords der Partei zu einem bestimmten Thema – hier dem Energiediskurs – ändert sich nur an dem letztgenannten lexikalischen Cluster etwas (Abbildung 2). Bei dem parteispezifischen Ideologievokabular zeigt sich deutlich die Position der Partei zur Energiewende, vergleicht man die Lexik der SVP mit der Lexik der Sozialdemokratischen Partei (SP, Abbildung 3) wird die Haltung der SVP zur Energiefrage umso auffälliger: Auf Seiten der SVP redet man von *KKW*, auf Seiten der SP von *AKW* bzw. *Atomkraftwerke*. Hier verbirgt sich hinter der Wahl einer Bezeichnung eine ganze politische Haltung: Die

SVP meidet den Begriff Atomkraft und favorisiert Kernkraft, die SP bevorzugt bzw. meidet den jeweils anderen Begriff.

Während die SVP *Kernenergie* (deutlich weniger *Atomkraft*) als Energieform nennt, ist es *Windenergie* (und Solarenergie) auf der SP-Seite. Kennzeichnend für die politische Haltung ist auch das Stichwort *CO₂-Abgabe* auf Seiten der SVP, eine Abgabe, die in der Schweiz für fossile Brennstoffe erhoben wird und welche nach Ansicht der SVP mit allen Mitteln vermieden werden soll, und *Einspeisevergütung* auf Seiten der SP, eine Maßnahme des Bundes, um erneuerbare Energien zu fördern, welche nach Meinung der SP nicht gedeckelt werden soll.

Die Haltung der SVP könnte nach den Keywords wie folgt zusammengefasst werden: Der *Wirtschaftsstandort* muss *bezahlbar*, bleiben, der *Wohlstand* ist durch eine *Verteuerung* bedroht, welche durch den Atomausstieg notwendigerweise entsteht. Die Schweiz muss *unabhängig* bleiben, die *Versorgungssicherheit* muss durch interne Energiepolitik gewährleistet bleiben. Während das Atomkraftwerk *Beznau* daher am Netz bleiben muss, ist das Atomkraftwerk *Mühleberg* ein Keyword der SP, die sich deutlich für dessen Abschaltung und die *Reduktion* des CO₂-Ausstoßes ausspricht.

Bewertet man nun die durch das korpuslinguistische Verfahren erhobenen Keywords der beiden Schweizer Parteien auf Grundlage der Schlagwortforschung, kann man die Ergebnisse in der Klassifikation von Burkhardt (1998) als parteiliche Schlagwörter im Sinne von Programm- (etwa *Cleantech-Initiative*) und Fahnenwörter (wie *Reduktion* (SP) und *Wohlstand* (SVP)) einstufen – Lexik, welche parteilich und wertend ist (vgl. Abbildung 4).

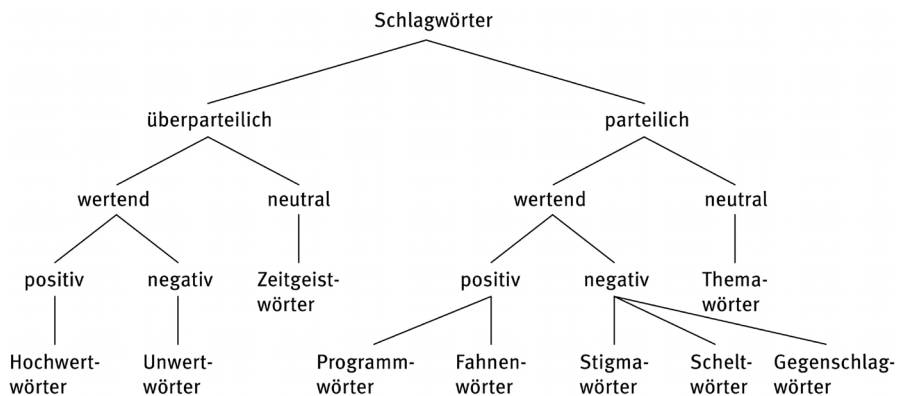


Abb. 4: Klassifikation der Schlagwörter nach Burkhardt (1998) (Abbildung von Wengeler übernommen, vgl. Wengeler 2017, 26)

2.2. Überparteiliche Keywords: Kollokationsanalysen

Deutlich schwieriger ist es, in der Gesamtmenge der im Energiediskurs verwendeten Lexik diejenigen Lexeme ausfindig zu machen, die einerseits frequent und andererseits in der Semantik bzw. in der Verwendung auffällig sind. Das Problem der Frequenz lässt sich zunächst wieder über eine Keywordanalyse lösen, indem man das gesamte Energiediskurskorpus mit einem Referenzkorpus vergleicht. Dann ist (wenig überraschend) *Energie* das wichtigste Keyword. Auffälligkeiten in der Semantik bzw. in der Verwendung zwischen verschiedenen Akteuren können dann typischerweise mit Hilfe von Kollokationsanalysen untersucht werden. Untersucht man das akteursabhängige Verständnis von *Energie*, fällt bei der Analyse von Kollokationen bei den Akteuren SP und SVP auf, dass *erneuerbar* eines der wichtigsten Kollokate ist – bei beiden Parteien. Wertet man jedoch die Kontexte aus, so wird deutlich, dass sich die SVP aus wirtschaftlichen Gründen gegen die erneuerbaren Energien ausspricht:

- (1) „[...] dass pro neuen Arbeitsplatz im Sektor der erneuerbaren Energien 2.5 Arbeitsplätze in der herkömmlichen Wirtschaft verlorengehen [...]“ (SVP)
- (2) „Für die Förderung der erneuerbaren Energien und der Energieeffizienz werden laufend zusätzliche Gebühren und Abgaben eingeführt.“ (SVP)

Eine komplementäre Position findet sich – bei gleichem Kollokat – bei der SP:

- (3) „Ein Umsteigen auf erneuerbare Energien schon nicht nur Umwelt und Klima, sondern führt durch die Schaffung von neuen Arbeitsplätzen auch zu Wertschöpfung und Wohlstand im Inland.“ (SP)
- (4) „Lenkungsabgaben auf nicht erneuerbare Energien stärken nicht zuletzt die Wettbewerbsfähigkeit der sauberen einheimischen Wasserkraft.“ (SP)

Das Keywordverfahren, in Kombination mit der Analyse von Kollokationen, kann für die Untersuchung salienter Lexik im Gesamtdiskurs folglich nur bedingt weiterhelfen:

1. Weil die Auswertung auf Keywords beschränkt ist, welche zufälligerweise bei verschiedenen Akteuren auffällig oft verwendet werden, und
2. weil ein qualitativer Analyseschritt zwingend notwendig ist, um zu zeigen, in welchem Kontext das Wort verwendet wird. Eine Beschränkung auf die Analyse der direkten Kollokationen führt hier nicht unbedingt weiter.

3. Überparteiliche Keywords: Modellierung im Vektorraum

Wie lassen sich aber Schlagwörter finden, die überparteilich verwendet werden und um deren Deutung ggf. sogar semantische Kämpfe geführt werden? Durch die Berechnung von Keywords innerhalb eines Korpus wird Lexik, die auch von anderen Akteuren, etwa in der Wirtschaft oder in den Medien verwendet wird, relativ zu dem untersuchten Subkorpus herausgefiltert, da diese insgesamt typisch für den Diskurs ist.

Eine Möglichkeit, mit bereits bekannten Verfahren Kandidaten für diese Art von Schlagwörtern zu ermitteln, ist der Vergleich von Keywords einzelner Akteure (wie SVP oder SP) bzw. Akteursklassen (Medien, Politik, Wirtschaft etc.) mit einem themenunabhängigen, großen Referenzkorpus: Berechnet man nun für die wichtigsten (oder interessantesten) Akteure die Keywords im Vergleich zu einem themenspezifischen Korpus (hier: ‚Swiss-AL‘), enthält jede Liste eben auch Lexik, die diskurspezifisch ist – diese würden durch den Vergleich innerhalb eines themenspezifischen Korpus nicht als signifikante Keywords erkannt. So enthalten die Keywordlisten von SVP und SP im Energiediskurskorpus nur den unterschiedlichen Sprachgebrauch der Parteien, nicht jedoch den für den gesamten Energiediskurs typischen Wortschatz.

Durch die getrennte Betrachtung der Keywordlisten nach Akteur wird vermieden, dass bestimmte Lexeme, die nur von wenigen, aber im Diskurs sehr präsenten Akteuren verwendet werden, als typisch für den Energiediskurs gelten. Im Gegensatz zur klassischen Analyse, wie sie in Kapitel 2 vorgestellt wurde, liegt der Fokus nun nicht auf dem unterschiedlichen Sprachgebrauch einzelner Akteure, sondern auf den Gemeinsamkeiten: Durch die einzelnen Keyword-Listen kann man einfach herausfinden, welche Lexeme von allen oder den meisten Akteuren verwendet werden – und in welcher Verteilung auf die Akteure diese vorkommen, eine Information, die durch den Abgleich des gesamten themenspezifischen Korpus (‚Swiss-AL-ED‘) mit einem themenspezifischen Korpus (‚Swiss-AL‘) verloren ginge.

Um eine Auswahl zu treffen, welche Lexeme weiterhin untersucht werden sollen, lohnt sich eine Filterung dieses Ergebnisses nach Wortarten und die Konzentration auf Nomen, Adjektive und Verben (vgl. dazu Klein 2017, 774): „Vielfach handelt es sich in semantischer Hinsicht um Abstrakta und in pragmatischer um Schlagwörter“ (ebd.).

Als Ergebnis hat man eine Liste von Schlagwort-Kandidaten, bei denen noch weiter überprüft werden muss, ob sie wirklich die folgenden Merkmale erfüllen:

1. Die Lexeme sind im Diskurs präsent, können also über Frequenz wie Signifikanztests ermittelt werden.
2. Die Lexeme stehen nicht für den typischen Sprachgebrauch eines Akteurs (z. B. einer Partei), sondern es sind überparteilich im Diskurs verwendete Lexeme, d. h. sie weisen eine weite Streuung über Akteure auf, was sich anhand der einzelnen Keywordlisten nachweisen lässt.
3. Die Lexeme werden zwar von (fast) allen Akteuren im Diskurs verwendet, die Semantik des Lexems kann jedoch je nach Akteur bzw. Akteursklasse (z. B. Medien, Politik) anders besetzt sein (sogenannte „Bedeutungskonkurrenz“, vgl. Klein 2017, 777–783).

Während es bei den parteilich verwendeten Schlagwörtern oftmals auffällig ist, dass ganz offen Bedeutungskämpfe oder Benennungskämpfe ausgetragen werden (z. B. auch durch Explizitmachen solcher Bezeichnungskonkurrenten oder durch durchgehende Verwendung einer Bezeichnung, siehe das obige Beispiel *Kernenergie* (SVP) vs. *Atomenergie* (SP)), ist an den überparteilich verwendeten Schlagwörtern interessant, dass die Bezeichnungsidentität suggeriert, dass auch die Semantik des Schlagworts die gleiche ist. Da es sich vielfach bei den überparteilich verwendeten nominalen Lexemen um Abstrakta handelt, umfasst die semantische Beschreibung häufig verschiedene Bedeutungsaspekte, die unterschiedlich hervorgehoben werden können. Dies grenzt Klein zum einen von der „ideologischen Polysemie“ nach Diekmann ab, die bei „klar gegeneinander abgrenzbare[n], ideologisch motivierte[n] Alternativbedeutungen“ (Klein 2017, 777) vorliegt. Und zum anderen grenzt er die semantische Vagheit solcher Lexeme, bei denen die Bedeutung aber noch konsensuell ausgehandelt wird, von der (mehrdimensionalen) Bedeutungskonkurrenz ab, bei denen von einzelnen Akteuren versucht wird, Begriffe inhaltlich zu konkretisieren: „Die Konkretisierung kann deskriptive, emotionale und deontische Bedeutung, Konnotationen und Referenzobjekte betreffen“ (ebd., 778). In diesen Fällen liegt Potenzial für mögliche Diskurskontroversen.

Um zu untersuchen, ob die Bezeichnungsidentität eines überparteilich verwendeten Schlagwortes auch mit einer ähnlichen Semantik einhergeht, können sogenannte Vektorraummodelle genutzt werden. Es handelt sich dabei um Verfahren der distributionellen Semantik, die auf der Annahme beruhen, dass sich Wörter, die in ähnlichen linguistischen Kontexten verwendet werden, auch semantisch ähnlich sind (Turney/Pantel 2010). Die Bedeutung eines Wortes wird im Firth'schen Sinne über seinen Kontext bestimmt: „You shall know a word by the company it keeps“ (Firth 1957, 11). Dafür wird für jedes Wort im Korpus zunächst erfasst, wie häufig es innerhalb einer vordefinierten Spanne (z. B. fünf

Wörter links bzw. rechts) mit anderen Wörtern des Korpus vorkommt (vgl. Tabelle 1).

	<i>erhöhen</i>	<i>senken</i>	<i>erneuerbar</i>	<i>nahezu</i>	<i>halten</i>	...
<i>konstant</i>	0	0	1	1	1	
<i>Kernkraft</i>	1	0	1	0	0	
<i>Anteil</i>	1	1	1	0	0	
...						

Tab. 1: Vektorrepräsentation von Worttypen (exemplarisch, mit absoluten Häufigkeiten). Die Zeilen repräsentieren die Worttypen im Korpus, die Spalten repräsentieren den Kontext. 0 = beide Wörter treten nicht gemeinsam innerhalb eines festgelegten Kontextfensters auf, 1 = beide Wörter treten im Kontext gemeinsam auf.

Jedes Wort erhält so ein ganz eigenes Profil und ist in Form eines Vektors von Zahlen (z. B. in Form von absoluten Frequenzen) darstellbar. Auf diese Weise lassen sich nun Wörter eines Korpus vergleichen, indem ihre Profile, d. h. ihre Vektoren, verglichen werden. Wörter mit einem ähnlichen Profil werden in ähnlichen Kontexten verwendet. Der Nachteil dieser Methode besteht darin, dass es sich um eine Aggregation über einzelne Belege handelt und folglich Worttypen modelliert werden. Polysemie, aber auch ganz allgemein konträre Muster im Gebrauch eines Lemmas können dabei nicht erfasst werden. So wird beispielsweise bei diesem Ansatz das Wort *Energie* mit einem Vektor erfasst, der sowohl Frequenzen für Kontexte aus dem Bereich physikalische Energie enthält als auch aus dem Bereich spirituelle Energie. Beide Lesarten von Energie sind aber durch unterschiedliche Gebrauchskontexte charakterisiert.

Basierend auf dem Ansatz von Schütze (1998) können typenbasierte Vektormodelle derart erweitert werden, dass einzelne Ausprägungen von Typen, d. h. Token, modelliert werden (in Form von Belegstellen, die das zu modellierende Token sowie Kontextinformation enthält) (vgl. auch Heylen/Wielfaert/Speelman/Geeraerts 2015 und Hilpert/Correia 2017). Solche tokenbasierten Vektormodelle beziehen zur Modellierung nicht nur die unmittelbaren Kontextwörter mit ein, sondern wiederum auch die Kontexte, in denen die Kontextwörter selbst auftreten (sogenannte Second Order Collocations, Kollokationen zweiter Ordnung). Auf diese Weise können einzelne Belegstellen (Konkordanzen) modelliert werden und Aussagen darüber getroffen werden, wie ähnlich oder unähnlich sie zu anderen Belegstellen desselben Wortes sind. Die fiktiven Beispielsätze in (1) bis (3) veranschaulichen dieses Prinzip:

- (1) Der Anteil der Windenergie bleibt stabil.
- (2) Der Beitrag der Windenergie ist konstant.
- (3) Die Schweiz steckt bei der Windenergie noch in den Kinderschuhen.

Die drei Belege von *Windenergie* teilen sich, bis auf den bestimmten Artikel, kein Vokabular. Dennoch sind sich (1) und (2) semantisch ähnlicher als beispielsweise (1) und (3), weil die Kontextwörter selbst an anderer Stelle im Korpus in ähnlichen Kontexten vorkommen. *Anteil* und *Beitrag* kommen zum Beispiel beide häufig mit *erhöhen*, *senken* oder *erneuerbar* vor, *stabil* und *konstant* häufig mit *halten*, *nahezu* oder *ziemlich*. Über die Oberflächenform der Belege (1) und (2) ist eine Ähnlichkeit folglich nicht erkennbar, aber durch Hinzuziehen der Second Order Collocations. Analog zu den typenbasierten Vektoren (vgl. Tabelle 1) können so Vektoren für einzelne Belege berechnet werden, um anschließend Aussagen über die Ähnlichkeit von Belegstellen machen zu können. Dafür wird für jedes Kontextwort des entsprechenden Belegs der Vektor aus dem zuvor berechneten Typenvektor genommen, anschließend werden die Spalten aufsummiert und durch die Anzahl Kontextwörter geteilt. Auf diese Weise erhält man einen Zahlenvektor, der den gesamten Beleg repräsentiert. Die Tabellen 2 und 3 verdeutlichen dieses Vorgehen für die Beispielsätze (1) und (2). Obwohl *Windenergie* mit ganz unterschiedlichen Kontextwörtern verwendet wird, sind die Vektoren für (1) und (2) gleich, weil die Kontextwörter über die gleichen Second Order Collocations verfügen.

	<i>erhöhen</i>	<i>senken</i>	<i>erneuerbar</i>	<i>nahezu</i>	<i>halten</i>
<i>Anteil</i>	1	1	1	0	0
<i>bleiben</i>	0	0	0	1	0
<i>stabil</i>	0	0	1	1	1
	$(0+0+1)/3$	$(1+0+0)/3$	$(1+0+1)/3$	$(0+1+1)/3$	$(0+0+1)/3$
(1)	0.33	0.33	0.66	0.66	0.33

Tab. 2: Token-Vektor für (1) „Der Anteil der Windenergie bleibt stabil.“
Die Zeilen entsprechen den Typenvektoren in Tab. 1.

	<i>erhöhen</i>	<i>senken</i>	<i>erneuerbar</i>	<i>nahezu</i>	<i>halten</i>
<i>Beitrag</i>	1	1	1	0	0
<i>sein</i>	0	0	0	1	0
<i>konstant</i>	0	0	1	1	1
	$(0+0+1)/3$	$(1+0+0)/3$	$(1+0+1)/3$	$(0+1+1)/3$	$(0+0+1)/3$
(2)	0.33	0.33	0.66	0.66	0.33

Tab. 3: Token-Vektor für (2) „Der Beitrag der Windenergie ist konstant.“
Die Zeilen entsprechen den Typenvektoren in Tab. 1.

Bezogen auf die Ausgangsfrage, ob die Bezeichnungsidentität eines überparteilich verwendeten Schlagwortes auch mit einer ähnlichen Semantik einhergeht, lässt sich die eben beschriebene Methode wie folgt anwenden:

1. Berechnung des typenbasierten Vektormodells, basierend auf dem gesamten Untersuchungskorpus („Swiss-AL-ED-DE“)
2. Extraktion von Konkordanzen, die das zu untersuchende Lemma enthalten, aus dem Korpus (inklusive Metainformationen über die Quelle)
3. Berechnung der Tokenvektoren für alle Konkordanzen
4. Darstellung der Tokenvektoren in einem zweidimensionalen Koordinatensystem (vgl. Abbildung 7), basierend auf einer Distanzmatrix, die die Ähnlichkeit aller Belege basierend auf der Cosinusähnlichkeit erfasst.³

4. Beispielanalysen: *Abgabe* und *Energieeffizienz*

Zwei ausgewählte Beispiele aus den Analysen von Schlagwörtern im Schweizer Energiediskurs sollen im Folgenden diskutiert werden. Die Visualisierung der Tokenvektoren wurde dabei nicht nur bei der Analyse der Schlagwörter genutzt, sondern auch bei deren Auswahl. Beide Beispiele weisen Auffälligkeiten bei der Verteilung auf: Während die Tokenverteilung bei *Abgabe* akteursabhängig besonders ist, zeigt sich bei *Energieeffizienz* deutlich eine verwendungsabhängige Verteilung.

4.1. Das Beispiel *Abgabe*

Obwohl das Wort *Abgabe* semantisch sehr unterschiedlich verwendet werden kann, ist durch die Einbettung in den Energiediskurs bei den folgenden Beispielen sichergestellt, dass es sich um Abgabeformen handelt, die für den Energiediskurs relevant sind. Aber selbst mit der Eingrenzung auf einen Diskurs wird deutlich, dass die Semantik dennoch sehr offen ist, was sich einerseits an den Keywords (Abbildungen 5 und 6) und andererseits an der Streuung der Tokenvektoren (Abbildungen 7 und 8) zeigt.

3 Die Berechnung der Typenvektoren erfolgte mit Hilfe von ‚Positive Pointwise Mutual Information‘ (PPMI) für die 50.000 frequentesten Types im Korpus (exkl. Stopwörter), Kontextfenster: zwei Wörter links/rechts. Für die Berechnung der Tokenvektoren wurde ein Kontextfenster von 10 Wörtern rechts/links gewählt, anschließend wurden alle Wörter entfernt, die nicht im Vokabular der Typenvektoren enthalten waren. Die Implementation erfolgte in R unter Verwendung der Pakete ‚text2vec‘ (Selivanov/Wang 2018) und ‚word-space‘ (Evert 2014). Für die Visualisierung erfolgte eine Dimensionsreduktion m. H. von ‚Multidimensional Scaling‘.

Auch bei den Tokenvektoren zeigen sich deutlich Differenzen im parteispezifischen Sprachgebrauch und ganz besonders die Sonderstellung der SVP zur Frage von *Abgaben* (vgl. Abbildung 7).

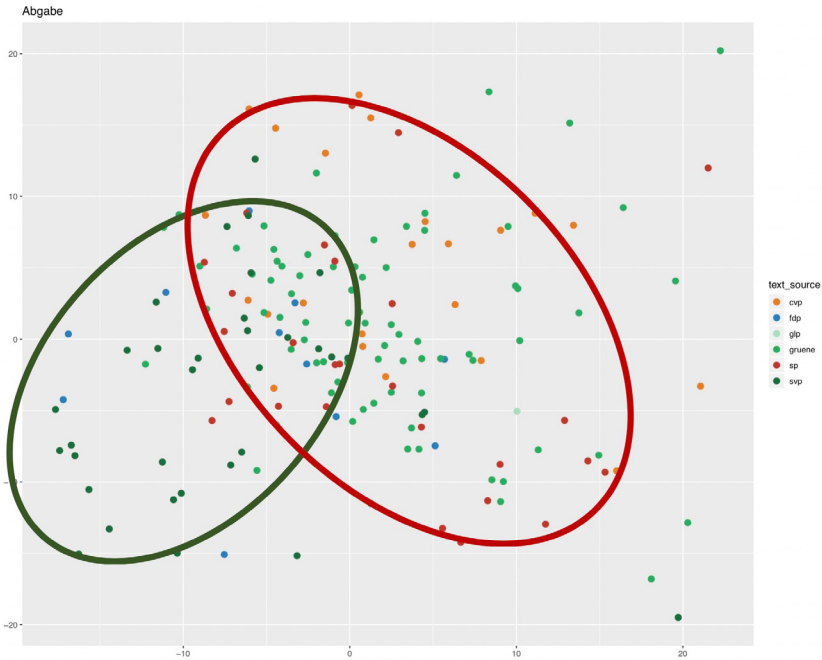


Abb. 7: Das Lexem Abgabe im Energiediskurs, beschränkt auf die wichtigsten Parteien der Schweiz. Hier die Darstellung der Tokenvektoren mit Angabe der Parteizugehörigkeit (im grünen Kreis viele Belege der SVP, im roten Kreis viele Belege der SP). Ein Punkt repräsentiert eine Belegstelle. Die Nähe von Punkten signalisiert Ähnlichkeit.

Während man an Abbildung 7 sieht, wie die verschiedenen Verwendungen (Token) von *Abgabe* in einem zweidimensionalen Raum abgebildet werden können, sieht man an Abbildung 8, die Zuordnung der einzelnen Token zu den Belegstellen im Korpus. Auf diese Weise lässt sich rekonstruieren, welche Merkmale der sprachlichen Umgebung zur Distanz bzw. Nähe der einzelnen Token in der Darstellung geführt haben.

Sowohl durch die Keywords der SVP wie auch durch die Belege zu den Tokenvektoren wird klar, dass sich die SVP ganz klar gegen Abgaben oder Steuern bzw. eine Erhöhung von Abgaben im Energiebereich ausspricht, die nach Ansicht der SVP den Wirtschaftsstandort Schweiz schwächen würde:

bfe_2253: „Steuern und Abgaben, welche den Strom künstlich verteuern, sind strikte abzulehnen.“ (SVP)

bfe_2259: „sich deshalb gegen jede weitere Verteuerung der Energie durch neue Abgaben und Steuern, als Unternehmer und Politiker muss ich [...]“ (SVP)

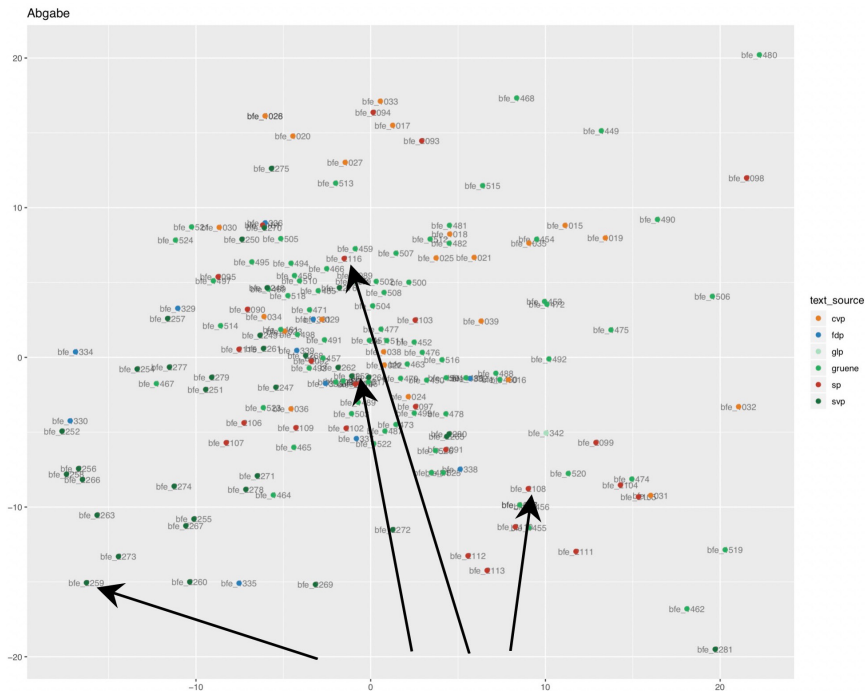


Abb. 8: Das Lexem Abgabe im Energiediskurs, beschränkt auf die wichtigsten Parteien der Schweiz. Hier die gleiche Darstellung wie Abbildung 7 mit zusätzlicher mit Angabe der Textstellen-ID. Ein Punkt repräsentiert eine Belegstelle. Die Nähe von Punkten signalisiert Ähnlichkeit.

Durch die Analyse der Belegstellen zeigt sich, dass sich die SVP-Belege auf einer Skala anordnen lassen: Während bfe_2253 als Beispiel für den Gebrauch von *Abgabe* innerhalb einer politischen Auseinandersetzung dienen kann, die noch relativ neutral gehalten ist, verschärft sich der Ton der Argumentation von bfe_2253 bis hin zu bfe_2259 (z. B. bfe_2263: „staatlicher Dirigismus“, „ideologische Wunschvorstellung“, bfe_2256: „schwächt massiv den Wirtschaftsstandort Schweiz im internationalen Wettbewerb“). Je mehr die SVP im eigenen Sprachgebrauch von dem der anderen Akteure abweicht, desto isolierter

werden auch die einzelnen Belege in der Grafik dargestellt. Die Belege bis zu bfe_2259 zeigen eine kontextuelle Einbettung des Lexems, welche die SVP im Energiediskurs isoliert dastehen lässt.

Dagegen zeigen Keywords und Belegstellen im Energiediskurs-Korpus für die SP, dass das Lexem *Abgabe* eher in Übereinstimmung mit dem Sprachgebrauch anderer Parteien verwendet wird (vgl. Abbildung 5). Bei der Analyse der verschiedenen Textstellen fällt außerdem auf, dass das Lexem *Abgabe* in verschiedenen Kontexten verwendet wird.

bfe_2116: „mit einer Lenkungsabgabe zu belegen. Zweitens braucht es eine Abgabe auf Strom aus nicht erneuerbaren Quellen [...]“ (SP)

Dagegen:

bfe_2108: „Die Km-Abgabe führt von einer pauschalen zu einer nutzungsabhängigen Abgabe, der km-Zuschlag zu einer Verkehrslenkung zu Hauptverkehrszeiten [...]“ (SP)

Während es hier also bei den Belegen um bfe_2116 um Abgaben geht, bei denen bestimmte Energieformen begünstigt bzw. benachteiligt werden, wird in den Textstellen um bfe_2108 eher über eine Abgabe im Kontext mit Verkehrsmitteln diskutiert (analog zu den Keywords *fahrleistungsabhängig* oder *-bezogen*) – und dies nicht nur von der SP, sondern auch von den anderen aufgeführten Parteien.

4.2. Das Beispiel *Energieeffizienz*

Anders als bei dem Lexem *Abgabe* zeigt das Beispiel *Energieeffizienz*, dass es hier in erster Linie nicht um Unterschiede im parteipolitischen Sprachgebrauch geht, sondern dass generell zwei Unterdiskurse zu *Energieeffizienz* existieren, was sich deutlich an der Verteilung der Token festmachen lässt (vgl. Abbildung 9).

Wie beim Beispiel *Abgabe* kann für eine erste Analyse der beiden Unterdiskurse (Cluster 1 und Cluster 2), an denen sich alle aufgeführten Parteien beteiligen, erneut die Angabe von Keywords zu den Clustern helfen (vgl. Abbildungen 10 und 11).



Abb. 11: Keywords zu den Belegstellen zu Energieeffizienz von Cluster 2.

Die Keywords zu Cluster 1 sowie die entsprechenden Textbelege verdeutlichen, dass es in diesem Diskurs um Maßnahmen geht, wie Energieeffizienz konkret umgesetzt werden kann (*Gebäudeprogramm, Geräte, elektrisch*). Als Beispiele sollen hier zwei Textbelege aus der Mitte der Cluster angeführt werden:

bfe_6083: „[...] die ungebrochene Zunahme von Elektrogeräten. Zusätzlich zur Steigerung der Energieeffizienz sind also Anstrengungen erforderlich, auch den individuellen Konsum auf [...]“ (SP)

bfe_5031: „[...] Konsumgeräte ist das Top-Runner-Prinzip einzuführen. Die Mindestanforderungen in Sachen Energieeffizienz sind laufend zu überprüfen und anzuheben. Ein Label für [...]“ (Grüne)

Bei Cluster 2 dagegen geht es um eine Positionierung im politischen Umfeld, um die Förderung bestimmter *Energieformen*, um *Aktionspläne*, *Zieldefinitionen* etc. – auch die eher unbestimmte Lexik zeigt, dass hier ein (bundes-)politischer Diskurs geführt wird.

bfe_4999: „Der Bund muss Energieeinsparungen, Energieeffizienz und erneuerbare Energien stärker fördern. Klimafreundlicher und landschaftsschonender Atomausstieg [...]“ (Grüne)

bfe_6039: „Nur ein konsequenter Weg mit Energieeffizienz und erneuerbaren Energien garantiert Energiesicherheit, schafft zukunftsfähige Arbeitsplätze, [...]“ (SP)

Mit Blick auf die bisher untersuchten Parteien SVP und SP lässt sich zu Energieeffizienz sagen, dass in dem Sample von 200 Textbelegen zu *Energieeffizienz*,

nur 7 Belege der SVP, dagegen 44 der SP enthalten waren, was ein Hinweis auf den Sprachgebrauch der beiden Parteien sein kann.

5. Fazit

Schlagwörter in einem politischen Diskurs zeichnen sich dadurch aus, dass sie in ihrer Verwendung und ihrer Semantik Besonderheiten aufweisen. Während Keyword- und Kollokationsanalyse einen ersten, einfachen Schritt darstellen, Bezeichnungskonkurrenzen im Sinne von Fahnen- und Programmwörtern zu eruieren, ist diese Methode für die Ermittlung überparteilich verwendeter Schlagwörter nicht ausreichend.

Der in diesem Beitrag vorgestellte methodische Zugang über maschinelles Lernen und Verfahren der distributionellen Semantik hat gezeigt, dass dieser sowohl ergänzend für die Bestimmung von überparteilichen Schlagwort-Kandidaten durch Keyword-Analysen, wie auch bei der Analyse von deren Verwendung im Diskurs eine wertvolle Hilfestellung bietet. Token-Vektoren können sowohl dazu genutzt werden, Schlagwortkandidaten zu identifizieren (über die Verteilung im Raum), als auch um akteurs- und kontextspezifische Verwendungen nachzuweisen. Auf diese Weise tragen Tokenvektoren zur Identifikation von Kontroversen im Diskurs bei.

Allerdings erstaunt nicht, dass aufgrund der semantischen und pragmatischen Komplexität von Schlagwörtern weiterhin ein qualitativer Analyseschritt notwendig ist, für den wiederum die mittels Korpuslinguistik erhobenen Belegstellen gute Dienste leisten.

6. Literatur

- Blei, David M./Ng, Andrew Y./Jordan, Michael I. (2003): Latent Dirichlet Allocation. In: *The Journal of Machine Learning Research*, 3, 993–1022.
- Burkhardt, Armin (1998): Deutsche Sprachgeschichte und politische Geschichte. In: Werner Besch, Anne Betten, Oskar Reichmann, Stefan Sonderegger (Hrsg.): *Sprachgeschichte. Ein Handbuch zur Geschichte der deutschen Sprache und ihrer Erforschung*. 1. Halbband. Berlin/New York: de Gruyter, 98–122.
- Dieckmann, Walther (1964): *Wortschatz und Wortgebrauch der politischen Werbung. Ein Beitrag zur Wortforschung am Beispiel der deutschen Sprache im 19. und 20. Jahrhundert*. Marburg: Elwert.

- Evert, Stefan (2014): Distributional Semantics in R with the wordspace Package. In: Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: System Demonstrations, 110–114.
- Firth, John Rupert (1957): A synopsis of linguistic theory. Studies in linguistic analysis. Oxford: Blackwell.
- Hermanns, Fritz (2007): Slogans und Schlagwörter. In: Jochen A. Bär, Thorsten Roelcke, Anja Steinhauer (Hrsg.): Sprachliche Kürze. Konzeptuelle, strukturelle und pragmatische Aspekte. Berlin [u. a.]: de Gruyter, 459–478.
- Heylen, Kris/Wielfaert, Thomas/Speelman, Dirk/Geeraerts, Dirk (2015): Monitoring Polysemy: Word Space Models as a Tool for Large-Scale Lexical Semantic Analysis. In: *Lingua*, 157, 153–72.
- Hilpert, Martin/Correia, Saavedra David (2017): Using Token-Based Semantic Vector Spaces for Corpus-Linguistic Analyses: From Practical Applications to Tests of Theoretical Claims. In: *Corpus Linguistics and Linguistic Theory*, ahead of print.
- Jung, Matthias (2005): Schlüsselwortforschung im Internet. Möglichkeiten, Beispiele, Grenzen. In: Dietrich Busse, Thomas Niehr, Martin Wengeler (Hrsg.): Brisante Semantik. Neuere Konzepte und Forschungsergebnisse einer kulturwissenschaftlichen Linguistik. (= Germanistische Linguistik; 259). Tübingen: Niemeyer, 355–368.
- Klein, Josef (2017): Um Begriffe kämpfen. In: Jörg Kilian, Thomas Niehr, Martin Wengeler (Hrsg.): Handbuch Sprache und Politik. Bd. 2. Bremen: Hempen Verlag, 773–793.
- Niehr, Thomas (2014): Einführung in die Politolinguistik. Göttingen: Vandenhoeck & Ruprecht.
- Schröter, Melani (2015): Besondere Wörter III: Schlagwörter in der öffentlich-politischen Auseinandersetzung. In: Ulrike Haß, Petra Storjohann (Hrsg.): Handbuch Wort und Wortschatz. (= Handbücher Sprachwissen; 3.). Berlin/Boston: de Gruyter, 394–412.
- Schütze, Hinrich (1998): Automatic Word Sense Discrimination. In: *Computational Linguistics*, 24/1, 97–123.
- Selivanov, Dmitriy/Wang, Qing (2018): text2vec: Modern Text Mining Framework for R. R package version 0.5.1.
- Storjohann, Petra/Schröter, Melani (2013): Präsenz und Absenz lokaler Diskursgebrauchsmuster am Beispiel des deutschen und britischen Krisendiskurses. In: Martin Wengeler, Alexander Ziem (Hrsg.): Sprachliche Konstruktionen von Krisen. Interdisziplinäre Perspektiven auf ein fortwährend aktuelles Phänomen. Bremen: Hempen Verlag, 185–208.

- Stücheli-Herlach, Peter/Ehrensberger-Dow, Maureen/Dreesen, Philipp (2018): Energiediskurse in der Schweiz. Anwendungsorientierte Erforschung eines mehrsprachigen Kommunikationsfelds mittels digitaler Daten. (Working Papers in Applied Linguistics; 16). Winterthur: ZHAW Zürcher Hochschule für Angewandte Wissenschaften.
- Turney, Peter. D./Pantel, Patrick (2010): From Frequency to Meaning: Vector Space Models of Semantics. In: *Journal of Artificial Intelligence Research*, 37, 141–188.
- Wengeler, Martin (2017): Wortschatz I: Schlagwörter, politische Leitvokabeln und der Streit um Worte. In: Kersten Sven Roth, Martin Wengeler, Alexander Ziem (Hrsg.): *Handbuch Sprache in Politik und Gesellschaft*. Berlin/Boston: de Gruyter, 22–46.