Aspekte einer automatischen Meinungsbildungsanalyse von Online-Diskussionen

Matthias Liebeck

Institut für Informatik
Heinrich-Heine-Universität Düsseldorf
Universitätsstr. 1
40225 Düsseldorf
liebeck@cs.uni-duesseldorf.de

Abstract: Heutzutage haben Menschen die Möglichkeit, ihre Meinung zu verschiedensten Themen in onlinebasierten Diskussionsplattformen zu äußern. Diese Meinungen können in Form einer Meinungsbildungsanalyse genauer untersucht werden. In diesem Beitrag werden verschiedene Aspekte einer automatisierten Diskussionsverfolgung untersucht. Dazu werden Analysekriterien definiert und die vorgestellten Ansätze auf zwei deutschsprachige Datensätze angewendet.

1 Einleitung

Das Internet ermöglicht es Menschen in onlinebasierten Plattformen ihre Meinung über Themen, z.B. politische Ereignisse oder Konsumgüter, in Form von Textbeiträgen, unter anderem in Foren, als Kommentare auf Nachrichtenportalen oder in sozialen Netzwerken, öffentlich preiszugeben. Die über ein Thema gemachten Äußerungen können individuell analysiert werden, um daraus im Rahmen einer Meinungsbildungsanalyse ein Meinungsbild zu erstellen. Aus Sicht einer Firma kann eine automatisierte Untersuchung von Social Media gegebenenfalls Hinweise auf die öffentliche Meinung über firmeneigene Produkte liefern. Im Rahmen eines Online-Partizipationsverfahrens, bei dem sich beispielsweise Bürger bei einer Diskussion über lokalkommunale Entscheidungen beteiligen können, kann die Haltung zu bestimmen Themen untersucht werden.

Durch die große Anzahl an Beiträgen ist eine automatisierte Analyse erstrebenswert. Unter Einsatz von Techniken der maschinellen Sprachverarbeitung können durch eine **Natural Language Processing (NLP) Pipeline** Meinungen in Textform analysiert und für eine weitere Verarbeitung aufbereitet werden. Anwendungsabhängig kann ermittelt werden, über welches Thema eine Aussage gemacht wird und welche **Tonalität** $t \in \{\text{positiv}, \text{neutral}, \text{negativ}\}$ dabei ausgedrückt wird. So erfolgt beispielsweise im Satz „*Peter liebt Schokolade.*“ eine positive Aussage über das Thema *Schokolade*. Im Rahmen dieser Arbeit werden Kommentare eines deutschsprachigen Nachrichtenportals und eines Online-Partizipationsverfahrens untersucht und themenspezifische Meinungsbilder in Hinblick auf einen zeitlichen Aspekt konstruiert.

Zur maschinellen Sprachverarbeitung deutschsprachiger Kommentartexte wird eine NLP-Pipeline eingesetzt, die sequenziell mehrere Verarbeitungsschritte durchläuft. Zu Beginn wird OpenNLP[1] zur Trennung eines Eingabetextes in mehrere Sätze, zur Zerlegung der Sätze in einzelne Wörter bzw. **Tokens** und für ein anschließendes POS-Tagging verwendet, bei dem für einzelne Wörter ein **Part-of-Speech-Tag (POS-Tag)** bzw. eine Wortart aus dem *Stuttgart-Tübingen-Tagset (STTS)* [STST99] bestimmt wird.

Um neben POS-Tags Informationen über weitere grammatikalische Eigenschaften aus einem Satz zu erhalten, werden **Dependenzen** bestimmt, die jeweils eine Abhängigkeit zwischen zwei Wörtern angeben. Dadurch kann für ein Wort beispielsweise bestimmt werden, ob es Teil des Subjekts oder des Objekts eines Satzes ist. Zur Bestimmung von Dependenzen im *TIGER Annotationsschema* [AAB$^+$03] wird Mate Tools [BBHN10] eingesetzt.

Zur Lemmatisierung eines Wortes (z.B. Türme $\rightarrow$ Turm, warte $\rightarrow$ warten) werden Mate Tools und TreeTagger [Sch94] eingesetzt, damit anschließend in dem Tonalitätslexikon SentiWS [RQH10] eine Tonalitätsangabe aus dem Intervall $[-1, 1]$ nachgeschlagen werden kann.

Im folgenden Kapitel werden die untersuchten Aspekte einer Diskussionsverfolgung erläutert. Danach werden zwei Datensätze vorgestellt, die anschließend auf ihren Inhalt und dessen Tonalität untersucht werden. In Kapitel 4 erfolgt ein Vergleich mit verwandten Arbeiten. Abschließend wird ein Fazit gezogen und Ideen für zukünftige Arbeiten aufgeführt.

2 Ansätze zur Diskussionsverfolgung

Bei einer automatisierten Diskussionsverfolgung sind, je nach Analyseziel, ein oder mehrere Aspekte zu berücksichtigen. Im Folgenden werden Überlegungen zur Bestimmung relevanter Sätze angestellt und Ansätze zur Themenextraktion, zum Vergleich semantischer Äquivalenzen und zur Tonalitätsanalyse vorgestellt, deren Ergebnisse anschließend als Argumentationsketten zusammengefasst werden.

2.1 Identifizierung von relevanten Sätzen

In einer Meinungsbildungsanalyse werden Meinungen von Personen in Form von Aussagen untersucht. Daher ist es sinnvoll, nur Sätze zu analysieren, in denen Aussagen gemacht werden. In der deutschen Sprache kann zwischen fünf Satzarten unterschieden werden: Aussagesätzen, Fragesätzen, Ausrufesätzen, Wunschsätzen und Aufforderungssätzen.

Der Inhalt eines Fragesatzes ist für eine Meinungsbildungsanalyse nicht geeignet, da der Autor des Satzes keine Aussage tätigt. Die Zuordnung eines Satzes zu den anderen vier Satzarten ist eine Aufgabe, deren Lösung tiefgehende grammatikalische Eigenschaften benötigt und nicht nur durch das Betrachten des satzbeendenden Zeichens zu lösen ist. Autoren können beispielsweise ein Ausrufezeichen als Satzende eines Aussagesatzes ver-

[1] `https://opennlp.apache.org/`

wenden, in der Absicht dadurch ihrer Aussage mehr Ausdruckskraft zu verleihen. Daher wird vereinfachend angenommen, dass nur mit einem Fragezeichen beendete Sätze ignoriert werden können.

Es gibt Aussagen, für die im gleichen Satz eine Begründung angegeben wird. Diese Begründungen können in einem Nebensatz stehen, der durch eine unterordnende Konjunktion eingeleitet wird. Genauer ausgedrückt durch kausale Konjunktionen (z.B. *weil, da*) bzw. in Kausalsätzen.

Beispiel: $\underbrace{\text{Ich mag dein Auto}}_{\text{Hauptsatz}}$, $\underbrace{\text{weil es blau ist.}}_{\text{Nebensatz}}$

Eine Möglichkeit, solche Aussagen zu suchen, ist nach Begründungen in einem Nebensatz zu suchen, um anschließend Aussagen aus dem Hauptsatz zu extrahieren. Dazu ist eine Klassifizierung der Satzteile eines gegebenen Satzes mit Kommata in Hauptsätze und Nebensätze hilfreich. Anhand von POS-Tags, Dependenzen und drei Regeln aus Grammatikbüchern (in Bezug auf eine Verberst-, Verbzweit- und Verbletztstellung oder dem Auftreten eines Infinitivs) können Muster identifiziert werden, die im Allgemeinen eine gute Trennung in Hauptsätze und Nebensätze ermöglichen. Anschließend kann nach Begründungen gesucht werden, indem Nebensätze identifiziert werden, die mit einer kausalen Konjunktion beginnen. Häufig vorkommende kausale Konjunktionen sind *da, weil, zumal* und *denn*.

Dieser Ansatz ermöglicht die Identifikation einiger Aussagen, jedoch existieren auch Aussagen in Sätzen, die nicht diesem Schema entsprechen, wie manuelle Stichproben aus den in Abschnitt 3.2 vorgestellten Datensätzen zeigen. Somit ist dieser Ansatz nicht erschöpfend, da er Aussagen, zu denen der Autor keine Begründung angibt, zu denen eine Begründung im nachfolgenden oder vorangestellten Satz erfolgt, Aussagen mit Begründungen innerhalb des Hauptsatzes (z.B. „*Wegen Bauarbeiten ist die Straße blockiert.*“) oder Aussagen mit fehlerhafter Kommasetzung übersieht.

Ohne eine explizite Filterung nach Aussagesätzen muss daher die Annahme getroffen werden, dass in allen Sätzen, die nicht auf ein Fragezeichen enden, Aussagen stecken können und sie folglich potenziell relevant für eine Meinungsbildungsanalyse sind.

2.2 Themenerkennung auf Satzebene

Eine große Aufgabe der maschinellen Sprachverarbeitung ist die Beantwortung der Fragestellung, was das Thema in einem Satz ist bzw. welche Handlung in einem Satz beschrieben wird. Im Folgenden werden zwei Methoden zur Themenextraktion beschrieben.

POS-Tag basiert: Zur Themenerkennung werden alle Wörter eines Satzes extrahiert, deren POS-Tags Nomen (Tag *NN*) oder Eigennamen (Tag *NE*) sind. Dies liefert die Information, wer in einem Satz agiert und womit agiert wird. Dabei werden direkt aufeinanderfolgende Wörter zu einer Fundstelle zusammengefasst, wenn es sich dabei um Eigennamen (Tag *NE*), Nomen aus derselben Nominalphrase (Dependenz *NK*) oder um Namensbestandteile (Dependenz *PNC*) handelt. Ein schlichtes Zusammenfassen von aufeinander-

folgenden Wörtern kann zu Problemen führen, wie der Satz „*Peter mag Kinder, die mit Feuerwehrwagen Rettungen nachspielen.*“ zeigt, da dort die aufeinanderfolgenden Nomen *Feuerwehrwagen* und *Rettungen* fälschlicherweise zusammengefasst werden.

Dependenzbasiert: Da bei dem POS-Tag basierten Ansatz nicht die Information berücksichtigt wird, welche Aktion im Satz beschrieben wird, bezieht der zweite Ansatz zur Themenextraktion die Rolle des Verbs mit ein. In der grundlegendsten Version wird dabei nach SVO-Tripeln, die jeweils aus drei Wortgruppen bestehen und damit zweistellige Valenzen darstellen, gesucht. Die erste Wortgruppe repräsentiert das Subjekt im Satz, die zweite das Verb und die dritte das Objekt. Identifiziert wird ein SVO-Tripel durch ein Verb, von dem eine Dependenz für das Subjekt $s \in \{\text{SB,SBP,SP}\}$ und eine Dependenz für das Objekt $o \in \{\text{OA,OA2,OG,OP,DA}\}$ ausgehen. Aus dem Beispielsatz „*Peter kritisiert Anne.*“, dessen Dependenzen in Abbildung 1 visualisiert sind, wird das SVO-Tripel *(Peter, kritisiert, Anne)* extrahiert.

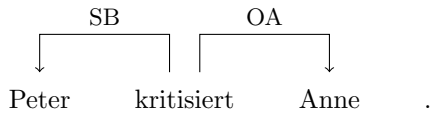

Abbildung 1: Beispiel für die Dependenzen eines SVO-Tripels

Bei SVO-Tripeln sollte darauf geachtet werden, dass Subjekte und Objekte grammatikalische Funktionen sind und nicht semantische Rollen darstellen. In Aktivsätzen gilt in der Regel, dass das Subjekt mit dem Agens (der handelnden Entität) und das Objekt mit dem Patiens (der betroffenen Entität) übereinstimmen. In Passivsätzen kann eine Rollenvertauschung auftreten, wie z.B. im Satz „*Markus wird von Peter bestohlen.*“, in dem *Markus* als Subjekt und als Patiens agiert und *Peter* das Objekt und den Agens verkörpert. Ebenfalls müssen Negationen gesondert behandelt werden. Verben mit einer anderen Valenzordnung werden Gegenstand zukünftiger Forschung sein.

2.3 Semantische Äquivalenzen erkennen

Nachdem ermittelt wurde, worüber in einzelnen Sätzen gesprochen wird, erfolgt anschließend eine Überprüfung, ob sich verschiedene Wörter auf den gleichen Sachverhalt beziehen bzw. ob semantische Äquivalenzen bestehen, da sprachlich vielfältige Möglichkeiten bestehen, denselben Inhalt auszudrücken. Wurde als Thema eines Satzes ein Nomen extrahiert, so gilt zu identifizieren, ob das Thema synonym zu einem anderen Thema ist, um Aussagen über dasselbe Thema gruppieren zu können. Dazu kann ein Thesaurus verwendet werden, in dem üblicherweise sogenannte Synsets enthalten sind. Ein Synset kann aus mehreren Wörtern bestehen, welche die gleiche semantische Bedeutung tragen wie beispielsweise {*Orange, Apfelsine*}.

Als deutschsprachiger Thesaurus wird GermaNet [HF97] verwendet, in dem Synsets für Nomen, Verben und Adjektive vorhanden sind. Für Nomen sind in GermaNet 9.0 insgesamt 71575 Synsets vorhanden, mit denen überprüft werden kann, ob zwei Nomen syn-

onym zueinander sind.

Es ist geplant, weiterführende Techniken der *Koreferenzanalyse* (Zusammenfassung verschiedener Bezeichnungen für dieselbe Person bzw. dasselbe Objekt) und der *word sense disambiguation* (Erkennung der kontextabhängigen Wortbedeutung) in nachfolgenden Arbeiten zu berücksichtigen.

2.4 Tonalitätsbestimmung

Für die extrahierten Themen wird anschließend untersucht, mit welchen Wörtern der jeweilige Sachverhalt beschrieben bzw. welche Tonalität dabei ausgedrückt wird. Dazu werden von den gefundenen Nomen und den voranstehenden Adjektiven die wortbasierten Tonalitätsangaben in SentiWS nachgeschlagen und zu einem Mittelwert zusammengefasst. Bei SVO-Tripeln wird zusätzlich das Verb einbezogen. Adjektive werden unter anderem für die Tonalitätsbestimmung bei Produktrezensionen [HL04] berücksichtigt. Im Rahmen dieser Arbeit wird untersucht, inwiefern Adjektive zur Tonalitätsbestimmung von Diskussionsbeiträgen beitragen.

Sprachliche Negationen können auf verschiedene Arten ausgedrückt werden. Negationen durch die Verwendung eines Affixes (z.B. *glücklich* und *unglücklich*) werden teilweise durch SentiWS abgedeckt, wenn für beide Wörter Einträge vorhanden sind. Bei einer Negation durch das Wort *„nicht“* wird vereinfachend angenommen, dass die Polarität eines Adjektivs umgedreht wird, wenn das Wort *„nicht“* direkt vorangestellt ist. Alternativ kann überprüft werden, ob für ein negiertes Adjektiv in GermaNet per Antonym-Beziehung ein entgegengesetztes Adjektiv existiert, das in SentiWS eine Tonalitätsangabe besitzt.

2.5 Argumentationsketten

Nachdem bereits Themen identifiziert und dazu Tonalitätsangaben ermittelt wurden, erfolgt eine Gruppierung in Argumentationsketten, die sich jeweils auf genau ein Thema, wie eine Person oder einen Begriff, beziehen. Eine Argumentationskette kann als ein Zeitstrahl betrachtet werden, auf dem die zu einem Thema gemachten Äußerungen mit den zugehörigen Tonalitäten zeitlich geordnet sind.

Die Konstruktion von Argumentationsketten aus Sätzen erfolgt in mehreren Schritten (vgl. Abbildung 2):

(1) Für den durch die NLP-Pipeline aufbereiteten Eingabesatz erfolgt eine Themenerkennung.

(2) Für alle gefundenen Themen wird eine Tonalität bestimmt.

(3) Für jede Fundstelle wird überprüft, ob eine Argumentationskette existiert, deren Thema textuell mit der Fundstelle übereinstimmt oder dazu per GermaNet synonym ist. Kann keine übereinstimmende Argumentationskette gefunden werden, wird eine

neue Argumentationskette begonnen.

(4) Der in Schritt (3) bestimmten Argumentationskette wird ein neues Element angefügt, in dem das Thema, die dazugehörige Tonalität und Metadaten (Erstellungsdatum, Autorname, Referenz auf den Satz) vermerkt werden.

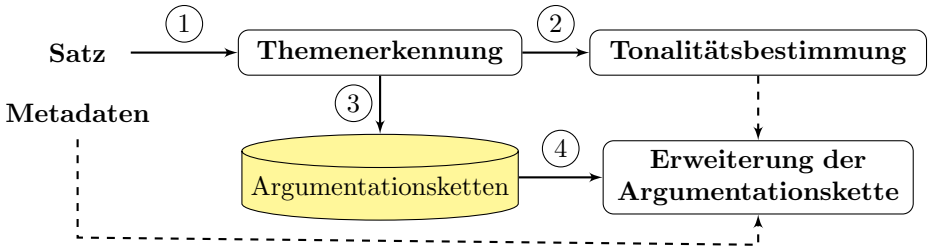

Abbildung 2: Konstruktion von Argumentationsketten

3 Auswertung

Die genannten Ansätze zur Diskussionsverfolgung werden im Rahmen einer prototypischen Implementierung auf zwei Datensätze angewendet. Dabei werden exemplarisch mehrere Analysekriterien definiert.

3.1 Analysekriterien

Abhängig von der Aufgabenstellung einer Meinungsbildungsanalyse können unterschiedliche Betrachtungen stattfinden. Eine Betrachtung kann globale Eigenschaften der überwachten Texte in Hinblick auf eine Kommentarfrequenz und auf durchschnittliche Tonalitätswerte untersuchen oder auf themenspezifischen Eigenschaften der konstruierten Argumentationsketten erfolgen. Bei einer längerfristigen Überwachung eines Themas ist eine zeitliche Gruppierung von Texten sinnvoll. Eine abschnittweise Untersuchung ermöglicht einen Vergleich zwischen Zeitabschnitten und dadurch die Analyse einer zeitlichen Entwicklung. Die zeitliche Gruppierung kann z.B. per fester Zeitspanne, per fester Anzahl an Kommentaren oder einer Kombination beider Kriterien erfolgen.

Bei einer Meinungsbildungsanalyse ist ebenfalls interessant zu bestimmen, über welche Themen in den analysierten Texten gesprochen wird. Anhand der Länge aller Argumentationsketten können häufig besprochene Themen identifiziert werden. Bei der Betrachtung eines konkreten Themas kann untersucht werden, wann es häufig besprochen wird und wie sich die verwendete Tonalität entwickelt.

Firmen können beispielsweise aus der automatisierten Analyse von Social Media einen Vorteil ziehen, wenn sie maschinell Meinungen über ihre eigenen Produkte untersuchen.

Wird über ein bestimmtes Produkt innerhalb kürzester Zeit vermehrt Kritik geäußert, so können die Entwicklungs- und Marketingabteilungen auf ein mögliches Problem des Produkts hingewiesen werden.

3.2 Datensätze

Die Ansätze zur Diskussionsverfolgung werden auf zwei Datensätze angewendet. Der erste Datensatz wird als *HHU Normsetzungskorpus* bezeichnet. Er besteht aus 434 Kommentaren (1444 Sätzen), die im Rahmen der *Online-Fakultätsratssitzung „Neugestaltung der Promotionsordnung“*[2] [EST+14] auf einer Instanz der Plattform *Adhocracy*[3] abgegeben wurden. Der zweite Datensatz wurde durch einen Crawling-Prozess aus Kommentaren eines deutschsprachigen Nachrichtenportals zusammengestellt und wird als *Krim-Korpus* bezeichnet. Er besteht aus 62362 Kommentaren (354476 Sätzen), die aus Nachrichten der Kategorie *Politik* stammen, die mit dem Schlagwort *Krim* versehen sind und im Zeitraum vom 1.2.2014 bis zum 31.7.2014 abgegeben wurden.

3.3 Analyseergebnisse

Die konstruierten Argumentationsketten wurden auf Gesprächsthemen und auftretende Tonalitäten untersucht. Dazu wurden die Kommentare des Krim-Korpus wochenweise gruppiert. In der Woche des 17.3.2014, in der ein Referendum über die Zugehörigkeit der Halbinsel Krim zu Russland oder zur Ukraine erfolgte, wurden mit über 9000 Kommentaren die meisten Kommentare in einer Woche gemacht. Für jede untersuchte Woche ist der durchschnittliche Tonalitätswert leicht negativ.

Gesprächsthemen: In Tabelle 1 werden die zehn häufigsten Themen, die über den POS-Tag basierten Ansatz ermittelt wurden, mit der Länge der jeweiligen Argumentationskette angegeben.

Platzierung	HHU Normsetzungskorpus	Krim-Korpus
1	Überprüfung (106)	Russland (23436)
2	Arbeit (105)	Ukraine (18758)
3	Promotion (95)	Putin (13720)
4	Publikation (88)	Krim (11374)
5	Betreuer (67)	USA (11029)
6	Gebiet (67)	Grund (10811)
7	Vorschlag (62)	Westen (8995)
8	Beispiel (51)	EU (8487)
9	Note (51)	Mensch (6405)
10	Gutachter (47)	Regierung (6271)

Tabelle 1: Liste der zehn häufigsten Themen der beiden Datensätze

[2] https://normsetzung.cs.uni-duesseldorf.de/
[3] https://adhocracy.de/

Eine zu Beginn gemachte, intuitive Annahme, dass reale Sachverhalte von Ereignissen aus der Region Ukraine im Krim-Korpus häufig durch das gleiche SVO-Tripel ausgedrückt werden, konnte nicht bestätigt werden. Anstatt dass in häufig genannten Tripeln die Beziehung zwischen zwei Personen, wie z.B. *(Peter, kritisiert, Anne)*, ausgedrückt wird, besitzen die am häufigsten auftretenden Tripel das Wort „*Ich*“ als Subjekt oder Verben der Meinungsäußerung wie „*denken*“ oder „*meinen*“.

Tonalität: Zunächst ist anzumerken, dass für beide Korpora durch das Tonalitätslexikon SentiWS nur wenige Tonalitätsangaben pro Satz bestimmt werden konnten. Im Krim-Korpus gibt es lediglich 0,9 Tonalitätsangaben pro Satz. Beim HHU Normsetzungskorpus sind es 1,2 annotierte Tokens pro Satz. Dies erschwert die Tonalitätsbestimmung einzelner Themen, da durch die niedrige Anzahl an Angaben oft eine neutrale Tonalität angenommen werden muss. Wird als Schwellwert $\alpha = 0{,}1$ zur Einordnung eines numerischen Tonalitätswerts in $t \in \{\text{positiv, neutral, negativ}\}$ verwendet, so tritt für das im Krim-Korpus am häufigsten auftretende Thema *Russland*, unter Berücksichtigung von vorangestellten Adjektiven, in 99,5% eine neutrale Tonalität auf.

Da vergleichbare Verteilungen auch bei anderen Themen entstehen, ist anzunehmen, dass ein größeres Tonalitätslexikon benötigt wird und dass Tonalitäten im Krim-Korpus auch durch andere sprachliche Mittel als Adjektive ausgedrückt werden. Daher werden sich zukünftige Arbeiten mit der Frage beschäftigen, welche sprachlichen Mittel im Rahmen von Online-Diskussionen besonders viel Tonalität tragen und wie diese automatisiert erfasst werden können.

Zur approximativen Untersuchung einer zeitlichen Tonalitätsentwicklung wird daher auf die Annahme zurückgegriffen, dass die Tonalität einer Fundstelle durch eine Mittelwertbildung aller Tonalitätsangaben des gesamten Satzes angegeben wird. Die zeitliche Entwicklung der Tonalitäten wird nur anhand des Krim-Korpus untersucht, da der HHU Normsetzungskorpus zu klein ist. Für das Thema *Russland* verschiebt sich die Verteilung der Tonalität zu 9,7% positiv, 67,7% neutral und 22,6% negativ. In Abbildung 3 werden exemplarisch die Tonalitäten der Argumentationskette über das Thema *Wahl* des Krim-Korpus (jeweils als Mittelwert einer wochenweisen Gruppierung) dargestellt. Die dort auftretenden Vorzeichenwechsel zwischen zwei Zeitpunkten können im Rahmen einer automatisierten Themenüberwachung interessant sein.

4 Verwandte Arbeiten

In [KG12] wird ein verborientierter Ansatz zur Tonalitätsbestimmung von Meinungen in englischsprachigen Diskussionen vorgestellt. Die Autoren erstellen ein Tonalitätswörterbuch aus 440 Verben, die eine Meinung ausdrücken können, wie z.B. *agree, hate, like* und *love*, zu denen jeweils die Größe des Valenzrahmens und eine numerische Tonalität angegeben ist. Anhand eines Verbs wird entschieden, wie sich die Tonalitäten der Wörter im Valenzrahmen aufeinander auswirken. Die Autoren bemerken, dass in den untersuchten Kommentaren 47% der Tonalität durch Verben und nur 33% der Tonalität durch Adjektive ausgedrückt werden. Sie kommen ebenfalls zu dem Fazit, dass die Übertragung von Tech-

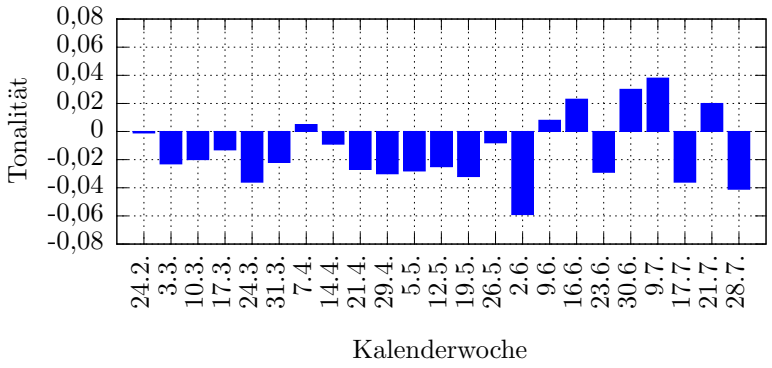

Abbildung 3: Tonalitätsentwicklung der Argumentationskette über das Thema *Wahl*

niken zur Tonalitätsbestimmung von Produktrezensionen auf Online-Diskussionen problematisch ist.

Eine POS-Tag basierte Themenerkennung, die Nomen als Themen identifiziert, wird in [CZKL10] verwendet. Die Autoren extrahieren politische Haltungen aus Aussagen von amerikanischen Senatoren und benutzen für die Tonalitätsbestimmung Adjektive, Verben und Adverbien.

5 Fazit und Ausblick

Im Rahmen dieser Arbeit wurden verschiedene Aspekte einer maschinellen Analyse von deutschsprachigen Online-Diskussionen vorgestellt. Dabei wird das Fazit gezogen, dass ein valenzorientierter Ansatz, der sich auf Verben als zentrales Element eines Satzes bezieht, detaillierter untersucht werden sollte. Dazu wird ein elektronisches Valenzwörterbuch, in dem pro Verb die Größe des Valenzrahmens angegeben ist, erforderlich sein. Zur Tonalitätsanalyse wird in zukünftigen Arbeiten ein individuell auf die Analyse von Online-Diskussionen angepasstes Verfahren entwickelt werden, das vor allem die Rolle von Verben der Meinungsäußerung stärker einbezieht und dazu auf ein noch zu entwickelndes Tonalitätslexikon zurückgreifen wird.

Darüber hinaus werden Textbeiträge aus einem Online-Partizipationsverfahren annotiert werden, auf denen die vorgestellten Aspekte genauer evaluiert werden können. In Zukunft wird geprüft werden, ob OpenThesaurus[4] als weiterer deutschsprachiger Thesaurus und Wiktionary[5] als zusätzliches Wörterbuch zur Grundformreduktion eingesetzt werden können. Darüber hinaus wird die Zuordnung von Aussagen zu einer konkreten Argumentationskette durch Techniken der Koreferenzanalyse, der Anaphorikauflösung und der word sense disambiguation verbessert werden.

[4] `https://www.openthesaurus.de/`
[5] `https://de.wiktionary.org/`

Literatur

[AAB+03] S Albert, J Anderssen, R Bader, S Becker, T Bracht, S Brants, T Brants, V Demberg, S Dipper, P Eisenberg, S Hansen, H Hirschmann, J Janitzek, C Kirstein, R Langner, L Michelbacher, O Plaehn, C Preis, M Pussel, M Rower, B Schrader, A Schwartz, Smith G und H Uszkoreit. TIGER-Annotationsschema. Bericht, Universität Potsdam, Universität Saarbrücken, Universität Stuttgart, 2003.

[BBHN10] Anders Björkelund, Bernd Bohnet, Love Hafdell und Pierre Nugues. A High-Performance Syntactic and Semantic Dependency Parser. In *COLING (Demos)*, Seiten 33–36. Demonstrations Volume, 2010.

[CZKL10] Bi Chen, Leilei Zhu, Daniel Kifer und Dongwon Lee. What Is an Opinion About? Exploring Political Standpoints Using Opinion Scoring Model. In *AAAI*. AAAI Press, 2010.

[EST+14] Tobias Escher, Jost Sieweke, Ulf Tranow, Simon Dischner, Dennis Friess, Philipp Hagemeister und Katharina Esau. Internet-Mediated Cooperative Norm Setting in the University: Design and Evaluation of an Online Participation Process to Redraft Examination Regulations. In *IPP2014: Crowdsourcing for Politics and Policy*, 2014.

[HF97] Birgit Hamp und Helmut Feldweg. GermaNet - a Lexical-Semantic Net for German. In *Proceedings of ACL workshop Automatic Information Extraction and Building of Lexical Semantic Resources for NLP Applications*, Seiten 9–15, 1997.

[HL04] Minqing Hu und Bing Liu. Mining and Summarizing Customer Reviews. In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '04, Seiten 168–177. ACM, 2004.

[KG12] Mostafa Karamibekr und Ali A. Ghorbani. Sentiment Analysis of Social Issues. In *Proceedings of the 2012 International Conference on Social Informatics*, SOCIALINFORMATICS '12, Seiten 215–221. IEEE Computer Society, 2012.

[RQH10] R. Remus, U. Quasthoff und G. Heyer. SentiWS – a Publicly Available German-language Resource for Sentiment Analysis. In *Proceedings of the 7th International Language Resources and Evaluation (LREC'10)*, Seiten 1168–1171, 2010.

[Sch94] Helmut Schmid. Probabilistic Part-of-Speech Tagging Using Decision Trees. In *Proceedings of the International Conference on New Methods in Language Processing*, 1994.

[STST99] Anne Schiller, Simone Teufel, Christine Stöckert und Christine Thielen. Guidelines für das Tagging deutscher Textcorpora mit STTS (kleines und großes Tagset). Bericht, Universität Stuttgart, Universität Tübingen, Stuttgart, Germany, 1999.