

Auf den Weg zur Visualisierung von linguistischen Mustern der Deliberation: eine Fallstudie anhand der Daten von S21

Tina Bögel Valentin Gold Annette Hautli-Janisz Christian Rohrdantz Sebastian Sulger
Miriam Butt Katharina Holzinger Daniel A. Keim

Fachbereich Sprachwissenschaft Fachbereich Politikwissenschaft Fachbereich Informatik
Universität Konstanz

1 Hinführung und Motivation

Dieser Beitrag berichtet über Fortschritte im interdisziplinären Projekt *VisArgue*¹, das sich mit der automatischen linguistischen und visuellen Analyse von politischem Diskurs befasst. Der Schwerpunkt des Projekts liegt auf der Untersuchung des Konzeptes der deliberativen Kommunikation, einer Theorie, die hauptsächlich von Habermas (1981), Dryzek (1990, 2000) Bohman (1996) und Gutmann and Thompson (1996) propagiert wird. Gemäß dieser Theorie sollten Interessenvertreter in der Kommunikation mit anderen Parteien ihre Positionen extensiv rational begründen und sich schlussendlich dem besseren Argument fügen. Die automatische Messung der deliberativen Qualität einer Gesprächssituationen erfordert die Identifikation von linguistischen Einheiten, die Aufschluss geben über Faktoren wie objektive versus subjektive Kommunikation oder dem Rückbezug auf das Allgemeinwohl bzw. demokratischen Werten als Teil der Argumentation. Konzepte wie die Haltung des Sprechers zum Gesagten und der Wahrscheinlichkeitsgehalt des Gesagten sind als rhetorische Mittel imminent wichtig, da sie konventionelle Implikaturen generieren.

Im Folgenden wird ein erster Ansatz zur Analyse von Argumentativität mittels automatischem Verfahren präsentiert. Neben einer umfassenden linguistischen Analyse der relevanten Parameter stellen wir eine computerlinguistische Implementation vor, die die Daten automatisch mit den von Seiten der Pragmatik relevanten Merkmalen annotiert. Diese Implementation kombiniert ein regelbasiertes System, das eine tiefgreifende linguistische Analyse auf die Daten anwendet, mit einem visuellen Analysesystem, das neben der visuellen Darstellung der pragmatischen Informationen eher oberflächliche Sprachmerkmale wie die Identifikation von Schlüsselworten, Modellierung der Topics der Unterhaltung, Standardberechnungen zur Länge von Sprechbeiträgen sowie die Anzahl der Wortwechsel einfließen lässt.

2 Daten

Die für die Analyse herangezogenen Daten basieren auf den Schlichtungsgesprächen zu “Stuttgart 21” (S21), das 2010 als Antwort auf den massiven öffentlichen Druck auf den Bau des Tiefbahnhofs und der damit verbundenen Stadtentwicklung in Stuttgart eingesetzt wurde. Diese als Beispiele für die Anwendung der deliberativen Kommunikation angesehen Verfahren kommen seit den frühen 1980er Jahren in Deutschland zunehmend zum Einsatz. Die Datenbasis ist offen im Netz verfügbar² und besteht aus den transkribierten Protokollen dieser Verhandlung an neun Verhandlungstagen, jeweils mit einer Dauer von etwa sieben Stunden. Insgesamt enthalten die Transkripte etwa 265.000 Token in mehr als 1330 Redebeiträgen von 70 Personen. Die Transkripte umfassen Gespräche und Diskussionen zwischen dem Mediator, Experten, Projektbefürwortern und Projektgegnern in deutscher Sprache und wurden in ein XML-lesbares Format umgewandelt, um eine spätere automatische Verarbeitung und Annotation zu ermöglichen. Basierend auf den Informationen der online verfügbaren Transkripte wurden die Transkripte außerdem mit Sprecherinformation und Gesprächsthema annotiert.

Um eine feine Analyse des Diskurses zu erreichen, wurden alle Redebeiträge in elementare Diskurseinheiten (“elementary discourse units” — EDUs) unterteilt (Marcu 2000). Obwohl in der Literatur

¹www.visargue.uni-konstanz.de

²<http://stuttgart21.wikiwam.de/Schlichtungsprotokolle>

kein Konsensus bezüglich der genauen Eigenschaften einer EDU herrscht, wird generell angenommen, dass jede Diskurseinheit ein einzelnes Event beschreibt (z.B. Polanyi et al. 2004). Im vorliegenden Fall approximieren wir diese Annahme, indem wir alle lexikalischen Einheiten zwischen zwei Interpunktionszeichen als Diskurseinheit angesehen werden.

3 Linguistischer Hintergrund

Ein zentraler Aspekt unserer Arbeit ist die linguistisch-motivierte Operationalisierung der Kriterien die die deliberative Qualität von Kommunikation markieren, insbesondere bezüglich der Art der Realisierung und der kommunikativen Funktion von Argumenten im Diskurs und die Einstellung des Sprechers zum Gesagten. Die vorliegende Arbeit zeigt, dass dazu zwei linguistische Parameter, und zwar kausale Diskurskonnectoren und Modalpartikeln und deren Interaktion, hochrelevant sind.

Kausale Diskurskonnectoren wie *da*, *weil*, *dann*, *zumal* leiten generell die Begründung eines Sprechers ein (z.B. Prasad et al. 2008). Diese Konnectoren und deren begründungseinleitende Phrase können zwar automatisch extrahiert werden, allerdings fehlt zu einer umfassenden linguistischen Interpretation eine Antwort auf die Frage, wie die Aussage getönt ist, beziehungsweise wie forciert die Aussage kommuniziert ist und wie die Haltung des Sprechers zum Gesagten ist. Diese Faktoren werden im Deutschen insbesondere mit Hilfe von Modalpartikeln ausgedrückt (z.B. Zimmermann 2011, Karagjosova 2004). Zum Beispiel leiten *halt* und *eben* eine konventionelle Implikatur ein, die eine vom Sprecher angenommenen unabänderlichen Einschränkung durch externe Fakten ausdrückt. Diese Verwendung wird in Beispiel (1) gezeigt. Im Gegensatz dazu signalisiert *ja* wie in (2) illustriert, dass der Inhalt der Argumentation Teil des gemeinsamen Wissens der Gesprächsteilnehmer ist.

- (1) (...) weil halt in dem Bereich die meisten Autos unterwegs sind.
(...) as HALT in Art area Art most car.Pl underway be.3.Pl
'(...) because most cars are underway in this area.' (Dr. Heiner Geissler, S21, Nov. 4th 2010)
- (2) (...) da Sie ja gesagt haben, dass (...)
(...) as Pron.2.Sg.Pol JA say.Past.Part have.Inf that (...)
'(...) as you JA said that (...)' (Tanja Gönner, S21, Nov. 4th 2010)

3.1 Ambiguität

Ambiguitäten stellen eine große Herausforderung für die automatische Extraktion und Identifikation von Kausalkonnectoren und Modalpartikeln dar. Insbesondere ist dies der Fall für den Konnektor *da*, der neben dem kausalen Gebrauch auch als temporales und lokatives Pronomen fungieren kann. Allerdings kann ein Großteil dieser Ambiguitäten aufgelöst werden, indem linguistische Indikatoren wie die Position des Konnectors, seine angrenzenden Elemente und die generelle Struktur des Satzes mit einbezogen werden. Dies alles fließt in eine Disambiguierungsregel ein, wie in (3) dargestellt.

- (3) IF *da* nicht gefolgt von einem Verb AND
kein anderer Partikel oder Konnektor vor *da* AND
finales Verb ist ein Infinitiv THEN
da ist ein Kausalkonnektor.

Dieselbe Herangehensweise wird für Modalpartikeln angewendet, zum Beispiel kann *eben* neben seiner Verwendung als Modalpartikel auch noch als Fokuspartikel und temporales Adverb auftreten. Als Modalpartikel signalisiert *eben* die resignierte Zustimmung zu einer Sache aufgrund unabänderlicher Randbedingungen (Kwon 2005).

3.2 Inferenzregeln

Während die zwei oben genannten Parameter für sich genommen schon wichtig für die Interpretation des Diskurses sind, entsteht ein zusätzlicher Nutzen durch die Kombination der zwei Dimensionen. Das in (1) gezeigte Beispiel erhält durch die Inferenzregel in (4) die in Abbildung 1 gezeigte Annotation.

- (4) IF Kausalkonnektor gefunden AND
Kausalkonnektor von einer Partikel, der unabänderlichen Randbedingung markiert, gefolgt wird
THEN
annotiere den start tag der Diskurseinheit mit
<DiscRel="justification" CI="immutable_constraint">

```
<discourse_unit id="17" DiscRel="justification" CI="immutable_constraint">
  <lexeme id="1" connector="causal">weil</lexeme>
  <lexeme id="2" particle="resignation_acceptance">halt</lexeme>
  <lexeme id="3">in</lexeme>
  <lexeme id="4">dem</lexeme>
  <lexeme id="5">Bereich</lexeme>
  <lexeme id="6">auch</lexeme>
  <lexeme id="7">die</lexeme>
  <lexeme id="8">meisten</lexeme>
  <lexeme id="9">Autos</lexeme>
  <lexeme id="10">unterwegs</lexeme>
  <lexeme id="11">sind</lexeme>
</discourse_unit>
```

Abbildung 1: Annotation Beispiel (1)

Im Gegensatz dazu annotiert die Annotationsregel in (5) die Kombination des Kausalkonnektors *da* mit der Modalpartikel *ja*; daraus folgt die Annotation des Beispiels aus (2) in Abbildung 2.

- (5) IF *da* ein Kausalkonnektor AND
da gefolgt von einer Partikel, der Gemeinsamkeit markiert THEN
annotiere den start tag der Diskurseinheit mit
<discrel="justification" CI="common_ground">.

```
<discourse_unit id="2" DiscRel="justification" CI="common_ground">
  <lexeme id="1" connector="causal">da</lexeme>
  <lexeme id="2">Sie</lexeme>
  <lexeme id="3" particle="common_ground">ja</lexeme>
  <lexeme id="4">gesagt</lexeme>
  <lexeme id="5">haben</lexeme>
</discourse_unit>
```

Abbildung 2: Annotation Beispiel (2)

Trotz der relativ geringen Größe des Korpus ist es schwer, übergreifende Muster in der Argumentativität über den Diskursverlauf hinaus auf einen Blick zu sehen, ohne den Blick auf einzelne Annotationen zu verlieren. Um diesem Hindernis zu begegnen, werden die Annotationen in einem Visualisierungssystem dargestellt, das die Muster zugänglicher macht. Dieses wird im Folgenden kurz erläutert.

4 Visualisierung von Argumentativität

Die Visualisierung linguistischer Muster hat in vielen Fällen gezeigt, dass diese Art der Informationsverarbeitung große Vorteile in der Erkennung von Mustern darstellt: bei theoretisch motivierten Phänomenen wie phonologischen Mustern (Mayer and Rohrdantz 2013) und Bedeutungswandel (Rohrdantz et al. 2011, 2012), bis hin zu Fragestellungen des maschinellen Lernens im Bereich des automatischen Clusterings (Lamprecht et al. 2013). Das Ziel der Visualisierung im vorliegenden Fall ist zum einen die Darstellung der Annotation der oben genannten Inferenzregeln; zum anderen kann die Verteilung der Annotation über den Diskursverlauf hinweg Aufschlüsse über den Grad der Deliberation liefern.

Abbildung 3 zeigt die Visualisierung eines Teils der S21-Mediationssitzung vom 4. November 2010; hierbei füllt jeder Satz eine eigene Zeile und jeder Sprechbeitrag wird von einem grauen Rechteck umrandet. Die gelben Balken markieren diejenigen Diskurseinheiten, die Kausalkonnektoren und damit Begründungen beinhalten.

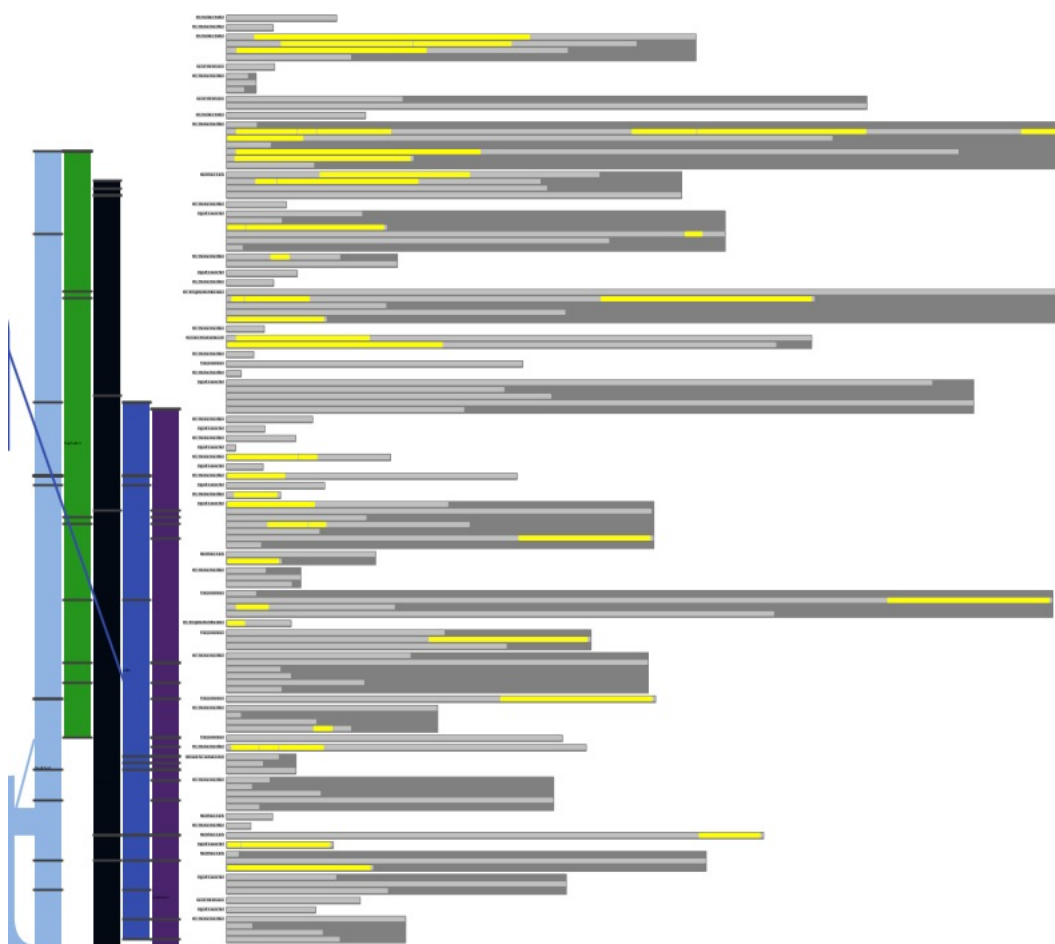


Abbildung 3: Visualisierung kausal begründeter Aussagen in der S21-Sitzung am 4. November 2010

Das entwickelte Werkzeug ist dahingehend interaktiv, als dass der Benutzer hinein- und herauszoomen und relevante Diskurseinheiten im Detail untersuchen kann, ohne die allgemeine Verteilung aus den Augen zu verlieren. Eine detaillierte Ansicht des Diskurses wird in Abbildung 4 gezeigt.

5 Zusammenfassung und weitere Fragestellungen

Dieses Papier präsentiert eine Methode der Operationalisierung des Begriffs der Deliberation durch Diskurskonnektoren und Modalpartikeln mit dem Ziel, die Mittel und Wege zu beleuchten, mit denen Argumente ausgetauscht und unter Sprechern und Zuhörern auf sie Bezug genommen wird. Durch die Verwendung eines Visualisierungsansatzes können die annotierten Datensätze über den gesamten Diskurs

Dr. Volker Kefer

nein nicht verzweifeln.

Ich gebe hier nur den Überblick ich gehe auf die einzelnen Punkte gleich nochmal ein.

Also maschineller Vortrieb ist das zweite Thema.

Das Dritte ist Reduktion der Mehrkosten gegenüber 2004 bei eisenbahntechnischer Ausrüstung.

Auch dazu hab ich ein Bild mitgebracht.

Und eine Herausnahme von Rückbaukosten für eine bestimmte Fläche weil das Bestandteil eines Immobilienprojektes ist.

Hier hab ich keine Folie dabei weil der Sachverhalt ja ein sehr einfacher ist.

Es geht nicht darum dass wir den Rückbau nicht machen wollen sondern es geht darum dass wir ihn schlichtweg

Abbildung 4: Detaillierte Visualisierung von begründenden Diskurseinheiten

hinweg inspiziert werden, was eine Interpretation der Rolle von Argumentativität im Meinungsbildungsprozess ermöglicht.

Künftige Ziele umfassen die Miteinbeziehung von zusätzlichen diskursrelevanten linguistischen Mustern sowie Multiword-Elementen. Bei steigender Anzahl der Annotationsebenen soll die Visualisierung entsprechend erweitert werden, sodass Interaktionen zwischen verschiedenen Ebenen aufgezeigt und tiefere Einsichten in Diskursstruktur und letztendlich Deliberation gewonnen werden können.

Literatur

- Bohman, James. 1996. *Public Deliberation: Pluralism, Complexity and Democracy*. Cambridge, MA: The MIT Press.
- Dryzek, John S. 1990. *Discursive Democracy: Politics, Policy, and Political Science*. Cambridge, MA: Cambridge University Press.
- Dryzek, John S. 2000. *Deliberative Democracy and Beyond: Liberals, Critics, Contestations*. Oxford: Oxford University Press.
- Gutmann, Amy and Thompson, Dennis Frank. 1996. *Democracy and Disagreement. Why moral conflict cannot be avoided in politics, and what should be done about it*. Cambridge, MA: Harvard University Press.
- Habermas, Jürgen. 1981. *Theorie des kommunikativen Handelns*. Frankfurt am Main: Suhrkamp.
- Karagjosova, Elena. 2004. *The Meaning and Function of German Modal Particles*. Saarbrücken Dissertations in Computational Linguistics and Language Technology.
- Kwon, Min-Jae. 2005. *Modalpartikeln und Satzmodus: Untersuchungen zur Syntax, Semantik und Pragmatik deutscher Modalpartikeln*. Ph. D.thesis, LMU München.
- Lamprecht, Andreas, Hautli, Annette, Rohrdantz, Christian and Bögel, Tina. 2013. A Visual Analytics System for Cluster Exploration. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 109–114, Sofia, Bulgaria: Association for Computational Linguistics.
- Marcu, Daniel. 2000. *The Theory and Practice of Discourse Parsing and Summarization*. MIT Press, Cambridge, Mass.
- Mayer, Thomas and Rohrdantz, Christian. 2013. PhonMatrix: Visualizing co-occurrence constraints in sounds. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics: System Demonstrations*.
- Polanyi, Livia, Culy, Chris, van den Berg, Martin, Thione, Gian Lorenzo and Ahn, David. 2004. Sentential structure and discourse parsing. In *Proceedings of the 2004 ACL Workshop on Discourse Annotation*, pages 80–87.
- Prasad, Rashmi, Dinesh, Nikhil, Lee, Alan, Miltsakaki, Eleni, Robaldo, Livio, Joshi, Aravind and Webber, Bonnie. 2008. The Penn Discourse Treebank 2.0. In *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008)*, pages 2961–2968.
- Rohrdantz, Christian, Hautli, Annette, Mayer, Thomas, Butt, Miriam, Keim, Daniel A. and Plank, Frans. 2011. Towards Tracking Semantic Change by Visual Analytics. In *Proceedings of the 49th Annual*

- Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL-HLT '11): shortpapers*, pages 305–310, Portland, Oregon: Association for Computational Linguistics.
- Rohrdantz, Christian, Niekler, Andreas, Hautli, Annette, Butt, Miriam and Keim, Daniel. 2012. Lexical Semantics and Distribution of Suffixes — A Visual Analysis. In *Proceedings of the EACL 2012 Joint Workshop of LINGVIS and UNCLH*, pages 7–15.
- Zimmermann, Malte. 2011. Discourse Particles. In Paul Portner, Claudia Maienborn and Klaus von Heusinger (eds.), *Semantics (Handbücher zur Sprach- und Kommunikationswissenschaft)*, pages 2011–2038, Mouton de Gruyter.