

Towards visualizing linguistic patterns of deliberation: a case study of the S21 arbitration

Bögel, Tina

Universität Konstanz

tina.boegel@uni-konstanz.de

Gold, Valentin

Universität Konstanz

valentin.gold@uni-konstanz.de

Hautli-Janisz, Annette

Universität Konstanz

annette.hautli@uni-konstanz.de

Rohrdantz, Christian

Universität Konstanz

christian.rohrdantz@uni-konstanz.de

Sulger, Sebastian

Universität Konstanz

sebastian.sulger@uni-konstanz.de

Butt, Miriam

Universität Konstanz

miriam.butt@uni-konstanz.de

Holzinger, Katharina

Universität Konstanz

katharina.holzinger@uni-konstanz.de
<https://papers.lirm.fr/paper/1381217065.145.169.16.html>

Keim, Daniel A.

Universität Konstanz

daniel.keim@uni-konstanz.de

Short Paper

Towards visualizing linguistic patterns of deliberation: a case study of the S21 arbitration

- Bögel, Tina
Universität Konstanz
tina.boegel@uni-konstanz.de
- Gold, Valentin
Universität Konstanz
valentin.gold@uni-konstanz.de
- Hautli-Janisz, Annette
Universität Konstanz
annette.hautli@uni-konstanz.de
- Rohrdantz, Christian
Universität Konstanz
christian.rohrdantz@uni-konstanz.de
- Sulger, Sebastian
Universität Konstanz
sebastian.sulger@uni-konstanz.de
- Butt, Miriam
Universität Konstanz
miriam.butt@uni-konstanz.de
- Holzinger, Katharina
Universität Konstanz
katharina.holzinger@uni-konstanz.de
- Keim, Daniel A.
Universität Konstanz
daniel.keim@uni-konstanz.de

Summary

This paper presents on-going work on analyzing strategies for argumentation via automatic means, using public data from a German arbitration. In addition to a thorough linguistic analysis of the relevant parameters, we provide a computational implementation that automatically annotates the corpus with respect to pragmatically relevant features. The annotated data is used in a visual analysis system which shows overall patterns of the discourse as well as detailed views on the data.

1. Introduction

This paper reports on the interdisciplinary project VisArgue which is concerned with the automatic linguistic and visual analysis of political discourse with a particular focus on the concept of deliberative communication [1] [2] [3] [4] [5]. According to the theory of deliberative argumentation, stakeholders participating in a multilog, i.e. a multi-party conversation, should justify their positions truthfully and rationally and should eventually defer to the better argument.

Automatically measuring the deliberative quality of a multilog calls for an identification of linguistic cues that shed light on issues such as objective vs. subjective argumentation, invocation of the common good or democratic notions as part of the argument. Notions such as speaker stance, speaker belief/certainty are also immediately relevant, as is an analysis of rhetorical devices known to trigger conventional implicatures [6]. In short, a promising way of arriving at an operationalization of the indications for the deliberative quality of a multilog is a linguistic analysis of the linguistic cues present in the multilog.

This paper presents on-going work on analyzing strategies for argumentation via automatic means, using public data from a German arbitration. In addition to a thorough linguistic analysis of the relevant parameters, we provide a computational implementation that automatically annotates the corpus with respect to pragmatically relevant features. This implementation combines a rule-based system that reflects deep linguistic analysis with a visual analysis system that also provides results with respect to more shallow natural language processing methods such as keyword identification, topic modeling, and standard calculations with respect to length of utterances, amount and type of turn-taking, etc.

2. Data

In Germany, the method of deliberative discourse has been increasingly applied to the resolution of large-scale public conflicts since the early 1990s. One recent well-known example is the public arbitration process on "Stuttgart 21" (S21), a new railway and urban development project in the city of Stuttgart. In response to massive public criticism of the project, a public arbitration procedure was established. The data for our initial investigation are the transcribed minutes of the S21 arbitration process, which consist of nine days of sessions, each lasting for around seven hours, with a total of about 70 different speakers. The transcripts consist of spoken German conversation between mediator, experts, project supporters and opponents and are converted into an XML-readable format in order to facilitate later processing and annotation. Based on the information contained in the web transcripts, the XML transcripts are annotated with speaker information and the general topic of the session. Overall, the transcripts contain around 265.000 tokens in 1330 utterances.

In order to arrive at a more fine-grained analysis of the discourse, all utterances have to be split up into elementary discourse units (EDUs) [7]. Although there is no consensus in the literature on what EDUs are, in general, each DU is assumed to describe a single event (e.g., [8]). In the case at hand, we approximate this by treating all lexical items between two punctuation marks as belonging to one DU.

3. Linguistic background

A central aspect of our work is a linguistically motivated operationalization of features that indicate the deliberative quality of a multilog, important parameters being the realization and the communicative function of arguments in the discourse as well as speaker stance and speaker belief. We have decided to initially focus on just two of the relevant linguistic parameters found in German: the interaction between causal discourse connectors and modal particles.

Causal discourse connectors (e.g., *da*, *weil*, *denn*, *zumal* 'because (of)/due to/as') generally introduce a justification of a speaker's statement (e.g., [9]). These connectives and the justification they indicate can be extracted automatically. However, the precise shade and force of the argument being made, including speaker stance and speaker belief are modulated in spoken German by a heavy use of modal particles (e.g., [10] [11]). For instance *halt* or *eben* indicate a conventional implicature that the speaker believes the argument to refer to an immutable constraint imposed by the outside world, exemplified in (1). The particle *ja* in (2), in contrast, signals that the speaker assumes that the content of the argument is part of the common ground of the multilog participants.

(1) [...] weil halt in dem Bereich die meisten Autos unterwegs sind.
[...] as HALT in Art area Art most car.PI underway be.3.PI
'[...] 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. Ambiguity

Ambiguity presents a serious problem for the automatic extraction and identification of both causal connectors and modal particles. For example, especially *da* 'as' presents a challenge for automatic processing, because of its multi-functional usage as either the temporal or locative pronoun 'there' or as a connector meaning 'because'. However, such ambiguities can be largely resolved by taking linguistic factors such as the position of the connector, its neighboring elements and the general structure of the carrier sentence into account. In (3), we schematize the identification rule for the German causal connector *da* 'as'.

(3) IF *da* not followed by verb AND
 da not preceded by a particle or another causal connector AND
 final verb is an infinitive THEN

da is a causal connector.

The same procedure is followed with respect to modal particles such as *eben*, which can be a focus particle, a temporal adverbial meaning 'just', or a modal particle that indicates the speaker's resigned acceptance of a fact due to an immutable constraint [12].

3.2. Inference rules

While these two dimensions are by themselves important for the interpretation of a given discourse, the additional benefit for measuring deliberation results from a combination of the two dimensions. Taking the example in (1), the inference rule in (4) yields the annotation in Figure 1.

- (4) IF causal connector found AND
causal connector followed by particle denoting immutable constraint THEN
annotate the DU start tag with <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>
```

Fig. 1: Annotation of example (1).

On the other hand, the rule in (5) deals with the combination of the causal connector *da* and the modal particle *ja*, rendering the annotation of example (2) in Figure 2.

- (5) IF *da* is used as causal connector AND
da is followed by particle denoting common ground THEN
annotate the DU start tag with <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>
```

Fig. 2: Annotation of example (2).

These inferences on the level of the discourse unit help to interpret the whole discourse and shed light on the way speakers and listeners interact.

Despite the comparatively small corpus, it is nevertheless difficult to see overall patterns of argumentativity at a glance, while still maintaining a detailed view on single annotations. In order to overcome this drawback, we introduce a visualization system which encodes those annotations visually and makes the patterns more

accessible.

4. Visualizing argumentativity

The visualization of linguistic patterns has been shown to shed light on a number of phenomena, from theoretically motivated topics like phonological patterns [13] and lexical semantic change [14] [15] to machine learning issues with respect to clustering [16]. The goal of visualizing the structure of argumentativity across the discourse is twofold: First, patterns of argumentation that have been identified through the linguistic inference rules can be analyzed in their context. Second, the distribution of arguments over the course of the conversation may reveal additional knowledge on the deliberative quality of different parts of the overall discourse.

Figure 3 shows a visualization of parts of the S21 arbitration session on Nov. 4th 2010, each sentence occupying one line, each speaker turn contained in a grey square. The bars marked in yellow represent discourse units containing justifying statements. The tool is interactive in that the user can zoom in and out of the discourse and can investigate the relevant discourse units in detail without losing the overall distribution. A detailed view on the data is shown in Figure 4.

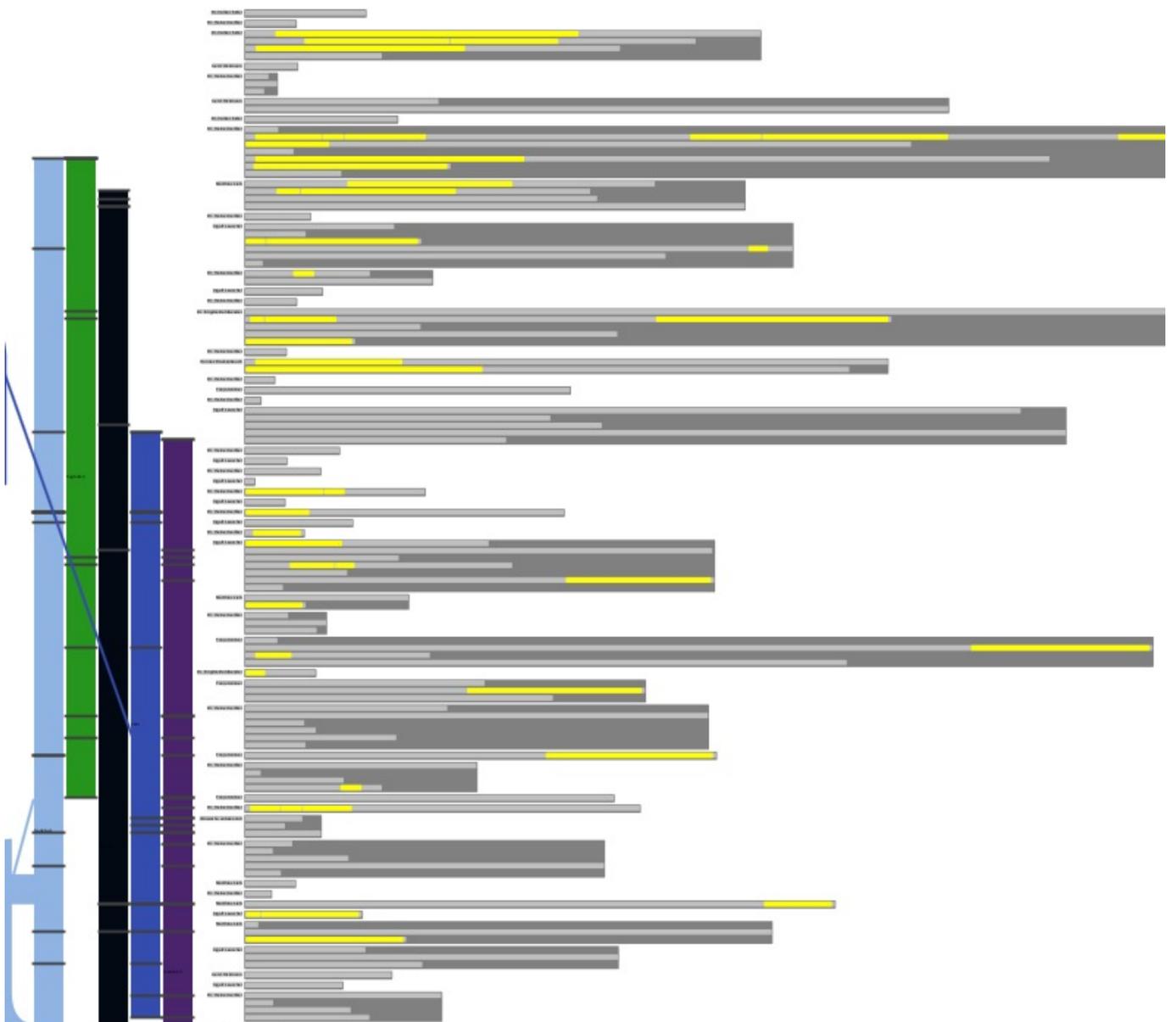


Fig. 3: Visualization of justifying statements in the S21 arbitration on November 4th, 2010.

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

Fig. 4: Detailed visualization of justificational discourse units.

5. Summary and future work

This paper presents an approach of operationalizing the notion of deliberation using discourse connectors and modal particles in order to shed light on the way arguments are exchanged and how speakers and listeners relate to them. By using a visualization approach, the annotated data can be inspected over the whole discourse, allowing for an interpretation of the role that argumentativity plays in the arbitration. In the future, we will incorporate more linguistic cues that are relevant for deliberative communication and also deal with multiword instances that are relevant on a number of levels. With the increasing number of annotation levels, the visualization will be extended to show the interactions between different levels, allowing for more insights into discourse structure and eventually deliberation.

Acknowledgements

We thank Mennatallah el Assady, Manuel Hotz and Rita Sevastjanova for their help with implementing the discourse visualization and the German Ministry for Education and Research (BMBF) for their funding under the eHumanities research grant 01UG1246.

References

1. Habermas, Jürgen. 1981. *Theorie des kommunikativen Handelns*. 2 Bände. Frankfurt am Main: Suhrkamp.
2. Habermas, Jürgen. 1992. *Faktizität und Geltung: Beiträge zur Diskurstheorie des Rechts und des demokratischen Rechtsstaates*. Frankfurt am Main: Suhrkamp.
3. Dryzek, John S. 1990. *Discursive Democracy: Politics, Policy and Political Science*. Cambridge: Cambridge University Press.
4. Bohman, James. 1996. *Public Deliberation: Pluralism, Complexity and Democracy*. Cambridge, MA: The MIT Press.
5. Gutmann, Amy and Dennis Thompson. 1996. *Democracy and Disagreement — Why moral conflict cannot be avoided in politics, and what should be done about it*. Cambridge, MA: Harvard University Press.
6. Potts, Christopher. 2012. Conventional implicature and expressive content. In Claudia Maienborn, Klaus von Heusinger, and Paul Portner, eds., *Semantics: An International Handbook of Natural Language Meaning*, Volume 3, 2516-2536. Berlin: Mouton de Gruyter.
7. Marcu, Daniel. 2000. *The Theory and Practice of Discourse Parsing and Summarization*. Cambridge, MA: The MIT Press
8. Polanyi, Livia, Martin van den Berg, Chris Culy, Gian Lorenzo Thione, David Ahn. 2004. Sentential Structure and Discourse Parsing. *Proceedings of the ACL2004 Workshop on Discourse Annotation*.
9. Prasad, Rashmi, Nikhil Dinesh, Alan Lee, Eleni Miltsakaki, Livio Robaldo, Aravind Joshi & Bonnie Webber. 2008. The Penn Discourse Treebank 2.0. *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008)*. Marrakech, Morocco.
10. Zimmermann, Malte. 2011. *Discourse Particles*. In P. Portner, C. Maienborn und K. von Heusinger (eds.), *Semantics*. (Handbücher zur Sprach- und Kommunikationswissenschaft HSK 33.2). Berlin, Mouton de Gruyter. 2011-2038.
11. Karajosova, Elena. 2004. *The Meaning and Function of German Modal Particles*. Saarbrücken Dissertations in

Computational Linguistics and Language Technology.

12. Min-Jae Kwon (2005). Modalpartikeln und Satzmodus. Untersuchungen zur Syntax, Semantik und Pragmatik der deutschen Modalpartikeln. Dissertation, LMU Muenchen.
13. Mayer, Thomas, Christian Rohrdantz. 2013. PhonMatrix: Visualizing co-occurrence constraints of sounds. In Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, pages 73-78.
14. Rohrdantz, Christian, Annette Hautli, Thomas Mayer, Miriam Butt, Daniel A. Keim and Frans Plank. 2011. Towards Tracking Semantic Change By Visual Analytics. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics, pages 305-310.
15. Rohrdantz, Christian, Andreas Niekler, Annette Hautli, Miriam Butt and Daniel A. Keim. 2012. Lexical Semantics and Distribution of Suffixes - A Visual Analysis. In Proceedings of the EACL 2012 Joint Workshop of LINGVIS & UNCLH, pages 7-15.
16. Lamprecht, Andreas, Annette Hautli, Christian Rohrdantz and Tina Bögel. 2013. A Visual Analytics System for Cluster Exploration. In Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, pages 109-114.