

Visual linguistic analysis of political discussions: Measuring deliberative quality

Valentin Gold

Department of Politics and Public Administration, University of Konstanz, Germany

Mennatallah El-Assady

Department of Computer and Information Science, University of Konstanz, Germany

Annette Hautli-Janisz

Department of Linguistics, University of Konstanz, Germany

Tina Bögel

Department of Linguistics, University of Konstanz, Germany

Christian Rohrdantz

Department of Computer and Information Science, University of Konstanz, Germany

Miriam Butt

Department of Linguistics, University of Konstanz, Germany

Katharina Holzinger

Department of Politics and Public Administration, University of Konstanz, Germany

Daniel Keim

Department of Computer and Information Science, University of Konstanz, Germany

Abstract

Correspondence:

Valentin Gold, Department of Politics and Public Administration, PO Box 90, Universität Konstanz, 78457 Konstanz, Germany.

E mail:

valentin.gold@uni-konstanz.de

This article reports on a Digital Humanities research project which is concerned with the automated linguistic and visual analysis of political discourses with a particular focus on the concept of deliberative communication. According to the theory of deliberative communication as discussed within political science, political debates should be inclusive and stakeholders participating in these debates are required to justify their positions rationally and respectfully and should eventually defer to the better argument. The focus of the article is on the novel interactive

visualizations that combine linguistic and statistical cues to analyze the deliberative quality of communication automatically. In particular, we quantify the degree of deliberation for four dimensions of communication: Participation, Respect, Argumentation and Justification, and Persuasiveness. Yet, these four dimensions have not been linked within a combined linguistic and visual framework, but each single dimension helps determining the degree of deliberation independently from each other. Since at its core, deliberation requires sustained and appropriate modes of communication, our main contribution is the automatic annotation and disambiguation of causal connectors and discourse particles.

1 Introduction

For the last two decades, notions of deliberative democracy have been intensively debated within political science and related fields. In recent years, deliberation research has experienced an empirical turn (Chambers, 2003). In particular, the deliberative quality of communication and its consequences on the overall political decision-making process has attracted attention, partly in light of highly public resistance to political agendas with respect to major development projects (e.g. airports, train stations, fracking). Some natural questions that arise with respect to this are whether the deliberative quality of political discussions has an impact on the final decisions that are taken and whether a higher deliberative quality leads to greater acceptance of the final decision that is taken. Particularly, does a higher deliberative quality of a political discussion lead to greater acceptance of the arguments that have been brought forward and of the final decision that is taken?

A crucial component for finding an answer to these questions is the determination of the deliberative quality of a discussion. The tools that have been developed so far within political science to measure the deliberative quality of a given discussion are based on the manual coding of deliberative categories and a subsequent statistical analysis of the categories coded. The Discourse Quality Index, developed by Steenbergen *et al.* (2003) is thus far the most prominent manual coding scheme (Hangartner *et al.*, 2007; Thompson, 2008; Lord and Tamvaki, 2013). This procedure, however, is beset with several difficulties. One difficulty is that manual coding is comparatively (labor) expensive and takes a long time (e.g. Black *et al.*, 2010, p. 329).

Another is the lack of inter-annotator agreement. The categories developed so far often fall prey to subjective judgments of the human annotators, thus leading to a problematic amount of disagreement among annotators (Black *et al.*, 2010, p. 330; Dacombe, 2013).

A desideratum in research on deliberative democracy is thus the automatic coding and analysis of political discussions according to criteria which reflect the deliberative nature of a discussion. In this article, we present an approach that draws on a combination of linguistics and visual analytics in the creation of an automatic annotation system that can be used for the analysis of deliberative quality. The approach is interdisciplinary and falls under the purview of ‘Digital Humanities’.

From the political science perspective, deliberation is defined as a communicative process that is based on an inclusive and constructive debate between the participating stakeholders (Habermas, 1981; Gutmann and Thompson, 1996). Hence, this definition refers to the notion of the procedural importance of how decisions are made. Since deliberative decision-making is complex, we assume that deliberative quality is a latent construct (i.e. not a directly observable variable) consisting of several observable measures that can be used to approximate the overall deliberative quality of a discussion.

From the linguistic perspective, these observable measures mainly consist of linguistic features present in the communication. Examples are rhetorical devices designed to make the communication as persuasively effective as possible, e.g. the use of linguistic features to establish a common ground/ understanding between the discussants (rhetorical questions, use of inclusive pronouns such as

‘we/us’ rather than exclusive ones such as ‘I/you’, discourse particles, tag questions, etc.), the presentation of justified arguments (identified, for example, by the linguistic feature of causal connectors, e.g. ‘because’), verbs signaling speaker stance with respect to a certain topic (‘think, believe’ versus ‘know’ or ‘accept, reject’), but also indications of respect and politeness conveyed by one speaker to another (e.g. interruptions are generally known to signal disrespect and tend to be employed by more dominant speakers, see Brown and Levinson (1987) for more details on linguistic markers of politeness).

Taking together the political science and the linguistic perspectives, we have identified four broad areas for which to calculate deliberative quality: Participation (Section 4.1), Respect (Section 4.2), Argumentation and Justification (Section 4.3), and Persuasiveness (Section 4.4). Among these areas, we focus mainly on the area of argumentation and justification in order to demonstrate our overall linguistic visual analytics approach. In particular, we provide a computational implementation that automatically annotates corpora of deliberative communications with respect to linguistic and meta-linguistic features in each of these areas. Our implementation combines a rule-based system that reflects deep linguistic analysis with more shallow natural language processing (NLP) approaches that include standard strategies such as keyword identification, topic modeling, and calculations of utterance length, but also innovative perspectives on the data. The annotated data are further processed in the visual analytics system that (1) depicts structures in the data through adapted textual data mining algorithms; and (2) allows an explorative and interactive access to the underlying data.

Details on this are presented in Section 4, after a brief look at related work in Section 2 and a description of our underlying methodology and data in Section 3. Section 5 concludes the article.

2 Related Work

There are several interdisciplinary approaches which focus on understanding argumentation in

discussions. Here, we briefly discuss the ones closest to our research interests and distinguish them from our approach.

A Digital Humanities cooperation between communication sciences, computer science, and computational linguistics is currently looking at how the exchange of arguments plays out in unmoderated exchanges within social media such as Twitter.¹ This collaboration does not contain a political science component, nor does it contain a visual analytics component. However, like the work described in this article, a considerable part of the overall effort is directed at identifying, understanding, quantifying, and analyzing linguistic features of argumentation found in the data. Twitter data are quite different from the political discussions investigated here; however, both our efforts focus on discussions in which the language of communication is German, and we have found that both our efforts have so far identified similar linguistic features as being relevant for an overall analysis of argumentation (e.g. we have both identified rhetorically significant interactions between discourse particles and other parts of the grammar, see Scheffler, 2014).

The ‘eIdentity’ project is also a Digital Humanities project.² It involves a collaboration between political science and computational linguistics. However, it does not include a visual analytics component, and its research topic is quite different. The overall goal of the ‘eIdentity’ project is to identify collective identities: how they are formed and how these change over time. The project works with large amounts of text and seeks to identify semantic fields that reflect complex concepts within these large corpora (Blessing *et al.*, 2013). The semantic fields—semantic co-occurrences of words related to collective identities—are meant to provide an automated assistance to the work conducted by a human researcher.

Additionally, the goal of the ‘epol’ project³ is to measure neo-liberal argumentation and trace its impact over time (Wiedemann *et al.*, 2013). This project is again a Digital Humanities project and represents a collaboration between political science and computer science. The latter includes visual computing as well as NLP. ‘epol’ is also concerned with the identification of arguments in a political

context. One difference to our work lies in the type of data they use. While we work oral negotiations, ‘epol’ uses newspaper articles on certain predefined political topics. In contrast to oral discussions, newspaper articles represent written and edited language. Another difference lies in the depth to which the data are processed. Rather than aiming at a deeper linguistic understanding of features or argumentation, the project focuses on text mining, i.e. on using shallow approaches to extracting information from a given text. We use these approaches as well as deeper linguistic knowledge.

Finally, DocuScope (Kaufer *et al.*, 2006) provides a text analysis environment to determine rhetorical effects. The software allows classifying over 100 categories of rhetorical effects, e.g. emotions or confidence. In general, DocuScope allows incorporating any dictionary and visualizes the categories. DocuScope provides a good starting point for any dictionary-based approach. Our approach, however, goes beyond visualizing dictionaries and provides various tools determining the degree of deliberative quality.

3 Data and Methodology

In this section, we briefly describe our data and overall methodology. The linguistic visual analysis is applied to two different types of data: (1) multi-party negotiations which are the result of simulations conducted in an experimental setting; and (2) real-world examples of political communication. In what follows, we provide details for each of the disciplines involved in the work.

3.1 Political science

One of the main challenges in the analysis of deliberation is the collection and analysis of data with regard to oral communications. Most of the work conducted does not study the effects of synchronic face-to-face communications and instead tends to analyze asynchronous communication via digital means (see also the discussion in Section 2 on arguments in Twitter data). For instance, Sulkin and Simon (2001) allow 200 s of computer-based communication in order to analyze the effects on

decision-making processes. Persson *et al.* (2013) allow face-to-face deliberation but do not analyze the deliberative quality of communication. Instead, they focus on the individual effects with regard to legitimacy of the decision-making outcome.

In order to overcome this shortcoming, we have run a large number of simulation-gaming experiments. In these simulations, experimental subjects were asked to discuss the pros and cons of fracking and to decide unanimously whether fracking should be allowed in general or not. Each experimental subject had to argue either in favor of or against fracking. To allow for a comparative analysis, the experimental subjects were provided with a predefined set of arguments. Moreover, the experimental subjects had to answer surveys before and after the discussion. Overall, we have conducted thirty-four experiments. Each of the experiments lasted about 2.5 h, with a total time of 1 h of group discussion. In most simulations, the maximum of 1 h of discussion was fully made use of by the subjects. This provides us with the necessary amount of comparative data to test and evaluate the deliberative structure and content of political discussions. It also provides us with more data than is feasible to annotate manually.

For the purposes of this article, we demonstrate our automatic visual linguistic approach not with respect to the fracking simulations, but with respect to a real-world example of political discussion. Yet, we have used the experimental data to identify relevant deliberative features which can be further extracted and analyzed in real-world discussion. In this article, we work with the publicly available (transcribed) data of the public arbitration that took place with respect to ‘Stuttgart21 (S21)’, a railway and urban development project in Southern Germany. The project includes the restructuring of the central station in Stuttgart. Ever since the project was officially announced in the late 1980s, criticism was expressed. It was not until the late 2000s, however, that large demonstrations and protests with over 100,000 participants took place. The main aim of the protests was directed against the demolition of the existing central railway station. On 30 September 2010, hundreds of protesters were injured when the police tried to secure the

beginning of the construction work. This triggered massive public outrage (and a change in the government). In response, the (new) government agreed to establish a public arbitration procedure to discuss the facts of the project with both supporters and opposition. Between 22 October and 27 November 2010, the public arbitration took place. Within eight rounds of arbitration, supporters and opposition discussed the merits of the project. The discussions were broadcasted live. The data we use to demonstrate the automated methods are the official transcripts that are available online.⁴ Overall, this provides us with a corpus of around 6,000 utterances.

3.2 Linguistics and computational linguistics

The need for an automated annotation of relevant linguistic markers in the political discussions poses a challenge for linguistics and computational linguistics. The challenge for linguistic analysis lies in identifying and understanding the linguistic markers that are relevant for measuring the deliberative quality of a discussion. In particular, while much work has been done on understanding the pragmatic import of linguistic features within English, there is very little previous work to draw on for German. For example, while in English polar questions and hedges are known to be used to signal a broad range of speech acts and speaker stance (e.g. Lakoff, 1975; Asher and Reese, 2005), these do not feature prominently in our corpora. Instead, an interaction between discourse particles (which English does not have) and other parts of the grammar, such as causal connectors, appears to play a large role in conveying pragmatic speech acts. In our work, we have thus concentrated on these.

Once relevant linguistic features have been identified, a further challenge must be surmounted with respect to computational linguistics. Computational Linguistics is concerned with the automatic extraction of linguistic information of a given data set. While fairly reliable tools exist for the annotation of German data with respect to morphological analysis (Schiller, n.d.), Part-of-Speech (POS) annotation⁵ and syntactic analysis (Schmid, 1995; Dipper, 2003), it is notoriously difficult to automatically

and reliably identify information at the semantic and pragmatic level. In our work, we have thus concentrated on finding those linguistic markers that can be identified reliably via automatic methods and have implemented programs to annotate the corpora automatically with the relevant information.

In a first step, the data sets to be analyzed are converted into an XML-readable format. This is to guarantee the exchange of data across different platforms (interoperability) and in order to facilitate the annotation and subsequent extraction of linguistic information, as we can make use of the hierarchical organization of information that XML facilitates. In a next step, the data sets are organized in terms of elementary discourse units (EDUs) (Marcu, 2000). This step also bears its own challenges—for our purposes and conforming with general current practice within computational linguistics, all lexical items between two punctuation marks are treated as belonging to one discourse unit. The data are then further annotated with morphological and POS information. These annotations at a very basic level of linguistic analysis then provide the input for our more sophisticated annotation layer. For example, where possible, we annotate EDUs as to what kind of speech act is being performed by the speaker. Consider here just the example of ‘justification’. The primary linguistic marker for this is taken to be causal connectors such as ‘weil’, ‘deshalb’, ‘da’ (‘because’). However, most of the causal connectors used in German are ambiguous between a complementizer reading and a different POS. For example, ‘da’ is also used as a spatial term meaning ‘there’. In order to disambiguate the occurrences, we integrate information about position in the clause and the POS of the elements surrounding the word in question. The overall work represents a fairly deep linguistic analysis of each of the EDUs.

The overall result of the linguistic and computational linguistic work is a corpus that is annotated with several different types of linguistic information. Some of this information can be used as is, some of it can be used together with other information contained in the corpus as the basis for calculating the effect of further interactions between different elements in an utterance (concrete examples are

presented in Section 4.3), or with respect to the overall patterns found in the data. This is exactly what the visual analytics component does.

3.3 Visual analytics

Visual Analytics has been defined as ‘the science of analytical reasoning facilitated by interactive visual interfaces’ (Thomas and Cook, 2006). Within the field of digital humanities, the approaches of distant reading (Moretti, 2013) and algorithm criticism (Ramsay, 2003, Ramsay, 2011) share fundamental principles with Visual Analytics. In contrast to the longer-standing field of Information Visualization, where data (typically numerical data) are directly transformed into visualizations, Visual Analytics involves automated algorithmic analyses of the data before and after visualization. This procedure is described by the Visual Analytics Mantra ‘Analyse First—Show the Important—Zoom, Filter and Analyse Further—Details on Demand’ (Keim *et al.*, 2008). It has been shown that Visual Analytics approaches can be very beneficial to the analysis of language and linguistic data. First, statistical and algorithmic analyses are performed on text data and then suitable visual representations are designed to show the outcomes of the analyses. Illustrative examples are visualizations of vowel harmonic constraints within languages (Mayer *et al.*, 2010), cross-linguistic comparisons of linguistic features (Rohrdantz *et al.*, 2012), or approaches for tracking semantic change (Rohrdantz *et al.*, 2011).

Only few approaches exist that closely relate to our goals and tasks. First, for the analysis of conversation content, Angus *et al.* (2012a,b) suggest Conceptual Recurrence Plots. All utterances of a multiparty conversation are displayed as rectangles along the diagonal of a triangle. The rectangles within the triangle indicate for each pair of utterances how much they relate in content. Different patterns within the triangle indicate different kinds of concept recurrence, e.g. utterances that summarize the content of several previous utterances.

Second, Nguyen *et al.* (2013) introduce Argviz, a visualization system of the topical structure of multiparty conversations based on topic modeling. A strength of the system is that topic shifts can be spotted in topic columns, which are coordinated

with further standard views on the discourse. Their topic modeling strategy, however, requires a whole corpus of related multiparty conversations in order to be trained. Topics and text content are not contained in one single display, but distributed over coordinated views.

Both approaches help in obtaining insight into the topical structure of conversation. Our goal is to go beyond that and incorporate further characteristics that are of relevance for measuring deliberation into our analyses. We therefore adapt and extend established visualization technologies and introduce novel approaches in order to enable an interactive exploration of deliberative communication.

4 Quantifying Deliberative Quality

In this section, we will demonstrate how deliberative quality can be measured automatically using linguistically informed visual analyses. For each of the four pillars, we have identified as being significant for the measurement—Participation, Respect, Argumentation and Justification, Persuasiveness—we will first briefly introduce the assumed causal link to deliberation before the applied method and some examples are presented.

4.1 Role and structure of participation

One of the basic characteristics of deliberative communication is equality in participation. Within a deliberative discourse, each proponent should be treated equally, i.e. equality exhibits deliberation if all stakeholders are heard. Conversely, if some stakeholders manage to achieve conversational hegemony, this indicates inequality in participation and, consequently, no deliberative communication (e.g. Habermas, 1984; Steenbergen *et al.*, 2003; Edwards *et al.*, 2008).

A simple way of assessing participation is to calculate the share of each individual in a multiparty conversation. The amount of turns and the turn lengths can be measured—for a high deliberative quality, these should be equally distributed among the participants. Beyond numbers indicating an equal or unequal participation with respect to a whole conversation, it is further interesting to

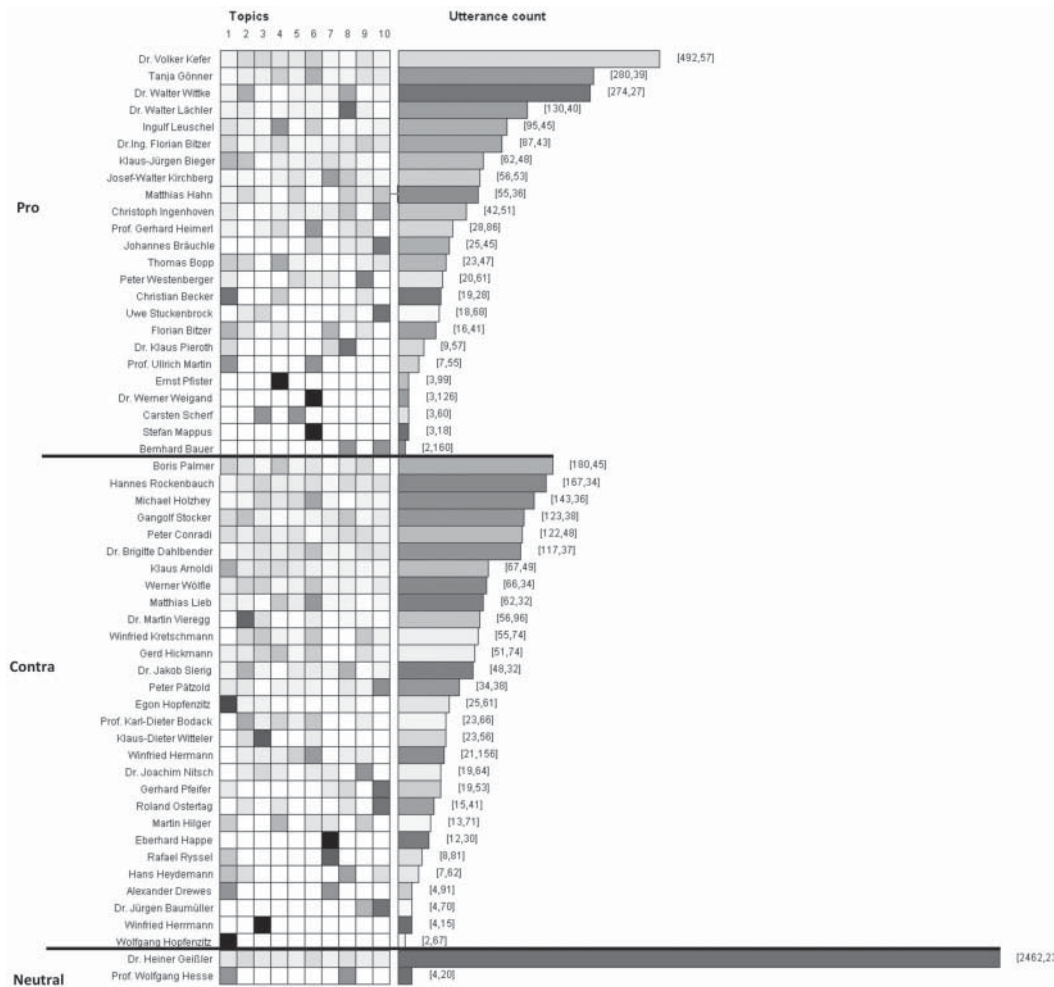


Fig. 1 Visualization demonstrating the topic distribution and basic statistics for each participant in the S21 arbitration. The saturation in the left matrix indicates the relative frequency of the topics as automatically learned using Mallet (McCallum, 2002). The bar chart at the right side of the figure indicates the amount of turns (length of the bar) as well as the average turn length from short (blue) to long (red). The figure is sorted for the participants' position toward the S21 project and additionally for the amount of turns.

inspect the course of the conversation more closely. Do individuals only participate in certain phases of the conversation? How does the turn taking evolve? Do certain persons tend to respond to certain others? Are there sections with dialog structure within a multilog, i.e. a dialog with more than two participants? This is something that cannot be easily grasped by merely computing numbers, and it can therefore profit from the strengths of visualization. In some cases, visualizations may also reveal

unexpected or unknown conversational patterns at a glance.

The topical structure of a discourse is also relevant when investigating participation. It may be the case that the participants of a multilog have an equal share in terms of turn numbers and speaking time, but that their contributions are distributed over quite different topics. We are also interested in cases where each participant perhaps tries to push their own topics of interest and does not

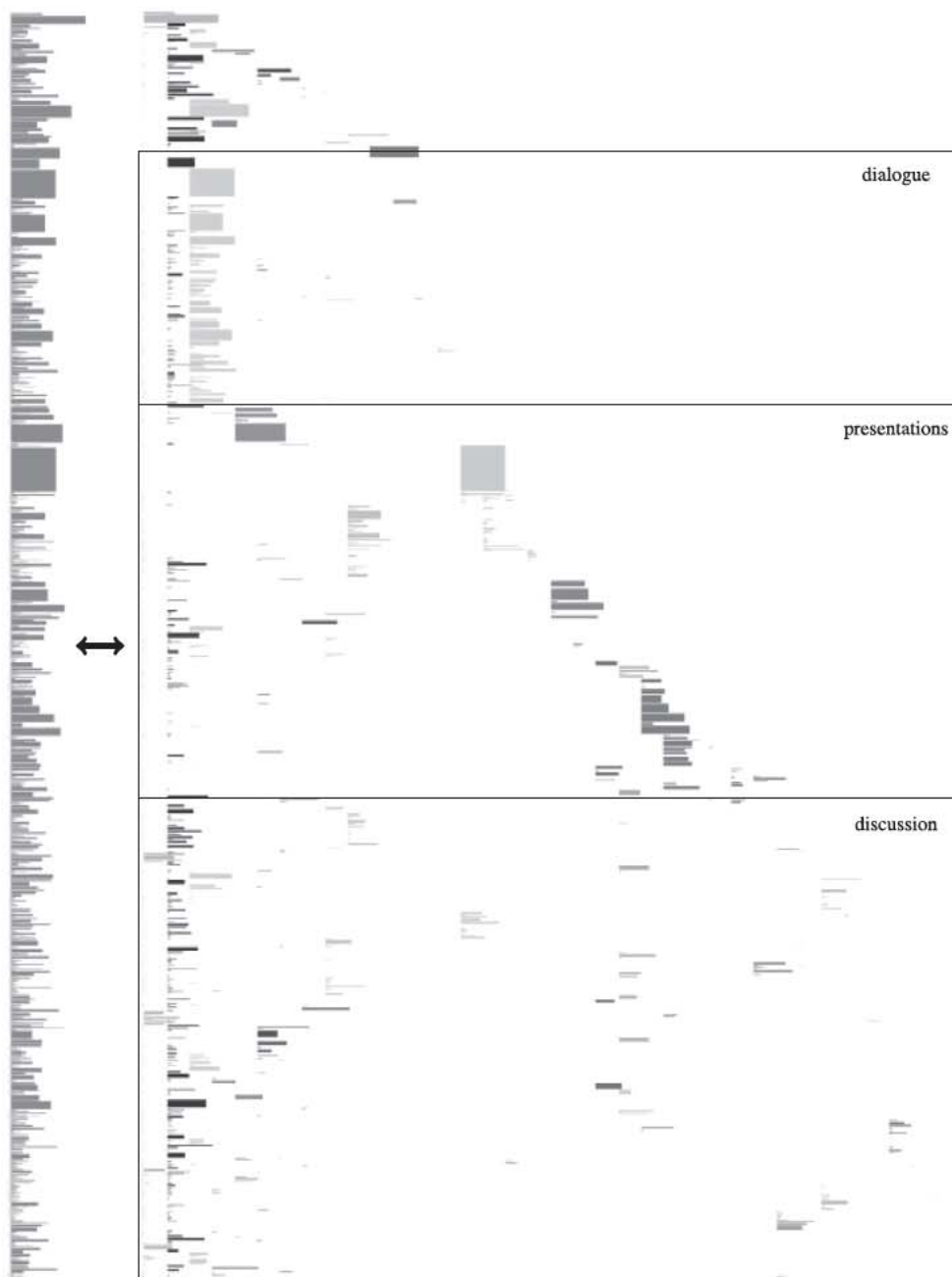


Fig. 2 Speaker participation in 1 day of the S21 arbitration. On the left side, each turn is represented as a gray text block. The height of the text block corresponds to the number of sentences contained in the turn, and the width corresponds to the length of the longest sentence of the turn. The left part of the figure shows that there are sections with larger blocks which typically indicate presentations (long turn). There are also sections of short turns indicating lively discussions or crosstalk. The right part of the figure shows the same situation with different colorings and different indentations for the speakers, i.e. one column and color represent one specific speaker. Besides granting an informative structural overview, different patterns can be readily spotted, e.g. segments of dialog, presentations, and discussion.

respond to topics raised by others. Instances of such a development in a multilog indicate a lower deliberative quality, as all topics should be treated equally.

How the topical preferences and the participation of different speakers go together is something that again is best to be analyzed using visualizations. For example, visualizations can show who participated to which extent in the elaboration of which topic. This may be analyzed aggregating proportions over the whole discourse. Not only topic proportions of individuals but also topic proportions of opposing camps are of interest, for example the sets of participants in favor and against the construction of the S21 train station. This can be achieved generating views with a matrix or table structure. For instance, in Fig. 1, the topic distribution as well as some basic statistics are shown. The visualization allows to determine—within each row—which topics the participants contributed the most to and—within each column—whether there are topics that are discussed by many participants or whether there are specific topics that are only mentioned by single participants. In particular, the participants that are on top of each category contribute mostly to all topics, the participants on the bottom only to single topics. In combination with the colored bar chart at the right side of the figure, our approach gives an overview over the amount of thematic equality in participation.

Going beyond aggregated views, representing the participation structure together with the topical structure over the course of a multiparty conversation is an even more interesting challenge for visualization research. In order to address this challenge, we have come up with several options.

First, Fig. 2 introduces a novel visualization, showing how turn lengths and speaker participation develop over the course of 1 day of the S21 arbitration. The rationale for this figure is to demonstrate patterns over the course of a discourse and to identify various types of communication. For instance, in Fig. 2, a segment of intense dialog between the arbitrator Heiner Geißler and the representative of the German railway company, Volker Kefer, can be identified. In the subsequent segment, various

blocks can be seen indicating long monologs (i.e. presentations) by external experts. Finally, after the presentations were given, the floor was opened for discussion.

Second, we automatically identify sections of the discourse where one topic is highly dominant and then label these sections in order to represent them visually. The label of a topic section contains up to five words, which correspond to the most frequent words belonging to the topic within the given section. Fig. 3 shows the statistically most significant topic sections of the complete S21 discourse. In a next step, the views of Figs 2 and 3 can be integrated in order to show how topics and speakers are connected (cf. Gold *et al.*, 2015). Fig. 4 provides an example. Again, the strategy is to let the computer automatically detect, structure, and display characteristics of the discourse in order to support the analysis process of the researcher.

4.2 Respect

Mutual respect in terms of reciprocity is seen as a prerequisite of deliberative communication (e.g. Gutmann and Thompson, 1996; Fishkin and Luskin, 2005; Gastil and Black, 2008). Reciprocity requires both speakers and listeners to treat one another with respect and equal concern—no matter how intensive or emotional the debate is. This includes listening to and respecting each other's arguments even though they may be inconsistent with one's own beliefs and interests. A number of linguistic markers can indicate (dis)respect and/or (im)politeness. The challenge lies in being able to identify these consistently via automatic means. For example, rhetorical questions containing focus particles such as 'even' as in 'Have you ever even done a real cost calculation?' signal that speaker seriously doubts the overall competence of the addressee in a manner that is disrespectful to the addressee (e.g. Guerzoni, 2004). However, it appears that at least with respect to Twitter data, it is nontrivial to extract this kind of information reliably by automatic means (Zymła, 2014). At this stage of our work, we have thus decided to first focus on easily detectable features such as patterns of interruptions. However, it is challenging to differentiate between (disrespectful) interruptions and



Fig. 3 Result of an automated analysis of the entire S21 arbitration. Topics are trained inserting the set of all turns into the standard topic modeling provided by Mallet (McCallum, 2002). After that, our algorithm identifies sections of the mediation where individual topics cluster heavily. In another step, the most frequent words used within a topic cluster are extracted and provided as labels to the left of the blocks representing topic clusters. As can be seen, at the beginning of the arbitration, there are fewer significant topic clusters, mostly related to the capacity of S21, environmental and security issues. Toward the end of the arbitration, a longer discussion with clearer topical focus evolved. The high costs of the project, which are the main issue, become more prominent, indicated by terms like *Milliarden* ‘billions’, *Kostenkalkulation* ‘cost accounting’, *Euro*, *Risikopuffer* ‘risk buffer’, *Millionen* ‘millions’, *Preissteigerung* ‘price hike’, *Preise* ‘prices’, *Finanzierung* ‘funding’, *Geld* ‘money’, or *Kosten* ‘costs’.

regular (deliberative) crosstalk. We have found that some of the relevant features for the identification of these different types are the length of utterances, the distribution of utterances, and the degree of recurrence (i.e. the degree of similarity between the

utterance and the previous utterances, see Angus *et al.* (2012a,b)).

To determine the effects of interruptions, we have developed a visual framework that mainly works with the length of utterances. Based on

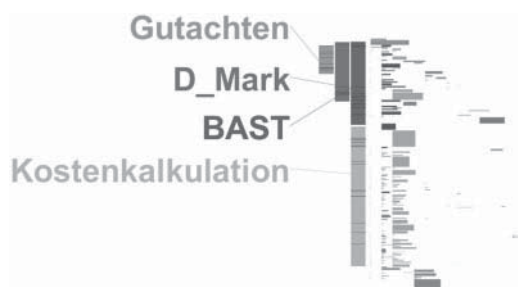


Fig. 4 More detailed view on highlighted part from Fig. 2. Both the turn structure (to the right) and the content structure (to the left) have been integrated into one view. In contrast to Fig. 3, the algorithm in this case has not searched for sections with topic clusters, but for sections with word clusters, which is another option of our method. During the dialog of Heiner Geißler (dark blue) and Volker Kefer (light green), the word *Kostenkalkulation* ‘cost calculation’ clusters highly significantly indicating that this is the main subject of their dialog section.

these lengths and the different colors for speakers, sections with interruptions can be identified. For instance, in Fig. 5, the lime green speaker Ms Gönner, who is in favor of the project interrupts her opponent, the orange speaker Mr Holzhey—even though the arbitrator Heiner Geißler has given Mr Holzhey the floor. As can be seen in the right panel, Mr Holzhey is irritated by this behavior and seeks to regain his turn: ‘Moment Moment!’ (‘Wait a minute wait a minute!’); ‘Ganz ruhig ganz ruhig!’ (‘Be calm be calm!’). Moreover, the green and gray participants also rise to speak out of turn. In the former case, Ms Gönner is interrupted by the green participant who demands an answer to his question (*...ich hätte darauf gerne eine Antwort* ‘I would like an answer to that’). However, Ms Gönner does not give up her turn and simply continues with her argument. In our interpretation, this is a show of speaker strength; however, the fact that an interruption was attempted is valued as a mark of low deliberative quality.

4.3 Argumentation and justification

At its core, deliberation requires sustained and appropriate modes of argumentation (e.g. Stromer-

Galley, 2007; Gastil and Black, 2008; Thompson, 2008). For one, arguments should be properly justified. For another, arguments should make reference to a common set of principles—indeed, it is more likely that an argument be successful if the speaker can appeal to a commonly agreed upon set of values or a commonly agreed upon understanding of the world (Habermas, 1984). Two of the linguistic features relevant for the determination of these aspects of deliberation are presented in this section: (1) causal connectors that support the position of the speaker, as opinions are not only stated, but are justified as well; and (2) discourse particles which provide information about speaker stance/attitude and/or which trigger conventional implicatures as to what common knowledge about the world should be assumed (common ground) and to which degree.

Causal connectors in German can be divided into two classes: ‘markers of reason’ introduce the cause of an effect, while ‘markers of conclusion/result’ introduce a clause describing the effect of previously stated cause. Both markers relate two parts of a sentence or several sentences: one part contains the reason for a specific statement and the other part contains the result. The following two sentences demonstrate this relationship, where (1) states a result followed by a reason (‘weil’), and (2) states a reason, followed by a result (‘daher’) (for previous computational work on these, see e.g. Dipper and Stede 2006; Schneider and Stede, 2012).

(1)
 er ist grün, weil ihm schlecht ist
 he.nom be.3.Sg green because he.dat feel.sick be.3.Sg
 ‘He is green, because he feels sick’.

(2)
 ihm ist schlecht daher ist er grün
 he.dat be.3.Sg feel.sick that.is.why be.3.Sg he.nom green
 ‘He is feeling sick, that’s why he is green’.

There are several challenges in the automatic analysis of these relations. As already mentioned in Section 3.2, some of the connectors are ambiguous and need to be disambiguated via the application of a rule-based system containing deep linguistic knowledge. A second challenge is the determination of scope: the reason/result relation can scope over



Fig. 5 Visual framework to identify sections of interruptions. Each colored row in the three panels belongs to one participant. The left panel provides an overview of the complete S21 arbitration. In order to allow a more refined overview, the visual framework provides zooming functionality. This is demonstrated in the middle and right panel.

several discourse units (EDUs) or even sentences and are thus not limited to the EDU containing the causal connector. In order to determine the scope (and to annotate the EDUs accordingly), further deep linguistic knowledge about the cues that delimit or license the relation is needed. An example of the type of algorithm used in our rule-based system for the automatic annotation of causal relations is given in (3).

(3)

IF *result connector* not in first EDU of sentence AND
result connector not preceded by other connector
 within same sentence THEN
 mark every EDU from sentence beginning to current
 EDU with **reason**.
 ELSIF *result connector* in first EDU of sentence THEN
 mark every EDU in previous sentence with **reason**
 UNLESS
 encountering another connector.

Starting from a (disambiguated) causal connector encoded in the text, rules of the type in (3) are used to annotate the preceding and following discourse units to indicate the speaker's use of justification. An evaluation of our rule-based system with respect to a manually annotated gold standard has yielded precision, recall, and f-score values of 0.84 (Bögel *et al.*, 2014). An error analysis showed that the system can be improved further in future work. However, the present results are already of a high enough quality so that we can include this fundamentally important feature as part of the measurement of deliberative quality.

Another relevant feature is the expression of common ground. In German, one of the very frequently encountered strategies for expressing whether a speaker considers information to be in the common ground (or whether they would like to have it be assumed as being in the common ground) is the use of discourse particles. German has an inventory of several different discourse particles, many of which are currently the subject of active linguistic research (see Zimmermann, 2011 for a recent overview). For example, by using the modal particle 'ja', the speaker indicates that they assume that a given statement/proposition is already known to the addressee or is general knowledge; i.e.

speaker and addressee share a common ground, and the speaker expects that the addressee will not contradict the statement (Karajosova, 2004; Zimmermann, 2011). An example is given in (4).

(4)

First brother to second brother:

Morgen	wird	Mama	ja	siebzig
tomorrow	be.3.Sg	mum	indeed	seventy
‘Tomorrow mum turns 70 (as you know)’.				

Rhetorically, this strategy can be used to put the addressee at a disadvantage—if the addressee does not want to acknowledge information as being commonly agreed upon knowledge (common ground), then they have to explicitly reject it, something that is difficult to do since the speaker conveyed their assumption only indirectly via a conventional implicature (Potts, 2005) in the first place.

A slightly different pragmatic import about the mutual common ground that is assumed is conveyed by *doch* ‘indeed’. In this case, the speaker assumes that the knowledge conveyed in utterance is already in principle also known by the addressee, but that it is not at present activated in the common ground. The use of ‘*doch*’ is thus a signal that the speaker wishes to reactivate information that is assumed to already be in the common ground. Other particles like *wohl* ‘apparently’ signal speaker attitude toward a given proposition—the use of ‘*wohl*’ conveys a weak commitment to the proposition uttered. In contrast, as a discourse particle *halt* ‘stop/well’ is used to indicate that the speaker considers the topic talked about to contain an immutable (world) constraint and also to express a certain degree of resignation in the face of how the world is (and cannot be changed). A study of these discourse particles showed that ‘*halt*’, ‘*doch*’, and ‘*ja*’ occur frequently in the S21 arbitration, whereas ‘*wohl*’ occurs only rarely Janka (2014). The particles can be used in interaction with one another and also in interaction with causal connectors. An example from the S21 arbitration is shown in (5).

This example also illustrates that while causal connectors and modal particles each separately already serve as indicators for the determination of deliberation, their interaction is also significant. Thus, a justification that also includes a particle

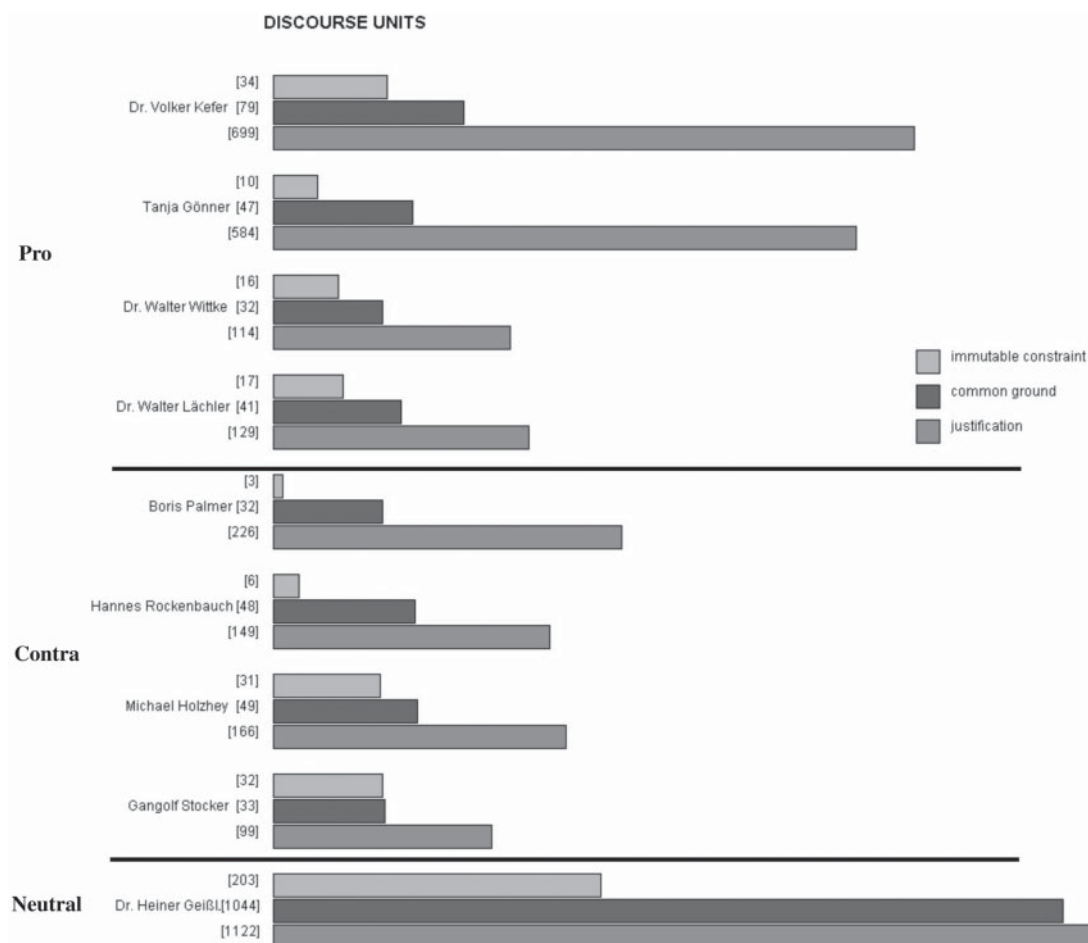


Fig. 6 Use of justification, immutable constraints, and common ground assumptions by some of the speakers of the S21 mediation process, normalized according to the number of words each speaker uttered during the process. The number indicates the absolute value of the discourse units.

representing an immutable constraint ('halt' in (5)) indicates that the speaker considers this justification to be irrevocable; i.e. the speaker has made a point that they are not expecting to be contradicted. In (5), the speaker (Heiner Geissler) states that most cars are present in a certain area. By using 'halt' in this context, he conveys that this point is absolutely true and does not need to be discussed any further.

(5)
 ... weil halt da Die meisten Autos unterwegs sind
 ... as HALT there Art most car.Pl underway be.3.Pl
 '... because most cars are underway in this area'.
 (Heiner Geissler, S21, 4 November 2010)

Note that like the causal connectors, discourse particles also tend to be highly ambiguous; e.g. 'halt' also means 'stop' and 'ja' is also the word for 'yes'. In order to achieve a successful identification of the discourse particles, a deep linguistic analysis is again necessary.

Fig. 6 shows a visual analysis based on the linguistic annotation with respect to causal connectors, discourse particles, and their interaction. The pragmatic import of these linguistic features is registered as 'justification', 'common ground', and 'immutable constraint'. The figure shows which speakers justify their arguments with which frequency and whether

they use discourse particles to convey that they consider certain information to be part of commonly agreed on knowledge (common ground) or to convey that they consider certain aspects to be hard, unchangeable facts (immutable constraints) about the world that cannot be discussed further and that thus make a solid point.

The speakers depicted in Fig. 6 are among the ones which spoke the most during the S21 arbitration. We have represented four speakers of the pro group and four speakers of the contra group. The bottom part of the figure shows an analysis of the arbitrator, Heiner Geissler. The visual analysis very clearly shows that Heiner Geissler makes the most use of common ground particles. A possible interpretation of the data is that Geissler's overall goal was to create a common ground for the two opposite groups—an attempt that is expected from a neutral arbitrator. He also brought in the most justifications and pointed out immutable facts about the world more than others. Again, these are strategies that are expected from an arbitrator who is trying to reach a consensus on the arguments that are exchanged.

Looking at the speakers in the pro versus contra groups, the visual analysis shows that the two top representatives of the pro group (Kefer and Gönner) use significantly more justification patterns than the other speakers. As the S21 mediation process was the result of an offensive against the pro S21 group, we can speculate that these representatives needed to justify their positions and decisions more during the arbitration.

4.4 Persuasiveness

Deliberation is a process whose aim is to exchange arguments and to find a common strategy. However, the process of political deliberation does not necessarily result in an agreement. From a theoretical perspective, deliberation has taken place if all the stakeholders have expressed their intention of coming to an agreement (even if none is reached) (e.g. Gastil, 2006; Mannarini and Talò, 2013). However, due to real-world pressures and the necessity that the problem at stake needs to be resolved, most deliberations do end in an agreement. Hence, the deliberative quality of a discourse can also be

measured in terms of the degree of persuasiveness, i.e. who convinced whom and how/why.

With regard to our experimental simulation-gaming experiments, we can evaluate information about persuasiveness since the experimental subjects had to note down their preferences after the discussion. For real-world conversations, analyzing who convinced whom is a more complex task since most agreements are based on a compromise between the contesting parties. This renders an analysis of the overall degree of persuasiveness difficult. Hence, we propose a procedural measure for persuasiveness.

Based on Holzinger and Landwehr (Holzinger, 2001, 2004; Landwehr and Holzinger, 2010), we propose to measure the deliberative intentions of the stakeholders based on the types of speech acts expressed by performative verbs (e.g. 'accept', 'threaten') and the information conveyed by epistemic or attitude verbs (e.g. 'believe, think, assume' versus 'know') about speaker stance. The idea is that this approach will reveal sequences within the discourse that are characterized by either extensive bargaining or intensive argumentation. Moreover, if these sequences are linked to specific topics, it will be possible to identify the argumentative quality of specific topics and to discern instances of persuasion within the discourse.

5 Summary and Future Work

This article presents work from an interdisciplinary research effort involving political science, linguistics, and visual analytics. The overall goal of our research is to find reliable indicators for the deliberative quality of a discussion. Our strategy is to identify linguistic markers that pertain to a political science-oriented analysis of deliberation and that can be identified automatically via computational linguistic methods. This computational linguistic component is rule based and draws on deep linguistic knowledge. Its outputs are automatically annotated corpora with relevant linguistic information. These corpora are used as the basis for the visual analytics component, which incorporates shallow NLP methods and other sophisticated statistical analyses of various features of the discussions. We provide

examples of visualizations with respect to the S21 arbitration process and demonstrate that our methods yield information that can ultimately be used to judge the deliberative quality of a discussion via the visual integration of very different types of information.

In the future of this research project, several steps are necessary in order to automate and refine the measures for deliberative communication. First, more features for quantifying the deliberative degree of communication need to be extracted and evaluated. For instance, similar to the deep linguistic analysis of Argumentation and Justification, we intent to apply some automated procedures also to reveal patterns of persuasiveness. Second, to achieve a single automated measure for the degree of deliberation, a combination of the four deliberative dimensions is required. An evaluation will be conducted to determine the validity of the automated measure. Overall, the combination of automated measures and visual analytics proves to not only be conducive to measuring the deliberative quality of communication but also to understanding the relevant features leading to deliberative decision-making.

References

- Angus, D., Smith, A. E., and Wiles, J. (2012a). Conceptual recurrence plots: Revealing patterns in human discourse. *IEEE Transactions on Visualization and Computer Graphics*, 18(6): 988–97.
- Angus, D., Smith, A. E., and Wiles, J. (2012b). Human communication as coupled time series: Quantifying multi participant recurrence. *IEEE Transactions on Audio, Speech & Language Processing* 20(6): 1795–807.
- Asher, N. and Reese, B. (2005). Negative Bias in Polar Questions. In Maier, E. Bary, C., and Huitink, J. (eds.), *Proceedings of Sinn and Bedeutung (SuB) 9*, Nijmegen: Nijmegen Centre of Semantics (NCS), pp. 30–43.
- Black, L. W., Burkhalter, S., Gastil, J., and Stromer Galley, J. (2010). Chapter 17: Methods for analyzing and measuring group deliberation. In Bucy, E. P. and Holbert, R. L. (eds), *Sourcebook of Political Communication Research: Methods, Measures, and Analytic Techniques*. New York: Routledge, pp. 323–45.
- Blessing, A., Sonntag, J., Kliche, F., Heid, U., Kuhn, J., and Stede, M. (2013). *Towards a Tool for Interactive Concept Building for Large Scale Analysis in the Humanities. Proceedings of the 7th Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities. Association for Computational Linguistics (ACL)*, Sofia, Bulgaria, pp. 55–64.
- Bögel, T., Hautli Janisz, A., Sulger, S., and Butt, M. (2014). *Automatic Detection of Causal Relations In German Multitlogs, Proceedings of the EACL 2014 Workshop on Computational Approaches to Causality in Language (CAtoCL)*, Gothenburg, Sweden, pp. 20–7.
- Brown, P. and Levinson, S. C. (1987). *Politeness: Some Universals in Language Use*. Cambridge, MA: Cambridge University Press.
- Chambers, S. (2003). Deliberative democracy theory. *Annual Review of Political Science*, 6(1): 307–26.
- Dacombe, R. (2013). Thinking about the quality of deliberative politics: A critical look at the discourse quality index. *Paper presented at the SSPP Annual Research Conference*, King’s College London, June 14, 2013.
- Dipper, S. (2003). *Implementing and Documenting Large Scale Grammars German LFG*. Ph.D. thesis, IMS, University of Stuttgart.
- Dipper, S. and Stede, M. (2006). Disambiguating Potential Connectives. In Butt, M. (ed.), *Proceedings of KONVENS (Conference on Natural Language Processing)*, Konstanz, pp. 167–73.
- Edwards, P. B., Hindmarsh, R., Mercer, H., Bond, M., and Rowland, A. (2008). A three stage evaluation of a deliberative event on climate change and transformation energy. *Journal of Public Deliberation*, 4(1): article 6.
- Fishkin, J. S. and Luskin, R. C. (2005). Experimenting with a democratic ideal: Deliberative polling and public opinion. *Acta Politica*, 40(3): 284–98.
- Gastil, J. (2006). How balanced discussion shapes knowledge, public perceptions, and attitudes: A case study of deliberation on the los alamos national laboratory. *Journal of Public Deliberation*, 2(1): article 4.
- Gastil, J. and Black, L. W. (2008). Public deliberation as the organizing principle of political communication research. *Journal of Public Deliberation*, 4(1): 1–47.
- Gold, V., Rohrdantz, C., and El Assady, M. (2015). Exploratory text analysis using lexical episode plots. In Bertini, E., Kennedy, J., and Puppo, E., (eds.), *Eurographics Conference on Visualization (EuroVis) Short Papers*. The Eurographics Association. 2015.

- Guerzoni, E.** (2004). EVEN NPIs in yes/no questions. *Natural Language Semantics*, 12: 319–43.
- Gutmann, A. and Thompson, D. F.** (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, J.** (1981). *Theorie des Kommunikativen Handelns*. Frankfurt am Main: Suhrkamp.
- Habermas, J.** (1984). *The Theory of Communicative Action*. Boston, MA: Beacon.
- Hangartner, D., Bächtiger, A., Grünenfelder, R., and Steenbergen, M. R.** (2007). Mixing Habermas with Bayes: Methodological and Theoretical advances in the study of deliberation. *Swiss Political Science Review*, 13(4): 607–44.
- Holzinger, K.** (2001). Verhandeln statt Argumentieren oder Verhandeln durch Argumentieren? Eine empirische Analyse auf der Basis der Sprechaktheorie. *Politische Vierteljahresschrift*, 42(3): 414–46.
- Holzinger, K.** (2004). Bargaining through arguing: An empirical analysis based on speech act theory. *Political Communication*, 21(2): 195–222.
- Janka, M.** (2014). *Schattierungen in der Argumentation Modalpartikeln und kausale Konnektoren*. BA thesis, University of Konstanz.
- Karajosova, E.** (2004). *The Meaning and Function of German Modal Particles*. Saarbrücken Dissertations in Computational Linguistics and Language Technology.
- Kaifer, D., Geisler, C., Vlachos, P., and Ishizaki, S.** (2006). Chapter 9: Mining textual knowledge for writing education and research: The DocuScope project. In van Waes, L., Leijten, M., and Neuwirth, C. M. (eds), *Writing and Digital Media. Language and Linguistic Special*. Oxford: Elsevier, pp. 115–29.
- Keim, D. A., Mansmann, F., Schneidewind, J., Thomas, J., and Ziegler, H.** (2008). Chapter: Visual analytics: Scope and challenges. In Simoff, S. J., Bohlen, M. H., and Mazeika, A. (eds), *Visual Data Mining*. Berlin, Heidelberg: Springer Verlag, pp. 76–90.
- Lakoff, R.** (1975). *Language and Woman's Place*. New York, NY: Harper & Row.
- Landwehr, C. and Holzinger, K.** (2010). Institutional determinants of deliberative interaction. *European Political Science Review*, 2(3): 373–400.
- Lord, C. and Tamvaki, D.** (2013). The politics of justification? Applying the ‘discourse quality index’ to the study of the European parliament. *European Political Science Review*, 5: 27–54.
- Mannarini, T. and Talò, C.** (2013). Evaluating public participation: Instruments and implications for citizen involvement. *Community Development*, 44(2): 239–56.
- Marcu, D.** (2000). *The Theory and Practice of Discourse Parsing and Summarization*. Cambridge, MA: MIT Press.
- Mayer, T., Rohrdantz, C., Butt, M., Plank, F., and Keim, D. A.** (2010). Visualizing vowel harmony. *Linguistic Issues in Language Technology*, 4(2): 1–33.
- McCallum, A. K.** (2002). *Mallet: A Machine Learning for Language Toolkit*. <http://mallet.cs.umass.edu>
- Moretti, F.** (2013). *Distant Reading*. London: Verso.
- Nguyen, V. A., Hu, Y., Boyd Graber, J., and Resnik, P.** (2013). Argviz: Interactive visualization of topic dynamics in multi party conversations. *Human Language Technologies: The 2013 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, 10: 36.
- Persson, M., Esaiasson, P., and Gilljam, M.** (2013). The effects of direct voting and deliberation on legitimacy beliefs: An experimental study of small group decision making. *European Political Science Review*, 5(3): 381–99.
- Potts, C.** (2005). *The Logical of Conventional Implicatures*. Oxford: Oxford University Press.
- Ramsay, S.** (2003). Special section: Reconceiving text analysis toward an algorithmic criticism. *Literary and Linguistic Computing* 18(2): 167–74.
- Ramsay, S.** (2011). *Reading Machines: Toward an Algorithmic Criticism*. Urbana: University of Illinois Press.
- Rohrdantz, C., Hautli, A., Mayer, T., Butt, M., Plank, F., and Keim, D. A.** (2011). *Towards tracking semantic change by visual analytics, Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (Short Papers)*. Portland, OR: Association for Computational Linguistics, pp. 305–10.
- Rohrdantz, C., Hund, M., Mayer, T., Wlchli, B., and Keim, D. A.** (2012). The world's languages explorer: Visual analysis of language features in genealogical and areal contexts. *Computer Graphics Forum*, 31(3): 935–44.
- Scheffler, T.** (2014). Meaning Variations in German Tag Questions. Talk accepted at the DGfS (Deutsche Gesellschaft für Sprachwissenschaft) 2015 Workshop on “The prosody and meaning of (non)canonical questions across languages”.

- Schiller, A.** (1994) *DMOR User's Guide*. Technical Report, Universitat Stuttgart, Institut für maschinelle Sprachverarbeitung.
- Schmid, H.** (1995). *Improvements in Part of Speech Tagging with an Application to German, Proceedings of the ACL SIGDAT Workshop*, Dublin, Ireland.
- Schneider, A. and Stede, M.** (2012). Ambiguity in German Connectives: A Corpus Study. In Butt, M. (ed.), *Proceedings of KONVENS (Conference on Natural Language Processing)*. Konstanz, pp. 254–8.
- Steenbergen, M. R., Bächtiger, A., Spörndli, M., and Steiner, J.** (2003). Measuring political deliberation: A discourse quality index. *Comparative European Politics*, 1(1): 21–48.
- Stromer Galley, J.** (2007). Measuring deliberation's content: A coding scheme. *Journal of Public Deliberation*, 3(1): article 12.
- Sulkin, T. and Simon, A. F.** (2001). Habermas in the lab: A study of deliberation in an experimental setting. *Political Psychology*, 22(4): 809–26.
- Thomas, J. J. and Cook, K. A.** (2006). A visual analytics agenda. *IEEE Computer Graphics and Applications*, 26(1): 10–13.
- Thompson, D. F.** (2008). Deliberative democratic theory and empirical political science. *Annual Review of Political Science*, 11(1): 497–520.
- Wiedemann, G., Lemke, M., and Niekler, A.** (2013). Postdemokratie und Neoliberalismus – Zur Nutzung neoliberaler Argumentationen in der Bundesrepublik Deutschland 1949–2011. Ein Werkstattbericht. *Zeitschrift für Politische Theorie*, 4(1): 99–115.
- Zimmermann, M.** (2011). Discourse particles. In Portner, P., Maienborn, C., and von Stechow, K. (eds), *Semantics (Handbücher zur Sprach und Kommunikationswissenschaft)*. Mouton de Gruyter, pp. 2011–38.
- Zymła, M. M.** (2014). *Extraction and Analysis of Non Canonical Questions from a Twitter Corpus*. MA thesis, University of Konstanz.

Notes

- 1 <http://www.social-media-analytics.org/en/>, last accessed 2 October 2014.
- 2 <https://www.ling.uni-potsdam.de/acl-lab/Eidentity/main.html>, last accessed 2 October 2014.
- 3 <http://www.epol-projekt.de>, last accessed 2 October 2014.
- 4 <http://www.schlichtung-s21.de/dokumente.html>, last accessed 25 September 2014.
- 5 <http://www.cis.uni-muenchen.de/schmid/tools/TreeTagger/>, last accessed 2 October 2014.