

Interactive Visual Analysis of Transcribed Multi-Party Discourse

Mennatallah El-Assady¹, Annette Hautli-Janisz¹, Valentin Gold²,
Miriam Butt¹, Katharina Holzinger¹, and Daniel Keim¹

¹University of Konstanz, Germany

²University of Göttingen, Germany

Abstract

We present the first web-based Visual Analytics framework for the analysis of multi-party discourse data using verbatim text transcripts. Our framework supports a broad range of server-based processing steps, ranging from data mining and statistical analysis to deep linguistic parsing of English and German. On the client-side, browser-based Visual Analytics components enable multiple perspectives on the analyzed data. These interactive visualizations allow exploratory content analysis, argumentation pattern review and speaker interaction modeling.

1 Introduction

With the increasing availability of large amounts of multi-party discourse data, the breadth and complexity of questions that can be answered with natural language processing (NLP) is expanding. Discourses can be analyzed with respect to what topics are discussed, who contributes to which topic to what extent, how the turn-taking plays out, how speakers convey their opinions and arguments, what Common Ground is assumed and what the speaker stance is. The challenge presented for NLP lies in the automatic identification of relevant cues and in providing assistance towards the analysis of these primarily pragmatic features via the automatic processing of large amounts of discourse data. The challenge is exacerbated by the fact that linguistic data is inherently multidimensional with complex feature interaction being the norm rather than the exception. The problem becomes particularly difficult when one moves on to compare multi-party discourse strategies across different languages.

In this paper we present a novel Visual Analytics framework that encodes various layers of discourse properties and allows for an analysis of

multi-party discourse. The system combines discourse features derived from shallow text mining with more in-depth, linguistically-motivated annotations from a discourse processing pipeline. Based on this hybrid technology, users from political science, journalism or digital humanities are able to draw inferences regarding the progress of the debate, speaker behavior and discourse content in large amounts of data at-a-glance, while still maintaining a detailed view on the underlying data. To the best of our knowledge, our *VisArgue* system offers the first web-based, interactive Visual Analytics approach of multi-party discourse data using verbatim text transcripts.¹

2 Related work

Discourse processing A large amount of work in discourse processing focuses on analyzing discourse relations, annotated in different granularity and style in RST (Mann and Thompson, 1988) or SDRT (Asher and Lascarides, 2003). While a large amount of work is for English and based on landmark corpora such as the Penn Discourse Treebank (Prasad et al., 2008), the parsing of discourse relations in German has only lately received attention (Versley and Gastel, 2012; Stede and Neumann, 2014; Bögel et al., 2014).

Another strand of research is concerned with dialogue act annotation, to which end several annotation schemes have been proposed (Bunt et al., 2010, inter alia). Those have also been applied across a range of German corpora (Jekat et al., 1995; Zarisheva and Scheffler, 2015). Another area deals with the classification of speaker stance (Mairesse et al., 2007; Danescu-Niculescu-Mizil et al., 2013; Sridhar et al., 2015).

Despite the existing variety of previous work in discourse processing, our contribution is novel. For one, we combine different levels of analysis

¹Accessible at <http://visargue.inf.uni.kn/>. Accounts (beyond the demo) are available upon request.

and integrate information that has not been dealt with intensively in discourse processing before, for instance regarding rhetorical framing. For another, we provide an innovation with respect to the type of data the system can handle in that the system is designed to deal with noisy transcribed natural speech, a genre underresearched in the area.

Visual Analytics Visualizing the features and dynamics of communication has been gaining interest in information visualization, due to the diversity and ambiguity of this data. Erickson and Kellogg (2000) introduce a general framework for the design of such visualization systems. Other approaches attempt to model the social interactions in chat systems, e.g. Chat Circles (Donath and Viégas, 2002) and GroupMeter (Leshed and et al., 2009). Conversation Clusters (Bergstrom and Karahalios, 2009) and MultiConVis (Hoque and Carenini, 2016) group the content of conversations dynamically. Overall, the majority of these systems are designed to model the dynamics and changes in the content of conversations and do not rely on a rich set of linguistic features.

3 Computational linguistic processing

Our automatic annotation system is based on a linguistically-informed, hand-crafted set of rules that deals with the disambiguation of explicit linguistic markers and the identification of spans and relations in the text. For that, we divide all utterances into smaller units of text in order to work with a more fine-grained structure of the discourse. Although there is no consensus in the literature on what exactly these units have to comprise, it is generally assumed that each discourse unit describes a single event (Polanyi et al., 2004). Following Marcu (2000), we term these units *elementary discourse units* (EDUs).

For German, we approximate the assumption made by Polanyi et al. (2004) by inserting a boundary at every punctuation mark and every clausal connector (conjunctions, complementizers). For English we rely on clause-level splitting of the Stanford PCFG parser (Klein and Manning, 2003) and create EDUs at the SBAR, SBARQ, SINV and SQ clause level. The annotation is performed on the level of these EDUs, therefore relations that span multiple units are marked individually at each unit.

We were not able to use an off the shelf parser for German. For instance, an initial experiment using the German Stanford Dependency parser (Rafferty and Manning, 2008) showed that 60% of parses are incorrect due to interruptions, speech repairs and multiple embeddings. We therefore hand-crafted our own rules on the basis of morphological and POS information from DMOR (Schiller, 1994). For English, the data contained less noise and we were able to use the POS tags from the Stanford parser.

Levels of analysis With respect to *discourse relations*, we annotate spans as to whether they represent: reasons, conclusions, contrasts, concessions, conditions or consequences. For German, we rely on the connectors in the Potsdam Commentary Corpus (Stede and Neumann, 2014), for English we use the PDTP-style parser (Ziheng Lin and Kan, 2014).

In order to identify relevant *speech acts*, we compiled lists of speech act verbs comprising agreement, disagreement, arguing, bargaining and information giving/seeking/refusing. In order to gauge *emotion*, we use EmoLex, a crowdsourced emotion lexicon (Mohammad and Turney, 2010) available for a number of languages, plus our own curated lexicon of *politeness* markers. With re-

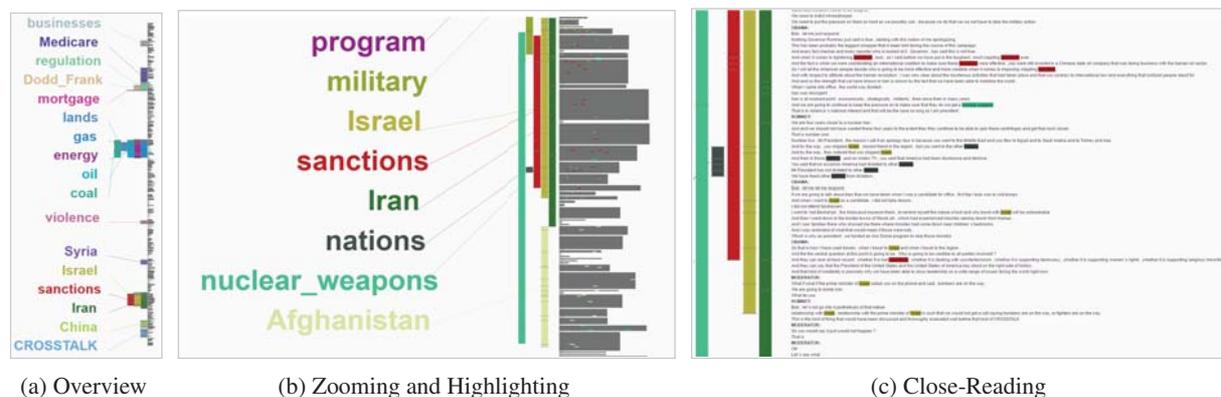


Figure 1: Lexical Episode Plots

spect to *event modality*, we take into account all modal verbs and adverbs signaling obligation, permission, volition, reluctance or alternative. Concerning *epistemic modality* and speaker stance we use modal expressions conveying certainty, probability, possibility and impossibility. Finally, we added a category called *rhetorical framing* (Hautli-Janisz and Butt, 2016), which accounts for the illocutionary contribution of German discourse particles. Here we look at different ways of invoking Common Ground, hedging and signaling accommodation in argumentation, for example.

Disambiguation Many of the crucial linguistic markers are ambiguous. We developed hand-crafted rules that take into account the surrounding context to achieve disambiguation. Important features include position in the EDU (for instance for lexemes which can be discourse connectors at the beginning of an EDU but not at the end, and vice versa) or the POS of other lexical items in the context. Overall, the German system features 20 disambiguation rules, the English one has 12.

Relation identification After disambiguation is complete, a second set of rules annotates the spans and the relations that the lexical items trigger. In this module, we again take into account the context of the lexical item. An important factor is negation, which in some cases reverses the contribution of the lexical item, e.g. in the case of ‘possible’ to ‘not possible’.

With respect to discourse connectors, for instance the German causal markers *da*, *denn*, *darum* and *daher* ‘because/thus’, we only analyze relations within a single utterance of a speaker, i.e., relations that are expressed in a sequence of

clauses which a speaker utters without interference from another speaker. As a consequence, the annotation system does not take into account relations that are split up between utterances of one speaker or utterances of different speakers. For causal relations (reason and conclusion spans), we show in Bögel et al. (2014) that the system performs with an F-score of 0.95.

4 Visual Analytics Framework

The web-based Visual Analytics framework is designed to give analysts multiple perspectives on the same datasets. The transcripts are uploaded through the web interface to undergo the previously discussed linguistic processing and other visualization-dependent processing steps. The visualizations are classified into four categories.

(1) **Basic Data Exploration Views**, which enable the user to explore the annotations and dynamically create statistical charts using all computed features. (2) **Content Analysis Views** are designed to allow the user to explore *what* is being said. (3) **Argumentation Analysis Views** rely on the linguistic parsing to address the question of *how* it is being said. (4) **Speaker Analysis Views** are focused on giving an insight into the speaker dynamics to answer the question *by whom* it is being said. In the following, we will discuss a sample of the visualization components using the transcripts of the three televised US presidential election debates from 2012 between Obama and Romney. In the visualizations, the three speakers in the debate are distinguished through their set colors and icons: Obama as 🇺🇸 **Democrat** (blue); Romney as 🇺🇸 **Republican** (red) and all moderators combined as 🗣️ **Moderator** (green).

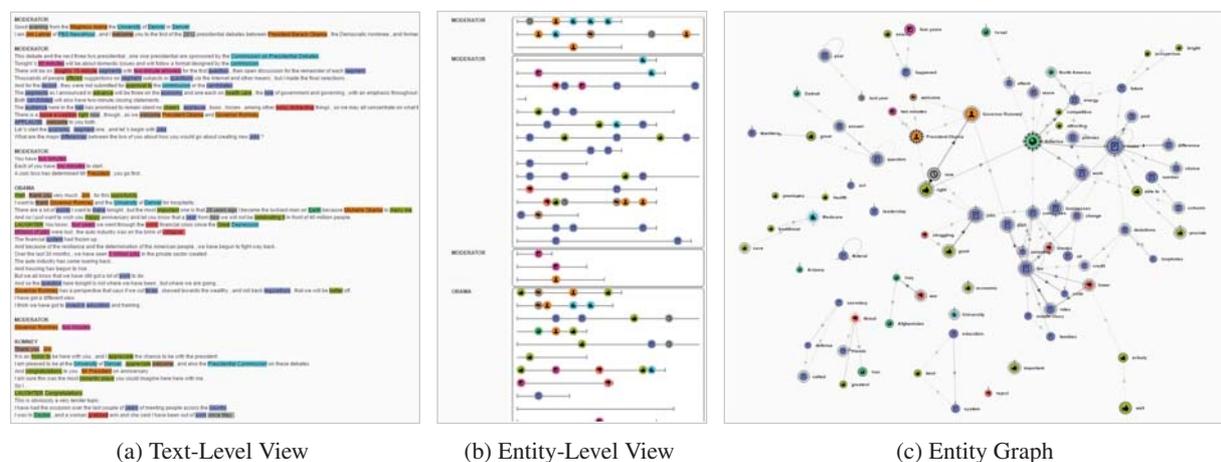


Figure 2: Named-Entity Relationship Explorer

4.1 Content Analysis Views

Lexical Episode Plots This visualization is designed to give a high-level overview of the content of the transcripts, based on the concept of lexical chaining. For this, we compute word chains that appear with a high density in a certain part of the text and determine their importance through the compactness of their appearance. Lexical Episodes (Gold et al., 2015) are defined as a portion of the word sequence where a certain word appears more densely than expected from its frequency in the whole text. These episodes are visualized as bars on the left-hand side of the text (Figure 1). The text is shown on the right and each utterance is abstracted by one box with each sentence as one line. This visualization supports a smooth uniform zooming from the text level to the high-level overview, which enables both a close-reading (Figure 1c) of the text and a distant-reading using the episodes. The user can also select episodes which are then highlighted in the text (Figure 1b). The level of detail is adjusted by changing the significance level of the episode detection. Figure 1a shows an overview of the three presidential debates, with a high significance level selected to achieve a high level of detail.

Named-Entity Relationship Explorer This visualization (El-Assady et al., 2017) enables the analysis of different concepts and their relation in the utterances. We categorize relevant named-entities and concepts from the text and abstract them into ten classes: 🧑 Persons, 🌍 Geo-Locations, 🏢 Organizations, 🕒 Date-Time, 📏 Measuring Units, 📊 Measures, 🗣️ Context-Keywords, 🟢 Positive- and 🟡 Negative-Emotion Indicators, and 🗣️ Politeness-Keywords. We then abstract the text from the Text-Level View (Figure 2a) to the Entity-Level View (Figure 2b) to allow a high-level overview of the entity distribution across utterances.



Figure 4: Data-type color mapping for glyphs.

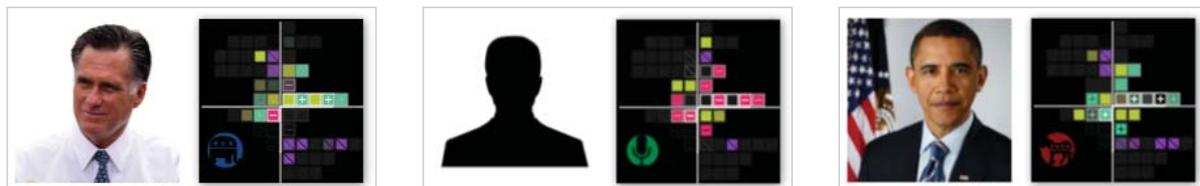


Figure 3: Speaker Profiles with Argumentation Glyphs

In order to extract their relations, we devise a tailored distance-restricted entity-relationship model to comply with the often ungrammatical structure of verbatim transcripts. This model relates two entities if they are present in the same sentence within a small distance window defined by a user-selected threshold. The concept map of the conversations, which builds up as the discourse progresses, can then be explored in the Entity Graph (Figure 2c). All views support a rich set of interactions, e.g., linking, brushing, selection, querying and interactive parameter adjustment.

4.2 Argumentation Analysis Views

Argumentation Feature Fingerprinting In an attempt to measure the deliberative quality of discourse (Gold and Holzinger, 2015), we use the annotations discussed in Section 3 and create a fingerprint of all utterances, the Argumentation Glyph. The glyph maps the four theoretic dimensions of deliberation in its four quadrants which are separated by the axes: NW (Accommodation), NE (Atmosphere & Respect), SE (Participation), SW (Argumentation & Justification). In each row, we group features that are thematically related, e.g. speech acts of information-giving/seeking/refusing. Each feature is represented as a small rectangular box. The strength of each value is encoded via a divergent color mapping, with each type of data (binary, numerical, bipolar) having a different color scale (Figure 4). The small circular icon at the bottom left shows the average length of each utterance.

This glyph-based fingerprinting of discourse features can be used to analyze sets of aggregated utterances, e.g. Figure 3 displays one glyph for every speaker representing the average of all their utterances. These speaker profiles are used for the identification of individual behavior patterns. In addition, the glyphs can be aggregated for topics, speaker parties, and combinations of these.

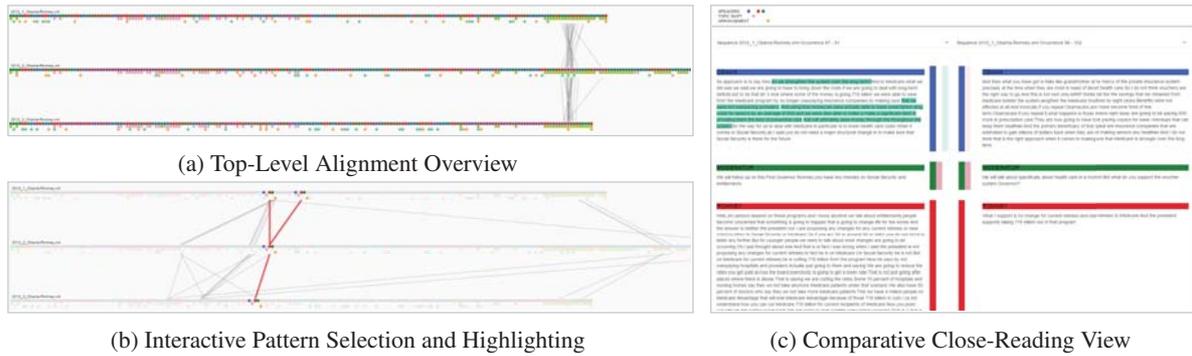


Figure 5: Argumentation Feature Alignment for Discourse Pattern Detection

Argumentation Feature Alignment The user can also form hypotheses about the occurrences of these discourse features in the data. To facilitate their verification across multiple conversations we use sequential pattern mining to create feature alignment views (Jentner et al., 2017) based on selected features. Figure 5 shows alignment views created using the following three features: **Speakers** (● Obama, ● Romney, ● Moderator); **Topic Shift** (● Progressive, ● Recurring); and **Arrangement** (● Agreement, ● Disagreement). The sidefigure shows the pattern of Obama making a statement, followed by a topic shift and a turn of Romney and the moderator, followed by an agreement. This pattern can be found across all three presidential debates, shown in Figure 5b. For further analysis, the user can switch to a comparative close-reading view to investigate two occurrences of the found pattern on the text level, as shown in Figure 5c.

4.3 Speaker Analysis Views

Topic-Space Views In this visualization, we model the interactions between speakers using the metaphor of a closed discussion floor. We designed a radial plot, the topic space, in which the speakers interact over the course of a discussion. Using this metaphor, we created a set of different (static and animated) views to highlight the various aspects of the speaker interactions. Figure 6 displays one time-frame of the utterance sedimentation view (El-Assady et al., 2016) of the accumulated presidential debates. In this animation, all discussed topics (ordered by their similarity to a selected base-topic at 12 o’clock) span the radial topic space. The length of the arch representing a topic is mapped to the size of the topic. All currently active speakers are displayed as moving

dots with motion chart trails. A gradual visual-decay function blends out non-active speakers over time. Using a sedimentation metaphor, all past utterances are pulled to their top topic by a radial gravitation.

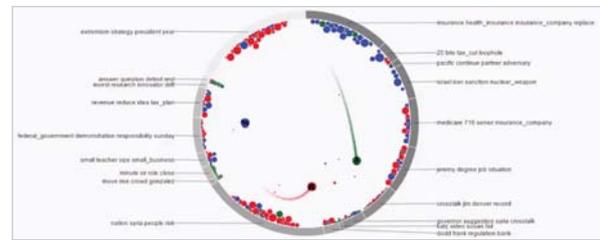


Figure 6: Topic Space View

5 Summary

The VisArgue framework provides a novel visual analytics toolbox for exploratory and confirmatory analyses of multi-party discourse data. Overall, each of the presented visualizations support disentangling speaker and discourse patterns.

Acknowledgments

The research carried out in this paper was supported by the Bundesministerium für Bildung und Forschung (BMBF) under grant no. 01461246 (eHumanities VisArgue project).

References

Nick Asher and Alex Lascarides. 2003. *Logics of Conversation*. Cambridge: Cambridge University Press.

Tony Bergstrom and Karrie Karahalios. 2009. Conversation clusters: grouping conversation topics through human-computer dialog. In *Proc. of the SIGCHI Conf. on Human Factors in Computing Systems*, pages 2349–2352.

Tina Bögel, Annette Hautli-Janisz, Sebastian Sulger, and Miriam Butt. 2014. Automatic Detection of Causal Relations in German Multilog. In *Proc. of the EACL 2014 CATCL Workshop*, pages 20–27.

- Harry Bunt, Jan Alexandersson, and Jean Carletta et al. 2010. Towards an ISO standard for dialogue act annotation. In *Proc. of LREC'10*. pages 2548–2555.
- Cristian Danescu-Niculescu-Mizil, Moritz Sudhof, Dan Jurafsky, Jure Leskovec, and Christopher Potts. 2013. A computational approach to politeness with application to social factors. In *Proc. of ACL'13*. pages 250–259.
- Judith Donath and Fernanda B Viégas. 2002. The chat circles series: explorations in designing abstract graphical communication interfaces. In *Proc. of the 4th conference on Designing interactive systems*. pages 359–369.
- Mennatallah El-Assady, Valentin Gold, Carmela Acevedo, Christopher Collins, and Daniel Keim. 2016. ConToVi: Multi-Party Conversation Exploration using Topic-Space Views. *Computer Graphics Forum* 35(3):431–440.
- Mennatallah El-Assady, Rita Sevastjanova, Bela Gipp, Daniel Keim, and Christopher Collins. 2017. NEREx: Named-Entity Relationship Exploration in Multi-Party Conversations. *Computer Graphics Forum*.
- Thomas Erickson and Wendy A Kellogg. 2000. Social translucence: an approach to designing systems that support social processes. *ACM transactions on computer-human interaction (TOCHI)* 7(1):59–83.
- Valentin Gold and Katharina Holzinger. 2015. An Automated Text-Analysis Approach to Measuring Deliberative Quality. Annual Meeting of the Midwest Political Science Association (Chicago).
- Valentin Gold, Christian Rohrdantz, and Mennatallah El-Assady. 2015. Exploratory Text Analysis using Lexical Episode Plots. In E. Bertini, J. Kennedy, and E. Puppo, editors, *Eurographics Conference on Visualization (EuroVis) - Short Papers*.
- Annette Hautli-Janisz and Miriam Butt. 2016. On the role of discourse particles for mining arguments in German dialogs. In *Proc. of the COMMA 2016 FLA workshop*. pages 10–17.
- Enamul Hoque and Giuseppe Carenini. 2016. Multi-convis: A visual text analytics system for exploring a collection of online conversations. In *Proc. of Intelligent User Interfaces*. ACM, IUI, pages 96–107.
- Susanne Jekat, Alexandra Klein, Elisabeth Maier, Ilona Maleck, Marion Mast, and J. Joachim Quantz. 1995. Dialgue acts in verbmobil. Technical report, Saarländische Universitäts- und Landesbibliothek.
- Wolfgang Jentner, Mennatallah El-Assady, Bela Gipp, and Daniel Keim. 2017. Feature Alignment for the Analysis of Verbatim Text Transcripts. *EuroVis Workshop on Visual Analytics (EuroVA)*.
- Dan Klein and Christopher D. Manning. 2003. Accurate unlexicalized parsing. In *Proc. of ACL 2003*. pages 423–430.
- Gilly Leshed and Diego Perez et al. 2009. Visualizing real-time language-based feedback on teamwork behavior in computer-mediated groups. In *Proc. of the SIGCHI Conf. on Human Factors in Computing Systems*. pages 537–546.
- Francois Mairesse, Marilyn A. Walker, Matthias R. Mehl, and Roger K. Moore. 2007. Using Linguistic Cues for the Automatic Recognition of Personality in Conversation and Text. *Journal of Artificial Intelligence Research* 30:457–500.
- William C. Mann and Sandra A. Thompson. 1988. Rhetorical structure theory: Towards a theory of text organization. *Text* 8(3):243–281.
- Daniel Marcu. 2000. *The Theory and Practice of Discourse Parsing and Summarization*. MIT Press, Cambridge, Mass.
- Saif M. Mohammad and Peter D. Turney. 2010. Emotions evoked by common words and phrases: Using mechanical turk to create and emotion lexicon. In *Proc. of the NAACL 2015 WASSA Workshop*. pages 26–34.
- Livia Polanyi, Chris Culy, Martin van den Berg, Gian Lorenzo Thione, and David Ahn. 2004. Sentential structure and discourse parsing. In *Proc. of the ACL'04 Workshop on Discourse Annotation*. pages 80–87.
- Rashmi Prasad, Nikhil Dinesh, Alan Lee, Eleni Miltasakaki, Livio Robaldo, Aravind Joshi, and Bonnie Webber. 2008. The Penn Discourse Treebank 2.0. In *Proc. of LREC'08*. pages 2961–2968.
- Anne N. Rafferty and Christopher D. Manning. 2008. Parsing three german treebanks: Lexicalized and unlexicalized baselines. In *Proc. of the ACL-08 PaGe-08 workshop*. pages 40–46.
- Anne Schiller. 1994. DMOR - User's Guide. Technical report, IMS, Universität Stuttgart.
- Dhanya Sridhar, James Foulds, Marilyn Walker, Bert Huang, and Lise Getoor. 2015. Joint models of disagreement and stance in online debate. In *Proc. of ACL 2015*. pages 26–31.
- Manfred Stede and Arne Neumann. 2014. Potsdam commentary corpus 2.0: Annotation for discourse research. In *Proc. of LREC'14*. pages 925–929.
- Yannick Versley and Anna Gastel. 2012. Linguistic Tests for Discourse Relations in the Tüba-D/Z Corpus of Written German. *Dialogue and Discourse* 1(2):1–24.
- Elina Zarisheva and Tatjana Scheffler. 2015. Dialogue act annotation for twitter data. In *Proc. of SIGDIAL 2015*. pages 114–123.
- Hwee Tou Ng Ziheng Lin and Min-Yen Kan. 2014. A PDTB-Styled End-to-End Discourse Parser. *Natural Language Engineering* 20:151–184.