

Argumentation, Ideology, and Issue Framing in Parliamentary Discourse

Graeme Hirst,* Vanessa Wei Feng,* Christopher Cochrane,[†] and Nona Naderi*

*Department of Computer Science and [†]Department of Political Science

University of Toronto, Toronto, Ontario, Canada

*{gh, weifeng, nona}@cs.toronto.edu

[†]christopher.cochrane@utoronto.ca

Abstract

In argumentative political speech, the way an issue is framed may indicate the unstated assumptions of the argument and hence the ideological position of the speaker. Our goal is to use and extend our prior work on discourse parsing and the identification of argumentation schemes to identify specific instances of issue framing and, more generally, ideological positions as they are expressed in text. We are using annotated historical and contemporary proceedings of the British, Canadian, and Dutch parliaments, looking in particular at speech on the topic of immigration.

1 Introduction

A key aspect of any argument is the unstated assumptions and beliefs that underlie it. At bottom, all naturally occurring arguments are enthymematic. Our research in argumentation has the long-term goal of identifying these unstated elements, both at the micro level — the specific unstated premises of an argument — and at the macro level — the belief system or ideology within which the entire argument is constructed, which may in turn contribute to its unstated premises (and also to any unstated conclusions).

Our past research has concerned analysis of argumentation, and the related issue of determining the rhetorical structure of discourse, at the micro level. In this paper, we briefly describe this work. We then describe our present and planned research on ideology-based argumentation, including, in particular, the identification of specific kinds of issue framing and their role in ideological disagreement.

Our research is part of the project Digging Into Linked Parliamentary Data (“Dilipad”), an interdisciplinary tri-national project that is collecting

and richly annotating historical and contemporary parliamentary proceedings of the U.K., Canada, and the Netherlands for use in studies in political science, political history, and other areas of social science and linguistics.¹ The project includes two case studies on the identification of ideology, ideological frameworks, and argumentation in the data, which we will describe below.

2 Argumentation analysis

The context for our initial research on argumentation (presented in detail by Feng and Hirst (2011)) was the early work of Mochales and Moens (2008; 2009a; 2009b), who focused on automatic detection of arguments in legal texts. With each sentence represented as a vector of shallow features, they trained a multinomial naïve Bayes classifier and a maximum entropy model on the Araucaria corpus. In their follow-up work, they trained a support vector machine to further classify each argumentative clause into a premise or a conclusion. In addition, they developed a context-free grammar for argumentation structure parsing. Our work is “downstream” from that of Mochales and Moens. Assuming the eventual success of their, or others’, research program on detecting and classifying the components of an argument, we sought to determine how the pieces fit together as an instance of an argumentation scheme. This, in turn, would be used, in future work, to understand the argument and recover the unstated assumptions. Figure 1 shows the structure of a complete posited system, with our work addressing the part inside the red dashed line.

Of Walton’s set of 65 argumentation schemes (Walton *et al.*, 2008), we focused on the five that are most frequent in the Araucaria dataset (Reed and Rowe, 2004; Rowe and Reed, 2008): ar-

¹For more details of the project, including the other participating institutions and researchers, see <http://dilipad.history.ac.uk>

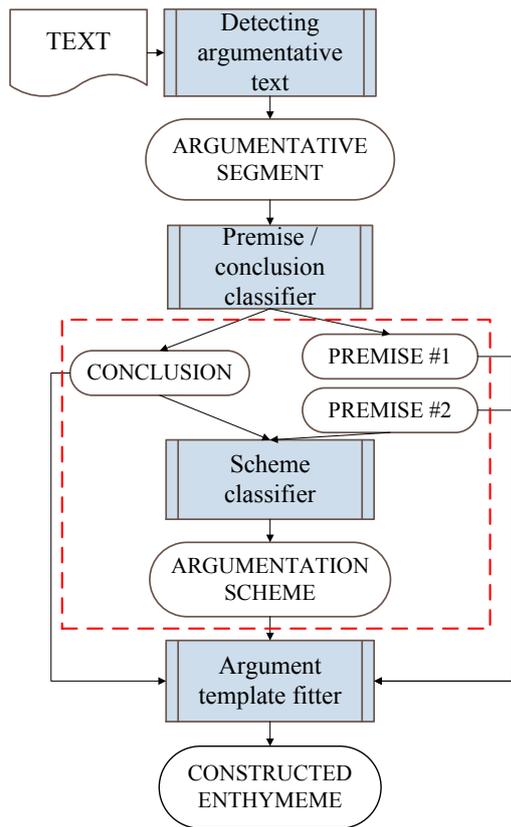


Figure 1: Overall framework of our research on argumentation schemes.

argument from example, argument from cause to effect, practical reasoning, argument from consequences, and argument from verbal classification. Casting the problem as one of text classification, we built a pruned C4.5 decision tree (Quinlan, 1993) for both one-against-others classification of each scheme and for pairwise classification of each possible pairing of schemes. We used a variety of textual features, some of them specific to a particular argument scheme and others identical across schemes. They ranged from specific keywords and phrases to word-pair similarity between the premise and the conclusion, the starting point of the premise or conclusion in its sentence, and various syntactic dependency relations. Additionally, we used one feature that cannot at present be automatically derived from text, but which we assume may be determined by cues such as discourse relations: whether the argument is linked or convergent; that is, whether or all just one of the premises suffice for the conclusion.

Using Araucaria for both training and testing, we achieved high accuracy in one-against-others

classification for argument from example and practical reasoning: 90.6% and 90.8% (baseline is 50%). The accuracy of classification of argument from cause to effect was just over 70%. However, with the other two schemes (argument from consequences and argument from verbal classification), accuracy was only in the low 60s. This is probably due at least partly to the fact that these schemes do not have such obvious cue phrases or patterns as the other three schemes, and therefore may require more world knowledge, and also because the available training data for each in Araucaria was relatively small (44 and 41 instances, respectively). In pairwise classification, we were able to correctly differentiate between most of the scheme pairs, with accuracies as high as 98% (baseline is again 50%). Performance was poor (64.0%) only for argument from consequences against argument from verbal classification — perhaps not coincidentally the two schemes for which performance was poorest in the one-against-others task.

3 Discourse analysis for argumentation analysis

The rhetorical or discourse structure of an argumentative text contributes to (or is, in part, determined by) the structure of the argument that it expresses. Consequently, much of our recent work has focused on **discourse parsing**, that is, determining the hierarchical rhetorical structure of the text: the logical relationships between sentences. Following the tenets of **Rhetorical Structure Theory (RST)** (Mann and Thompson, 1988), this is a tree structure that covers the text whose leaves are the elementary discourse units (EDUs) of text (roughly speaking, clauses and clause-like constituents) and whose edges are the RST relations that hold between EDUs or spans of related text. The set of relations include many that are pertinent to the structure of argumentation, such as CONTRAST, CAUSE, SUMMARY and ENABLEMENT. Also, as we noted above, an analysis of discourse structure may help us to discriminate convergent from linked arguments. So while an RST structure is not an argumentation structure per se, it clearly contains information that contributes to building an argumentation structure.

Our research on discourse parsing has three facets: improving the initial segmentation of text into EDUs (Feng and Hirst, 2014b); improving the parsing itself by using rich linguistic fea-

tures (Feng and Hirst, 2012); and technically improving the parser both in accuracy and in efficiency by separating the parsing of intra-sentence and multi-sentence structures into separate processes (following Joty *et al.* (2013)), and adding a post-editing pass to each process (Feng and Hirst, 2014a). Bringing the improvements together, and training and testing in the RST Discourse Treebank (Carlson *et al.*, 2001), we achieved an F_1 score of 92.6% on discourse segmentation, and an accuracy of 58.2% (against a baseline of 29.6%)² on recognizing discourse relations on a gold-standard segmentation.

Our next task will be to combine our discourse parser with our earlier work on identifying argumentation schemes. We will augment our classifier with new features derived from the discourse structure in order to improve its accuracy. We will also use discourse structure features to improve the upstream classification that feeds into the argumentation scheme classifier, and to begin the task of further downstream analysis. In particular, this will include analysis of arguments to determine the underlying ideology of a text.

4 Ideology and issue framing

Social scientists usually define ideology as a belief system: “a configuration of ideas and attitudes in which the elements are bound together by some form of constraint or functional interdependence” (Converse, 1964, p. 207). The **left / right** political divide is a systematic and enduring ideological cleavage that divides “the world of political thought and action” in democratic countries (Bobbio, 1996). Systematic left / right differences appear in the voting records of politicians in legislative assemblies (Hix *et al.*, 2006), in the election platforms of political parties (Budge *et al.*, 2001; Klingemann *et al.*, 2006), and in the patterns of public opinion (Jost, 2006). The left / right divide is so pervasive and enduring that many now wonder whether these political differences are manifestations of deeply rooted, and perhaps heritable, psychological traits (Alford *et al.*, 2005; Carney *et al.*, 2008; Haidt, 2012).

Several computational studies have looked at the question of whether a political speaker’s ideological position on the left / right spectrum can

be determined just from a quantitative analysis of the vocabulary that they use — both from the way they talk about particular topics and (in some contexts) from the topics that they tend to talk about (Lin *et al.*, 2006; Mullen and Malouf, 2006; Yu *et al.*, 2008; Diermeier *et al.*, 2012; Zirn, 2014). Typically, these studies attempt to induce a classifier from word-frequency vectors. Results have been mixed; for example, extreme positions in the U.S. Congress can be distinguished from those of the other side — sometimes by the use of topic-dependent shibboleths such as *gay* (liberal Democrat) or *homosexual* (conservative Republican) — but more-moderate positions cannot be (Yu *et al.*, 2008).

In our earlier work (Hirst *et al.*, 2010; Hirst *et al.*, 2014), we showed that the U.S. results do not apply to the Canadian Parliament. On one hand, we were able to classify party membership more reliably overall than the U.S. research did, but on the other hand we also showed that distinctions in the vocabulary of the speakers depend far more upon whether their party was in government or in opposition than upon their ideological position. The differences reflect primarily defence (government) and attack (opposition), a feature inherent to parliamentary governments in general, and especially to the Canadian parliament where party discipline is particularly strict (Savoie, 1999). When we applied classification methods based on word-frequency to the proceedings of the European Parliament, in which the factor of government–opposition status is absent, we achieved a more-accurate ideological classification of speakers from the five major parties across the left / right spectrum (Hirst *et al.*, 2014). This confounding role of institutions on left / right differences align with what others have recently uncovered in cross-national analysis of legislative voting patterns (Hix and Noury, 2013).

Casual observers of politics recognize left / right differences when they see them, but even experts struggle to define these terms. The root of the problem is the effort to define left and right by reducing each side to a single idea or “essential core”. The morphology of left and right is inconsistent with such a specification. Rather, left and right describe “family resemblances” between the systems of political ideas that actors on each side advance on the questions of political disagreement (Cochrane, 2014). Although no single idea de-

²This is the majority baseline of always labeling the resulting subtree with the relation ELABORATION with the current span as the nucleus and the next span as the satellite.

finer the left or the right, ideas are more or less central to one of these resemblances to the extent that they are more common among the belief systems of actors that are inside each category than they are among the beliefs systems of actors that are outside each category. From this vantage point, the central ideas on the political left are commitments to equality, pacifism, and, more recently, the environment. The distinguishing ideas on the right are support for capitalist economic orthodoxy, law and order, and patriotic militarism (Cochrane, 2014). The differences between political parties in their support for these ideas explain more than two-thirds of the variation in how citizens and experts position the parties on a left / right dimension (Cochrane, 2014).

The “content” of a belief system is the set of preferences that an actor harbours about political issues. The “structure” of a belief system is the way in which an actor puts different political issues together into bundles of constrained preferences. Actors that think about politics from the vantage point of altogether different ideas not only disagree in their positions on issues, they also disagree in their views of how different issues fit together logically in the political world around them. Thus, the content and the structure of belief systems varies on the left and the right (Cochrane, 2013).

Because of these differences, individuals from different ideological positions will often **frame** things differently in argumentation on any particular issue. For example, on the issue of how much immigration should be allowed into their country, one person might frame the argument as one of economic benefit or detriment, a second person as one of the benefits or problems of multiculturalism, and a third person as one of social justice.³ These differences will be reflected in the vocabulary that each of these people uses, which accounts for the results presented above on identifying ideology based on vocabulary alone; in the absence of confounding factors, as we saw most clearly in the case of the European Parliament, vocabulary is a strong indicator all by itself.

So we see that the framing of an issue by a speaker in an argumentative text is not, ultimately, a linguistic entity; it’s an ideological viewpoint or perspective: a set of beliefs, assumptions and pre-

³Immigration is in fact the particular topic on which we will conduct our case study on the framing of arguments; see section 5 below.

compiled arguments.⁴ Nonetheless, for automatic text analysis, quantifiable semantic characteristics of the speaker’s presentation of a position are indicators or proxies of the framing, which can then be interpreted qualitatively (by a human). In a simple analysis, this might be a statistical analysis of the key concepts of the text, as denoted by content words, significant collocations of words, and syntactic structures, much as in the simple text-classification-based ideology studies mentioned above, or a topic-model-based analysis, as in the work of Nguyen *et al.* (2013).

In our research, however, we are also proposing a novel, more-sophisticated analysis in which we also look at the actual argumentation structures and discourse relationships of the text and how the concepts adduced by the lower-level linguistic components are used in these structures. We will describe these proposals in the next section.

5 Argumentation and issue framing in parliamentary speech

Left / right speech is a subset of ideological speech more generally. Ideological speech is a subset of political speech more generally. As we noted above, previous analyses of political speech attempt to induce left / right classifiers from analyses of vocabulary across all of the many topics of discussion in a dataset. But this approach disregards the results of an extensive body of political science research that analyzes left / right ideological disagreement in legislative voting records (Poole and Rosenthal, 2007; Hix and Noury, 2013), party election manifestos (Budge *et al.*, 2001; Klingemann *et al.*, 2006), and opinions (Jost, 2006). A key finding from these studies concerns the varying centrality of specific actors, ideas, and topics to left / right political disagreement. Some actors are more central to the left or to the right than are other actors. Some ideas are more central to the left or to the right than are other ideas. Left / right disagreements implicate some political issues and not others. This provides an informative prior for models that seek to uncover left / right differences from the patterns of vocabu-

⁴A fortiori, framing is a political action: “Framing essentially involves selection and salience. To frame is to select some aspects of a perceived reality and make them more salient in a communicating text, in such a way as to promote a particular problem definition, causal interpretation, moral evaluation, and/or treatment recommendation for the item described” (Entman, 1993). But here, we focus on the linguistic and argumentative aspects of framing.

lary and argumentation in political text. The likelihood that speech conveys information about left / right argumentation is a function of the speaker and the topic.

Thus, the goal of our work, broadly speaking, is to develop computational models for the automatic analysis of ideology and issue framing in political speech that are better informed than the simple vocabulary-based models and that draw on automatic discourse parsing and automatic analysis of argumentation as their primary mechanism. We would like to look more narrowly and more deeply at argumentation on specific issues by individuals across the left / right spectrum, and develop automatic methods of analysis that will identify, or help analysts to identify, different frames and ideological positions. Our “help to” hedge reflects the difficulty of the goal and the context of our research as part of a much-larger project that is building datasets and tools to assist political scientists and political historians in their analyses.

The primary data for our work is the annotated parliamentary proceedings, from the present back to the mid-1800s or earlier, that are being produced by the Dilipad project (see section 1 above), from which we will draw speech⁵ on specific topics for diachronic and cross-national analysis of argumentation and framing. Immigration is a topic of special interest here, as it has been an important and recurring issue since the nineteenth century in all three participating countries. We hope to identify national and temporal differences and similarities in the frames used to discuss the issue.

In our models, we will bring together, and extend, the work on discourse parsing and argumentation scheme identification described in sections 2 and 3 above. Although these techniques are far from perfect, we hypothesize that typical political speech contains a sufficiently well-cued discourse structure that the analyses that we can achieve, although still quite imperfect, will be usefully indicative of issue framing and other ideological signals, and will be more immune to confounding factors, such as the attack-and-defence dynamics of parliamentary debates, than simple vocabulary classification. In particular, we will use features from discourse units and rhetorical re-

lations to find claims and analyze the reasoning structure that is used to justify, support, and derive the claims. In addition, we will take into account how the concepts adduced by lower-level linguistic components — phrases, syntactic dependency structures — are used in the actual argumentation structures and discourse relationships of the text. We hope to be able to recognize instances of known frames in the text, and possibly even discover new ones. Because we will be developing deeper and hence more tentative methods of computational linguistic analysis, we do not expect to provide a complete automated analysis of text in the first instance, but rather to provide data that can then be interpreted by a human analyst.

In parallel with this approach, we will also develop text-classification methods for identifying ideological positions in speech that will look beyond vocabulary and also take into consideration frequent collocations and lexicalized syntactic dependency structures as features. This will allow us to include differences in the way that particular words are used (even where speakers use the word with the same frequency) as a feature of the classification. This will provide a new, higher baseline against which the results of the discourse- and argumentation-based analysis can be evaluated. It may also provide information that can itself be a component of that analysis. In addition, the words, collocations, and dependency structures that are most informative for classification will, as with our other methods, be available for human interpretation.

6 Conclusion

Our work focuses on the structure of discourse and arguments to better understand ideological positions and issue framing through their linguistic realizations. By applying discourse parsing and the analysis of argumentation to parliamentary debates, we hope to determine how speakers with various ideologies argue on a range of issues. Ideologies are manifested not only by the vocabularies used, but also by how the differing beliefs of political speakers lead to different framing of issues. Ideology detection can therefore benefit from argumentation and discourse analysis techniques.

⁵Although we refer to *political* and *parliamentary speech* and *speakers*, as is conventional, we are working only with the published textual transcriptions of the parliamentary debates. We are not using audio data or any kind of automatic speech recognition.

Acknowledgements

The Digging Into Linked Parliamentary Data project is funded through the Digging Into Data Challenge. The Canadian arm of the project is funded by the Natural Sciences and Engineering Research Council of Canada through its Discovery Frontiers program and by the Social Sciences and Humanities Research Council. We are grateful to Kaspar Beelen for helpful comments on an earlier draft of this paper.

References

- Alford, John R.; Funk, Carolyn L.; and Hibbing, John R. (2005). *American Political Science Review*, 99(2), 153–167.
- Bobbio, Noberto (1996). *Left and Right: The Significance of a Political Distinction*. Cambridge, UK: Polity Press.
- Budge, Ian; Klingemann, Hans-Dieter; Volkens, Andrea; and Bara, Judith (2006). *Mapping Policy Preferences: Estimates for Parties, Electors, and Governments 1945–1998*. Oxford, UK: Oxford University Press.
- Carlson, Lynn; Marcu, Daniel; and Okurowski, Mary Ellen (2001). Building a discourse-tagged corpus in the framework of Rhetorical Structure Theory. *Proceedings of Second SIGDial Workshop on Discourse and Dialogue (SIGDial 2001)*, Aalborg, 1–10.
- Carney, Dana R.; Jost, John T.; Gosling, Samuel D.; and Potter, Jeff (2012). The secret lives of liberals and conservatives: Personality profiles, interaction styles, and the things they leave behind. *Political Psychology*, 29(6), 807–840.
- Cochrane, Christopher (2013). The asymmetrical structure of left / right disagreement: Left-wing coherence and right-wing fragmentation in comparative party policy. *Party Politics*, 19(1), 104–121.
- Cochrane, Christopher (2014). *Left and Right: The Small World of Political Ideas*. MS under review.
- Converse, Philip E. (1964). The nature of belief systems in mass publics. In David E. Apter, ed. *Ideology and Discontent*. London, UK: Collier-MacMillan, 206–261.
- Diermeier, Daniel; Godbout, Jean-François; Yu, Bei; and Kaufmann, Stefan (2012). Language and ideology in Congress. *British Journal of Political Science*, 42(1), 31–55.
- Entman, Robert M. (1993). Framing: Toward clarification of a fractured paradigm. *Journal of Communication*, 43(4), 51–58.
- Feng, Vanessa Wei and Hirst, Graeme (2011). Classifying arguments by scheme. *Proceedings, 49th Annual Meeting of the Association for Computational Linguistics*, Portland, Oregon, 978–996.
- Feng, Vanessa Wei and Hirst, Graeme (2012). Text-level discourse parsing with rich linguistic features. *Proceedings, 50th Annual Meeting of the Association for Computational Linguistics*, Jeju, Korea, 60–68.
- Feng, Vanessa Wei and Hirst, Graeme (2014a). A linear-time bottom-up discourse parser with constraints and post-editing. *Proceedings, 52nd Annual Meeting of the Association for Computational Linguistics*, Baltimore.
- Feng, Vanessa Wei and Hirst, Graeme (2014b). Two-pass discourse segmentation with pairing and global features. Submitted.
- Haidt, Jonathan (2012). *The Righteous Mind: Why Good People are Divided by Politics and Religion*. New York, NY: Pantheon Books.
- Hirst, Graeme; Riabinin, Yaroslav; and Graham, Jory (2010). Party status as a confound in the automatic classification of political speech by ideology. *Proceedings, 10th International Conference on Statistical Analysis of Textual Data / 10es Journées internationales d'Analyse statistique des Données Textuelles (JADT 2010)*, Rome, 731–742.
- Hirst, Graeme; Riabinin, Yaroslav; Graham, Jory; Boizot-Roche, Magali; and Morris, Colin (2014). Text to ideology or text to party status? In: Kaal, Bertie; Maks, E. Isa; and van Elfrinkhof, Annemarie M.E. (editors), *From Text to Political Positions: Text analysis across disciplines*, Amsterdam: John Benjamins, 93–115.
- Hix, Simon; and Noury, Abdul (2013). Government-opposition or left-right? The institutional determinants of voting in legislatures. Working paper, retrieved from <http://personal.lse.ac.uk/hix/Research.HTM>, 2014-06-14.
- Hix, Simon; Noury, Abdul; and Roland, Gérard (2006). Democratic politics in the European Parliament. *American Journal of Political Science*, 50(2), 494–520.
- Jost, John T. (2006). The end of the end of ideology. *American Psychologist*, 61(7), 651–670.
- Joty, Shafiq; Carenini, Giuseppe; Ng, Raymond; and Mehdad, Yashar (2013). Combining intra- and multi-sentential rhetorical parsing for document-level discourse analysis. *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (ACL 2013)*, Sofia, Bulgaria, 486–496.
- Klingemann, Hans-Dieter; Volkens, Andrea; Bara, Judith; Budge, Ian; and McDonald, Michael (2006). *Mapping Policy Preferences II: Estimates for Parties, Electors, and Governments in Eastern Europe, European Union, and OECD 1990–2003*. Oxford, UK: Oxford University Press.

- Lin, Wei-Hao; Wilson, Theresa; Wiebe, Janyce; and Hauptmann, Alexander (2006). Which side are you on? Identifying perspectives at the document and sentence levels. *Proceedings of the 10th Conference on Natural Language Learning (CoNLL-X)*, 109–116.
- Mann, William and Thompson, Sandra (1988). Rhetorical structure theory: Toward a functional theory of text organization. *Text*, 8(3), 243–281.
- Mochales, Raquel and Moens, Marie-Francine (2008). Study on the structure of argumentation in case law. *Proceedings of the 2008 Conference on Legal Knowledge and Information Systems*, Amsterdam: IOS Press, 11–20.
- Mochales, Raquel and Moens, Marie-Francine (2009a). Argumentation mining: the detection, classification and structure of arguments in text. *ICAIL '09: Proceedings of the 12th International Conference on Artificial Intelligence and Law*, ACM, 98–107.
- Mochales, Raquel and Moens, Marie-Francine (2009b). Automatic argumentation detection and its role in law and the Semantic Web. *Proceedings of the 2009 Conference on Law, Ontologies and the Semantic Web*, Amsterdam: IOS Press, 115–129.
- Mullen, Tony and Malouf, Robert (2006). A preliminary investigation into sentiment analysis of informal political discourse. *Proceedings of the AAAI Symposium on Computational Approaches to Analyzing Weblogs*, 159–162.
- Nguyen, Viet-An; Boyd-Graber, Jordan; and Resnik, Philip (2013). Lexical and hierarchical topic regression. *Proceedings of Advances in Neural Information Processing Systems 26*, Lake Tahoe, NV, 1106–1114.
- Poole, Keith T. and Rosenthal, Howard L. (2007). *Ideology and Congress*. New Brunswick, NJ: Transaction Publishers.
- Quinlan, J. Ross (1993). C4.5: Programs for machine learning, *Machine Learning*, 16(3), 235–240.
- Savoie, Donald (1999). *Governing from the Centre: The Concentration of Power in Canadian Politics*. Toronto, ON: University of Toronto Press.
- Reed, Chris and Rowe, Glenn (2004). Araucaria: Software for argument analysis, diagramming and representation. *International Journal of Artificial Intelligence Tools*, 14, 961–980.
- Rowe, Glenn and Reed, Chris (2008). Argument diagramming: The Araucaria project. In *Knowledge Cartography: Software tools and mapping techniques*, edited by Alexandra Okada, Simon J. Buckingham Shum, and Tony Sherborne, London: Springer-Verlag, 163–181.
- Walton, Douglas; Reed, Chris; and Macagno, Fabrizio (2008). *Argumentation Schemes*. Cambridge University Press.
- Yu, Bei; Kaufmann, Stefan; and Diermeier, Daniel (2008). Classifying party affiliation from political speech. *Journal of Information Technology in Politics*, 5(1), 33–48.
- Zirn, Cécilia (2014). Analyzing positions and topics in political discussions of the German Bundestag. *Proceedings of the ACL 2014 Student Research Workshop*, Baltimore, 26–33.