Who Sides With Whom? Towards Computational Construction of Discourse Networks for Political Debates

Sebastian Padó[1], **André Blessing**[1], **Nico Blokker**[2],
Erenay Dayanik[1], **Sebastian Haunss**[2], and **Jonas Kuhn**[1]

[1]IMS, University of Stuttgart, Germany
[2]SOCIUM, University of Bremen, Germany

Abstract

Understanding the structures of political debates (which actors make what claims) is essential for understanding democratic political decision-making. The vision of computational construction of such *discourse networks* from newspaper reports brings together political science and natural language processing. This paper presents three contributions towards this goal: (a) a requirements analysis, linking the task to knowledge base population; (b) a first release of an annotated corpus of claims on the topic of migration, based on German newspaper reports; (c) initial modeling results.

1 Introduction

Democratic decision making can follow broadly two logics: In a technocratic, depoliticized mode, decision-making is carried out by administrative staff and experts. However, arguably most political decisions affecting large populations attract public attention and thus happen in a politicized mode, in which public debates accompany decision making (de Wilde, 2011; Zürn, 2014; Haunss and Hofmann, 2015). Understanding the structure and evolution of political debates is therefore essential for understanding democratic decision making. Recent innovations that combine political claims analysis (Koopmans and Statham, 1999) with network science under the name of *discourse network analysis* (Leifeld, 2016a) allow us to systematically analyze the dynamics of political debates based on the annotation of large newspaper corpora. So far, such studies have been carried out manually.

In this paper, we outline the road towards using *computational methods* from natural language processing for the construction of discourse networks – working towards an integrated methodological framework for Computational Social Science. We make three contributions: (a) a requirements analysis; (b) a manually annotated corpus of claims from
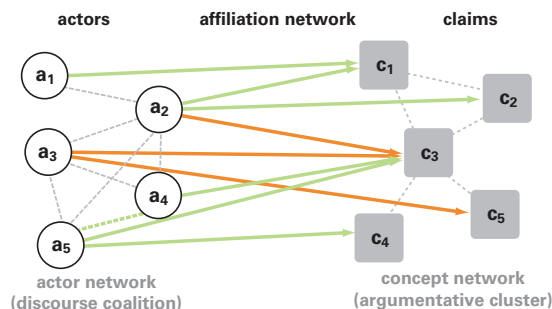

Figure 1: Actor, affiliation, and concept networks

debates about migration found in German newspaper reports; (c) initial modeling results that already demonstrate the usefulness of computational methods in this context.

2 Discourse Networks: Actors and Claims

Discursive interventions are one element among several that influence policy making (Schmidt and Radaelli, 2004). But the exact mechanisms of political discourse and under which condition discursive interventions do or do not translate into political decisions are largely unknown. At least there seems to be a general agreement that the formation and evolution of *discourse coalitions* is a core mechanism (Hajer, 1993; Sabatier and Weible, 2007).

A discourse coalition can be generally defined as "a group of actors who share a social construct" (Hajer, 1993, p. 43). Political Claims Analysis (Koopmans and Statham, 1999) provides a framework in which *claims*, that is demands, proposals, criticisms, or decisions in the form of statements or collective actions reported in newspaper articles are attributed to (groups of) *actors* and are *categorized*. Actors and claims can be represented as the two classes of nodes in a bipartite *affiliation network*. In Figure 1, actors are circles, claims are squares, and they are linked by edges that indicate support (green) or opposition (orange). A discourse coali-

tion is then the projection of the affiliation network on the actor side (dotted edges), while the projection on the concept side yields the argumentative clusters present in the debate.

3 NLP and Political Science

Our analytical goals have connecting points with a range of activities in NLP. There has been considerable work in Social Media Analysis using NLP – in particular sentiment analysis (e.g. Ceron et al. 2014), but also going into fine-grained analysis of groups of users/actors (Cesare et al., 2017). Nevertheless, most analyses in social media concern typically relatively broad categories, such as party preferences (see Hong et al. 2016 for a comparison of social media and news texts). NLP techniques are also used for stance classification (e.g. Vilares and He 2017) and measuring ideology in speeches (Sim et al., 2013), and there is a fair amount of work on agenda-setting and framing (e.g. Tsur et al. 2015; Field et al. 2018). To our knowledge, fine-grained distinctions both for actors and claims that are necessary for discourse network consideration (cf. Section 4) have not been explored in depth.

Also related is the growing field of argumentation analysis/mining (e.g. Peldszus and Stede 2013; Swanson et al. 2015; Stab and Gurevych 2017). However, a core interest there is analyzing the argument structure of longer pieces of argumentative text (i.e., claims and their (recursive) justifications), whereas we focus on the core claims that actors put forward in news coverage.

The aspect of dynamics in interaction among actors is shared with work on the extraction of actor/character networks from texts, which has been applied mostly to literary texts (Elson et al., 2010; Hassan et al., 2012; Iyyer et al., 2016).

4 Computational Construction of Discourse Networks

Seen as an end-to-end task, the computational construction of affiliation networks from newspaper articles as introduced in Section 2 represents a task that combines binary relation extraction (Doddington et al., 2004; Hendrickx et al., 2010) with ontologization (Pennacchiotti and Pantel, 2006; Hachey et al., 2013, i.a.). The task can be decomposed conceptually as shown in Figure 2. From bottom to top, the first task is to identify claims and actors in the text (Tasks 1 and 2). Then, they need to be mapped onto entities that are represented in the affiliation
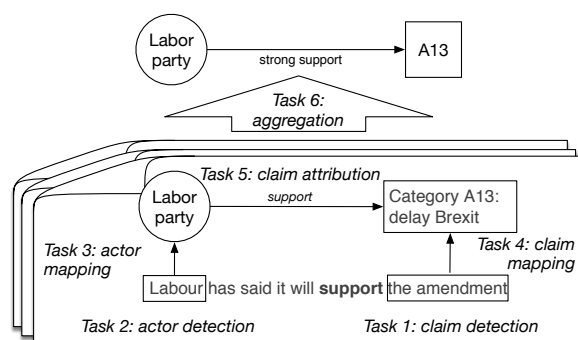

Figure 2: Construction of affiliation network construction (top) from text (bottom) as relation extraction

graph, that is, discourse referents for actors (Task 3: entity linking) and categories for claims (Task 4). Next, claims need to be attributed to actors and classified as support or opposition (Task 5). Finally, relations need to be aggregated across documents (Task 6).

This setup is related to Knowledge Base Population (McNamee et al., 2010) and presents itself as a series of rather challenging tasks:

Actor and claim ontologies. The actors and claims can either be known a priori (then Tasks 3 and 4 amount to classification) or can emerge from the data (then they become clustering tasks). We assume that there is a limited set of claims that structures public debates on a given topic (Koopmans and Statham, 2010). We thus build on an expert-defined ontology of claims (cf. Section 5). With regard to actors, the issue is less clear: knowledge bases such as Wikidata cover many persons in the public eye. However new actors can appear and take on importance at any time.

Discourse context. Tasks 3 and 4 regularly involves coreference resolution: in the example, the expression *the amendment* can only be mapped to the correct claim if its content can be inferred. Similarly, actors realized as pronouns have to be resolved. Coreference resolution is still a difficult problem (Martschat and Strube, 2014).

Dependencies among tasks. The various tasks are clearly not independent of one another, and joint models have been developed for a subset of the tasks, such as coreference and relation detection (Almeida et al., 2014) or entity and relation classification (Miwa and Sasaki, 2014; Adel and Schütze, 2017; Bekoulis et al., 2018). However, state-of-the-art models still struggle with sentence complexity, and there are no comprehensive models of the complete task including aggregation.

C1: Steuerung von Migration (Controlling Migration)
C2: Aufenthalt (Residency)
C3: Integration (Integration)
C4: Innere Sicherheit (Domestic Security)
C5: Aussenpolitik (Foreign Policy)
C6: Ökonomie, Arbeitsmarkt (Economy, Labor Market)
C7: Gesellschaft (Society)
C8: Verfahren (Procedures)

Table 1: Migration: Main categories in claim ontology

5 Claim Ontology and Corpus Annotation

We now demonstrate the first steps of computational discourse network construction in a concrete political context, namely the major topic of German politics of 2015: the domestic debate on (im-) migration precipitated by the war in Syria.

Claim Ontology. Following established approaches to content analysis from political science (Leifeld, 2016b), we chose an approach that combines deductive and inductive elements to identify an initial set of topic-specific claim categories. First, we review the literature, extract relevant categories, and validate and extend them based on an initial sample of newspaper articles from *Die Tageszeitung*, a large left-leaning German quality newspaper (www.taz.de). This results in eight superordinate categories (cf. Table 1) and 89 subcategories, capturing a variety of different political positions. These categories and their definitions form the codebook that the annotation is based on.[1]

Annotation Process. Annotation follows a procedure successfully used by Haunss et al. (2013) in the analysis of the German nuclear phase-out debate (2011). The analysis of articles is carried out in double, independent annotation by trained student research assistants. An example of a text passage and its corresponding annotation is presented in the following sentence:

(1) [Flüchtlinge zum Erlernen der deutschen Sprache [...] verpflichten]$_{Claim}$, will [die CDU in Niedersachsen]$_{Actor}$.
[Requiring refugees to learn the German language]$_{Claim}$ [...] is what [the CDU party in Lower Saxony]$_{Actor}$ wants.

Annotators mark the claim and the actor, classify the claim as (a subtype of) C3, integration, link them, and mark the position (support/opposition). That is, Tasks 1–5 from Section 4 are all carried

---

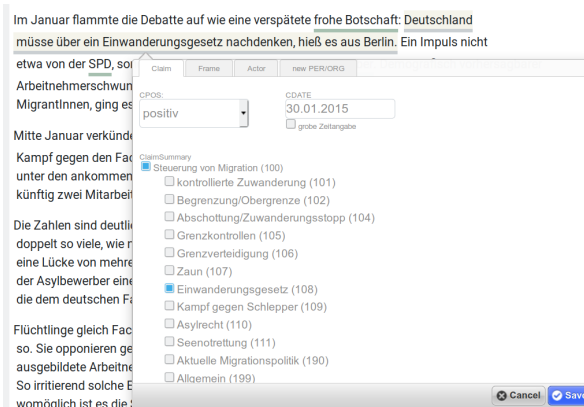[1]For the full codebook, see the supplementary material.

Figure 3: Screenshot of annotation platform, with text (back) and annotation window (front)

out. Crucially, cross-cutting ("*multi-label*") claims can instantiate multiple categories. In our annotation, about 17% of all claims carry multiple labels. Frequent combinations at the top level are C2+C8 (procedural aspects of residency) and C1+C5 (international perspective on migration control).

Building on experience and tool components from text annotation efforts in Digital Humanities projects (in particular the *Center for Reflected Text Analytics*, https://www.creta.uni-stuttgart.de/en/), we developed a web-based annotation tool, shown in Figure 3, which both streamlines annotation and encourages consistency. Annotation involves first marking claim and actor spans in the text and then selecting the correct categories for the claims and the correct referent for the actor from drop-down lists. See Blessing et al. (2019) for details.

Reliability and Adjudication. We compute annotation reliability of the original student annotators for the two initial and most immediate annotation steps (cf. Figure 2), namely claim detection (Task 1) and classification (Task 4). For claim detection, a classical single-label classification task, we use Cohen's Kappa: For each sentence, we compare whether the two annotators classified the sentence as part of a claim or not. We obtain a Kappa value of 0.58. For claim classification, a multi-label classification task, we cannot use Kappa. Instead, we compute Macro-F1 for all top level categories, and obtain an average F1 score of 63.5%.

These numbers, while still leaving room for improvement, indicate moderate to substantial agreement among the student annotators. The two sets of annotations per document are subsequently reviewed and adjudicated by senior domain experts

to create a reliable gold standard.

Dataset Release. With this paper, we publicly release 423 fully annotated articles from the 2015 *Tageszeitung*. 179 articles contain at least one claim. In total, 982 Claims in 764 different text passages have been annotated. This includes additional information such as actor attributes (name, party membership, etc.), date and position. This dataset – together with documentation and annotation guidelines is available for research purposes at https://github.com/mardy-spp/mardy_acl2019.

Remaining Challenges. A number of challenges remain. A technical one is the identification of relevant documents: keyword-based methods turn out to be insufficient. A conceptual one is that not all decisions made in the design of the claim ontology hold up to broad-coverage annotation. Political science has defined the ideal of 'multi-pass coding' (Leifeld, 2016b) according to which the researcher constantly reviews and updates annotation in an iterative process, adding and collapsing categories as needed. We perform such updates at regular intervals, but they can only be meaningfully applied to the adjudicated gold standard, not individual annotations. Thus, our reliability is likely underestimated by the analysis above.

6 Modeling results

Due to space restrictions, this paper only reports on first steps towards computational construction of discourse networks. Specifically, we present pilot models for Tasks 1 and 4 (claim identification and attribution), the two tasks for which we also presented reliability analyses in Section 5.

Data setup. We randomly sampled 90% of our dataset for training and evaluate on the other 10%; the split is published with the dataset. We discarded articles with no claims.

Claim Identification. We model claim identification as a sequence labeling task: The model labels each token in a sentence as B-Claim, I-Claim or Outside, adopting a BIO schema.

We experiment with two model architectures. The first one is BERT (Devlin et al., 2018), a state-of-the-art transformer-based neural network model, which we fine-tune on our training data. The second is a current architecture for sequence labeling that consists of an embedding layer, an LSTM layer,

and a CRF layer.[2] We use word embeddings from FastText (Bojanowski et al., 2017). In order to add task and domain specific representations and resolve Out-Of-Vocabulary (OOV) problem, we experiment with a second embedding approach, namely learning character-based embeddings from which we compute word-level embeddings by feeding the character embeddings through a CNN and max-pooling the out. Depending on the experimental condition (see below), we use either just the word-based or a concatenation of the word-based and character-based embeddings, and train the embeddings on different corpora.

All embeddings are fed to a bidirectional LSTM layer for contextualization. To jointly model the label sequence, we use a CRF layer on top. For a sequence with n words, we parameterize the distribution over all possible label sequences, $\mathcal{Y}$, as

$$p(\boldsymbol{y}|\mathbf{d}; \mathbf{W}) = \frac{\prod\limits_{i=1}^{n} \phi_i\left(y_{i-1}, y_i, \mathbf{d}\right)}{\sum\limits_{y' \in \mathcal{Y}} \prod\limits_{i=1}^{n} \phi_i\left(y'_{i-1}, y'_i, \mathbf{d}\right)} \quad (1)$$

where $d = [d_1, d_2, \ldots d_n]$ is the set of representation produced by BiLSTM for each input word and $\phi_i\left(y_{i-1}, y_i, \mathbf{d}\right)$ is a function calculating emission and transition potentials between the tags y_{i-1} and y_i. During training, we maximize the log-likelihood function over the training set

$$L(\mathbf{W}) = \sum_i \log p(\boldsymbol{y}|\mathbf{d}; \mathbf{W}) \quad (2)$$

During inference, the sequence with highest conditional probability is predicted by a Viterbi decoder:

$$\underset{y \in \mathcal{Y}}{\operatorname{argmax}} \, p(\boldsymbol{y}|\mathbf{d}; \mathbf{W}) \quad (3)$$

Experimentally, we compared BERT against versions of our own model which (a) do and do not include the CRF layer; (b) do or do not use the character-level embeddings; (c) train embeddings on different corpora. We measure performance as F1 scores per-class, and macro F1 scores overall.

We started with a simple model, (1), using the default Wikipedia FastText word-level embeddings and without CRF layer. Moving to in-domain TAZ embeddings, (2), improves performance by 4 points macro F1, with a slight further improvement of 0.5 points by adding character-level embeddings in

---

[2]See supplement for details and hyperparameters.

Method	B-C	I-C	O	Macro
(1) $\text{Emb}_{\text{Wiki:w}}$+BiLSTM	31.3	37.5	93.5	54.1
(2) $\text{Emb}_{\text{TAZ:w}}$+BiLSTM	38.5	43.9	93.6	58.7
(3) $\text{Emb}_{\text{TAZ:w,c}}$+BiLSTM	40.0	44.1	93.1	59.1
(4) $\text{Emb}_{\text{TAZ:w,c}}$+BiLSTM+CRF	49.4	**53.8**	**95.5**	**66.3**
(5) $\text{Emb}_{\text{Wiki:w,c}}$+BiLSTM+CRF	35.1	39.1	90.6	55.0
(6) BERT	**49.5**	52.4	94.7	65.5

Table 2: Claim identification scores on evaluation set: F1 for BIO labels and Macro average (Mac). $\text{Emb}_{\text{Corpus:type}}$: Corpus used to train embeddings and type (w: word, c: char). For example, $\text{Emb}_{\text{TAZ:w,c}}$ represents version for which character and word level embeddings trained on TAZ corpus.

Method	C1	C2	C3	C4	C5	C6	C7	C8	Macro
NB	46	50	0	0	43	0	29	0	21
MLP	**73**	53	0	0	**67**	0	57	46	37
BiLSTM	71	71	0	0	63	0	**78**	24	38
BERT	44	**82**	**54**	**29**	50	0	53	**57**	**46**

Table 3: Claim categorization performance of several models. Columns C1-C8 show F1 score for each category. Macro reports macro average F1 score. NB: Naive Bayes, MLP: Multi-Layer Perceptron.

(3). Adding a CRF layer to obtain the full model, (4), yields a further major increase by 7 points F-score and results in the best overall model with 66.3 macro F1. This model also outperforms BERT, (6), numerically in macro F1 and for the two classes (I-C) and (O). This model still profits substantially from the in-domain embeddings: replacing them by Wikipedia-trained ones in model (5) results in a drop of 11 points.

Claim Classification. For our experiments on claim classification, we assume that claims have already been detected. To each claim span, we assign one or more of the top categories from the claim ontology (cf. Section 5), i.e., we perform multi-class multi-label classification.

In terms of models, we evaluate a fine-tuned version of BERT against three standard classification architectures: a unigram Naive Bayes model and Multi-Layer Perceptron and BiLSTM architectures based on TAZ-trained FastText embeddings that performed well in the previous experiment. All models perform multi-class classification by making a binary decision for each class.

Table 3 shows the results, using the same F1 measures as before. BERT excels at this task, followed by the two embedding-based models; Naive Bayes comes last. Interestingly, the models differ in their performance across classes. BERT tends to make better predictions than the other models for small, homogeneous classes (C3: integration, C4: security) while MLP and BiLSTM do better on the larger and less clearly delineated classes (C1: migration control, C7: society).

7 Conclusion

In this paper, we have sketched the way towards a Computational Social Science (CSS) framework for the construction of discourse networks (claims and actors) from news coverage of political debates, which has great potential for expanding the empirical basis for research in political science. The complexity of the scenario (fine-grained categories, multi-category claims, complex relations, aggregation) suggests that an attempt at automating the construction in its entirety is currently not realistic at a quality that makes it useful for political scientists.

In the broader picture of a project that derives its motivation both from NLP and from CSS, scaling the computational component is an important objective, but one that should never come at the cost of reliability of the analytical components and methodological validity from the point of view of political science. A carefully laid out task analysis, as put forward in this paper, provides the basis for exploring more interactive "mixed methods" frameworks (see the discussion in Kuhn (to appear)): Computational models for a given set of claim categories can feed semi-automatic corpus annotation through manual post correction of predictions.

Finally, an interleaved cross-disciplinary collaboration may support the future research process further: the claim ontology for a new field of debate could be constructed in a bootstrapping process, combining the political scientists' analytical insights with (preliminary) predictions of computational seed models from partially overlapping fields. In our collaboration, systematic tool support has already made the process of codebook development considerably more effective.

References

Heike Adel and Hinrich Schütze. 2017. Global normalization of convolutional neural networks for joint entity and relation classification. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 1723–1729, Copenhagen, Denmark.

Mariana S. C. Almeida, Miguel B. Almeida, and André F. T. Martins. 2014. A joint model for quotation attribution and coreference resolution. In *Proceedings of the Conference of the European Chapter of the Association for Computational Linguistics*, pages 39–48, Gothenburg, Sweden.

Giannis Bekoulis, Johannes Deleu, Thomas Demeester, and Chris Develder. 2018. Joint entity recognition and relation extraction as a multi-head selection problem. *Expert Systems with Applications*, 114:34–45.

André Blessing, Nico Blokker, Sebastian Haunss, Jonas Kuhn, Gabriella Lapesa, and Sebastian Padó. 2019. An environment for the relational annotation of political debates. In *Proceedings of ACL System Demonstrations*, Florence, Italy.

Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. 2017. Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, 5:135–146.

Andrea Ceron, Luigi Curini, Stefano M. Iacus, and Giuseppe Porro. 2014. Every tweet counts? how sentiment analysis of social media can improve our knowledge of citizens' political preferences with an application to italy and france. *New Media & Society*, 16(2):340–358.

Nina Cesare, Christan Grant, and Elaine Okanyene Nsoesie. 2017. Detection of user demographics on social media: A review of methods and recommendations for best practices. *CoRR*, abs/1702.01807.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

George R. Doddington, Alexis Mitchell, Mark A. Przybocki, Lance A. Ramshaw, Stephanie Strassel, and Ralph M. Weischedel. 2004. The automatic content extraction (ACE) program – tasks, data, and evaluation. In *Proceedings of LREC*, Lisbon, Portugal.

David K Elson, Nicholas Dames, and Kathleen R McKeown. 2010. Extracting social networks from literary fiction. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, pages 138–147, Uppsala, Sweden.

Anjalie Field, Doron Kliger, Shuly Wintner, Jennifer Pan, Dan Jurafsky, and Yulia Tsvetkov. 2018. Framing and agenda-setting in Russian news: a computational analysis of intricate political strategies. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 3570–3580, Brussels, Belgium.

Ben Hachey, Will Radford, Joel Nothman, Matthew Honnibal, and James R. Curran. 2013. Evaluating entity linking with Wikipedia. *Artificial Intelligence*, 194:130 – 150.

Maarten A Hajer. 1993. Discourse Coalitions and the Institutionalization of Practice: The Case of Acid Rain in Britain. In Frank Fischer and John Forester, editors, *The Argumentative Turn in Policy Analysis and Planning*, pages 43–76. Duke University Press.

Ahmed Hassan, Amjad Abu-Jbara, and Dragomir Radev. 2012. Extracting signed social networks from text. In *Workshop Proceedings of TextGraphs-7 on Graph-based Methods for Natural Language Processing*, pages 6–14, Jeju, South Korea.

Sebastian Haunss, Matthias Dietz, and Frank Nullmeier. 2013. Der Ausstieg aus der Atomenergie. Diskursnetzwerkanalyse als Beitrag zur Erklärung einer radikalen Politikwende. *Zeitschrift für Diskursforschung*, 1(3):288–316.

Sebastian Haunss and Jeanette Hofmann. 2015. Entstehung von Politikfeldern – Bedingungen einer Anomalie. *dms – der moderne staat*, 8(1):29–49.

Iris Hendrickx, Su Nam Kim, Zornitsa Kozareva, Preslav Nakov, Diarmuid Ó Séaghdha, Sebastian Padó, Marco Pennacchiotti, Lorenza Romano, and Stan Szpakowicz. 2010. Multi-way classification of semantic relations between pairs of nominals. In *Proceedings of the International Workshop on Semantic Evaluation*, pages 33–38, Uppsala, Sweden.

Lingzi Hong, Weiwei Yang, Philip Resnik, and Vanessa Frías-Martínez. 2016. Uncovering topic dynamics of social media and news: The case of Ferguson. In *Proceedings of Social Informatics*, pages 240–256, Bellevue, WA.

Mohit Iyyer, Anupam Guha, Snigdha Chaturvedi, Jordan Boyd-Graber, and Hal Daumé III. 2016. Feuding families and former friends: Unsupervised learning for dynamic fictional relationships. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1534–1544, San Diego, CA.

Ruud Koopmans and Paul Statham. 1999. Political Claims Analysis: Integrating Protest Event And Political Discourse Approaches. *Mobilization*, 4(2):203–221.

Ruud Koopmans and Paul Statham. 2010. Theoretical Framework, Research Design, and Methods. In Ruud Koopmans and Paul Statham, editors, *The Making of a European Public Sphere*, pages 34–59. Cambridge University Press.

Jonas Kuhn. to appear. Computational text analysis within the humanities: How to combine working practices from the contributing fields? *Language Resources and Evaluation*.

Philip Leifeld. 2016a. Discourse Network Analysis: Policy Debates as Dynamic Networks. In Jennifer Nicoll Victor, Alexander H. Montgomery, and Mark Lubell, editors, *The Oxford Handbook of Political Networks*. Oxford University Press.

Philip Leifeld. 2016b. *Policy Debates as Dynamic Networks: German Pension Politics and Privatization Discourse*. Campus Verlag, Frankfurt/New York.

Sebastian Martschat and Michael Strube. 2014. Recall error analysis for coreference resolution. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 2070–2081, Doha, Qatar.

Paul McNamee, Hoa Trang Dang, Heather Simpson, Patrick Schone, and Stephanie Strassel. 2010. An evaluation of technologies for knowledge base population. In *Proceedings of the Seventh International Language Resources and Evaluation Conference*, Valletta, Malta.

Makoto Miwa and Yutaka Sasaki. 2014. Modeling joint entity and relation extraction with table representation. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 1858–1869, Doha, Qatar.

Andreas Peldszus and Manfred Stede. 2013. From argument diagrams to argumentation mining in texts: A survey. *International Journal of Cognitive Informatics and Natural Intelligence*, 7(1):1–31.

Marco Pennacchiotti and Patrick Pantel. 2006. Ontologizing semantic relations. In *Proceedings of the 21st International Conference on Computational Linguistics and the 44th Annual Meeting of the Association for Computational Linguistics*, pages 793–800, Sydney, Australia.

Paul A. Sabatier and Cristopher M. Weible. 2007. The Advocacy Coalition Framework: Innovations and Clarifications. In Paul A. Sabatier, editor, *Theories of the Policy Process*, pages 189–220. Westview Press.

Vivien A. Schmidt and Claudio M. Radaelli. 2004. Policy Change and Discourse in Europe: Conceptual and Methodological Issues. *West European Politics*, 27(2):183.

Yanchuan Sim, Brice D. L. Acree, Justin H. Gross, and Noah A. Smith. 2013. Measuring ideological proportions in political speeches. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 91–101, Seattle, WA.

Christian Stab and Iryna Gurevych. 2017. Parsing argumentation structures in persuasive essays. *Computational Linguistics*, 43(3):619–659.

Reid Swanson, Brian Ecker, and Marilyn Walker. 2015. Argument mining: Extracting arguments from online dialogue. In *Proceedings of the Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 217–226, Prague, Czech Republic.

Oren Tsur, Dan Calacci, and David Lazer. 2015. A frame of mind: Using statistical models for detection of framing and agenda setting campaigns. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, pages 1629–1638, Beijing, China.

David Vilares and Yulan He. 2017. Detecting perspectives in political debates. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 1573–1582, Copenhagen, Denmark.

Pieter de Wilde. 2011. No Polity for Old Politics? A Framework for Analyzing the Politicization of European Integration. *Journal of European Integration*, 33(5):559–575.

Michael Zürn. 2014. The politicization of world politics and its effects: Eight propositions. *European Political Science Review*, 6(01):47–71.