Part 3

Existing Resources and Benchmark Data

Outline

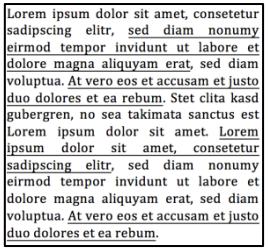
Part 3: Existing Resources and Benchmark Data

3.1 Existing Corpora

3.2 Debate Portals and Community Platforms

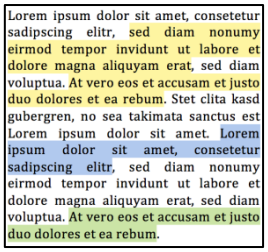
Existing Corpora

Tasks



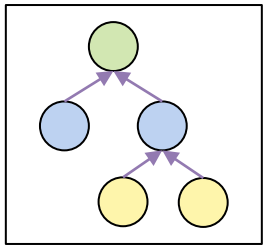
Argument component identification (CI)

- Requires presence of non-argumentative text units
- Possible on different granularities (e.g. sentence or clauses)



Argument component classification (CC)

- Commonly claim-premise-schemes
- Other component types include different types of evidence



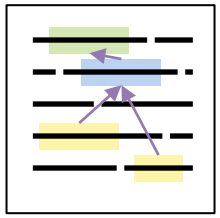
Structure Identification (SI)

- Argumentative relations between text units (e.g. t1 is-evidence-for t2)
- Note: pair classification systems can be used to identify (counter-)evidence for a given claim

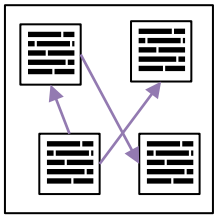
Existing Corpora

Granularity

Micro-level vs. Macro-level



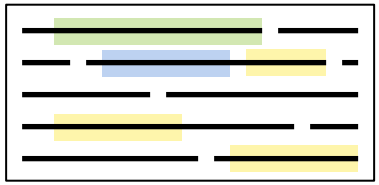
Micro-level:
internal structure of arguments



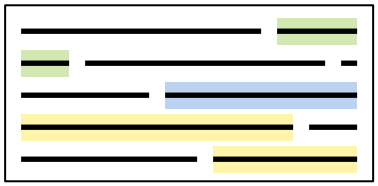
Macro-level:
connections between arguments

Granularity of argument components

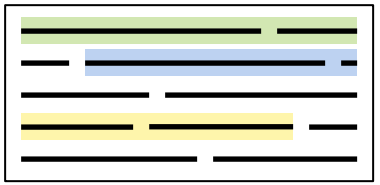
fine ⟷ coarse



clause-level components



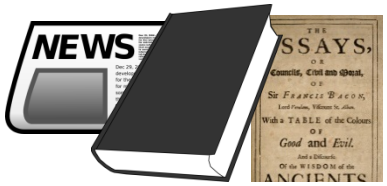
sentence-level components



multi-sentence components

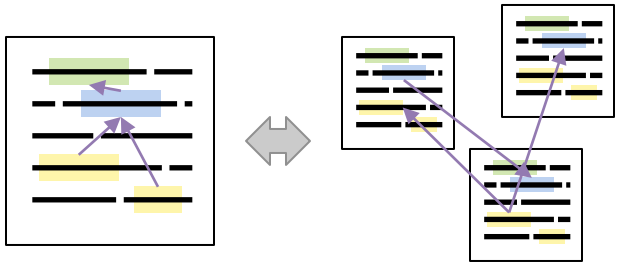
Existing Corpora

Additional Properties



document type

language



single- vs. multi-documents

κ α

reliability

Existing Corpora

Overview

Corpus	*Document types*	*Language*	*Tasks*	*Argument granularity*	*Component granularity*	*Size (#documents / #comps)*	*Single- / multi-docs*	*Reliability*	*Specials*
(Cabrio and Villata, 2014)	several	en	SI	macro	-	792 argument pairs	-	$\kappa = .71$	multiple sources including wiki revision
(Eckle-Kohler et al. 2015)	news	de	CI+CC	micro	multi	88 / ~1708	single	$\alpha_U = .40$	-
(Habernal and Gurevych., 2015)	user-generated web	en	CI+CC	micro	multi	340 / ~1k	single	$\alpha_U = .48$	several registers and topics
(Kirschner et al., 2015)	scientific articles	de	SI	micro	sentence	24 / ~2.7k	single	$\kappa = .43$	-
(Kwon et al., 2007)	online comments	en	CI+CC	micro	sentence	119 / 240	single	$\kappa = .62$	several claim types
(Mochales-Palau and Moens, 2009)	court cases	en	CI+CC	micro	sentence	47 / 1,067	single	$\kappa = .75$	-
(Peldszus and Stede, 2015)	micro texts	de / en	CC+SI	micro	clause	115 / 576	single	$\kappa = .83$	parallel corpus
(Reed et al., 2008)	various	en	CI+CC+SI	micro	clause	700 / 2k	single	-	argumentation schemes & enthymeme
(Rinott et al., 2015)	wiki articles	en	CI+CC+SI	micro	clause	547 / ~7.5k	multi	$\kappa = .39$	different types of evidence
(Stab et al., 2014)	student essays	en	CI+CC+SI	micro	clause	90 / 1.5k	single	$\alpha_U = .72$	-
(Walker et al., 2012)	online discussions	en	AT	macro	-	390k / -	-	$.22 < \kappa < .62$	several argument attributions

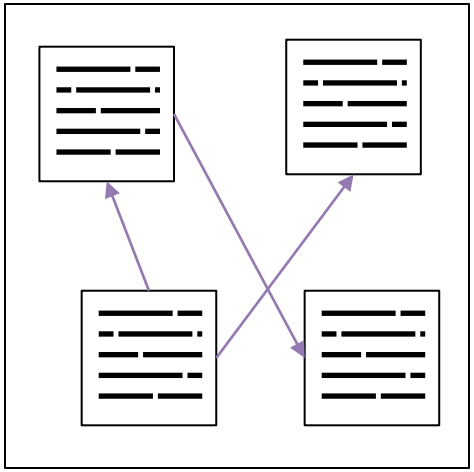
(CI = Component Identification; CC = Component Classification; SI = Structure Identification; AT = Argument Attribution)

Resources including arguments

NoDE - Natural language arguments in online Debates

Annotated macro-level relations between arguments (Cabrio and Villata, 2014)

- Three heterogeneous sources
  - Debatepedia
  - Script of a play (“12 Angry Men”)
  - Wikipedia revisions



Pairs of arguments annotated as

- Support and Attack

Inter-annotator agreement

- Between κ = .7 and κ = .74
- (depending on the source)

Resources including arguments

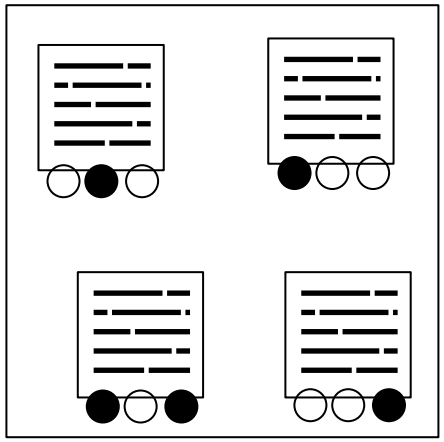
IAC - Internet Argument Corpus

Arguments from online debates (Walker et al., 2012)

- Source: 4forums.com
- Includes more than 390k posts

Annotated with several attributes

- Topic
- Degree of agreement
- Cordiality
- Audience-direction (previous poster or a wider audience)
- Combativeness
- Assertiveness
- Emotionality
- Sarcasm



Can be used for argument attribution tasks

Annotated argument components

ECHR

Micro argument components at the sentence-level

- Domain: legal (Documents from *European Court of Human Rights)*

Version 1 (Mochales-Palau and Moens, 2008)

- 10 documents (legal cases)
- Inter-annotator agreement: κ = .58

Version 2 (Mochales-Palau and Moens, 2009)

- 47 documents
- Inter-annotator agreement: κ = .75

Can be used for the following tasks

- Identification of argument components (1,067 argumentative and 1,449 non-argumentative sentences)
- Classification of argument components (304 conclusions; 763 premises)

Annotated argument components

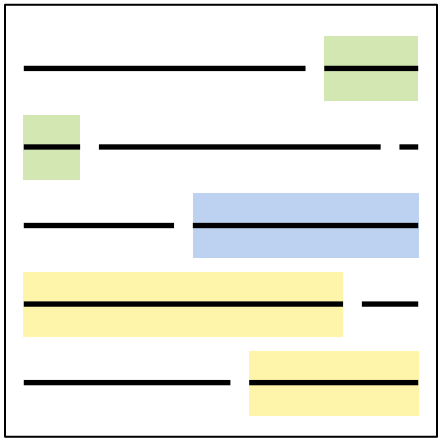
Kwon et al. (2007)

Annotation of different claim types (sentence-level)

- Data: Public comments about Environmental Protection Agency (EPA)
- 119 documents; 240 claims

Claim Types

- Supporting claim (7%)
- Opposing claim (59%)
- Proposing claim (34%)



Reliability

- Identification of claims: $\kappa = .62$
- Claim types: $\kappa = .80$

Annotated argument components

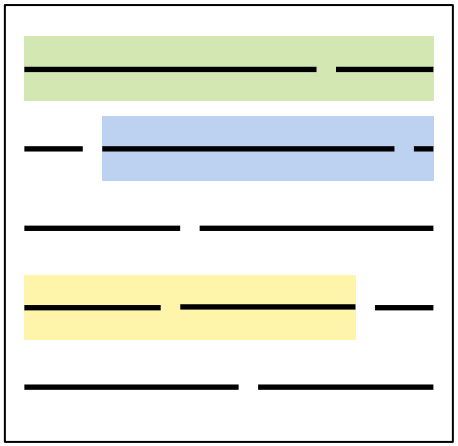
Eckle-Kohler et al. 2015

Multi-sentence argument components in news articles

- Data: focused crawl (current topics related to educational system in Germany)
- 88 documents; 3,863 sentences
- 74 % of the tokens are argumentative

Argument components

- Claim-premise scheme
- 1,708 argument components
- Supporting and attacking premises
- Target claim indicated by annotation scheme (pre, post)



Reliability

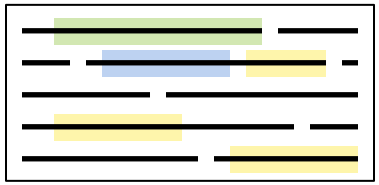
- Claim types: $\alpha_U = .402$

Annotated argument components

User-generated Web discourse

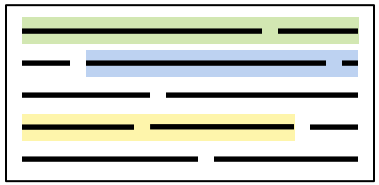
Micro argument components at multi-sentence level (Habernal & Gurevych, 2015)

- Domain: Six controversies in education
  - Redshirting; Single-sex education; Prayer in schools; Homeschooling; Mainstreaming; Public schools vs. private schools
- Four Web registers, user-generated content
  - Comments to newswire articles or blogs, discussion forum posts, blog posts, newswire editorials
- 340 documents, ~ 90,000 tokens, IAA $\alpha_U = 0.48$ (average)



Suitable for the following tasks

- Component identification
  - Free boundaries, mostly aligned to sentences
- Component classification (five types)



Specials

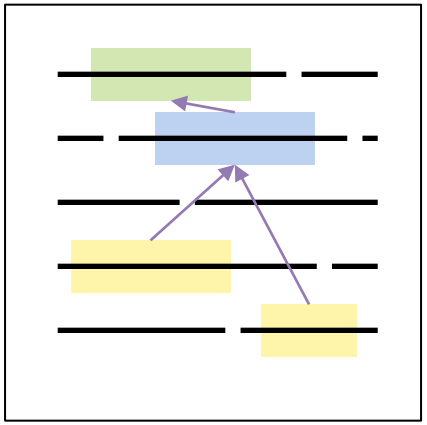
- Modified Toulmin's scheme, structure is implicitly encoded
- Includes implicit claim annotations

Micro-level argumentation structures

Araucaria

Micro argument structures at the clause level (Reed et al. 2008)

- Created using a graphical annotation tool (Reed and Rowe 2004)
- Various document types: e.g. newspaper editorials, parliamentary records, discussions, etc.
- ~700 documents / ~2,000 argument components



Suitable for the following tasks

- Component identification
  - Note that current releases do not include non-argumentative texts
- Component classification
- Argument structures

Specials

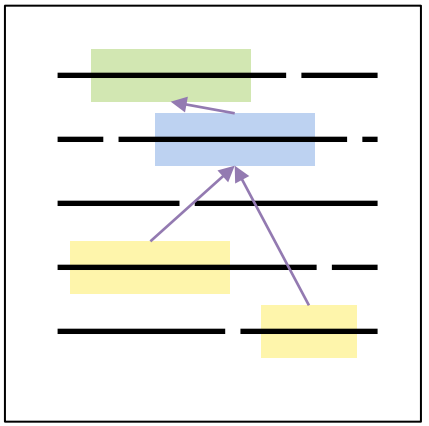
- Argumentation schemes (reasoning type)
- Includes implicit argument components (enthymeme)

Micro-level argumentation structures

Persuasive Essays

Micro argumentation structures in persuasive essays (Stab and Gurevych, 2014)

- Heterogeneous topics / prompts
- Collected from essayforum.com
- 90 persuasive essay (extended version includes 402 essays)



Usable for the following tasks

- Component identification (e.g. segmentation)
- Component classification (major claim, claim, premise)
- Structure identification (support and attack relations)

Inter-annotator agreement

- Argument components & type: $\alpha_U = .72$
- Argumentative relations $\alpha = .80$

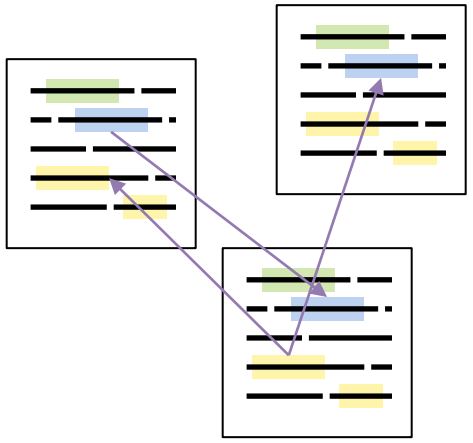
Extended version available at: https://www.ukp.tu-darmstadt.de/data/

Micro-level argumentation structures

Claims and Evidence from Wikipedia

Micro argumentation structures over multiple documents (Rinott et al. 2015; Aharoni et al. 2014)

- Argument components at the clause-level
- 58 topics from iDebate.org
- 547 Wikipedia articles
- 2,294 related claims and 4,960 associated evidences from Wikipedia
- http://www.research.ibm.com/haifa/dept/vst/mlta_data.shtml



Different types of evidence

- Study (quantitative analysis),
- Expert (testimony by a person)
- Anecdotal (specific events)

Inter-annotator agreement

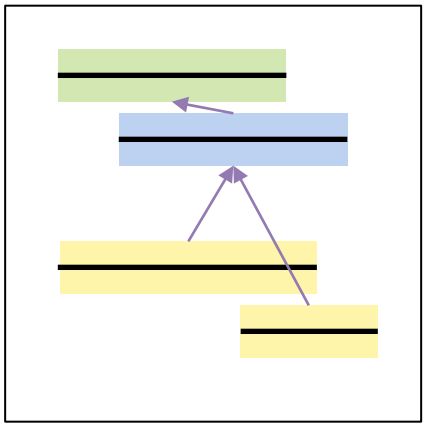
- Claims: κ = .39
- Evidence: κ = .40

Micro-level argumentation structures

Microtexts

Micro level argument structures (Peldszus and Stede, 2015)

- Controlled complexity (max of 5 argument components, no non-argumentative units)
- 112 documents (one argument per document)
- 576 argument components
- Annotation of opponent or proponent
- Several types of relations: rebuttal, support, undercut



Specials

- Originally created in German
- Professionally translated to English
- First parallel corpus in AM

Inter-annotator agreement

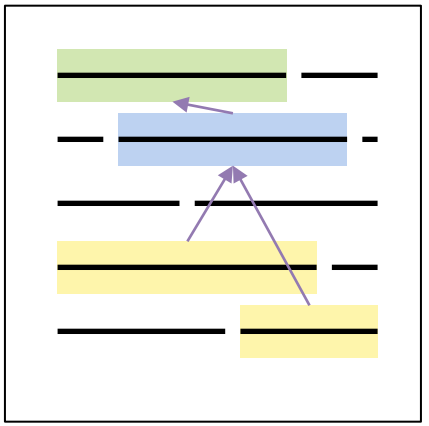
- Three expert annotators
- $\kappa = .83$

Micro-level argumentation structures

Scientific Articles

Micro argumentation structures at sentence-level (Kirschner et al., 2015)

- 24 German scientific articles
- Educational domain
- Annotations of introduction and discussion sections



Several argumentative relations

- Support
- Attack
- Detail
- Sequence

Inter-Annotator Agreement

- κ = .43 among four annotators

Existing Corpora

Overview

Corpus	*Document types*	*Language*	*Tasks*	*Argument granularity*	*Component granularity*	*Size (#documents / #comps)*	*Single- / multi-docs*	*Reliability*	*Specials*
(Cabrio and Villata, 2014)	several	en	SI	macro	-	792 argument pairs	-	$\kappa = .71$	multiple sources including wiki revision
(Eckle-Kohler et al. 2015)	news	de	CI+CC	micro	multi	88 / ~1708	single	$\alpha_U = .40$	-
(Habernal and Gurevych., 2015)	user-generated web	en	CI+CC	micro	multi	340 / ~1k	single	$\alpha_U = .48$	several registers and topics
(Kirschner et al., 2015)	scientific articles	de	SI	micro	sentence	24 / ~2.7k	single	$\kappa = .43$	-
(Kwon et al., 2007)	online comments	en	CI+CC	micro	sentence	119 / 240	single	$\kappa = .62$	several claim types
(Mochales-Palau and Moens, 2009)	court cases	en	CI+CC	micro	sentence	47 / 1,067	single	$\kappa = .75$	-
(Peldszus and Stede, 2015)	micro texts	de / en	CC+SI	micro	clause	115 / 576	single	$\kappa = .83$	parallel corpus
(Reed et al., 2008)	various	en	CI+CC+SI	micro	clause	700 / 2k	single	-	argumentation schemes & enthymeme
(Rinott et al., 2015)	wiki articles	en	CI+CC+SI	micro	clause	547 / ~7.5k	multi	$\kappa = .39$	different types of evidence
(Stab et al., 2014)	student essays	en	CI+CC+SI	micro	clause	90 / 1.5k	single	$\alpha_U = .72$	-
(Walker et al., 2012)	online discussions	en	AT	macro	-	390k / -	-	$.22 < \kappa < .62$	several argument attributions

(CI = Component Identification; CC = Component Classification; SI = Structure Identification; AT = Argument Attribution)

DKPro Argumentation

Unified type system for modeling argumentation

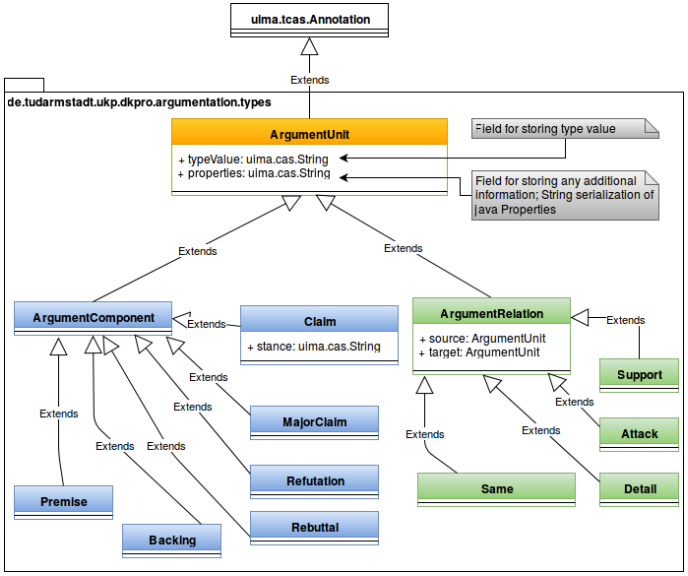
- Based on UIMA
- Easy to expand

Enables cross-resource-experiments

- Seamless integration in DKPro and DKPro-TC
- Several data sets using it

Supports all subtasks of AM

- Component identification on diff. granularities
- Component classification
- Relation identification



Available here: https://github.com/dkpro/dkpro-argumentation

Existing Resources

Summary

Existing resources are heterogeneous

- No one-fits-it-all-resource
- Different, tasks, genres, component types, granularities, etc.
- However, “claim-and-premise-scheme” seems to be common

Our taxonomy attempts to understand the differences w.r.t.:

- Argument detection tasks
- Argument granularities (micro- vs. macro-level)
- Argument component granularities (clause vs. sentence vs. multi-sentence)

➔ Facilitates the selection of existing benchmark resources

➔ Helps to define the focus of future annotation studies

References (1/2)

Aharoni, Ehud, Anatoly Polnarov, Tamar Lavee, Daniel Hershcovich, Ran Levy, Ruty Rinott, Dan Gutfreund, and Noam Slonim (2014): A benchmark dataset for automatic detection of claims and evidence in the context of controversial topics. In *Proceedings of the First Workshop on Argumentation Mining*, pages 64–68, Baltimore, MD, USA.

Eckle-Kohler, Judith, Roland Kluge, Iryna Gurevych (2015): On the Role of Discourse Markers for Discriminating Claims and Premises in Argumentative Discourse. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, pages 2236-2242, Lisbon, Portugal

Habernal, Ivan, and Iryna Gurevych (2015): Exploiting Debate Portals for Semi-Supervised Argumentation Mining in User-Generated Web Discourse. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (pp. 2127–2137). Lisbon, Portugal: Association for Computational Linguistics.

Kirschner, Christian, Judith Eckle-Kohler, and Iryna Gurevych (2015): Linking the thoughts: Analysis of argumentation structures in scientific publications. In *Proceedings of the 2nd Workshop on Argumentation Mining*, pages 1–11, Denver, CO, USA.

Kwon, Namhee, Liang Zhou, Eduard Hovy, and Stuart W. Shulman (2007): Identifying and classifying subjective claims. In *Proceedings of the 8th Annual International Conference on Digital Government Research: Bridging Disciplines & Domains*, pages 76–81, Philadelphia, PA, USA.

Meyer, Christian M., Margot Mieskes, Christian Stab, and Iryna Gurevych (2014): Dkpro agreement: An open-source java library for measuring inter-rater agreement. In *Proceedings of the 25th International Conference on Computational Linguistics: System Demonstrations (COLING)*, pages 105–109, Dublin, Ireland.

Mochales-Palau, Raquel and Marie-Francine Moens (2008): Study on the Structure of Argumentation in Case Law. In *JURIX the twenty-first annual conference on legal knowledge and information systems*, pages 11–20, Florence, Italy.

References (2/2)

Mochales-Palau, Raquel and Marie-Francine Moens (2009): Argumentation mining: The detection, classification and structure of arguments in text. In *Proceedings of the 12th International Conference on Artificial Intelligence and Law*, ICAIL '09, pages 98–107, Barcelona, Spain.

Peldszus, Andreas and Manfred Stede (2013b): Ranking the annotators: An agreement study on argumentation structure. In *Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse*, pages 196–204, Sofia, Bulgaria.

Andreas Peldszus, Manfred Stede. An annotated corpus of argumentative microtexts. To appear in: First European Conference on Argumentation: Argumentation and Reasoned Action, Portugal, Lisbon, June 2015.

Ruty Rinott, Lena Dankin, Carlos Alzate Perez, Mitesh M. Khapra, Ehud Aharoni, and Noam Slonim (2015): Show Me Your Evidence - an Automatic Method for Context Dependent Evidence Detection, In: *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, EMNLP '15, pp. 440–450, Lisbon, Portugal, 2015.

Reed, Chris and Glenn Rowe (2004): Araucaria: Software for argument analysis, diagramming and representation. *International Journal on Artificial Intelligence Tools*, 14(4):961–980.

Reed, Chris, Raquel Mochales-Palau, Glenn Rowe, and Marie-Francine Moens (2008): Language resources for studying argument. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation*, LREC '08, pages 2613–2618, Marrakech, Morocco.

Stab, Christian and Iryna Gurevych (2014): Annotating argument components and relations in persuasive essays. In *Proceedings of the 25th International Conference on Computational Linguistics (COLING 2014)*, pages 1501–1510, Dublin, Ireland, August.

Stab, Christian and Iryna Gurevych (2016): Parsing Argumentation Structures in Persuasive Essays. Under review in Computational Linguistics. arXiv preprint arXiv:1604.07370.