Towards segment-based recognition of argumentation structure in short texts

Andreas Peldszus
Supervisor: Manfred Stede

Applied Computational Linguistics, University of Potsdam

1st ACL WS on Argumentation Mining, June 26, 2014

What makes argumentation mining so hard?

- lots of text available, but only few arguments
- argumentative strategies vary across texts genres, topic, author
- understanding inferences may require very topic-specific background knowledge
- implicitness of argumentation
  - suppressed premisses
  - linguistic markedness
  - rhetorically gimmicks

Data: pro & contra commentaries

Source:

- pro & contra newspaper commentaries
- in Potsdam Commentary Corpus
  [Stede, 2004] [Stede and Neumann, 2014]

Properties:

+ lots of arguments
+ rather explicitly marked argumentation
– special background knowledge required
– main claim may be implicit
– full range of persuasive 'tricks' professional writers have to offer

DER TAGESSPIEGEL

16.04.2006 00:00 Uhr

Berlin

Hat Berlin zu viele Einkaufszentren?

Ohgottohgott! All diese Einkaufscenter, überall nur kalter Stahl und hartes Glas. Und die vielen Läden im Kiez – die haben doch keine Chance gegen diese Kommerzmonster.

So, jetzt aber Schluss mit solch jammernder Kleinstadt-Romantik. Berlin hat nicht zu viele, sondern noch lange nicht genug Einkaufszentren.

Nehmen wir die Wilmersdorfer Straße in Charlottenburg. Auch dort entsteht jetzt ein Einkaufszentrum mit vielen Stockwerken, und keinem wird das Ding schaden. Im Gegenteil. Endlich verschwinden die Ramschläden mit ihren Prepaid-Handykarten und Kochtopfdeckeln und das alte Parkhaus gleich mit. Ähnlich verhält es sich beim „Alexa" am Fernsehturm. Will denn wirklich jemand den alten Parkplatz zurück, der sich dort befand? Vermisst jemand den vermüllten Güterbahnhof, auf dem heute die „Spandau-Arcaden" stehen? Na also.

Doch nicht nur Lücken in der Stadt werden durch Einkaufszentren geschlossen, sondern auch ganze Viertel aufgewertet (wie die Gropiusstadt etwa durch die -passagen oder die Gegend am Savignyplatz durch das Stilwerk), wovon auch die kleineren Läden im Kiez wieder profitieren, weil mehr Kundschaft kommt.

Gekauft wird dort, wo es den besten Service gibt, die besten Preise. Wo es sauber ist, beleuchtet und die Wege kurz sind. Und: Wo nicht überall Hunde hinmachen. Das tun sie in Einkaufszentren jedenfalls nicht. André Görke

Data: microtexts

Source:

- 23 texts: hand-crafted, covering different arg. configurations
- 92 texts: collected in a controlled text generation experiment

Properties:

+ each segment is arg. relevant
+ explicit main claim
+ at least one possible objection considered

A (translated) example

[Energy-saving light bulbs contain a considerable amount of toxic substances.]$_1$ [A customary lamp can for instance contain up to five milligrams of quicksilver.]$_2$ [For this reason, they should be taken off the market,]$_3$ [unless they are virtually unbreakable.]$_4$ [This, however, is simply not case.]$_5$

Outline

1. Dataset Generation
2. Scheme
3. Annotation Study
4. Automatic Recognition

Outline

Generation of argumentative micro-texts: Collecting

Text generation experiment:

- 23 probands (of varying age, education and occupation)
- discuss a controversial issue (recent political, moral, everyday's life questions) in a short text
- max. 5 texts per proband

Requirements:

- length of five segments
- all segments argumentatively relevant
- at least one possible objection to be considered
- text understandable for readers without knowing the issue question

Generation of argumentative micro-texts: Collecting

Text generation experiment:

- 23 probands (of varying age, education and occupation)
- discuss a controversial issue (recent political, moral, everyday's life questions) in a short text
- max. 5 texts per proband

Requirements:

- length of five segments
- all segments argumentatively relevant
- at least one possible objection to be considered
- text understandable for readers without knowing the issue question

Generation of argumentative micro-texts: Dataset

Resulting Dataset:

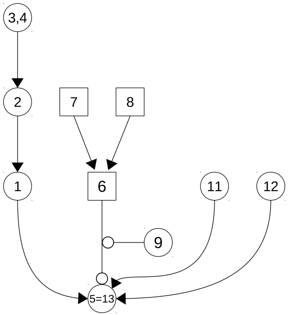
- 100 authentic texts
- 92 after cleanup
- plus 23 artificial texts

= 115 texts, 579 segments, now annotated with argumentation graphs!

Outline

1. Dataset Generation
2. Scheme
3. Annotation Study
4. Automatic Recognition

Scheme: A theory of argumentation structure

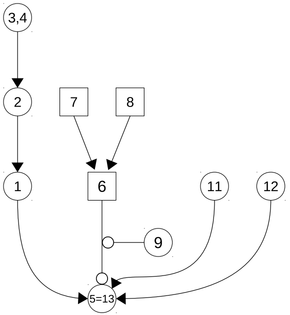


Freeman's theory, revised & slightly generalized:

[Freeman, 1991, 2011] [Peldszus and Stede, 2013b]

- node types = *argumentative role*
  proponent (presents and defends claims)
  opponent (critically questions)
- link types = *argumentative function*
  support own claims (normally, by example)
  attack other's claims (rebut, undercut)

Scheme: A theory of argumentation structure



Freeman's theory, revised & slightly generalized:

[Freeman, 1991, 2011] [Peldszus and Stede, 2013b]

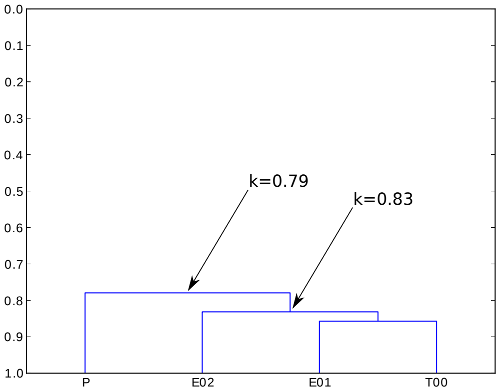
- node types = *argumentative role*
  proponent (presents and defends claims)
  opponent (critically questions)
- link types = *argumentative function*
  support own claims (normally, by example)
  attack other's claims (rebut, undercut)

Further complete annotation of authentic text:

- glue(3,4) – unitizing ADUs from EDUs
- skip(10) – arg. irrelevant segments
- join(5,13) – restatements

Outline

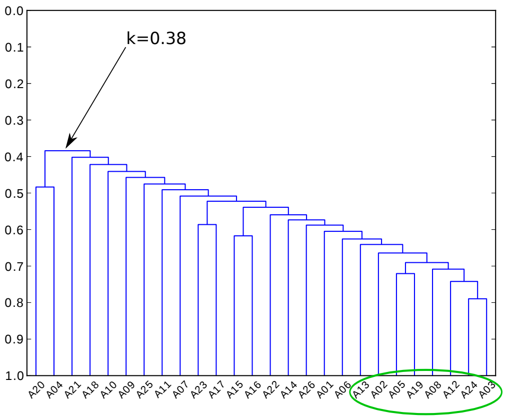
Annotation study



expert annotators: guideline authors + postdoc + student

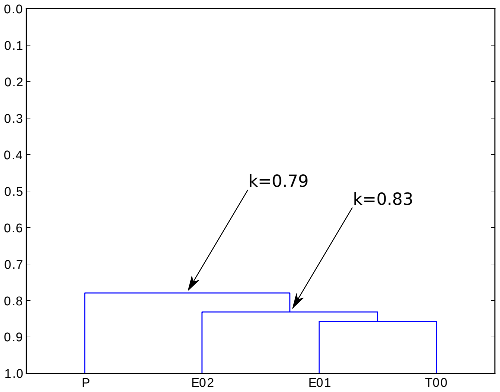
[This study]

Annotation study



naive, min. trained annotators: 26 undergrad students

[Peldszus and Stede, 2013a]



expert annotators: guideline authors + postdoc + student

[This study]

Outline

Modelling micro-texts: Segment-wise classification

prediction	text segment
	1: Energy-saving light bulbs ...
	2: A customary lamp can for ins...
	3: For this reason, they should ...
	4: unless they are virtually unbr...
	5: This, however, is simply not ...

Simple, supervised machine-learning approach, inspired by Argumentative Zoning models.

Modelling micro-texts: Segment-wise classification

prediction	text segment
PSN(n+2)	**1: Energy-saving light bulbs ...**
	2: A customary lamp can for ins...
	3: For this reason, they should ...
	4: unless they are virtually unbr...
	5: This, however, is simply not ...

Simple, supervised machine-learning approach, inspired by Argumentative Zoning models.

Modelling micro-texts: Segment-wise classification

prediction	text segment
PSN(n+2)	1: Energy-saving light bulbs ...
PSE(n-1)	**2: A customary lamp can for ...**
	3: For this reason, they should ...
	4: unless they are virtually unbr...
	5: This, however, is simply not ...

Simple, supervised machine-learning approach, inspired by Argumentative Zoning models.

Modelling micro-texts: Segment-wise classification

prediction	text segment
PSN(n+2)	1: Energy-saving light bulbs ...
PSE(n-1)	2: A customary lamp can for ins...
PT(0)	**3: For this reason, they shou...**
	4: unless they are virtually unbr...
	5: This, however, is simply not ...

Simple, supervised machine-learning approach, inspired by Argumentative Zoning models.

Modelling micro-texts: Segment-wise classification

prediction	**text segment**
PSN(n+2)	1: Energy-saving light bulbs ...
PSE(n-1)	2: A customary lamp can for ins...
PT(0)	3: For this reason, they should ...
OAU(r-3)	**4: unless they are virtually u...**
	5: This, however, is simply not ...

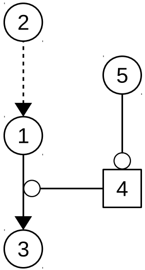
Simple, supervised machine-learning approach, inspired by Argumentative Zoning models.

Modelling micro-texts: Segment-wise classification

prediction	text segment
PSN(n+2)	1: Energy-saving light bulbs ...
PSE(n-1)	2: A customary lamp can for ins...
PT(0)	3: For this reason, they should ...
OAU(r-3)	4: unless they are virtually unbr...
PAR(n-1)	**5: This, however, is simply n...**

Simple, supervised machine-learning approach, inspired by Argumentative Zoning models.

Modelling micro-texts: Segment-wise classification



prediction	text segment
PSN(n+2)	1: Energy-saving light bulbs ...
PSE(n-1)	2: A customary lamp can for ins...
PT(0)	3: For this reason, they should ...
OAU(r-3)	4: unless they are virtually unbr...
PAR(n-1)	5: This, however, is simply not ...

Simple, supervised machine-learning approach, inspired by Argumentative Zoning models.

Modelling micro-texts: Features

- Lemma unigrams (with ± 1 window)
- Lemma bigrams
- First three lemma
- Part of speech tags (with ± 1 window)
- Main verb morphology, e.g. mood & tempus
- Dependency syntax triples, lemma-based
- Dependency syntax triples, POS-based
- Discourse markers and marked relations from DimLex [Stede, 2002] (with ± 1 window)
- Negation marker presence [Warzecha, 2013]
- Sentiment, sum of all pos. and neg. values, according to SentiWS [Remus et al., 2010]
- Segment position in text (relative)

Modelling micro-texts: Models

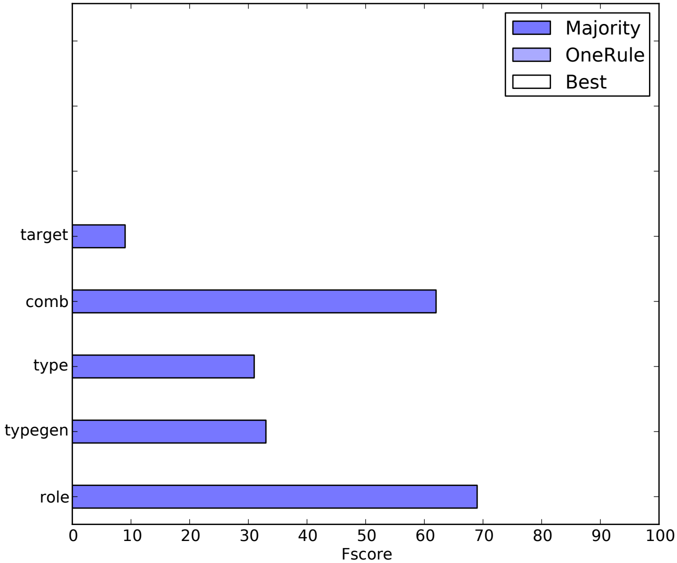
Baselines:

- **Majority**: Weka's [Hall et al., 2009] ZeroR
- **One Rule**: Weka's OneR with standard parameters

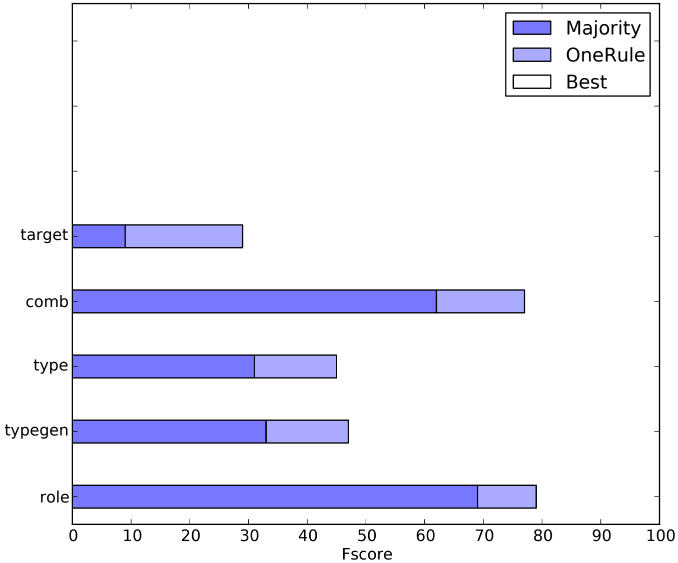
Simple Models:

- **Naïve Bayes**: Weka's Naïve Bayes, features with information gain $\ngtr 0$ are excluded
- **SVM**: Weka's wrapper to LibLinear [Fan et al., 2008] with the Crammer and Singer SVM type and standard wrapper parameters
- **MaxEnt**: MaxEnt toolkit [Zhang, 2004], 50 iterations, L-BFGS, no Gaussian prior
- **CRF**: Mallet [McCallum, 2002]. SimpleTagger interface with standard parameters

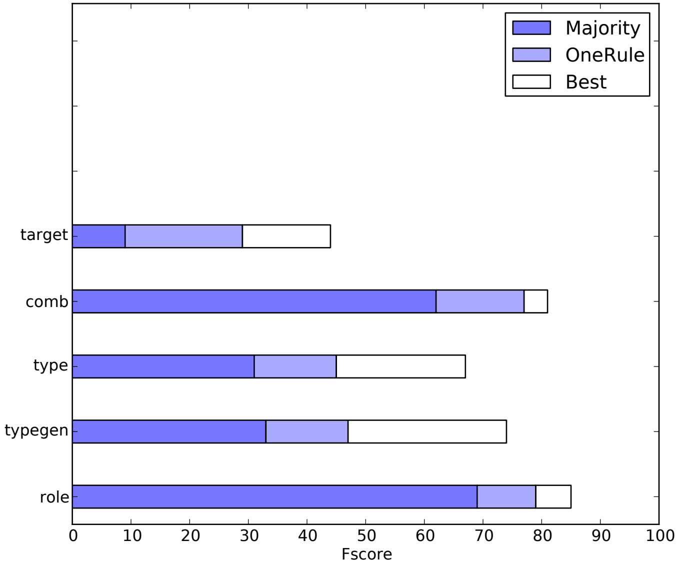
Modelling micro-texts: Results (F_1)



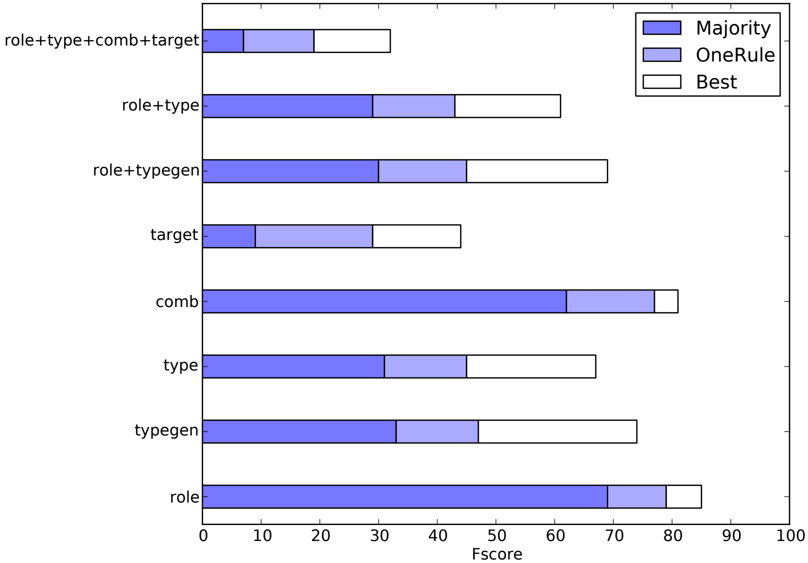
Modelling micro-texts: Results (F_1)



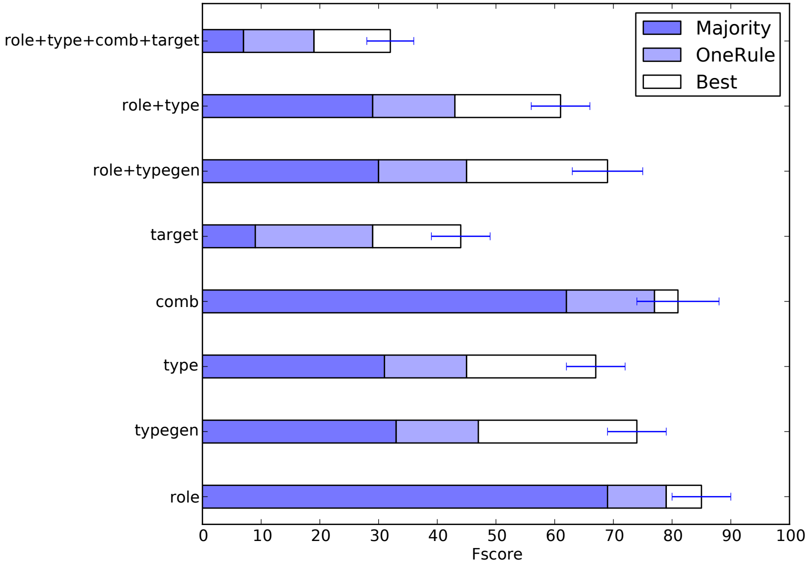
Modelling micro-texts: Results (F_1)



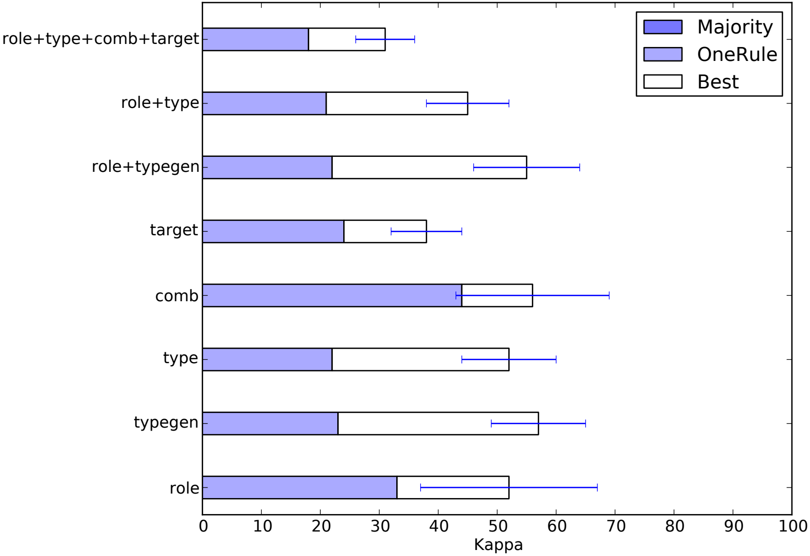
Modelling micro-texts: Results (F_1)



Modelling micro-texts: Results (F_1)



Modelling micro-texts: Results (κ)



Modelling micro-texts: Results - class wise

label	precision	recall	F1-score
PT	75 ± 12	74 ± 13	74 ± 11
PSN	65 ± 8	79 ± 7	71 ± 6
PSE	1 ± 6	1 ± 6	1 ± 6
PAR	12 ± 29	12 ± 27	11 ± 24
PAU	57 ± 26	49 ± 24	50 ± 22
OSN	1 ± 12	1 ± 12	1 ± 12
OAR	54 ± 18	42 ± 16	46 ± 13
OAU	8 ± 27	7 ± 23	7 ± 23

MaxEnt class-wise results on the ‘role+type’ level: Percent average and standard deviation in 10 repetitions of 10-fold cross-validation of Precision, Recall and F1-score.

Modelling micro-texts: Results - class wise

label	precision	recall	F1-score
PT	75±12	74±13	74±11
PSN	65±8	79±7	71±6
PSE	1±6	1±6	1±6
PAR	12±29	12±27	11±24
PAU	57±26	49±24	50±22
OSN	1±12	1±12	1±12
OAR	54±18	42±16	46±13
OAU	8±27	7±23	7±23

MaxEnt class-wise results on the ‘role+type’ level: Percent average and standard deviation in 10 repetitions of 10-fold cross-validation of Precision, Recall and F1-score.

Modelling micro-texts: Results - feature ablation

Features	SVM	
	F_1	κ
all	60±5	45±8
only lemma unigrams	55±5	37±8
only lemma bigrams	53±5	34±8
only discourse markers	53±6	34±9
only first three lemma	52±6	33±9
only dependencies lemma	47±4	27±6
only pos	45±6	24±9
only dependencies pos	41±6	18±8
only main verb morph	39±4	16±7
only segment position	25±10	4±7
only negation marker	19±8	0±4
only sentiment	15±11	-1±3

Feature ablation tests on the role+type level: Percent average and standard deviation in 10 repetitions of 10-fold cross-validation of micro averages of F1-scores, and Cohen's κ.

Outlook & Future Work

Data:

- generate more microtexts (crows-sourced text generation on the way)
- annotate more newspaper texts (new tool: GraPAT) [Sonntag and Stede, 2014]

Models:

- rerank predicted labels by graph-validity constraints (done)
- separate, individually tuned classifiers for different graph aspects
- ...

Features:

- automatic disambiguation of discourse markers
- use semantic similarity, contrastive word pairs for less-marked transitions

Outlook & Future Work

Data:

- generate more microtexts (crows-sourced text generation on the way)
- annotate more newspaper texts (new tool: GraPAT) [Sonntag and Stede, 2014]

Models:

- rerank predicted labels by graph-validity constraints (done)
- separate, individually tuned classifiers for different graph aspects
- ...

Features:

- automatic disambiguation of discourse markers
- use semantic similarity, contrastive word pairs for less-marked transitions

Outlook & Future Work

Data:

- generate more microtexts (crows-sourced text generation on the way)
- annotate more newspaper texts (new tool: GraPAT) [Sonntag and Stede, 2014]

Models:

- rerank predicted labels by graph-validity constraints (done)
- separate, individually tuned classifiers for different graph aspects
- ...

Features:

- automatic disambiguation of discourse markers
- use semantic similarity, contrastive word pairs for less-marked transitions

Thank You!

Literatur I

Rong-En Fan, Kai-Wei Chang, Cho-Jui Hsieh, Xiang-Rui Wang, and Chih-Jen Lin. Liblinear: A library for large linear classification. *J. Mach. Learn. Res.*, 9:1871–1874, June 2008. ISSN 1532-4435. URL `http://dl.acm.org/citation.cfm?id=1390681.1442794`.

James B. Freeman. *Dialectics and the Macrostructure of Argument*. Foris, Berlin, 1991.

James B. Freeman. *Argument Structure: Representation and Theory*. Argumentation Library (18). Springer, 2011.

Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H. Witten. The weka data mining software: An update. *SIGKDD Explor. Newsl.*, 11(1):10–18, November 2009. ISSN 1931-0145.

Andrew Kachites McCallum. Mallet: A machine learning for language toolkit. http://mallet.cs.umass.edu, 2002.

Andreas Peldszus and Manfred Stede. 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, August 2013a. Association for Computational Linguistics.

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

Robert Remus, Uwe Quasthoff, and Gerhard Heyer. SentiWS - A Publicly Available German-language Resource for Sentiment Analysis. In Nicoletta Calzolari (Conference Chair), Khalid Choukri, Bente Maegaard, Joseph Mariani, Jan Odijk, Stelios Piperidis, Mike Rosner, and Daniel Tapias, editors, *Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC'10)*, pages 1168–1171, Valletta, Malta, May 2010. European Language Resources Association (ELRA).

Jonathan Sonntag and Manfred Stede. GraPAT: a Tool for Graph Annotations. In *Proc. of the International Conference on Language Resources and Evaluation (LREC)*, Reykjavik, 2014.

Manfred Stede. DiMLex: A Lexical Approach to Discourse Markers. In Vittorio Di Tomaso Alessandro Lenci, editor, *Exploring the Lexicon - Theory and Computation*. Edizioni dell'Orso, Alessandria, Italy, 2002.

Manfred Stede. The potsdam commentary corpus. In *Proceedings of the 2004 ACL Workshop on Discourse Annotation*, DiscAnnotation '04, pages 96–102, Stroudsburg, PA, USA, 2004. Association for Computational Linguistics.

Literatur II

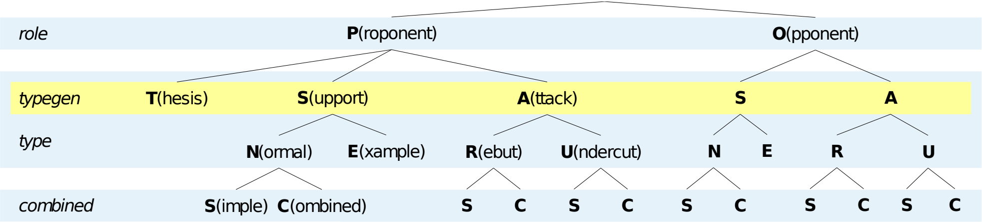
Manfred Stede and Arne Neumann. Potsdam commentary corpus 2.0: Annotation for discourse research. In *Proc. of the International Conference on Language Resources and Evaluation (LREC)*, Reykjavik, 2014.

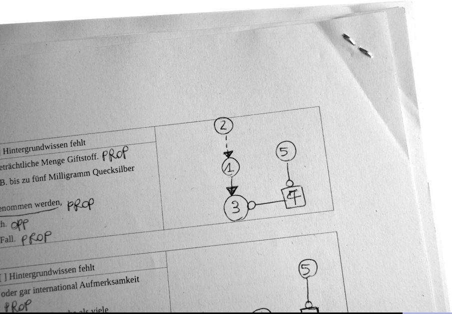
Saskia Warzecha. Klassifizierung und Skopusbestimmung deutscher Negationsoperatoren. Bachelor thesis, Potsdam University, 2013.

Le Zhang. *Maximum Entropy Modeling Toolkit for Python and C++*, December 2004.

Annotation study

Rewrite graphs as a list of (relational) segment labels






```
1:PSNS(3)
2:PSES(1)
3:PT()
4:OARS(3)
5:PARS(4)
```

Modelling micro-texts: Annotated corpus

level	role	typegen	type	comb	target	role+type
labels	P (454) O (125)	T (115) S (286) A (178)	T (115) SN (277) SE (9) AR (112) AU (66)	/ (115) S (426) C (38)	n-4 (26) n-3 (52) n-2 (58) n-1 (137) 0 (115) n+1 (53) n+2 (35) r-1 (54) r-2 (7) ...	PT (115) PSN (265) PSE (9) PAR (12) PAU (53) OSN (12) OSE (0) OAR (100) OAU (13)
# of lbls	2	3	5	3	16	9

Label distribution on the basic levels and for illustration on the complex ‘role+type’ level.

Generation of argumentative micro-texts: Topics

Top 5 chosen topics:

- Should the fine for leaving dog excrements on sideways be increased?
- Should shopping malls generally be allowed to open on Sundays?
- Should Germany introduce the death penalty?
- Should public health insurance cover treatments in complementary and alternative medicine?
- Should only those viewers pay a TV licence fee who actually want to watch programs offered by public broadcasters?