

Argumentation is key: a keyword-based study of arguments in online discourse

Natalie Dykes, Stefan Evert, Joachim Peters, Philipp Heinrich

FAU Erlangen-Nürnberg

www.linguistik.fau.de





Argument mining

- Automatic extraction and representation of arguments from texts
 - Identify structural components (premise, conclusion)
 - Map textual patterns to logical representation
- Classic argumentation schemes:
 - Modus Ponens (X → Y, X \vdash Y)
 - Modus Tollens ($X \rightarrow Y$, ¬Y ⊢ ¬X)





Related work

Challenges in everyday language:

- Implicit premises or conclusions (Bosc et al. 2016)
- "Defeasible" argumentation (see Walton et al. 2008)
 - Expert opinion
 - Ad hominem
 - Common folks ad populum
- Persuasion through rhetorical strategies (selection, arrangement, phrasing of argumentative units rather than strict logical implication (Wachsmuth et al. 2018)
- Non-standard language, especially on social media (Goudas et al. 2014)





PHILOSOPHISCHE FAKULTÄT

Hypotheses

- Traditional logical representation is not sufficient for capturing everyday argumentation
- Content-specific indicators can serve as proxies to less-structured arguments
- Combining grammatical templates with content-specific words
 will help to bridge the gap between logics and authentic language

Examples from Brexit tweets (Common folks ad populum):

- to stay , because <the average person doesn't need to be left in the hands of the brexit leaders> !! Are ppl really
- @MyronChristodou @vote_leave <ordinary folk will do worst from #Brexit> - except perhaps t
- @DrAlanGreene <I'm as against #Brexit as the next man> but this is nonsense





Related corpus approaches

- Degano (2007): starts from a predefined list of explicit markers for structural argumentative elements (cf. Levinson, 1983)
- O'Halloran (2011): manual coding of claims and challenges; keyness (words, POS, domains) in coded sequences (WMatrix; cf. Rayson, 2008)
- Content-centred approaches based on keyness: argumentation as one of various usage contexts (Partington, 2003; Baker, 2004; Al-Hejin, 2015)





PHILOSOPHISCHE FAKULTÄT

Data

- German web corpus on multidrug-resistant organisms (MDRO), clinical hygiene, antibiotics-induced diseases (collected with Bootcat, Baroni & Bernardini, 2004)
- State-of-the-art tools for tagging and lemmatisation (Proisl & Uhrig, 2016; Schmid, 1995; Schmid, Fitschen, & Heid, 2004)
- Manual annotation of text-level metadata: (actor group of author and intended readership, topic)

Sub-corpora for present study:

- Mass media articles (1.1k texts; 1.3M tokens)
- Online sources relating to alternative medicine (432 texts; 926k tokens)
- National, international and regional institutions disseminating information to the public (417 texts, 575k tokens)

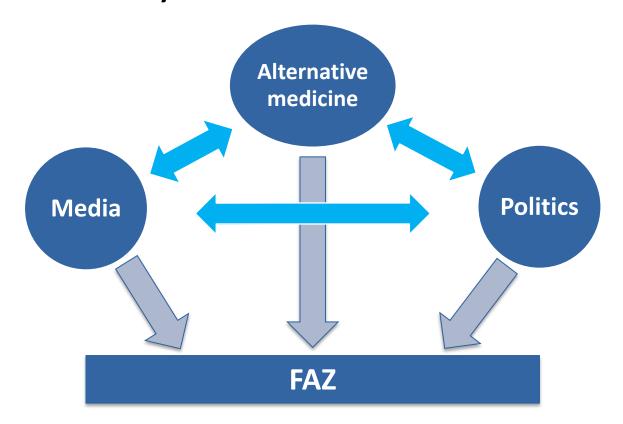
Reference corpus: 3 years of the widespread newspaper *Frankfurter Allgemeine Zeitung* (FAZ – 341k texts, 177.9M tokens)





PHILOSOPHISCHE FAKULTÄT

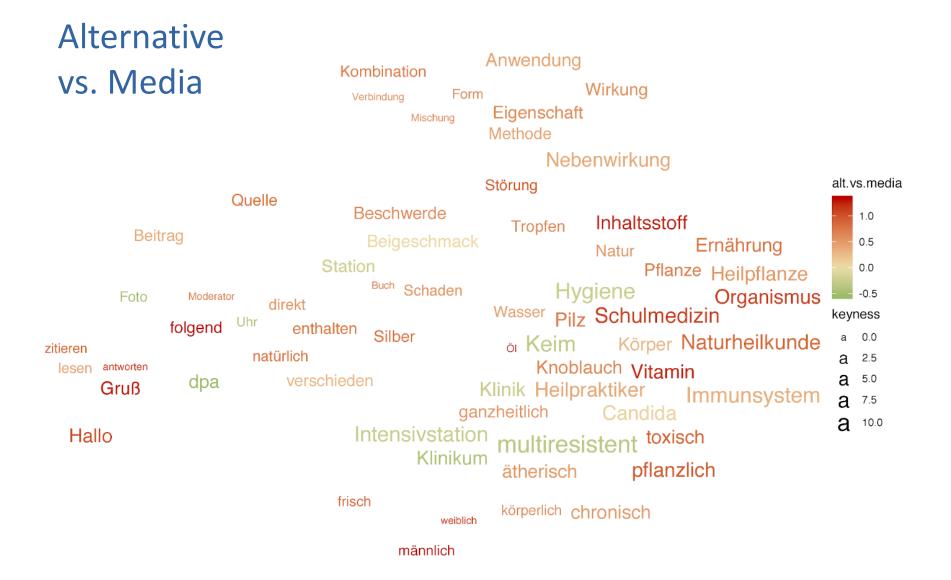
Keyword analysis



Keyness measure: **LRC** = conservative version of the effect-size based log ratio (Hardie, 2014), taking the lower end of a Bonferroni-adjusted 99% confidence interval (cf. Evert, Dykes, & Peters, 2018)



Making sense of keyness – visualisation





Making sense of keyness – visualisation

- Sorting via semantic similarity helps to identify false positives: greetings (Hallo 'hello') and artefacts from boilerplates on the websites (Beitrag 'post', zitieren 'cite') cluster towards the left
- Other clusters indicate topic complexes/ possible connection to discourse strategies, i.e. words relating to application of medical products (Wirkung 'effect', Mischung 'mixture', Anwendung 'application')
- Media articles: hospitals and multidrug resistance; focus on circumstances of contracting clinical infections (Hygiene 'hygiene', Intensivstation 'intensive care unit')



Making sense of keyness – annotation

Keyword annotation: text-linguistic gold standard – metaphor, lexis, topoi (Peters 2017, Dykes 2018)

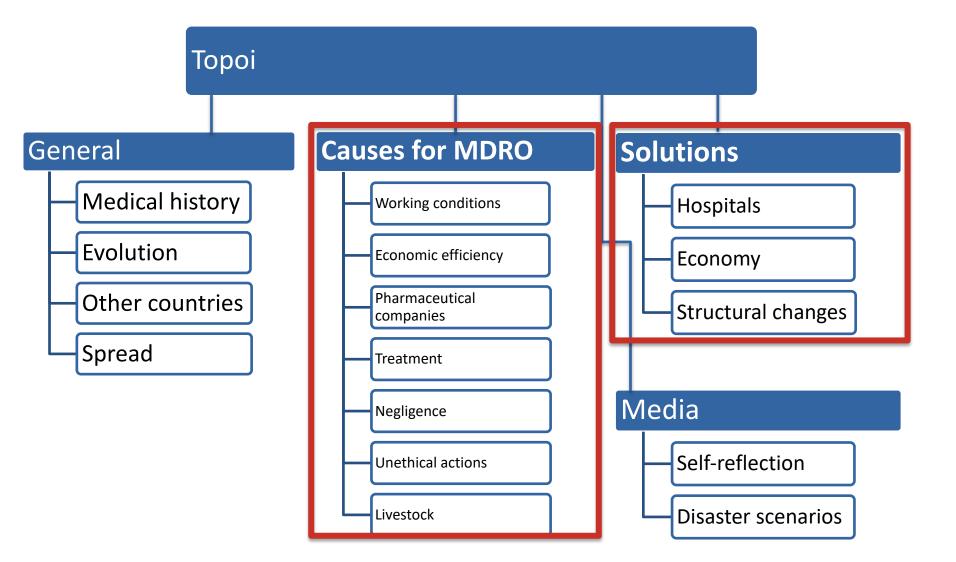
Topoi related to 2 argumentation schemes:

- argument from effect to cause (Walton, Reed, & Macagno, 2008, p. 172)
 - Major premise: Generally, if A occurs, then B will (might) occur.
 - Minor premise: In this case, B did in fact occur.
 - Conclusion: Therefore, in this case, A also presumably occurred.
- argument from positive consequences (Walton et al., 2008, p. 101)
 - Major premise: If A is brought about, then good consequences will occur.
 - Conclusion: Therefore, A should be brought about.



UND FACHBEREICH THEOLOGIE

Annotation scheme





UND FACHBEREICH THEOLOGIE

Making sense of keyness – annotation

Comparison	Actors	Effect – cause	Positive consequences	Other topoi	Not useful
Media / alternative	29%	12%	3%	12%	44%
Media/ politics	27%	10%	1%	8%	47%
Media / FAZ	38%	9%	6%	14%	33%
Alternative / media	7%	7%	18%	13%	55%
Alternative / FAZ	21%	9%	15%	15%	40%
Alternative/ politics	8%	6%	4%	19%	63%
Politics/alternative	32%	16%	5%	15%	32%
Politics/ media	19%	4%	9%	6%	56%
Politics/ FAZ	28%	12%	9%	15%	36%





ERLANGEN-NÜRNBERG

PHILOSOPHISCHE FAKULTÄT

UND FACHBEREICH THEOLOGIE

Actors across sub-corpora

Actors in MEDIA subcorpus

Frühchen-Station	Darmbakterie				
Krankenhauspatier	Superkeim				
•		Killerkeim	Pseudomonade		
Klinikpatient	Krankenhauserreger	MRSA-Keim			

Frühchen Intensivstation Hygieneexperte Frühgeborene Krankenhaushygieniker Hygienefachkraft Hygienefachkraft Hygienefachkraft Hygienefachkraft Hygienefach	Frühchenstation		STANNVIOKOKKAN	LRC	C.mrsa
Hygieneexperte Frühgeborene Krankenhaushygieniker Hygienefachkraft Hygieniker Landesgesundheitsamt Robert-Koch-Institut RKI RKI RKI RKI RCDC KH MRSA Darmkeim Infektionserreger Klebsiellen mikrobiologisch Enterokokken Enterokokken pathogen pathog		chen	Krankenhauskeim Krankheitserreger Mikrobe Bakterie Dakteriell		0
Frühgeborene Krankenhaushygieniker Hygienefachkraft Hygieniker Landesgesundheitsamt Robert-Koch-Institut WHO Koch-Institut Darmkeim infektionserfeger Klebsiellen mikrobiologisch Enterokokken Erreger Streptokokken Bakterienstamm MRSA aeruginosa Bakterium Klebsiella Enterobacter Clostridium Enterobacter Clostridium Enterobacter Clostridium MRSA aeruginosa Pseudomonas a hospital a med difficile aureus Salmonella EHEC Coli RKI CDC KH MRE MRGN MRE MRGN			Enterobakterien Asiastakasta	а	3
Krankenhaushygieniker Hygienefachkraft Hygieniker Landesgesundheitsamt Robert-Koch-Institut WHO Koch-Institut RKI CDC KH MRSA ESBL-Keim Erreger Streptokokken pathogen MRSA aeruginosa Bakterium Klebsiella al 12 Methicillin-resistente Bakterienstamm MRSA aeruginosa Pseudomonas pathogen baumannii Campylobacter Escherichia a med difficile aureus coli Salmonella EHEC Coli ESBL MRSA Aeruginosa Pseudomonas coli a pathogen a polit a pathogen a polit a pathogen a polit a polit BRKI CDC KH MRE MRGN			Darmkeim Intektionserreger Klebsiellen grampositiv	а	6
Hygieniker Hygieniker Landesgesundheitsamt Bremen-Mitte Bremen-Mitte Methicillin-resistente Bakterienstamm Clostridium MRSA aeruginosa Descudomonas a hospital baumannii Campylobacter Escherichia a med difficile aureus coli Salmonella WHO Koch-Institut Enterobacter Clostridium Clostridium Clostridium Compylobacter Escherichia a med a pat a pat pathogen EHEC Coli RKI CDC KH MRE MRGN MRGN MRE MRGN	Kanalia aharrahi ada di sa	rrungeborene	mikrobiologisch Enterokokken pathogen	а	9
Hygieniker Landesgesundheitsamt Bremen-Mitte Bremen-Mitte Gastmeier Robert-Koch-Institut Gastmeier WHO Koch-Institut CDC KH MRSA aeruginosa Descudomonas a hospital baumannii Campylobacter Escherichia a med difficile aureus coli Salmonella EHEC Coli ESBL MRGN ENTEROBACTETISTATII Clostridium Pseudomonas a hospital a pat a pat a pathogen a polit a science MRE MRGN	Krankennausnygleniker	Hygienefachkraft	ESBL-Keim Erreger Streptokokken Bakterium Klebsiella	а	12
Robert-Koch-Institut Gastmeier WHO Koch-Institut CDC KH RKI CDC KH KRSA aeruginosa baumannii Campylobacter Escherichia a med a pat a pathogen EHEC Coli ESBL Pseudomonas a hospital a med a pat a pathogen a polit a science	Hygien	iker	Clostridium Enterobacter		
Robert-Koch-Institut Gastmeier MHO Koch-Institut Koch-Institut RKI CDC KH KH KH KH KH KH KH KRGN KR	Landesgesundheitsamt	Bremen-Mitte	MRSA aeruginosa Pseudomonas	a	hospital
Robert-Koch-Institut WHO Koch-Institut EHEC Coli RKI CDC KH MRE MRGN ECDC	Landoogoodiidiiotodiiit		Eschenchia	a	med
WHO Koch-Institut EHEC Coli a pathogen a polit a science ESBL MRE MRGN ECDC	Gastmeier Robert-Koch-Institut		amono	a	pat
RKI CDC ESBL a science RKI KH MRE MRGN ECDC			Samonoia	a	pathogen
RKI CDC ESBL a science KH MRE MRGN ECDC	WHO	Koch-Institut	EHEC Coli	a	polit
KH MRE MRGN ECDC		C	ESBL ESBL	a	science
ECDC		HKI	MDE		
BVL KPC		ECDC	WINGN		
		BVL	KPC		





UNIVERSITÄT
ERLANGEN-NÜRNBERG
PHILOSOPHISCHE FAKULTÄT

UND FACHBEREICH THEOLOGIE

Actors across sub-corpora

Actors in POLITICS subcorpus

Frühchen-Station Darmbakterie

Krankenhauspatient

Krankenhauserreger Superkeim

MRSA-Keim

Krankenhaushygieniker H	rühchen Frühgeborene lygienefachkraft ieniker sundheitsamt Robert-Koch-Institut	Krankenha	Darmkeim mikrobiologi	ionserreger krank	St machend Enterokokk Bal Strepto te aeruginos A	aphylokokken Ken Enterob Kterienstamm Dkokken Bakteriuf Sa Clo Campylobacter aureus	aphylococcus grampositiv	а а а	12
WH	10			EHEC	ESBL			a	polit science
	Koch-Institut							а	science
	RKI	EFSA	MRE	MRGN					
ECDC									
	BVL								
	BfR								





UNIVERSITAT
ERLANGEN-NÜRNBERG
PHILOSOPHISCHE FAKULTÄT

UND FACHBEREICH THEOLOGIE

Actors across sub-corpora

Actors in ALTERNATIVE subcorpus

Krankenhauspatient

Krankenhauserreger

Krankenhauserreger

Krankenhauserreger

Krankenhauserreger

Krankenhauserreger

Krankenhauserreger

Krankenhauserreger

			Clostridien	Staphylokokken	acto	r
	Kranker	nhauskeim Krankheits	serreger Mikrobe	Bakterie bakteriell	а	hospital
Intensivstation		Keim Infektions		erEnterobakte Staphylococcus	a	med
		Darmkeim	krankmachend	grampositi Acinetobacter	a	pat
	m	nikrobiologisch ESBL-Keir	_m Bakterienst	pathogen	a	pathogen
Krankenhaushygieniker		Echinacea	Erreger Stre	eptokokken Bakterium Klebsiella	a	polit
Hygieniker		Enteroko	Woth Commit Tooloton to	Candida Haemophilus	a	science
,,		LINGIOKO	MRSA	ClostridiumPseudomonas	LRC	C.alt
			baumannii	Campylobacter Escherichia		0
Gesundheitsdienst Robert-Koch-Ins	stitut		difficile aure	US COII Salmonella	а	3
					а	6
WHO Koch-	nstitut		ESBL	Coli	а	9
	CDC	EHEC	LOBE		а	12
RKI	KH					
		MRE MI	RGN			
ECE	C	KPC				
BfR						



Keywords for Argumentation Mining

- Linguistic patterns, but not directly tied to word level
 - In aggregation, sizeable number of indicators for both schemes per sub-corpus
 - Corpus queries to bridge the gap between lexis and logical content
- Corpus-linguistic approach: CQP query language (Evert & Hardie 2011)
 - Phrase/ clause structure patterns defined by POS sequences
 - Word lists representing lexico-semantic categories (keywords)
 - Argument indicators from thesaurus (Dornseiff 2004)
 - Iterative development informed by regular concordance analysis



Query: Arg. from Positive Consequences

```
@1:[lemma = $solution_nouns
    lemma = $solution verbs
  word = "beste[rnms]?/ideal.*/perfekt.*"]+
  /np[] | /vp[] | /ac[] | /advp[] |
  [ word = "," | pos="KO.+|P.+|AP.+|ART"]
)*
@0:[lemma = $alt_solns]+
  /np[] | /vp[] | /advp[] | /ac[]
  [ word = "," | pos="KO.+|P.+|AP.+"]
)*;
```

Wordlists for semantic grouping





HILOSOPHISCHE FAKULTAT ND FACHBEREICH THEOLOGI

Positive consequence keywords per sub-corpus/ wordlist filler in the example query (unique results; altmed+media; pol+media):

```
[altmed]
Ätherisch ,aetheric' (28); Teebaumöl ,tea tree oil'(16); Silber ,silver'
(15); Probiotik/ Senföl/ Vitamin , probiotics/ mustard oil/ vitamin (9);
Bockshornklee ,fenugreek (8); Thymian ,thyme (7); Homöopathie/ Öl
,homeopathy/ oil' (6) ...
[media]
Bakteriophage ,bacteriophage', Krankenhaushygiene ,clinical hygiene',
desinfizieren ,disinfect' (10); Hygienemaßnahme ,hygiene measure', Phagen
,phages' (9); Desinfektion (7) ...
[pol]
Leitlinie ,guideline', antimikrobiell ,anti-microbial' (25);
```

```
Leitlinie ,guideline', antimikrobiell ,anti-microbial' (25);
Niedersachsen ,Lower Saxony' (16); signifikant ,significant' (15);
Umsetzung ,implementation' (14); Hygienemaßnahme ,hygiene measure' (13);
Intervention (12); Antibiotika-Resistenzstrategie ,antibiotic resistance strategy', Krankenhaushygiene ,clinical hygiene', regional (8);
Anforderung ,requirement', Minderung ,reduction', Therapietreue ,compliance' (5)
```

Examples





PHILOSOPHISCHE FAKULTÄT UND FACHBEREICH THEOLOGI

[altmed]

Da er zu diesem Zeitpunkt in einem homöopathischen Krankenhaus tätig war , kam er irgendwann auf die <**Idee**, diese Bakterienspuren **homöopathisch** zu verabreichen> mit sensationellen Heilerfolgen.

Because he was working in a homeopathic hospital at the time, he eventually came up with the **<idea** of administering these bacteria traces **homeopathically>** – with sensational success'

[media]

Schon seit Jahren <zeigen Holland und Dänemark wie man durch ein konsequentes Screening aller neuer Patienten und besonderer Behandlung der erkannten "Risikopatienten"> , den Anteil mit antibiotikaresistenten Keimen drastisch reduzieren kann

Holland and Denmark have been **<showing** for years how to drastically reduce the amount of multi-resistant organisms by consistent **screening** of all new patients and special treatment of "high risk patients">

[pol]

Die Publikation zeigt zum ersten Mal über einen mehrjährigen Zeitraum, dass im Vergleich zu anderen Regionen , die <**Umsetzung** einer **Präventionsstrategie** in allen Krankenhäusern einer Versorgungsregion> die absolute Anzahl von MRSA-Infektionen signifikant innerhalb von 2 Jahren senkt und niedrig halten kann.

The publication is the first one to show over a period of several years that the **<implementation** of a **prevention strategy** in all hospitals of a region**>** can reduce and control the absolute number of MRSA infections within 2 years



Conclusion

- Considerable overlap of solution keywords in media articles and politics
 - Media: overall KW candidates more on hospital level screenings, clinical hygiene, intensive care unit
 - Politics: stronger focus on infrastructure guideline, regional, resistance strategy
- Alternative medicine more focused on "unique" solutions barely reflected in the media/ political sub-corpora
- Logical perspective: very similar representations





Conclusion

Future work:

- Modularity enables transfer of logical/ grammatical templates to other corpora. Queries as templates to be "filled" with content-specific wordlists
- Consultation with professionals in the field (co-operation with department of Palliative Medicine at Erlangen's University Hospital)

Queries balance grammatical and semantic flexibility in patterns Each query: one linguistic instantiation of a given argumentation scheme

Combination of query pattern and content-specific lexical features to capture everyday argumentation

Qualitatively informed approach to handle variable and noisy data





References

Al-Hejin, Bandar (2015): Covering Muslim women. Semantic macrostructures in BBC News. In Discourse & Communication 9 (1), pp. 19–46. DOI: 10.1177/1750481314555262.

Baker, Paul (2004): 'Unnatural Acts'. Discourses of homosexuality within the House of Lords debates on gay male law reform. In Journal of Sociolinguistics 8 (1), pp. 88–106. DOI: 10.1111/j.1467-9841.2004.00252.x.

Baroni, Marco; Bernardini, Silvia (2004): BootCaT. Bootstrapping corpora and terms from the web. In Maria Teresa Lino, Maria Francisca Xavier, Fatima Ferreira, Rute Costa, Raquel Silva (Eds.): Proceedings of the IVth International Conference on Language Resources and Evaluation (LREC). Paris: ELRA (International Conference on Language Resources and Evaluation, 4), pp. 1313–1316. Available online at http://clic.cimec.unitn.it/marco/publications/lrec2004/bootcat_lrec_2004.pdf, checked on 5/6/2017.

Bosc, Tom Elena Cabrio, Serena Villata. Tweeties Squabbling: Positive and Negative Results in Applying Argument Mining on Social Media. 2016. Proceedings of the 6th International Conference on Computational Models of Argument.

Degano, Chiara (2007): Presupposition and Dissociation in Discourse. A corpus study. In Argumentation (21), pp. 361–378. DOI: 10.1007/s10503-007-9058-7.

Evert, Stefan & Andrew Hardie. 2011. Twenty-first century corpus workbench: Updating a query architecture for the new millennium. In *Proceedings of CL*.

Evert, Stefan; Dykes, Natalie; Peters, Joachim (2018): A quantitative evaluation of keyword measures for corpus-based discourse analysis. Lancaster.





References

Goudas, Theodosis, Christos Louizos, Georgios Petasis & Vangelis Karkaletsis. 2014. Argument Extraction from News, Blogs, and Social Media. In *SETN 2014*.

Hardie, Andrew (2012): CQPweb - combining power, flexibility and usability in a corpus analysis tool. In International Journal of Corpus Linguistics 17 (3), pp. 380–409. DOI: 10.1075/ijcl.17.3.04har.

Hardie, Andrew (2014): A single statistical technique for keywords, lockwords, and collocations. Internal CASS working paper no. 1, version 1.5 April 2014.

Levinson, Stephen (1983): Pragmatics. Cambridge: Cambridge University Press (Cambridge textbooks in linguistics, 8). O'Halloran, Keiran (2011): Investigating argumentation in reading groups. Combining manual qualitative coding and automated corpus analysis tools. In Applied Linguistics 32 (1), pp. 172–196. DOI: 10.1093/applin/amq041.

Owoputi, Olutobi, Brendan O'Connor, Chris Dyer, Kevin Gimpel, Nathan Schneider & Noah Smith. 2013. Improved Part-of-Speech Tagging for Online Conversational Text with Word Clusters. In *Proceedings of NAACL*.

Partington, Alan (2003): The linguistics of political argument: The spin-doctor and the wolf-pack at the White House. London: Routledge.

Proisl, Thomas; Uhrig, Peter (2016): SoMaJo: State-of-the-art tokenization for German web and social media texts. In: Proceedings of the 10th Web as Corpus Workshop.

Rayson, Paul (2008): From key words to key semantic domains. In International Journal of Corpus Linguistics 13 (4), pp. 519–549. DOI: 10.1075/ijcl.13.4.06ray.





References

Schmid, Helmut (1995): Treetagger |. A language independent part-of-speech tagger. In Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart 43.

Schmid, Helmut; Fitschen, Arne; Heid, Ulrich (2004): SMOR. A German computational morphology covering derivation, composition, and inflection. In Maria Teresa Lino, Maria Francisca Xavier, Fatima Ferreira, Rute Costa, Raquel Silva (Eds.): Proceedings of the IVth International Conference on Language Resources and Evaluation (LREC). Paris: ELRA (International Conference on Language Resources and Evaluation, 4).

Wachsmuth, Stede, El Baff, Al-Khatib, Skeppstedt, Stein. 2018. Argumentation Synthesis following Rhetorical Strategies. In Proceedings of the 27th International Conference on Computational Linguistics. 3753–3765.

Walton, Douglas, Chris Reed & Fabrizio Macagno. 2008. *Argumentation Schemes*. Cambridge: Cambridge University Press.

Appendix





PHILOSOPHISCHE FAKULTÄT UND FACHBEREICH THEOLOGIE

Given corpus

Reference corpus

Calculation of keywords through different measures

Group keywords using linguistic categories

Explore the various levels of granularity yielded by the different measures

Examine the categories through a sample of concordances

Goal: identification and manual deepening of discourse patterns in topic-specific corpora





Syntactic macros

```
IMPORT macro advp.txt
## A determiner phrase
MACRO dp(0)
[pos = "ART|PIS|PPOSAT"]?
  /advp[]
[pos = "N."] +
## A prepositional phrase
MACRO pp (0)
[pos = "APPR.*"]
/dp[]
# pronoun or pronoun + verb
MACRO pron(0)
```