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 ($X \rightarrow Y, X \vdash Y$)
  - Modus Tollens ($X \rightarrow Y, \neg Y \vdash \neg X$)
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)
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)

- Content-centred approaches based on keyness: argumentation as one of various usage contexts (Partington, 2003; Baker, 2004; Al-Hejin, 2015)
Data

- German web corpus on multidrug-resistant organisms (MDRO), clinical hygiene, antibiotics-induced diseases (collected with Bootcat, Baroni & Bernardini, 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)
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

Alternative vs. Media
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.
Annotation scheme

Topoi

General
- Medical history
- Evolution
- Other countries
- Spread

Causes for MDRO
- Working conditions
- Economic efficiency
- Pharmaceutical companies
- Treatment
- Negligence
- Unethical actions
- Livestock

Solutions
- Hospitals
- Economy
- Structural changes

Media
- Self-reflection
- Disaster scenarios
### Making sense of keyness – annotation

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

Actors in MEDIA subcorpus

- Frühchen-Station
- Krankenhauspatient
- Klinikpatient
- Krankenhausärzte
- Darmbakterie
- Superkeim
- Killerkeim
- Pseudomonade
- MRSA-Keim
- Frühchenstation
- Intensivstation
- Hygieneexperte
- Frühgeborene
- Krankenhaushygieniker
- Hygienefachkraft
- Hygieniker
- Landesgesundheitsamt
- Robert-Koch-Institut
- WHO
- Koch-Institut
- RKI
- CDC
- KH
- ECDC
- BVL
- BfR
- LRC.mrsa
  - 0
  - a 3
  - a 6
  - a 9
  - a 12
  - actor
    - hospital
    - med
    - pat
    - pathogen
    - polit
    - science
Actors across sub-corpora

Actors in POLITICS subcorpus
- Frühchen-Station
- Krankenhauspatient
- Krankenhausarbeiter
- Darmbakterie
- Superkeim
- MRSA-Keim
- Intensivstation
- Frühgeborene
- Hygienefachkraft
- Hygieniker
- Landesgesundheitsamt
- Gesundheitsdienst
- Robert-Koch-Institut
- RNDT
- WHO
- Koch-Institut
- RKI
- ECDC
- BVL
- BfR
- LRC.polit
  - a 3
  - a 6
  - a 9
  - a 12
- actor
  - a hospital
  - a med
  - a pat
  - a pathogen
  - a polit
  - a science
Actors across sub-corpora

Actors in ALTERNATIVE subcorpus

Krankenhauspatient
Krankenhauserreger
Superkeim
Killerkeim
MRSA-Keim
Chlamydia
Darmbakterie

Krankenhauskeim
Krankheitserreger
Infektionserreger
Keim
Darmkeim
Klebsiellen
mikrobiologisch

Clostridien
Staphylokokken
Bakterienstamm
Bakteriell
Enterobakter
Staphylococcus
Bakterium
Klebsiella
grampositiv
Acinetobacter
pathogen

Candida
Haemophilus
Clostridium
Pseudomonas
Campylobacter
Escherichia
difficile
coli
Salmonella
Coli

Robert-Koch-Institut
WHO
Koch-Institut
RKI
KH
ECDC
BfR
CDC
EHEC
ESBL
MRE
MRGN
KPC

actor
hospital
a
poli
pat
a
pathogen
a
science

LRC.alt
0
a
3
a
6
a
9
a
12
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

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); 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

[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


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


Appendix

Given corpus

Calculation of keywords through different measures

Reference corpus

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
A glimpse at the /np[] macro

Syntactic macros

```plaintext
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)
{
```

A glimpse at the /np[] macro