

Quantitative Approaches to Language: A Study on Controversial Concepts in Climate Change Discourse

Maria Becker & Johanna Meyer

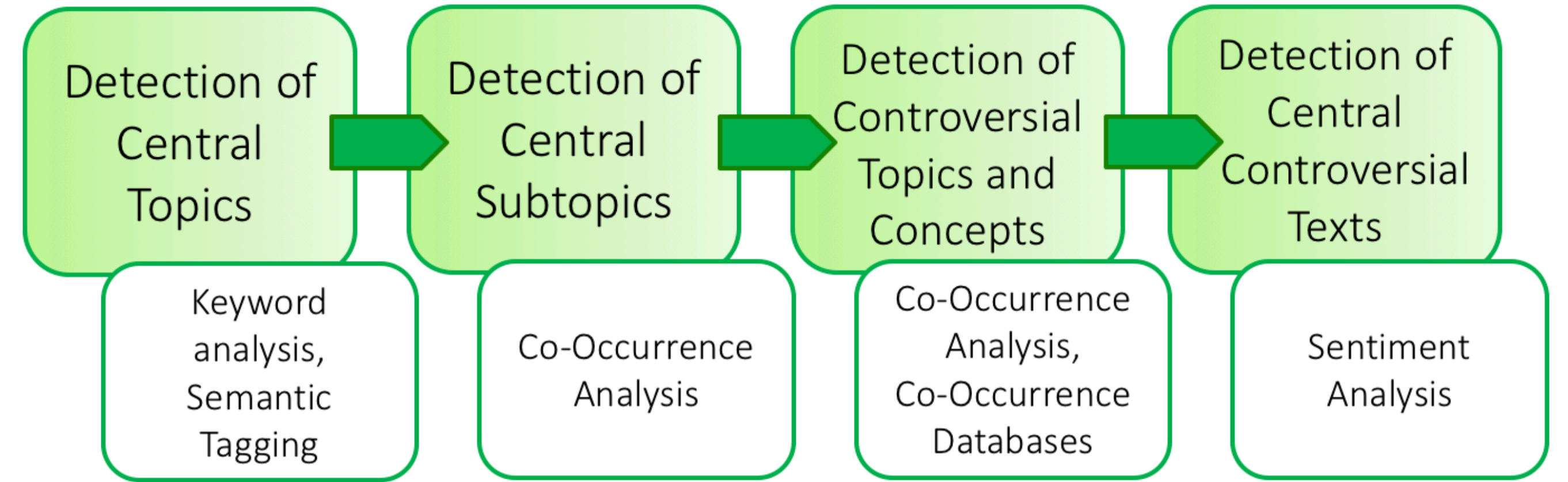
Discourse Lab

LIMO
Empirical Linguistics
and Computational Language Modeling

UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386

How to Detect Controversial Concepts within Corpora Using Quantitative Methods?

- We investigate a **corpus** of 72.000 German newspaper articles on **climate change** in order to find **controversial** topics and concepts
- Approach:**
 - Quantitative methods (keyword analysis, semantic tagging...)
 - Four steps with varying granularity
- Contributions:**
 - insights into central controversial topics, actors, foci, points of view and styles of discussion
 - method for detecting controversial concepts using quantitative methods → applicable to other discourses



STEP 1: Detecting Central Topics

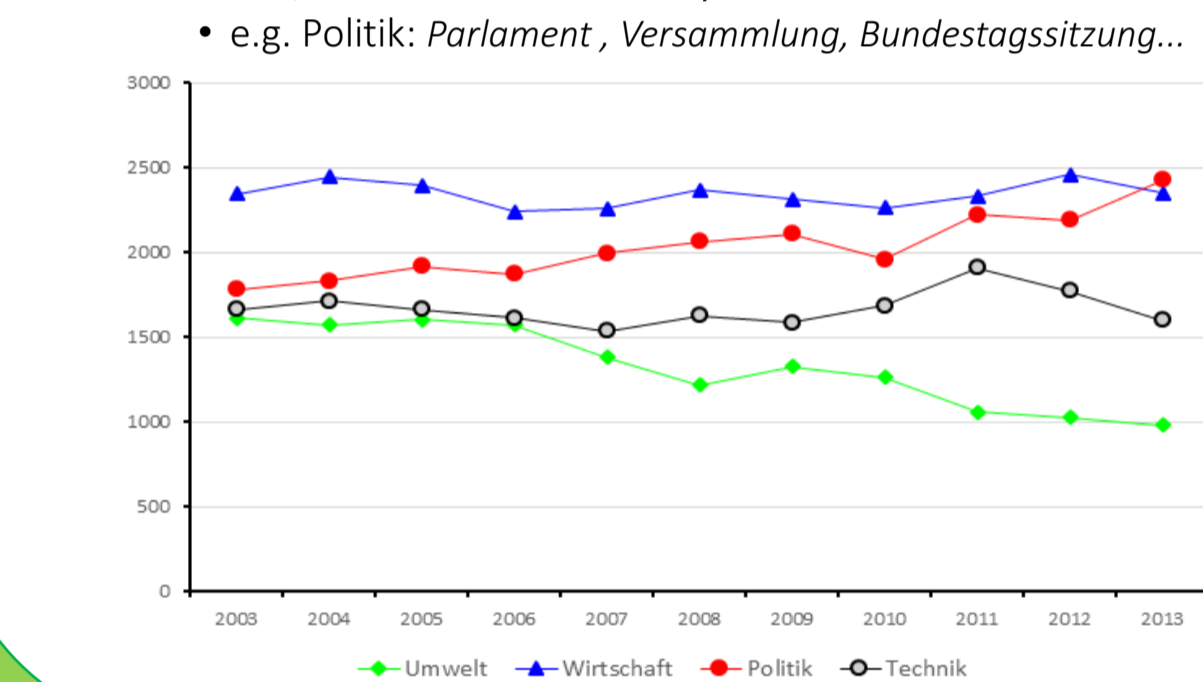
A) Keyword Analysis

- Keywords:** words whose frequency is exceptionally high in comparison with a reference corpus (Xiao/McEnery 2005)
- Reference corpus:** Smultron (subcorpus, various genres)
- Tool:** AntConc (toolkit for concordancing and text analysis)

- Keywords:**
Klimaschutz, Klimawandel, Energiewende, Deutschland, Emissionen, Forscher, Regierung, Technik, Erwärmung, Ausstoß, Energie...

B) Topic Progression Analysis

- Semantic tagging** with thematic word lists: When and how often do words related to specific topics occur?
- Dornseiff (1933):** Der deutsche Wortschatz nach Sachgruppen
- 21 thematic groups** (Religion, Science, Politics, Environment...)



STEP 2: Detecting Central Subtopics

Co-Occurrence Analysis

- Co-occurrence analysis:** statistical methods which allow the detection of significant regularities of word combinations within a corpus (Keibel/Belica 2007)
- Co-occurring words** occur more often together than can be expected if all words were distributed by chance
- Tool:** AntConc
- Co-occurrence analysis of the keywords detected in Step 1 reveal **central subtopics:**

- Klimaschutz:** *Treibgase, Umweltbehörde, EU, Deutschland, EU, Welt...*
- Klimawandel:** *Katastrophe, Kohlendioxid, Unternehmen, Menschen, Finanzen...*
- Energiewende:** *Konfliktthema, lahmen, leistungsorientiert, Belastung, Ausstieg, Debatte...*
- Deutschland:** *Übergang, überstanden, kränkeln, Zuversichtswerte, zusammenbrechen...*
- Emissionen:** *Ökoflächen, Wärmeaktivität, vermeidbar, Werkschließungen, verweigern...*

STEP 3: Detecting Controversial Topics & Concepts

Agonal Connectors & Prepositions

- Hypothesis:** Controversial topics can be found within the context of words which mark **controversy/agonality** (Felder 2013)
- Controversy marking words:
 - Concessive Connectors:** *obgleich, wünschon, obwohl, dennoch, trotzdem, wobei...*
 - Adversative Connectors:** *aber, jedoch, doch, wiederum, dagegen, allerdings...*
 - Substitutive Connectors:** *anstatt, vielmehr, anstelle dessen, sondern...*
 - Prepositions:** *wider, trotz, entgegen, statt, anstelle...*

Lexicon of Agonal Expressions

- Construction of a **lexicon of agonal expressions** using CCDB (Corpus-Linguistic Research & Development Workbench, Belica 2001ff.)
- CCDB provides information about words with similar collocation profiles extracted from the German Reference Corpus
- Words are similar if their **collocation profiles** are similar (Perkuhn et al. 2012)
- **Hypothesis:** words with similar collocation profiles as **agonal** words are also considered to be **agonal**
- Lexicon:** 907 entries based on 30 agonal search expressions

Discussion	Debatte	Debatte	Debatte	Debatte	Debatte
... Diskussion Debatte Debatte Debatte Debatte Debatte ...

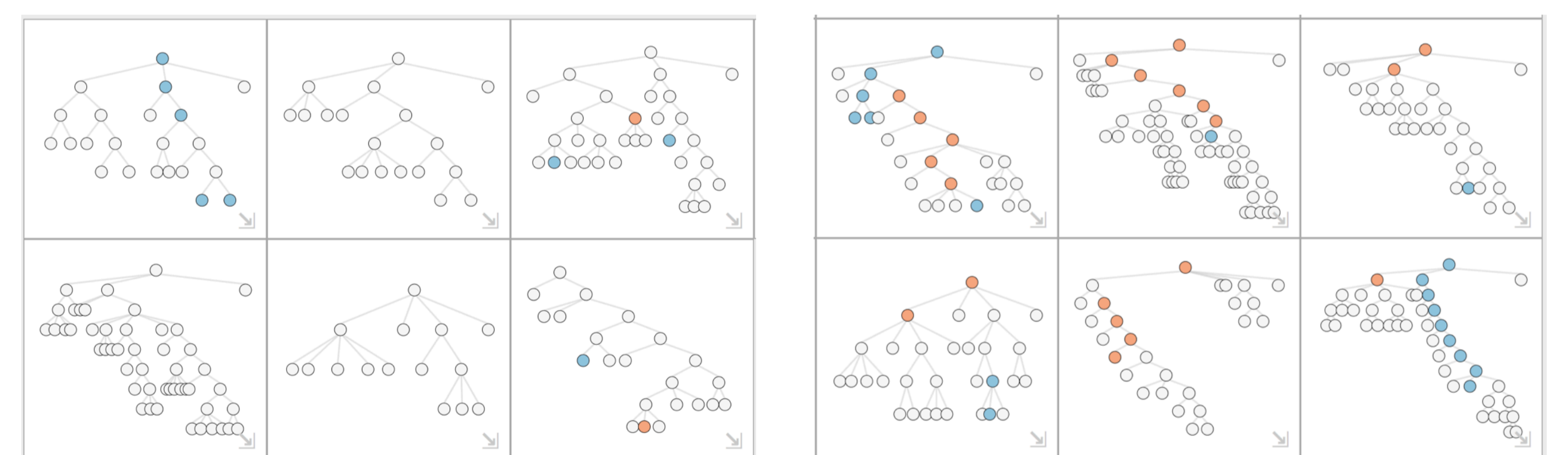
Words with similar collocation profiles as Diskussion

Co-Occurrence Analysis

- Tool:** CQPweb (web-based corpus analysis system)
- Set-wise **input** of agonal connectors, prepositions and entries of the agonal lexicon
- Analysis of the **Output**

STEP 4: Detecting Central Controversial Texts

- Hypothesis:** Controversial Topics & Concepts are more frequent within emotional than neutral texts
- Tool:** Stanford Sentiment Treebank
 - predicts the **sentiment** of phrases and sentences
 - distinguishes between **5 classes**, very negative (blue) to very positive (red): --, -, 0, +, ++
- Sentiment Analysis of a **neutral vs. emotional** text:



- Approach:**
 - extract** emotional texts with the help of Sentiment Analysis (texts which include many phrases/sentences labelled as (very) positive and/or (very) negative)
 - analyze** those texts in more detail
- Analysis reveals **more detailed information, e.g.** about
 - arguing actors** (*Obama, Europa, Regierung, Merkel, Bundesinnenminister, Seehofer, Rösler, Gabriel, CDU, Grüne...*)
 - controversial methods** (*Emissionskontrollen, Fracking, Gaspreise, Atomausstieg, Klimaschutzgesetz, Energiewende, Gesetzesbeschluss...*)

Results: Controversial Topics & Concepts within the Climate Change Discourse

- (I) **Actors**
- Politics** (*Gemeinderat, Bürgermeister, Regierung, SPD-Politiker, Ministerpräsident, Fraktionschef, Merkel, Gabriel, Obama...*) vs. **Science/Experts** (*Klimaforscher, Wissenschaftler, Energieexperte, Meteorologe, Experten, Forscher, Professor...*) vs. **Companies** (*Projektleiter, Unternehmenssprecher, Vorstandschef, Abteilungsleiter...*) vs. **Initiatives** (*Anti-Atom-Initiative, Umweltverbände, Umweltschützer, Greenpeace, Verbände, Naturschutzbund, Umweltorganisation, Umweltverbände...*)
- (II) **Foci and Points of View**
- Solution** (*Lösung, Zukunft, Bewältigung, angegangen, Ergebnisse, Entwurf, Vorschlag, Umsetzung, Möglichkeit, überwinden...*) vs. **Problem** (*monatelang, unlösbar, ungelöst, gravierend, fundamental, Schranken, verschärft, schwerste, erheblich, schwerwiegend...*)
 - Global** (*global, kategorisch, öffentlich, Afrika, weltweit, systematisch, gesellschaftlich...*) vs. **local** (*intern, innenpolitisch, Deutschland, ideologisch, Einzelfall, ich, er, wir, sie...*)
 - Causes** (*Klimawandel, Emissionen, Umweltveränderungen, Erderwärmung, Klimakatastrophe...*) vs. **Goals** (*Atomausstieg, Pläne, Klimaschutzziele, Vorschlag, Zukunftsprogramm, CO2-Grenzwerte, Klimaschutzkonzept, Gesetzentwurf, Maßnahmen...*)
- (III) **Styles of Discussion (Metalinguistic Level)**
- Consensus** (*Wille, Entscheidung, Beschluss, Abkommen, Gesetzentwurf, entscheiden, akzeptieren...*) vs. **Dissent** (*Auseinandersetzung, Widersacher, Widerstand, Gegenwind, Konkurrenzkampf, Machtkampf, Tauziehen, Kontroverse, Streit, kämpfen, streiten...*)
 - Objective** (*Debatte, Diskussion, Kontroverse, friedlich, Vorschläge, diskutieren, sachlich, konstruktiv, ergebnisoffen*) vs. **emotional** (*emotional, heftig, erbittert, hitzig, leidenschaftlich, entbrannt, Sprechchöre, Polizei, skandieren, Ärger, Angriffe, provoziert, stören...*)

References

- Xiao, Richard/McEnery, Tony (2005): Two approaches to genre analysis: three genres in modern American English. In: Journal of English Linguistics (33/1.), p. 62-82.
 Dornseiff, Franz (1933): Der deutsche Wortschatz, synonymisch geordnet. Berlin: de Gruyter.
 Belica, Cyril (2001ff.): Kookkurrenzdatenbank CCDB. Eine korpuslinguistische Denk- und Experimentierplattform für die Erforschung und theoretische Begründung von systemisch-strukturellen Eigenschaften von Kohäsionsrelationen zwischen den Konstituenten des Sprachgebrauchs. IDS, Mannheim.
 Keibel, Holger/Belica, Cyril (2007): CCDB: A Corpus-Linguistic Research and Development Workbench. Proceedings of the 4th Corpus Linguistics conference, Birmingham.
 Felder, Ekkehard (2013): Faktizitätsherstellung mittels handlungsleitender Konzepte und agonaler Zentren. Der diskursive Wettkampf um Geltungsansprüche. In: Felder, Ekkehard (Ed.): Faktizitätsherstellung in Diskursen. Die Macht des Deklarativen. Berlin/Boston: de Gruyter, p. 13-28
 Perkuhn, Rainer/Keibel, Holger/Kupietz, Marc (2012): Korpuslinguistik. Paderborn: Fink, 2012.