

Verbindung von manueller Inhaltsanalyse mit automatisierten Techniken zur Erkennung von Argumentationsstrukturen

Matthias Liebeck und Katharina Esau

Heinrich-Heine-Universität Düsseldorf

Berlin, 10.10.2015



Fachhochschule
für öffentliche Verwaltung
NRW



NRW-FORTSCHRITTSKOLLEG
ONLINE-PARTIZIPATION



HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

Einleitung

- Im Rahmen des Fortschrittskollegs beschäftigen wir uns unter anderem mit Online-Partizipationsverfahren.
- Der Aufwand einer manuellen Analyse kann sehr hoch sein, vor allem bei mehreren Tausend Textbeiträgen.
- Interessante Analyseergebnisse können sein:
 - Worüber reden die Beteiligten? Gibt es Handlungsvorschläge oder Entscheidungsoptionen?
 - Wie werden Vorschläge begründet?
 - Gibt es Emotionsäußerungen, die den Diskurs beeinflussen?
⇒ Moderation in einem laufenden Verfahren möglich
- Ziel: (Politische) Entscheidungsträger sollen über die diskutierten Veränderungen entscheiden können.
- Für solch eine Analyse werden zeitliche und personelle Ressourcen benötigt, für die Geld vorhanden sein sollte.



Interdisziplinäre Zusammenarbeit

- Das NRW Fortschrittskolleg Online-Partizipation bietet die Möglichkeit zur interdisziplinären Forschung.
- In unserem Fall: Kombination von Herangehensweisen und Erfahrung aus der Kommunikations- und Medienwissenschaft mit der Informatik
- Forschungsbereiche:
 - Argumentationsstrukturen
 - Kommunikationsmodi
 - Emotionsäußerungen
- Unser aktueller Fokus liegt auf Argumentationsstrukturen.

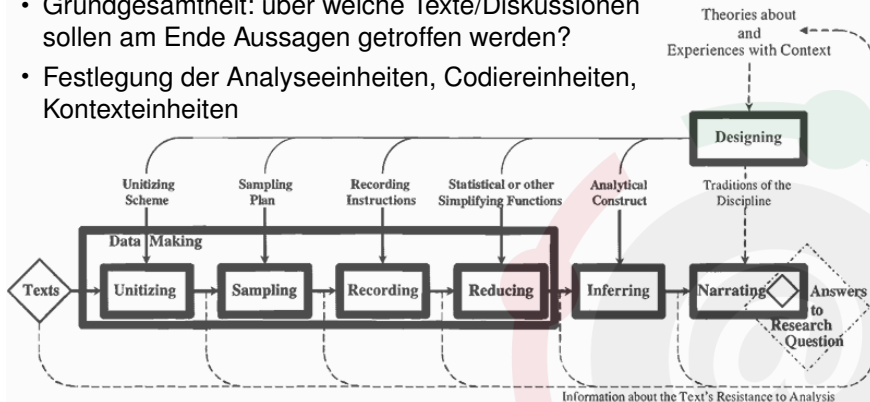
Beispiel: *Ich mag dein Auto, weil es blau ist.*

- Aussage: Ich mag dein Auto.
- Begründung: weil es blau ist.



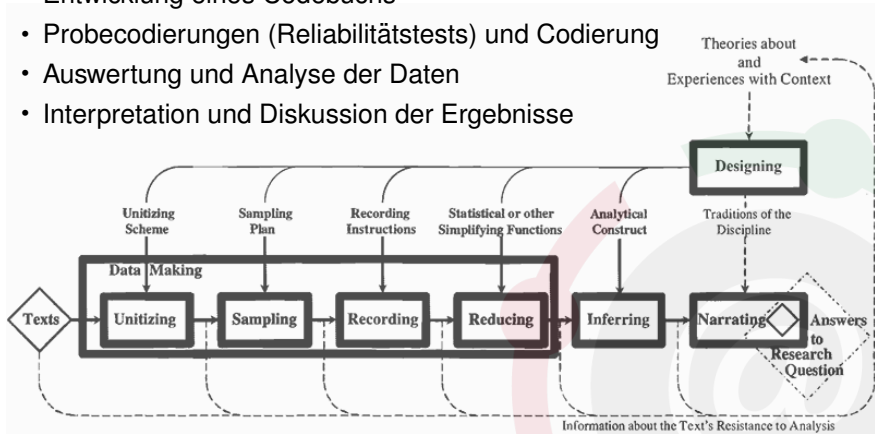
Quantitative Inhaltsanalyse in der Kommunikationswissenschaft

- Ausgangspunkt: Forschungsfrage
- Auswahl der Fälle, die untersucht werden sollen
- Grundgesamtheit: über welche Texte/Diskussionen sollen am Ende Aussagen getroffen werden?
- Festlegung der Analyseeinheiten, Codiereinheiten, Kontexteinheiten



Quantitative Inhaltsanalyse in der Kommunikationswissenschaft

- Kategorienbildung und Operationalisierung
- Entwicklung eines Codebuchs
- Probecodierungen (Reliabilitätstests) und Codierung
- Auswertung und Analyse der Daten
- Interpretation und Diskussion der Ergebnisse



Wahl eines Argumentationsmodells

- Die Frage, wie Argumentationen in Texten modelliert werden können, ist nicht neu. Im Bereich der Informatik ist in den letzten Jahren der Forschungsbereich *Argument Mining* entstanden.
- Im Bereich des Argument Minings gibt es kein einheitliches Argumentationsmodell, auf das sich die Forscher einigen können.
- Die Wahl des Modells hängt von der Textart (z.B. Nachrichtenartikel, persuasive essays etc.) ab.
- Zitat: *There is no one-size-fits-all argumentation theory to be applied to realistic data on the Web.*¹

¹Habernal et al., Argumentation Mining on the Web from Information Seeking Perspective, 2014

Wahl eines Argumentationsmodells

- Wir haben uns bestehende Argumentationsmodelle angeschaut. Dazu gehören:
 - Toulmin² (claim, grounds, warrant, backing, rebuttal/reservation, qualification)
 - claim-premise family
- Fragestellung: Ist eins der bestehenden Modelle für eine automatisierte Analyse aus unserer Sicht geeignet?
- Ziel: Idealerweise ein Argumentationsmodell finden, dass projektübergreifend anwendbar ist.
 - Dabei fokussieren wir uns momentan auf Texte, in denen pro und contra Positionen relativ eindeutig bestimmt werden können.
 - In offenen Diskussionstexten ist diese Bestimmung eher schwierig.
- Toulmin: Toulmins Modell ist für argumentativ gut strukturierte Texte anwendbar. Es ist schwer auf kurze Webinhalte zu übertragen.

²Stephen Toulmin, The Uses of Argument, 1958

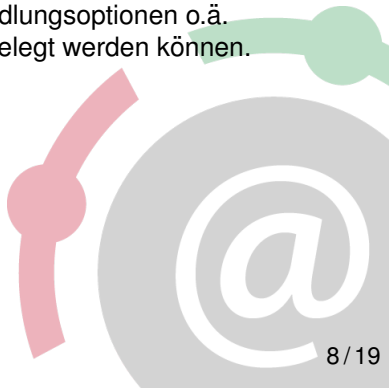
Wahl eines Argumentationsmodells

- Wir haben uns die claim-premise family genauer angeschaut:³
 - **claim**: Zentrale Komponente eines Arguments, dass nicht ohne Begründung akzeptiert werden sollte
 - **premise**: Begründung, die den Leser vom claim überzeugt (oder vom Gegenteil überzeugt)
- Wir sehen die Formulierung „*nicht ohne Begründung akzeptiert werden sollte*“ als zu subjektiv an, um sie praktisch anzuwenden, da fast jeder Satz mit „Warum?“ hinterfragt werden kann.
- Wir folgen der Empfehlung von Habernal et al. 2014: *selection of argumentation model should be based on the data at hand and the desired application*

³Stab & Gurevych, Identifying Argumentative Discourse Structures in Persuasive Essays, 2014

Wahl eines Argumentationsmodells

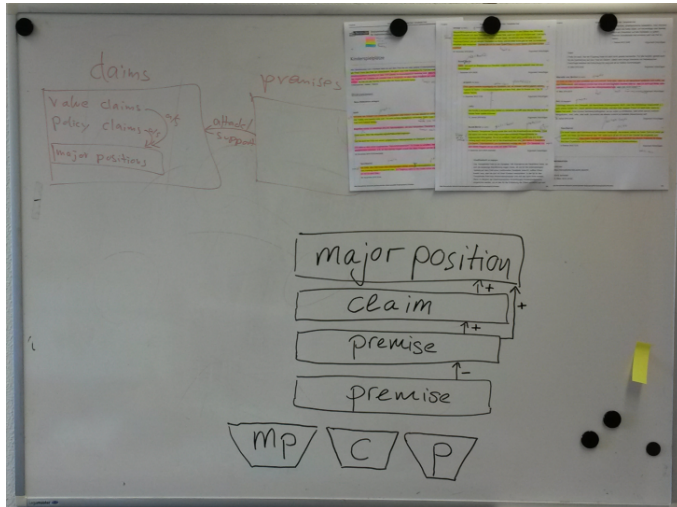
- In Bezug auf Online-Partizipationsverfahren soll die Wahl des Argumentationsmodells zielgerichtet auf die folgenden Fragen sein:
 - Was wollen die Beteiligten? Was soll umgesetzt werden?
 - Wie begründen die Beteiligten das?
- Unser Argumentationsmodell soll dabei berücksichtigen, dass politischen Entscheidungsträgern Handlungsoptionen o.ä. (idealerweise mit Begründungen) vorgelegt werden können.



Wahl eines Argumentationsmodells

- Wir haben uns für eine Modifikation in ein dreiteiliges Modell entschieden:
 - **major position:** Eine major position ist eine Handlungs- oder Entscheidungsoption, die erstmalig in der Diskussion auftritt.
Beispiel: *Es soll eine Oper gebaut werden.*
 - **claim:** Äußerung, in der vom Autor eine pro oder contra Position in Bezug auf eine major position bezogen wird.
Beispiel: *Ja, finde ich auch!*
Beispiel: *Super Vorschlag, ich bin auch dafür eine Oper zu bauen.*
 - **premise:** Begründung, die einen claim angreift oder unterstützt.
Beispiel: *Das wird den Tourismus in unserer Stadt ankurbeln.*
Beispiel: *Eine Oper wird das Kulturangebot bereichern.*
- Problem: Die Unterscheidung zwischen claim und premise ist kontextabhängig!
Beispiel: *Hunde sind dumm.*

Wahl eines Argumentationsmodells



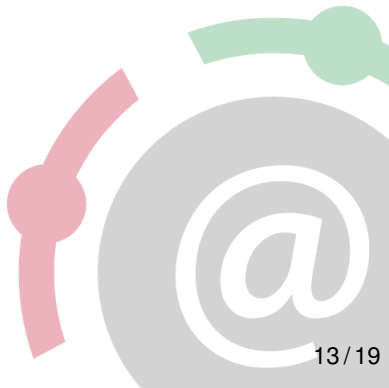
- Online-Partizipationsplattform zur Gestaltung des Flughafengeländes Berlin Tempelhof
- 7 Themenbereiche: Bewirtschaftung, Erinnerung, Freizeit, I ♥ THF, Mitmachen, Natur, Was vergessen?
- Größe des Datensatzes - Stand 7.7.2015:
 - 340 Proposals (\approx 2148 Sätze)
 - 604 Kommentare auf erster Hierarchie-Stufe
 - 1389 Kommentare insgesamt (\approx 4843 Sätze, durchschnittlich 3,487 Sätze pro Kommentar)
 - Vergleich mit HHU Normsetzungskorpus: 25 Proposals und 434 Kommentaren (1444 Sätze)
- Textinhalte unter Creative Commons-by-SA-Lizenz

Idee 50 - Kinderspielplätze

- major position 1: *Wir hoffen, dass Kinderspielplätze zukünftig mit dem THF-Gesetz im Aussenbereich vereinbar sind.*
 - claim 1: *Ich sehe das Anlegen von einfachen Spielplätzen eher kritisch und das obwohl ich selber Kinder habe. (contra MP1)*
 - premise 1: *Im Umkreis des Feldes sind bereits viele zum Teil sehr schöne Spielplätze vorhanden. Dafür muss meiner Ansicht nach das Feld nicht bebaut werden. (pro C1)*
 - claim: *ich finde die idee mit dem spielplatz gut und sinnvoll, finde aber, dass es ein besonderer spielplatz sein sollte, der zum thema passt. (pro MP1)*
- major position 2: *zum beispiel könnte man eine art mini flughafen machen oder so.*
- major position 3: *was ich auch gut fände, wäre zum beispiel eine kletterwand in form des luftbrückendenkmals.*

Maschinelle Sprachverarbeitung

- Zur Verarbeitung deutscher Texte werden sequenziell folgende Aspekte abgearbeitet:
 - 1) Satztrennung
 - 2) Wortarten
 - 3) Abhängigkeiten
 - 4) Lemmatizer
 - 5) Polaritäten



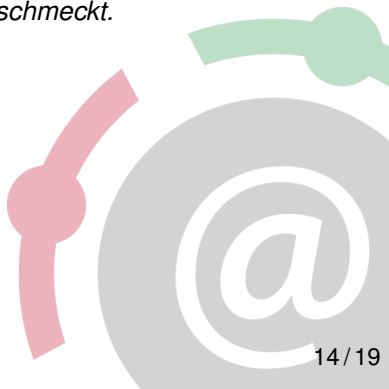
Maschinelle Sprachverarbeitung

Sentence Splitter

Ziel: Eingabetext in mehrere Sätze auftrennen

Beispiel: „*Heute hat das Essen in der Mensa geschmeckt. Morgen soll die Sonne scheinen.*“

- 1) *Heute hat das Essen in der Mensa geschmeckt.*
- 2) *Morgen soll die Sonne scheinen.*



Maschinelle Sprachverarbeitung

Part-of-Speech Tagger

Ziel: Jedem Wort bzw. Token einen POS-Tag bzw. eine Wortart zuweisen

11 Hauptwortarten (*Stuttgart-Tübingen-Tagset (STTS)*): Nomina (*N*), Verben (*V*), Artikel (*ART*), Adjektive (*ADJ*), Pronomina (*P*), Kardinalzahlen (*CARD*), Adverbien (*ADV*), Konjunktionen (*KO*), Adpositionen (*AP*), Interjektionen (*ITJ*), Partikeln (*PTK*)

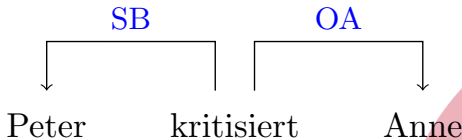
Beispieltagging: Autofahren *NN* bei *APPR* Glatteis *NN* ist *VAFIN* gefährlicher *ADJD* als *KOKOM* bei *APPR* Sommerwetter *NN* . \$.

Maschinelle Sprachverarbeitung

Dependenzen

Ziel: Grammatikalische Eigenschaften (Subjekt, Objekt usw.) aus einem Satz extrahieren

Beispiel: „*Peter kritisiert Anne.*“



Maschinelle Sprachverarbeitung

Lemmatizer

Ziel: Lemmatisierung, d.h. Entfernung von Flexionsformen bzw. Rückführung auf eine Grundform bzw. Lemma

Beispiele:

- Hauses → Haus
- Stifte → Stift
- Türme → Turm
- warte → warten

Ergänzung bestehender Software durch IWNL⁴

⁴Liebeck & Conrad, IWNL: Inverse Wiktionary for Natural Language Processing, 2015

Maschinelle Sprachverarbeitung

Polarität

Ziel: Numerische Polaritätsangaben für Wörter

Einsatz von Tonalitätslexikon *SentiwordSchatz* (kurz *SentiWS*)

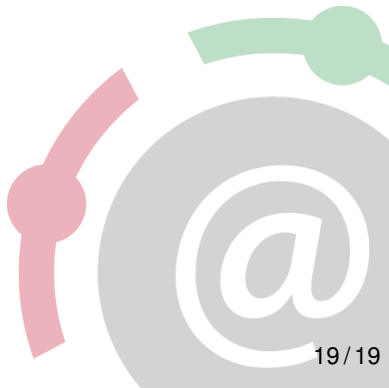
- Tonalitätsangaben $\in [-1, 1]$
- ≈ 3500 Einträge für Nomen, Adjektive, Verben und Adverbien

Beispiel: *Zerstörung* -0.3578



Weiteres Vorgehen

- Codierung des Tempelhofer Feld Datensatzes
- Entwicklung automatisierter Techniken um die manuellen Codierungen automatisiert vorzunehmen
- Zwei Publikationen in 2016
- Veröffentlichung des codierten, anonymisierten Datensatzes



Vielen Dank für Ihre Aufmerksamkeit!



Offene Punkte zur allgemeinen Diskussion

1. Wie relevant sind die Handlungsoptionen als Analyseziel?
2. Was wird Entscheidungsträgern bzw. den Auftraggebern von Online-Partizipationsverfahren als Ergebnis vorgelegt?
3. Wie wird in der Praxis mit Beiträgen ohne konkrete Entscheidungsoptionen umgegangen?
Beispiel: *Was ist in der Stadt wichtig?*



Bildquellen

- Inhaltsanalyse: Krippendorff 2004, Content Analysis: An Introduction to Its Methodology
- Logo Berlin: www.berlin.de

