UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386

Sprache im Netz

Wege zu einer soziolinguistischen Korpusannotation

Jöran Landschoff – Universität Heidelberg

Linguistisches Tagging

- POS-Tagging mittlerweile Standard
- Auch semantische Annotation schreitet voran
- Syntaktische Annotationen noch schwierig (z.B. Dependenz-Tagging)
- Meistens befasst mit dem Text selbst (Gries/Behrens 2017)
- Wozu aber Tagging?

Ziele für linguistische Arbeit

- Auffinden von Sprachgebrauchsmustern (Bubenhofer 2009, 23ff.)
  - POS-Tagging ermöglicht Abstraktion → z.B. gezieltes Auffinden von Kookkurrenzen bestimmter Ausdrücke mit Verben/finiten Verben/Eigennamen/etc.
- Diskurslinguistische Analyse
  - Gesellschaftliches Wissen (Langer/Nonhoff/Reisigl 2019)
  - Dominante, übersubjektive Bedeutungen (Müller 2012)
  - Umstrittene Perspektivierungen (Felder/Müller/Vogel 2012)
- Bei soziolinguistischem Interesse wird die „Wer“-Frage entscheidend

„Kulturwissenschaftliche (und damit auch sprachwissenschaftliche) Daten sind also schon in ihrer Entstehung nicht eindeutig oder objektiv gegeben, sondern immer schon zeichenhaft. Ihre Interpretation muss daher in Form dichter Beschreibung erfolgen, die versucht, den Sinn zu erfassen, den die Handelnden diesen Daten selbst zugeschrieben haben. Maschinelle, insbesondere datengeleitete Verfahren resultieren aber fast immer in ‚dünnen Beschreibungen'." (Scharloth 2018, 67)

„Doch mit welchem Instrument arbeitet man am Besten, in welcher Verbindung steht etwa eine Topos-Analyse mit der historischen Schlagwortforschung, was unterscheidet z. B. die Framesemantik von Akteursanalysen usw.? Aus unserer Erfahrung sind diese Fragen weit weniger konstruiert, als sie es vielleicht auf den ersten Blick erscheinen mögen." (Spitzmüller/Warnke 2011, 198)

- Welche Texte?
- Welche Akteure?

Anforderungen an Annotationen für soziolinguistisch interessierte Diskursforschung

- Streng linguistische Annotationen zum Auffinden von Sprachgebrauchsmustern (POS-tagging, semantic tagging, etc.)
- DIMEAN (Spitzmüller/Warnke 2011, 197ff, besonders 201) folgend:
  - Wer spricht? → Rückbindung des Sprachgebrauchsmusters an den Akteur (Akteursebene)
  - Worauf wird Bezug genommen? → Welche Texte stehen in einer Verbindung zueinander im Sinne eines Diskurses als „zerdehntes Gespräch“ (Müller 2012, 59)? (intertextuelle Ebene)
    - Akteursverbindungen werden hieraus ableitbar

Beispiel: Erkenntnisinteresse sozial divergente Semantik

- Bedeutung verstanden als sozial verhandelte Inhaltsfestsetzung sprachlicher Zeichen (Bubenhofer/Müller/Scharloth 2014, 4)
  - Soziale Gruppen erreichen unterschiedliche Konsense über Form-Inhalts-Paare, soziale Bedeutung von Ausdrücken usw.
  - Bestimmte Gruppen dürften daher
    - Spezifische Sprachgebrauchsmuster und
    - Spezifische Kontextualisierung dieser Muster entwickeln (Labov 1976)

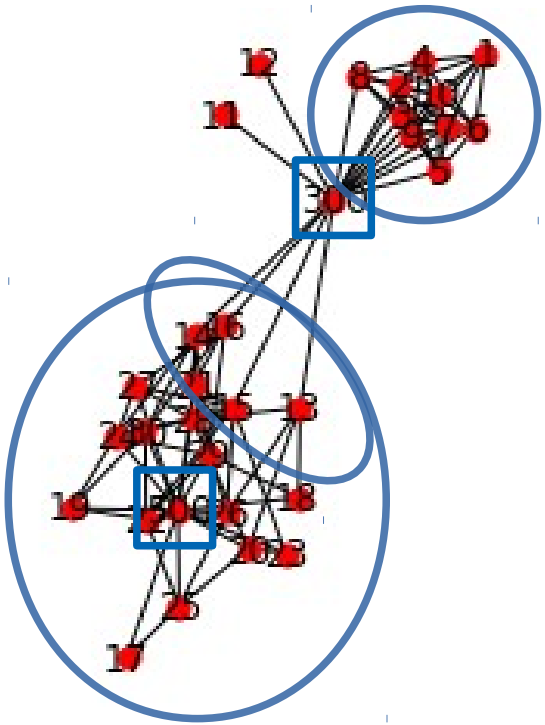
➔ Welche Sprachgebrauchsmuster sind für welche Gruppen spezifisch?

➔ Wie entstehen, verändern, verteilen sich diese Sprachgebrauchsmuster im Sozialen Gebrauch?

Methode und Daten

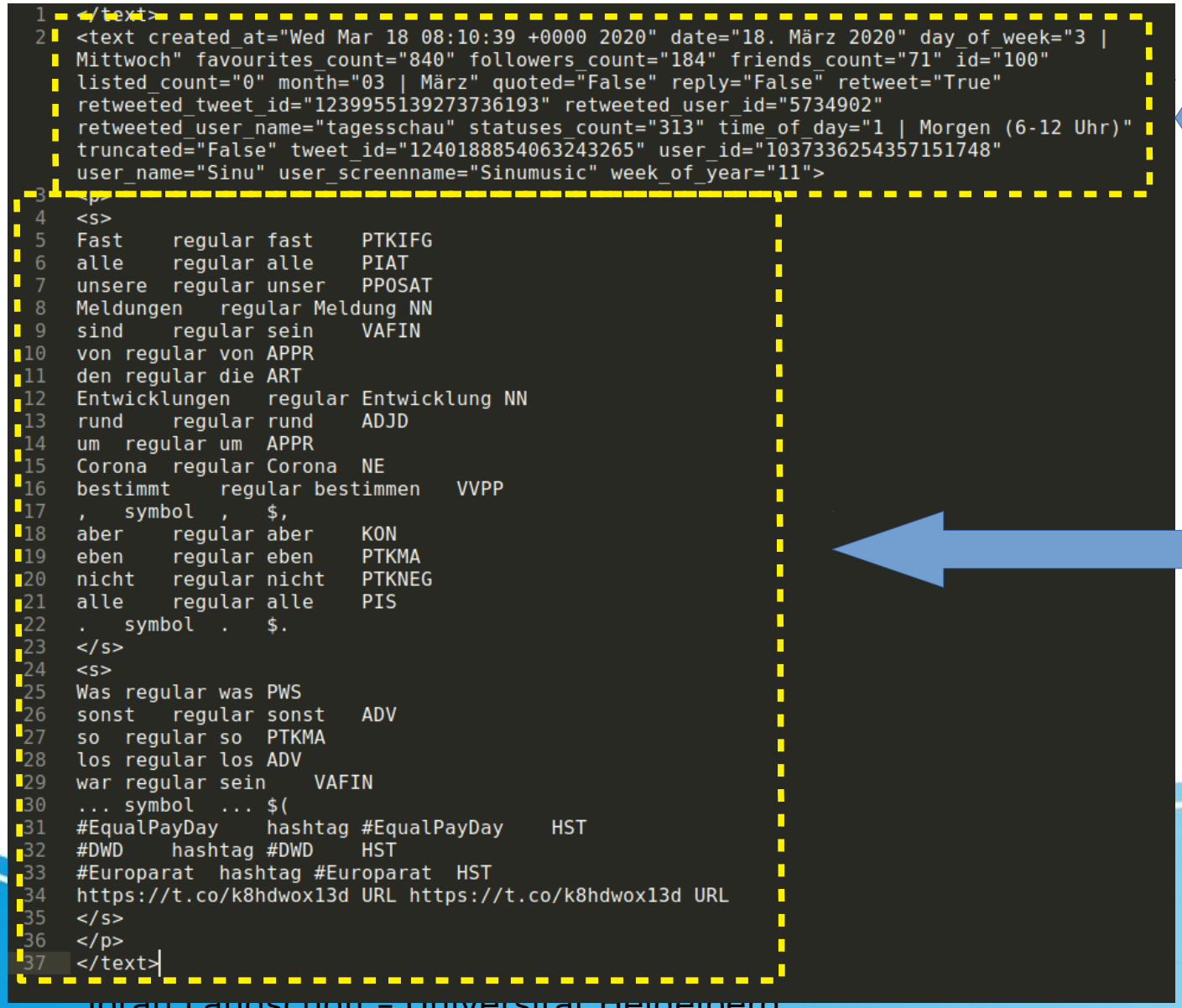
- Twitter-Daten
  - Relativ natürliche Kommunikation (Scheffler 2017, 127-128)
  - Relativ gut zugänglich (Pfaffenberger 2016, 111)
  - Intrinsische Relationsinformationen (Metadaten)
- Netzwerkanalyse (siehe z.B. Stegbauer 2016)
  - Darstellung von Beziehungen
  - Flexible Datenverarbeitung
  - Strukturale Analysemöglichkeiten

Beispielnetzwerk



- Zwei Kommunikationsgruppen
- Zwei Sprachgebrauchsmuster
- Eine Gruppe innerhalb Kommunikationsgruppe 2, die Sprachgebrauchsmuster von Kommunikationsgruppe 1 verwendet

Twitter-Daten und die Corpus Work Bench (CWB)



Text metadata

Nicht mit Suchsyntax ansteuerbar

Pos-tagged text

mit Suchsyntax ansteuerbar

Twitter-Daten und die Corpus Work Bench (CWB)

- Bereits alle nötigen Daten enthalten, allerdings immer in Bezug auf `text`–Elemente

```
<text created_at="Wed Mar 18 08:10:39 +0000 2020"
date="18. März 2020"
day_of_week="3 | Mittwoch"
favourites_count="840"
followers_count="184"
friends_count="71"
id="100"
listed_count="0"
month="03 | März"
quoted="False"
reply="False"
retweet="True"
retweeted_tweet_id="1239955139273736193"
retweeted_user_id="5734902"
retweeted_user_name="tagesschau"
statuses_count="313"
time_of_day="1 | Morgen (6-12 Uhr)"
truncated="False"
tweet_id="1240188854063243265"
user_id="1037336254357151748"
user_name="Sinu"
user_screenname="Sinumusic"
week_of_year="11">
```

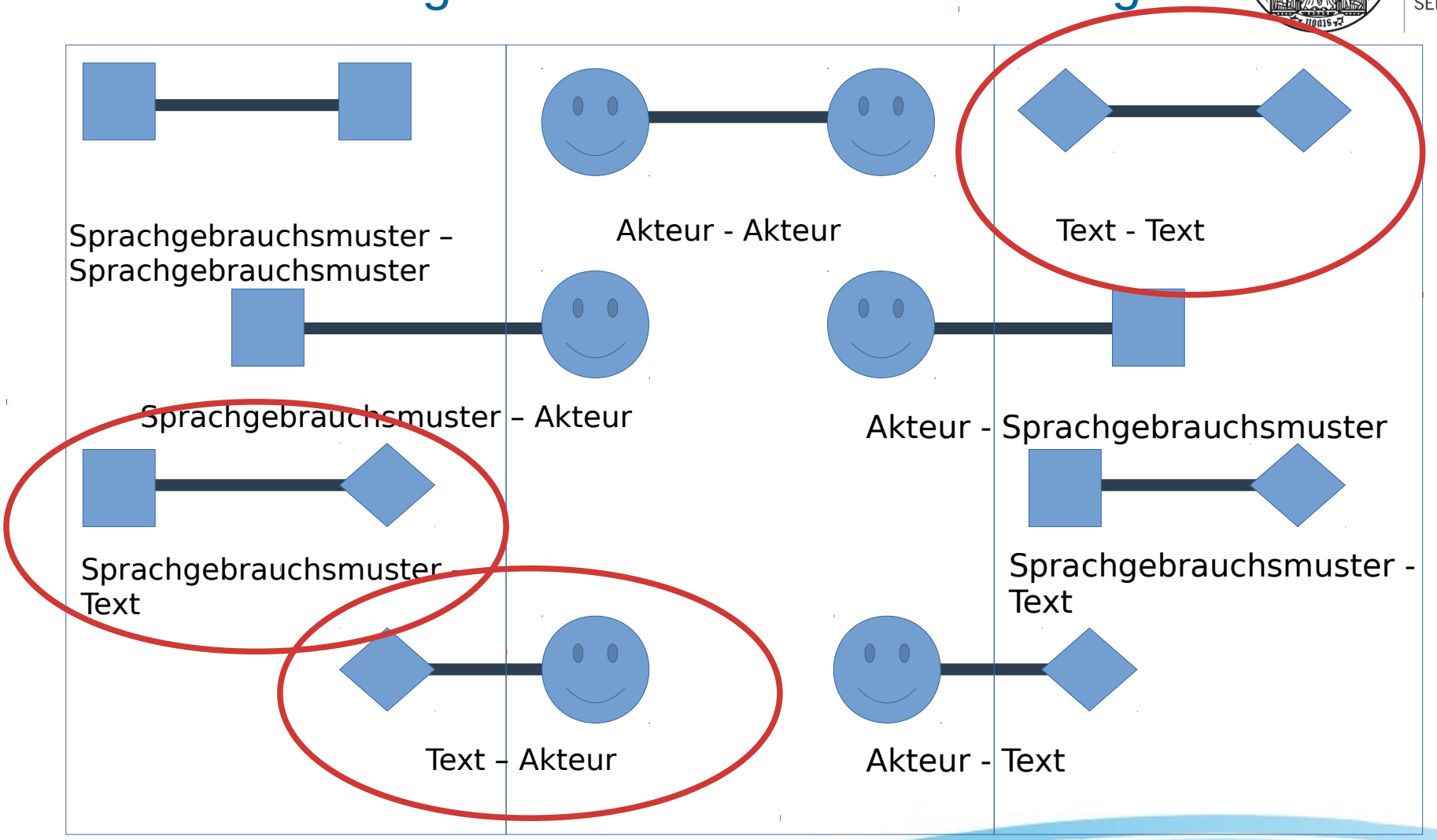
Weitere twitterbezogene Informationen über den User

Interaktions-Informationen → wird auf anderen Tweet Bezug genommen? Wenn ja, auf welchen, von wem?

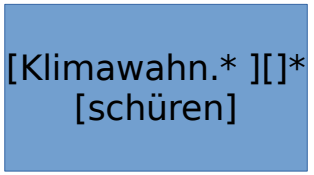
Text-Identifikation

User-Identifikation

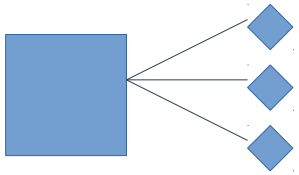
Vorüberlegung für Annotation: Möglichkeiten der Relationierung



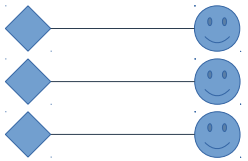
Annotation aufgefundener Sprachgebrauchsmuster



1. Sprachgebrauchsmuster (plural) korpuslinguistisch ermitteln

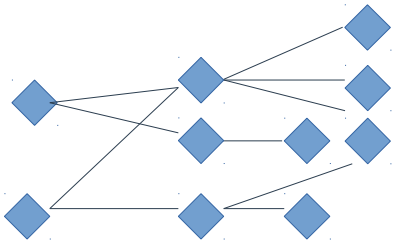


2. Texte ermitteln, in denen das SGM vorkommt

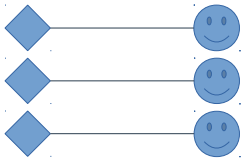


3. Von den Texten auf die Akteure schließen, um SGMs mit Akteuren in Verbindung zu bringen

Netzwerk1 – Akteure - SGMs



4. Intertextuelle Beziehungen auffinden (entspricht Kommunikationsstruktur)



5. Texte Akteuren zuweisen

Netzwerk2 – Akteure – Akteure, Text – Akteur, Text – Text

6. Netzwerke zusammenlegen

Analysemöglichkeiten

- Kommunikationsstränge durch Text-Text-Verbindungen erkenntlich
  - Gesprächsanalytische Möglichkeiten
- SGMs auf Akteure zurückzuführen
  - Korrelation von Kommunikationsnetzwerk und SGM-Netzwerk?
  - Korrelation von SGM-Netzwerken?
  - Korrelation von Bündeln von SGM-Netzwerken und Kommunikationsnetzwerken?
- Weitere Akteurs-Informationen
  - Gewichtung im Netzwerk
    - Eingangs-, und Ausgangsgrad
    - Followerzahl als Parameter
  - Position im Netz als Indikator für Einfluss

Literatur

Bubenhofer, Noah (2009): Sprachgebrauchsmuster. Korpuslinguistik als Methode der Diskurs- und Kulturanalyse. Berlin/New York: de Gruyter.

Bubenhofer, Noah / Nicole Müller / Joachim Scharloth (2014): Narrative Muster und Diskursanalyse: Ein datengeleiteter Ansatz. In: Zeitschrift für Semiotik. Band 35, Heft 3-4 (2013), S. 419-444. (hier zitiert das Pre-Print: http://www.scharloth.com/files/narrative_preprint.pdf

Felder, Ekkehard; Müller, Marcus; Vogel, Friedemann (2012): Korpuspragmatik. Paradigma zwischen Handlung, Gesellschaft und Kognition. In: Ekkehard Felder, Marcus Müller und Friedemann Vogel (Hg.): Korpuspragmatik : thematische Korpora als Basis diskurslinguistischer Analysen. Berlin [u.a.]: De Gruyter (Linguistik - Impulse & Tendenzen), S. 1–30.

Gries, Stephan Th./Berez, Andrea L. (2017): Linguistic Annotation in/for Corpus Linguistics. In: Ide, Nancy/Pustejovsky, James (Hg.): Handbook of Linguistic Annotation, Dodrecht: Springer, S. 379-409.

Labov, William (1971): The Study of Language in its Social Context. In: Fisburn, Joshua A. (Hg.): Advances in the Sociology of Language. Volume 1. Basic Concepts, Theories and alternative Approaches. De Gruyter.

Langer, Antje / Nonhoff, Martin / Reisigl, Martin (2019): Diskursanalyse und Kritik. Einleitung. In: Langer, Antje / Nonhoff, Martin / Reisigl, Martin (Hrsg.): Diskursanalyse und Kritik. Wiesbaden: Springer Fachmedien Wiesbaden, 1–11.

Müller, Marcus (2012): Vom Wort zur Gesellschaft: Kontexte in Korpora. Ein Beitrag zur Methodologie der Korpuspragmatik. In: Ekkehard Felder, Marcus Müller und Friedemann Vogel (Hg.): Korpuspragmatik : thematische Korpora als Basis diskurslinguistischer Analysen. Berlin [u.a.]: De Gruyter (Linguistik - Impulse & Tendenzen), S. 33–82.

Pfaffenberger, Fabien (2016): Twitter als Basis wissenschaftlicher Studien. Eine Bewertung gängiger Erhebungs- und Analysemethoden der Twitter-Forschung. Wiesbaden: Springer.

Spitzmüller, Jürgen/Warnke, Ingo H. (2011): Diskurslinguistik. Eine Einführung in Theorien und Methoden der transtextuellen Sprachanalyse. Berlin/Boston: de Gruyter.

Scheffler, Tatjana (2017): Conversations on Twitter. In: D. Fišer/M. Beißwenger, Investigating Computer-Mediated Communication: Corpus-Based Approaches To Language In The Digital World, Ljubljana: University Press.

Stegbauer, Christian (2016): Grundlagen der Netzwerkforschung. Wiesbaden: Springer Fachmedien.