
Universität Stuttgart
Nils Reiter
Sarah Schulz
Reflected Text Analysis:
Computational Linguistics
Meets
Digital Humanities

About this Talk

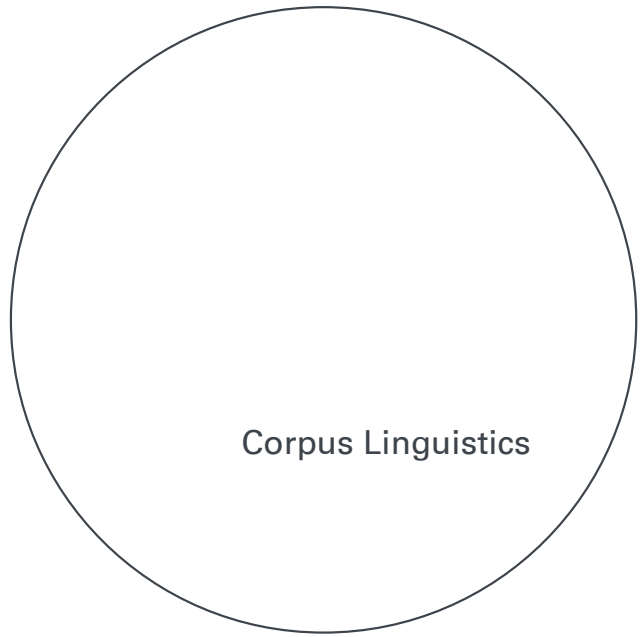
- Computational Linguistics and Digital Humanities
- DH in Stuttgart
- Coreference Resolution

If you should have questions, feel free to ask…

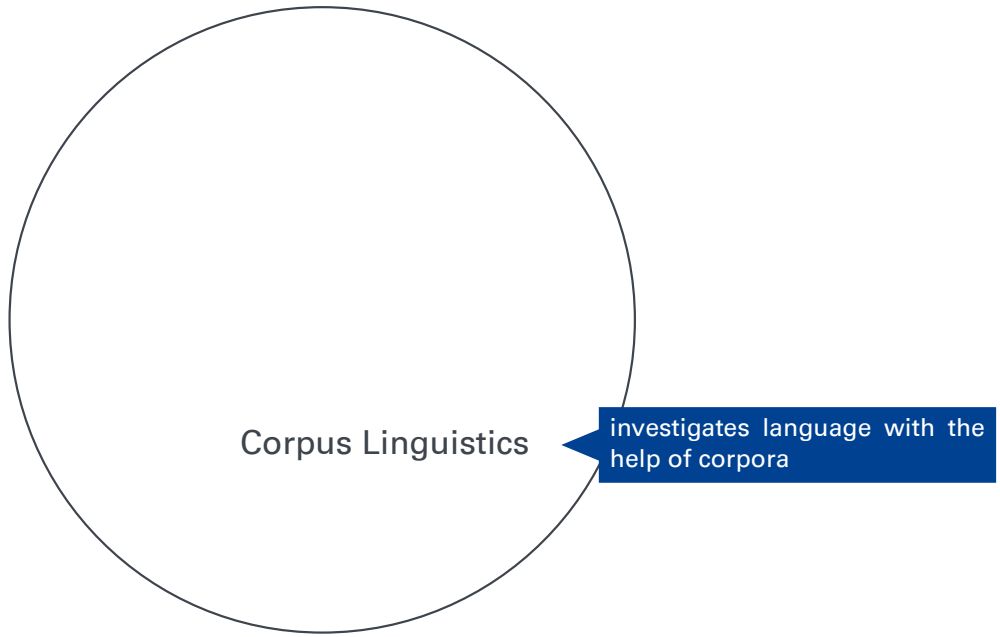
Computational Linguistics and Digital Humanities

1

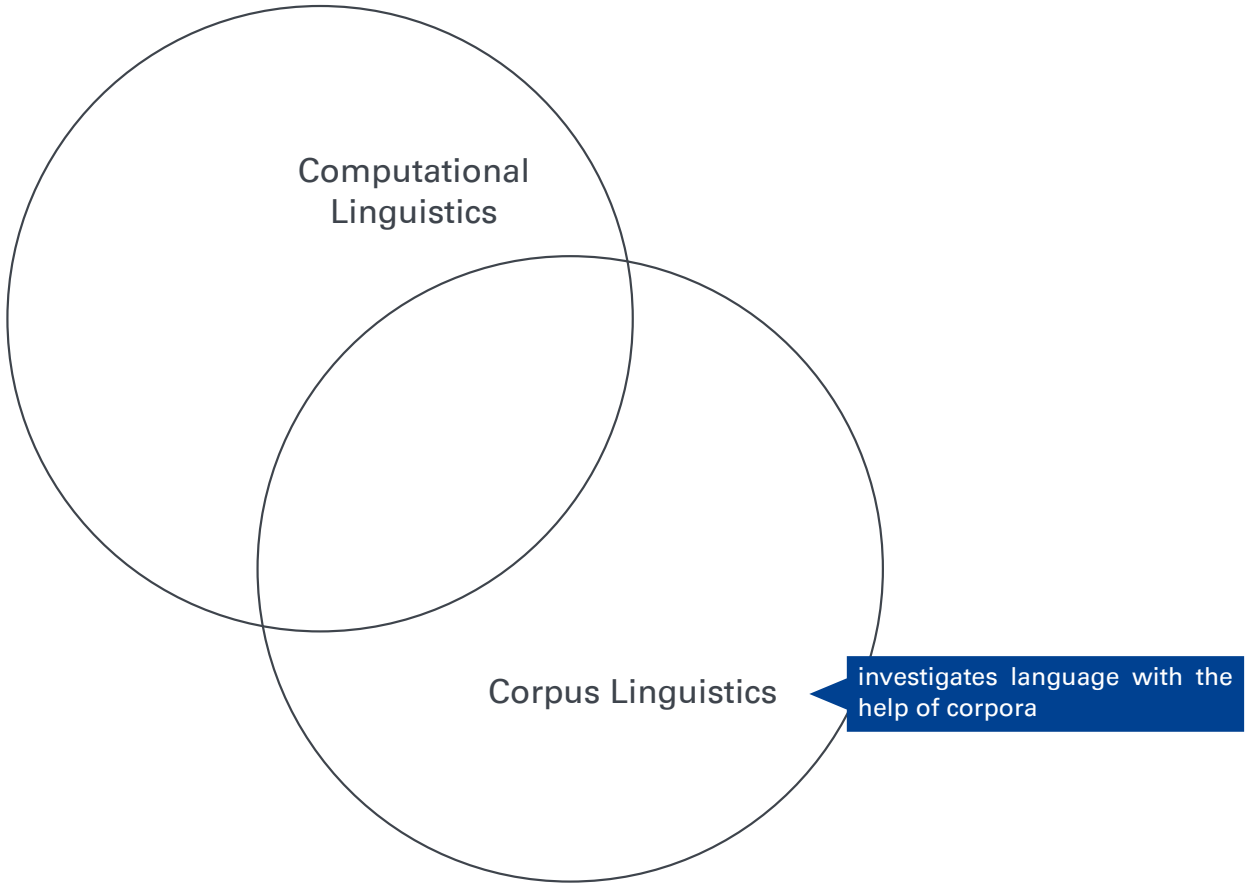
CL is DH



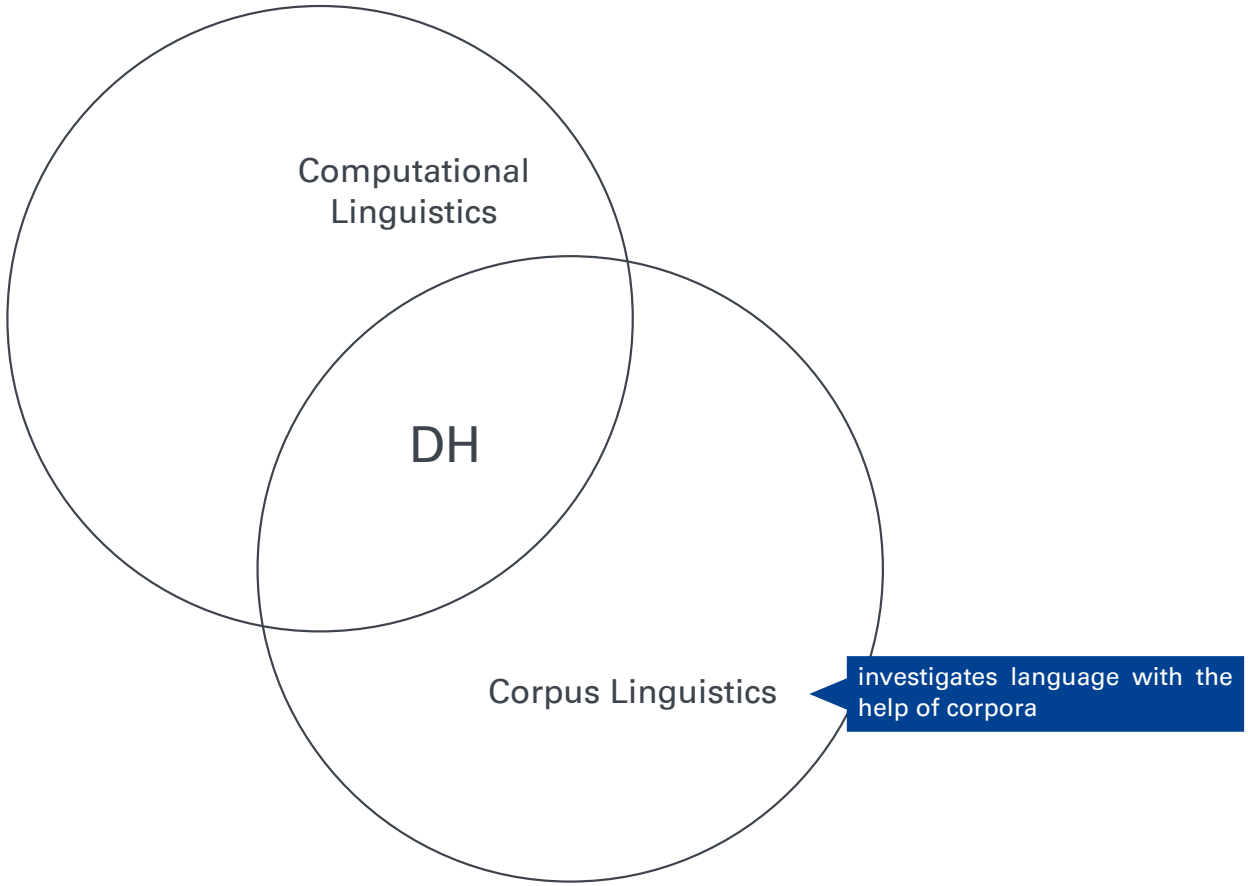
CL is DH


Corpus Linguistics
investigates language with the help of corpora

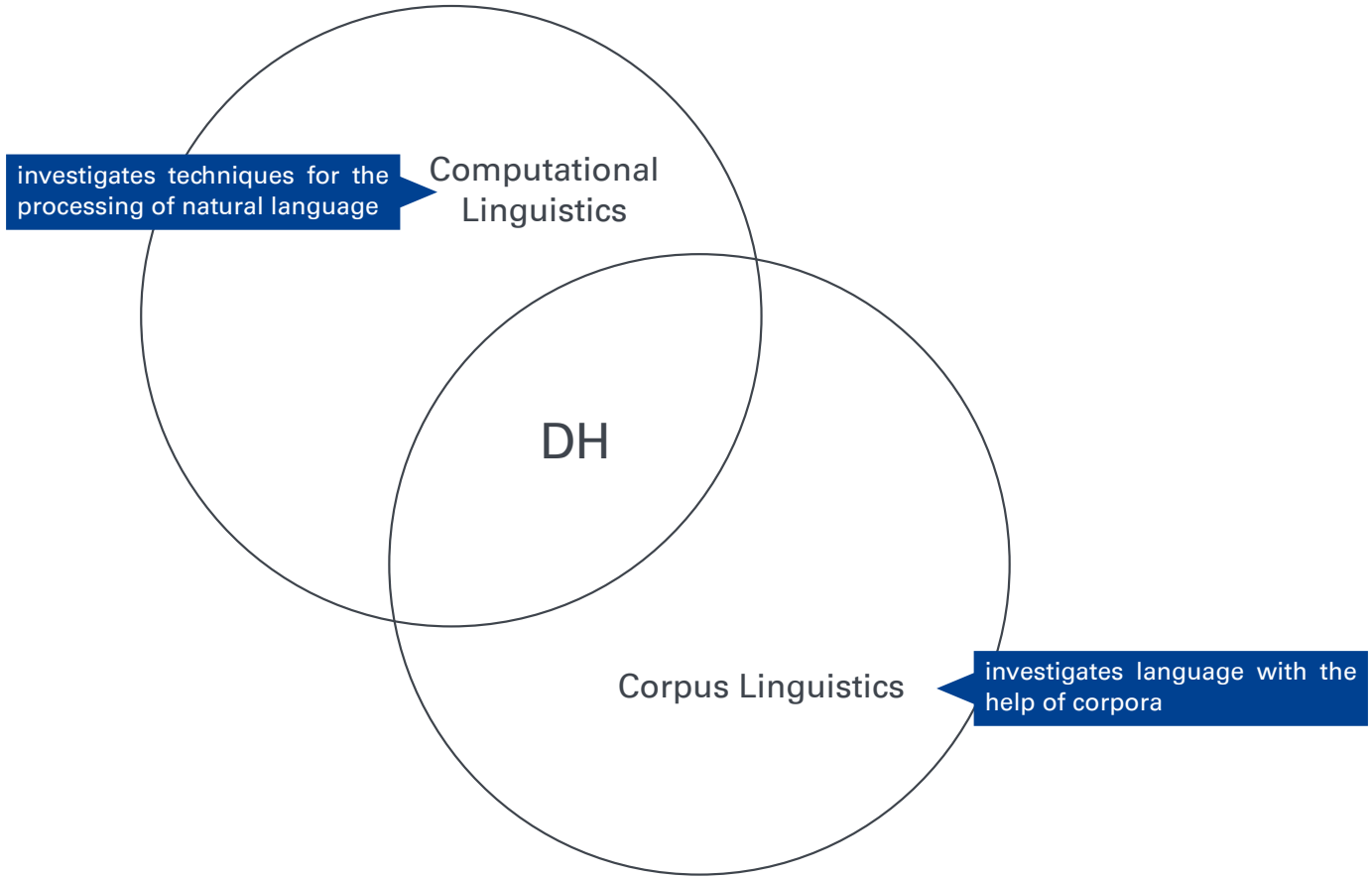
CL is DH


Computational Linguistics
Corpus Linguistics
investigates language with the help of corpora

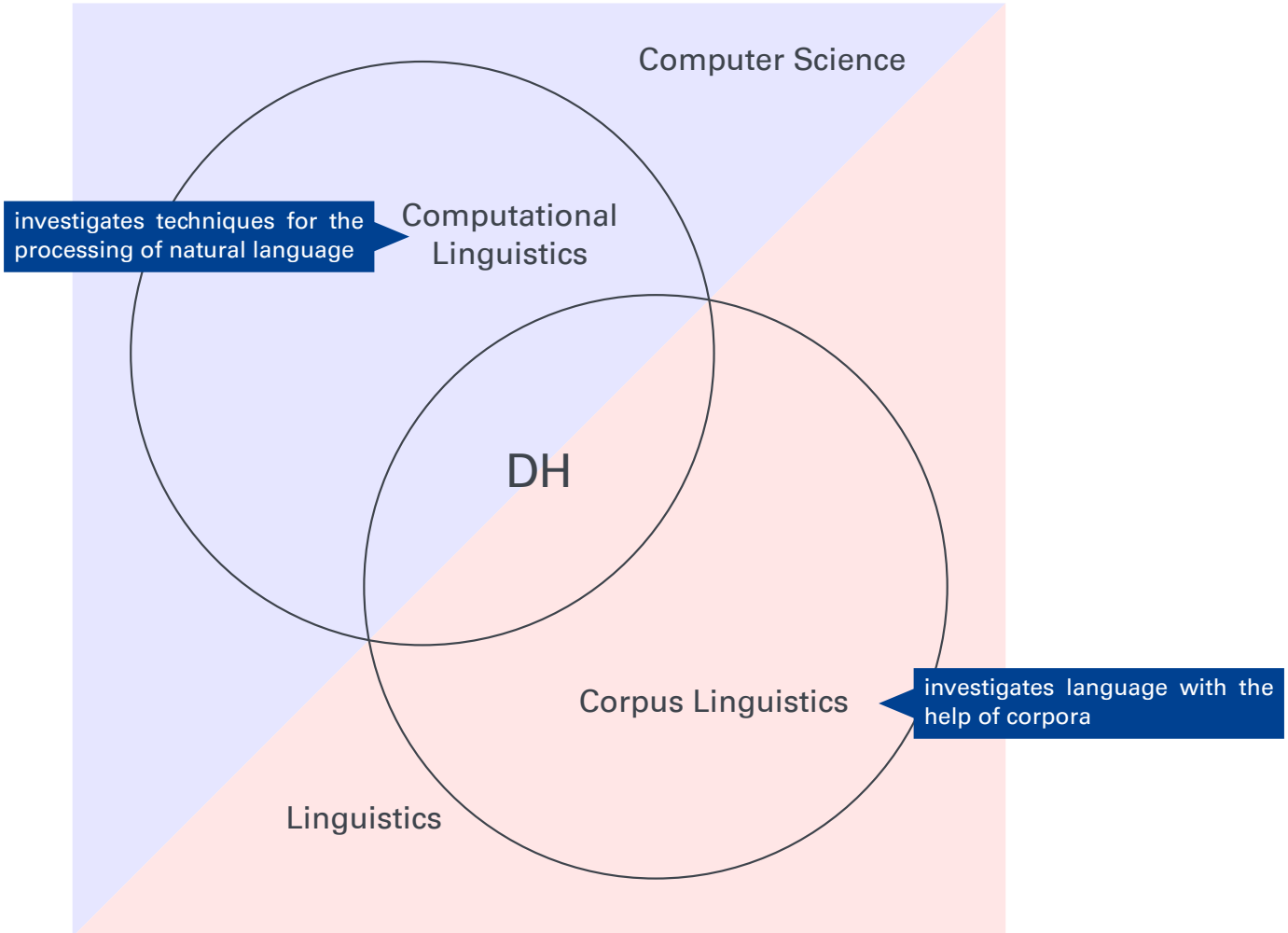
CL is DH


Computational Linguistics
DH
Corpus Linguistics
investigates language with the help of corpora

CL is DH


investigates techniques for the processing of natural language
Computational Linguistics
DH
Corpus Linguistics
investigates language with the help of corpora

CL is DH


Computer Science
investigates techniques for the processing of natural language
Computational Linguistics
DH
Corpus Linguistics
investigates language with the help of corpora
Linguistics

Computational Linguistics and Digital Humanities

- Why do the humanities want to work with CL?
  - many humanities projects are text-based
    - $\rightarrow$ tools and workflows from CL can be adapted to DH projects
  - larger corpora can be analysed

Computational Linguistics and Digital Humanities

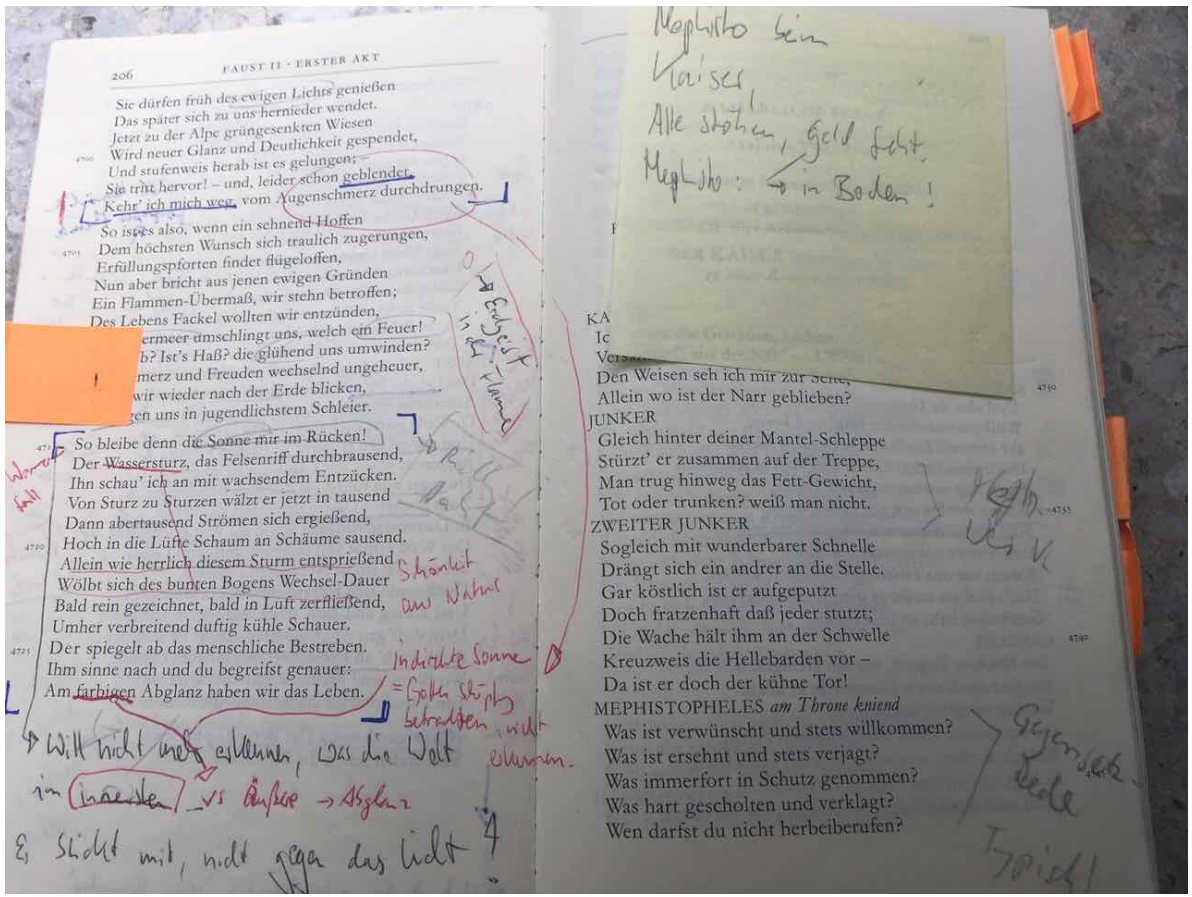
- Why do the humanities want to work with CL?
  - many humanities projects are text-based
    - $\rightarrow$ tools and workflows from CL can be adapted to DH projects
  - larger corpora can be analysed
- Why does CL want to work with the humanities?
  - non-standard language, text structure, applications and research questions from DH are new for CL
  - humanities scholars provide expert knowledge
  - research question can limit the scope of what a tool has to provide

Computational Linguistics and Digital Humanities

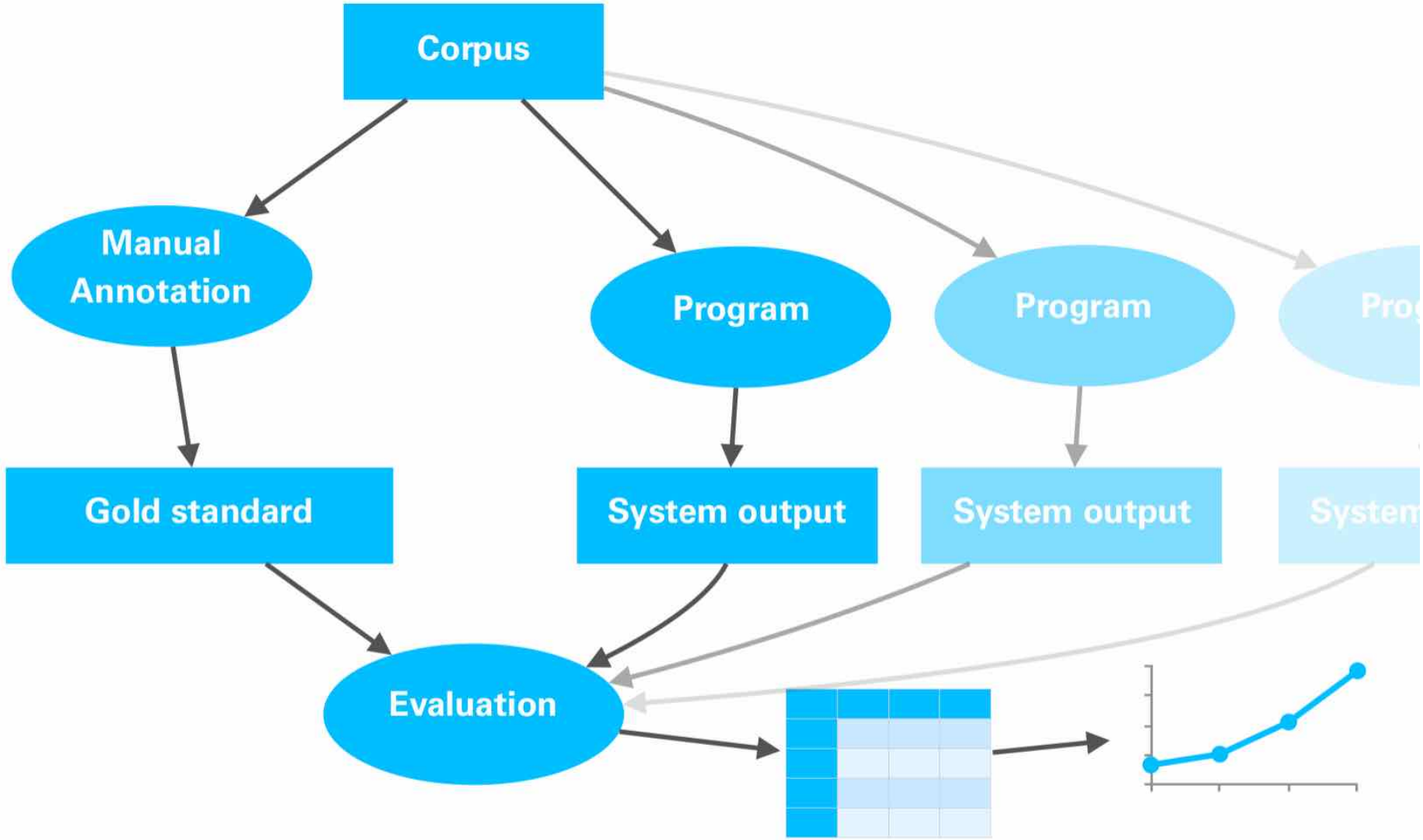
- Why do the humanities want to work with CL?
  - many humanities projects are text-based
    - $\rightarrow$ tools and workflows from CL can be adapted to DH projects
  - larger corpora can be analysed
- Why does CL want to work with the humanities?
  - non-standard language, text structure, applications and research questions from DH are new for CL
  - humanities scholars provide expert knowledge
  - research question can limit the scope of what a tool has to provide
- both: collaboration is a chance and challenge

Annotation as Basis for… Everything

- explicit assignment of categories to text spans
  - text spans are explicitly bounded (begin, end)
- annotation by humans vs. annotation by computers



The Day-to-Day Work of a Computational Linguist



Why do we annotate?

- empirical validation of theories
  - discovering phenomena not covered by a theory
  - strengthening definitions in a theory
    - often confused categories might be overlapping or at least unclear
  - uncovering implicit assumptions
- data creation
  - manually annotated data can be analysed
    - which categories are how frequent in what context?
  - automatic tools can be evaluated
    - How well do machines do this task?
  - supervised tools can be trained

Annotationworkflow

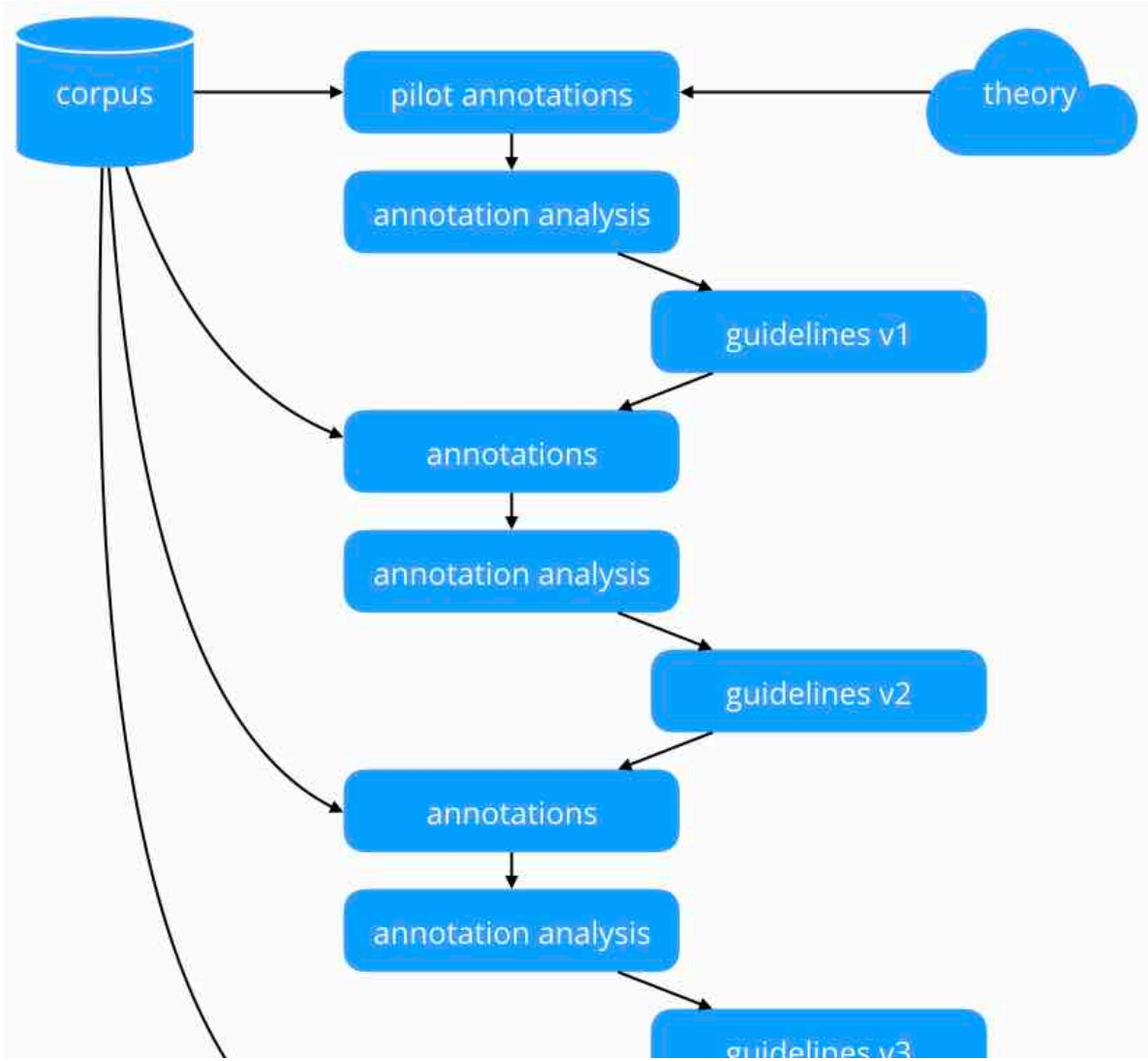


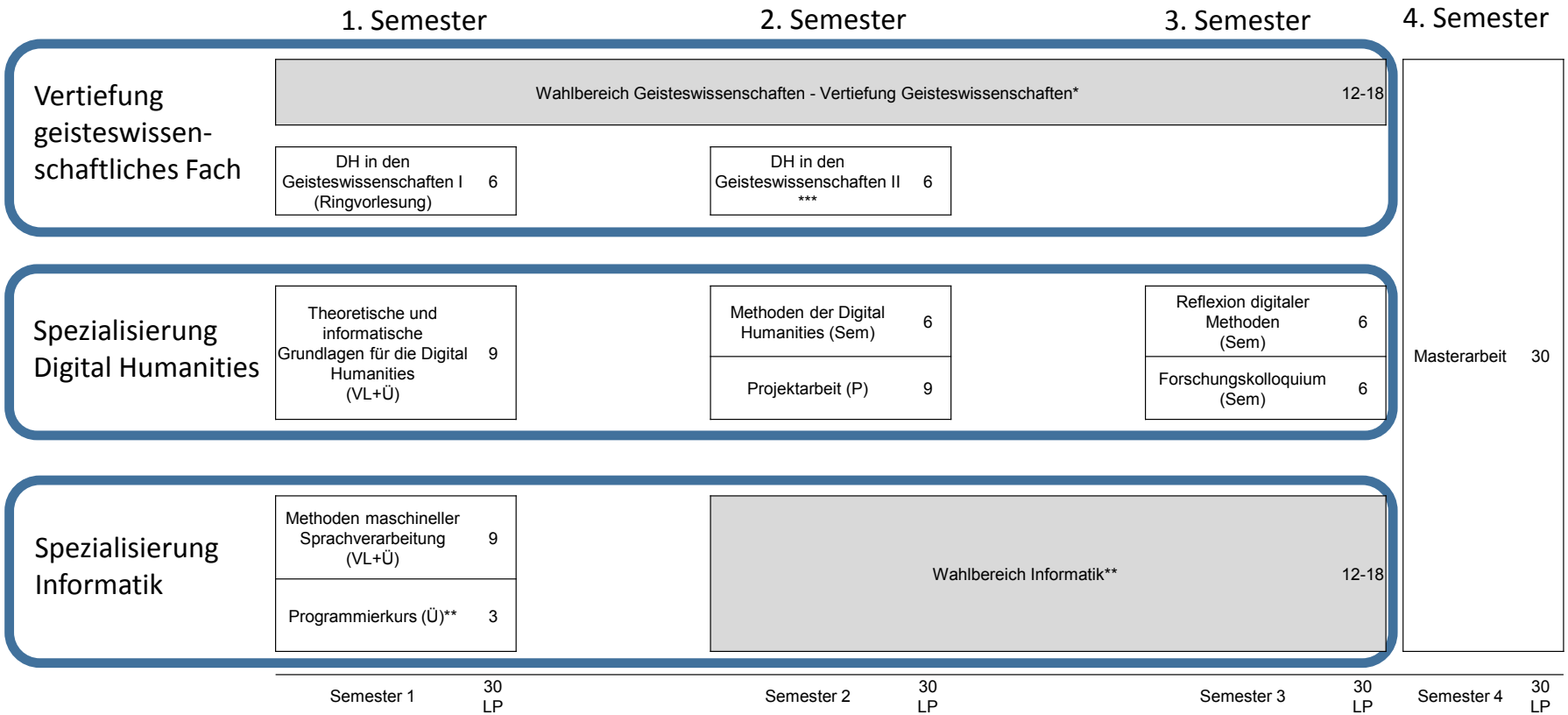
Figure: Annotation workflow schema. Arrows indicate (rough) temporal sequence.

Inter-subjective Annotations in DH?

- Linguistic annotations: Inter-subjective annotations
  - E.g., part of speech
  - Different annotators create the same annotations
  - Inter-Annotator Agreement (IAA)
- Annotation guidelines
  - Mediator between theory and annotation
  - Guidebook for annotators (who may not be experts)
    - What is to be annotated?
    - Which categories are used in which cases?
    - How to deal with borderline cases?
    - How have we dealt with difficult cases in the past?
    - Tons of examples

CL and DH in Stuttgart

Masterstudiengang *Digital Humanities*, Universität Stuttgart



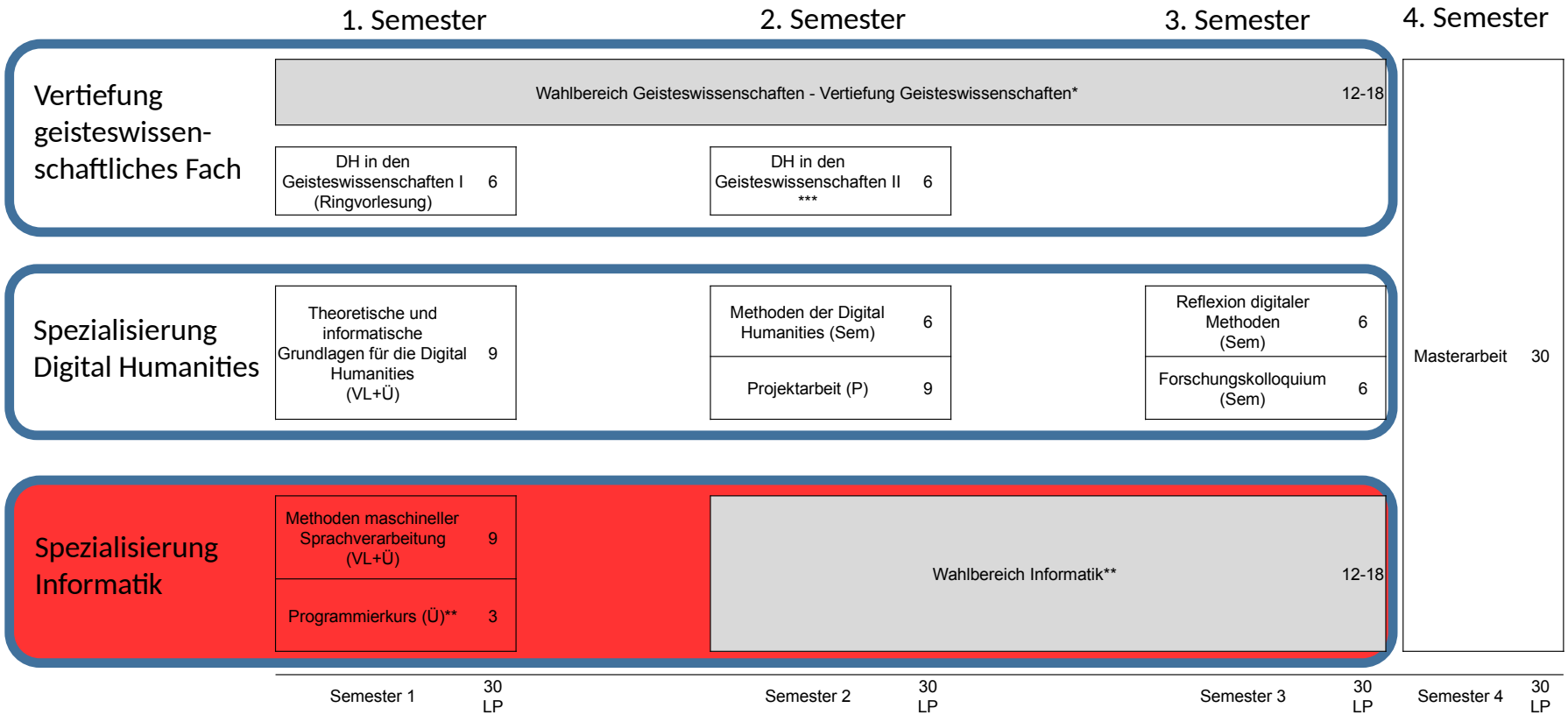
* Importmodule aus den Geisteswissenschaften

** Importmodule Informatik

*** DH-Veranstaltung in den Geisteswissenschaften - Veranstaltung wird vom geisteswissenschaftlichen Fach angeboten

14.04.2016

Masterstudiengang *Digital Humanities*, Universität Stuttgart



* Importmodule aus den Geisteswissenschaften

** Importmodule Informatik

*** DH-Veranstaltung in den Geisteswissenschaften - Veranstaltung wird vom geisteswissenschaftlichen Fach angeboten

14.04.2016

NLP Methods for DH

Two main ideas

- cover some basic NLP tasks
- cover conceptual workflow

	PoS tagging	Named Entity Recognition	Coreference Resolution
Annotation	STTS, IAA, Kappa	NoSta-D, BIO	TueBa + NoSta-D
Modelling	Decision Trees	HMM, CRF	Clustering
Evaluation	Accuracy	Precision/Recall	Specific metrics (CEAF, B3, BLANC)

QuaDramA[1]

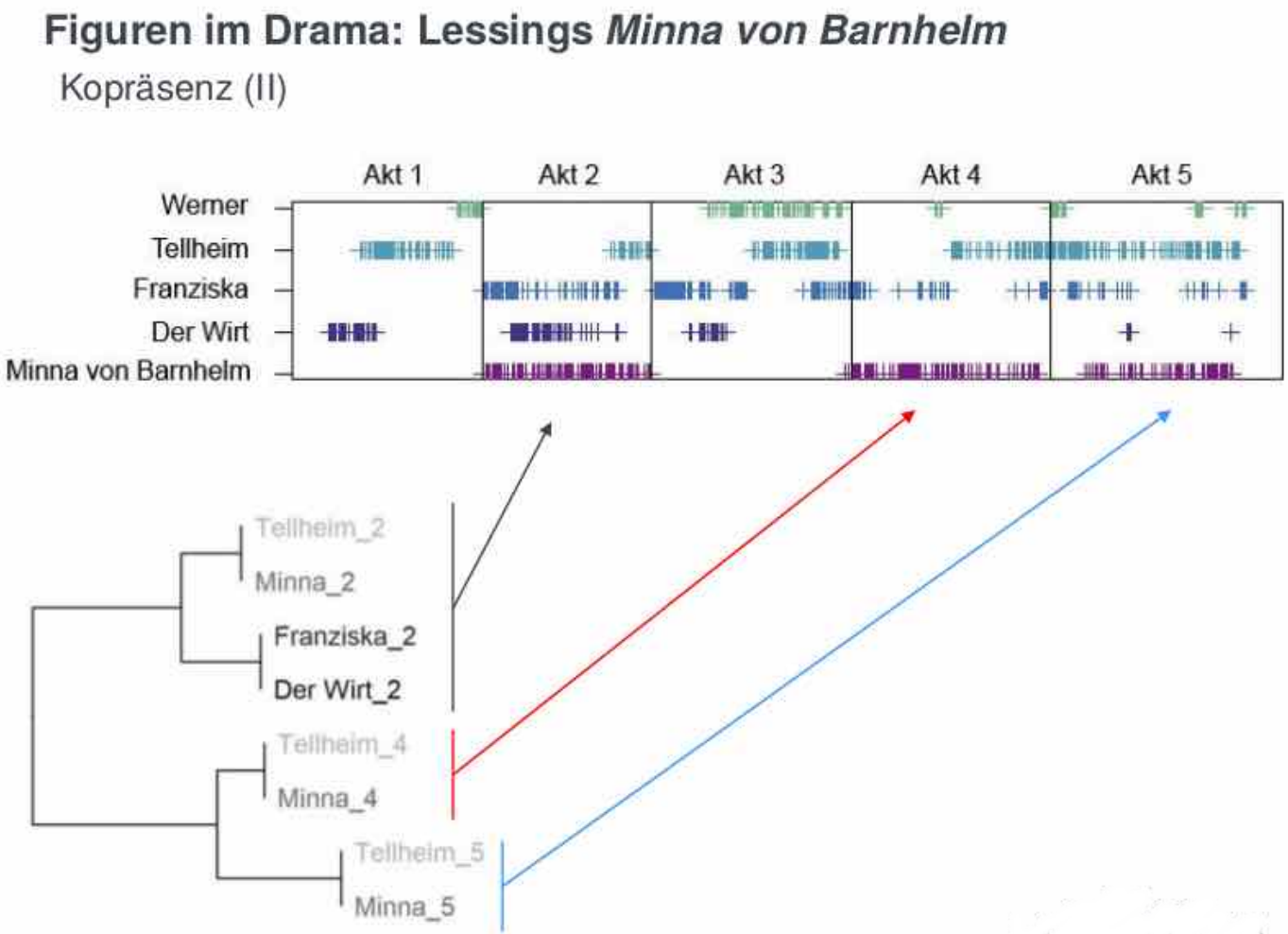
Test hypotheses about dramatic characters on large data sets using NLP methods

Dimensions:

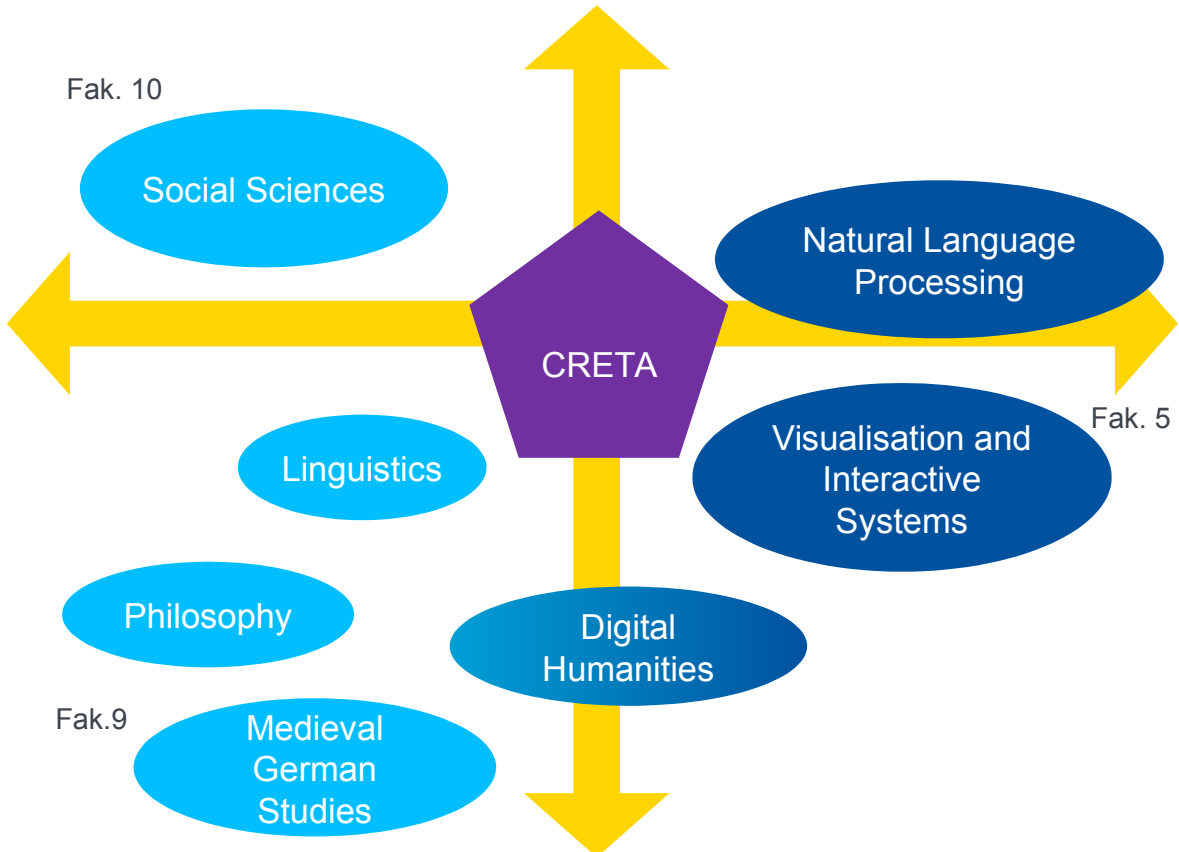
- Character classes by gender, action, social class, stock character
  - How are fathers marked as being ‘tender’?
- Relations between characters
  - What are shared topics/emotions?
- Character (type) development
  - When are characters types appearing/disappearing?

[1] Project members: Nils Reiter, Marcus Willand, Benjamin Krauter, Janis Pagel

Co-presence of Characters

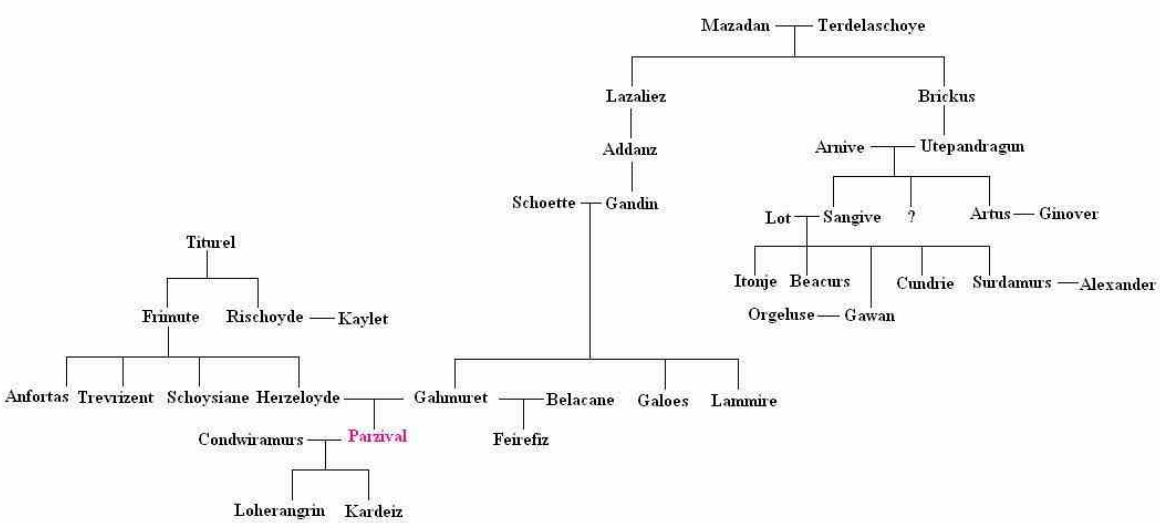


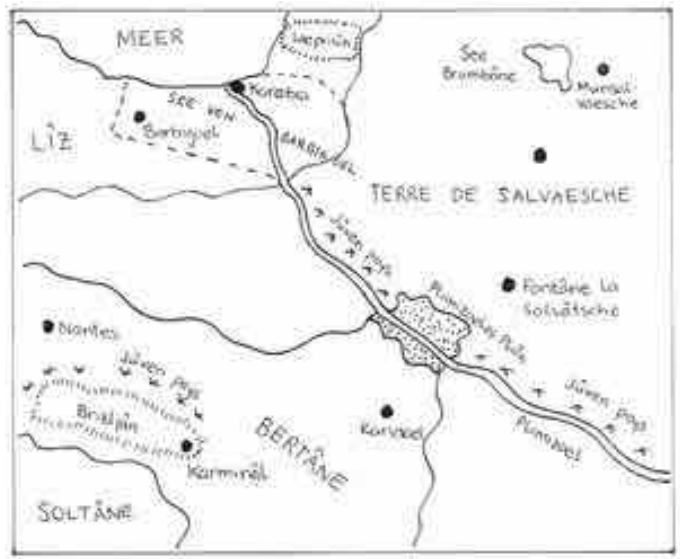
Center for Reflected Text Analytics (CRETA)



CRETA
CENTER FOR REFLECTED TEXT ANALYTICS

Narrative Structure of Arthurian Romance (Example Parzival)[1]





[1] Project members: Nora Ketschik, André Blessing, Manuel Braun

Wertheradaptionen[1]

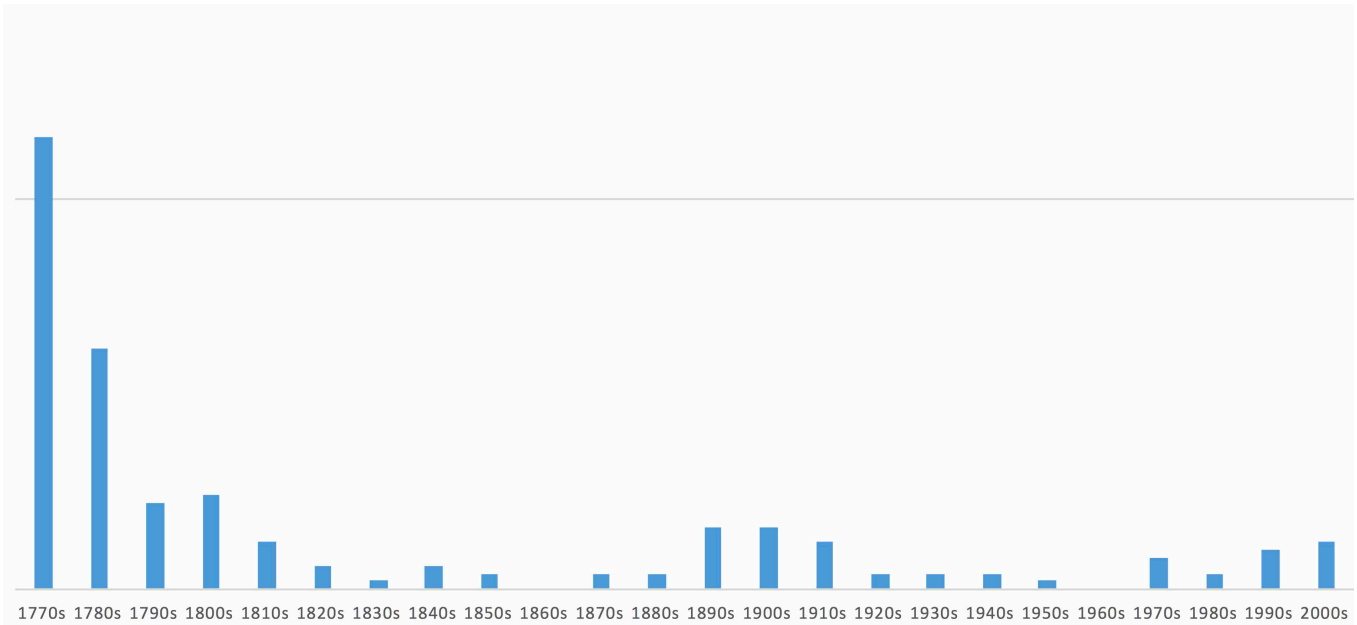
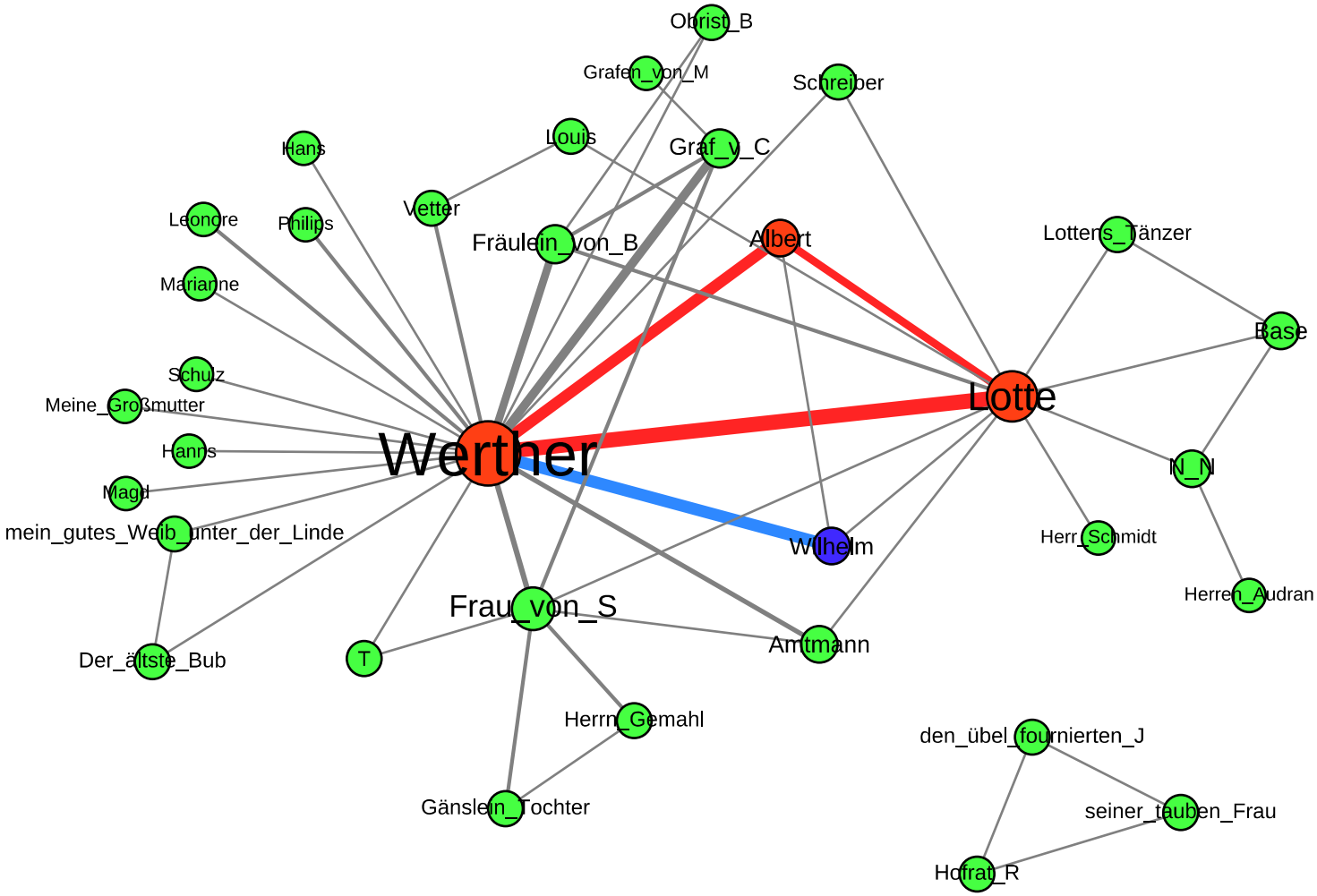


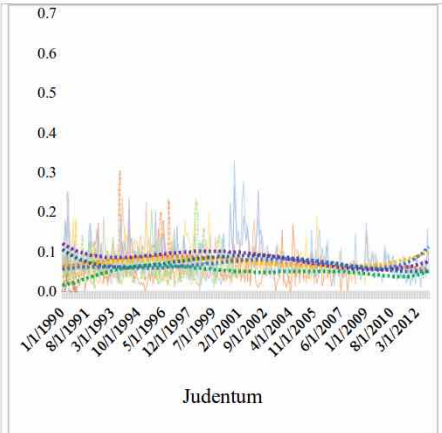
Figure: Anzahl publizierter Wertheriaden seit der Publikation des Originals in 1774.

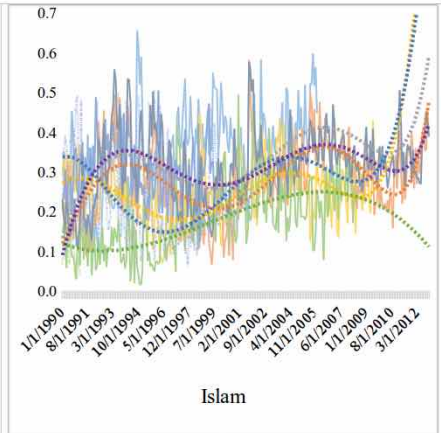
[1] Projektbeteiligte: Sandra Murr, Sandra Richter

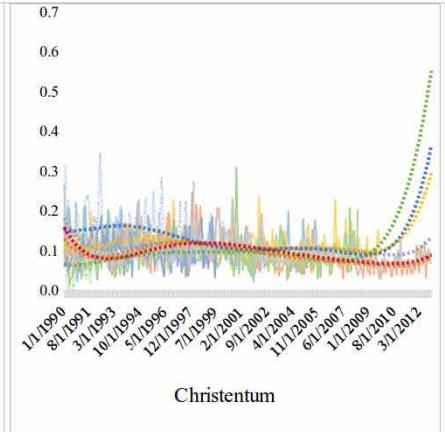
Love Triangle



Religion in International Newspaper Discourse[1]







Deutsche Presse

No.	Search result	Freq.
1	Muslime	9094
2	islamischen	7641
3	Islam	6953
4	Islamisten	6109
5	muslimischen	4593
6	islamische	4168
7	Schiiten	3271
8	Muslimen	3247
9	islamistischen	2886
10	Dschihad	2617
11	muslimische	2510
12	Moschee	2015
13	schiitischen	1995
14	islamistische	1871
15	Sunniten	1708
16	islamischer	1302
17	sunnitischen	1299
18	schiitische	1283
19	Moscheen	1103
20	Islams	977

US-Amerikanische Presse

No.	Search result	Freq.
1	Islamic	45276
2	Muslim	43293
3	Muslims	28619
4	Shiite	28498
5	Islam	12208
6	mosque	9625
7	Shiites	8860
8	Jihad	7669
9	Sunnis	6512
10	Islamist	5675
11	mosques	3599
12	Islamists	2689
13	Imam	1989
14	Muslim-led	1902
15	jihadists	1249
16	jihadist	1197
17	Muslim-Croat	1066
18	of Ramadan	983
19	Allah	931
20	Mecca	862

Französische Presse

No.	Search result	Freq.
1	Musulmans	10251
2	islamistes	10144
3	islamique	8540
4	islamiste	5873
5	chiites	5489
6	chiite	5319
7	islam	4852
8	sunnites	3702
9	djihad	3412
10	Musulman	3402
11	sunnite	2850
12	musulmane	2704
13	mosquée	2144
14	islamiques	1649
15	djihadistes	1579
16	mosquées	1284
17	imam	1234
18	musulmanes	1186
19	islamisme	1179
20	Allah	761

Britische Presse

No.	Search result	Freq.
1	Muslim	15300
2	Islamic	12368
3	Muslims	10973
4	Shia	5743
5	Islam	4193
6	Islamist	4169
7	jihad	3294
8	mosque	2821
9	Islamists	1552
10	Sunnis	1307
11	mosques	1054
12	Shias	995
13	Imam	640
14	Allah	493
15	Muslim-Croat	416
16	jihadists	390
17	jihadis	342
18	Jihadi	337
19	Mecca	336
20	jihadist	332

[1] Project mebers: Maximilian Overbeck, Cathleen Kantner

Concepts and their Connections in Adorno[1]

"Interessant ist, daß sich mir bei der Arbeit aus dem Inhalt der Gedanken gewisse Konsequenzen für die Form aufdrängen, die ich längst erwartete, aber die mich nun doch überraschen. Es handelt sich ganz einfach darum, daß aus meinem Theorem, daß es philosophisch nichts 'Erstes' gibt, nun auch folgt, daß man nicht einen argumentativen Zusammenhang in der üblichen Stufenfolge aufbauen kann, sondern daß man das Ganze aus **einer Reihe von Teilkomplexen montieren muß, die gleichsam gleichgewichtig sind und konzentrisch angeordnet**, auf gleicher Stufe; deren Konstellation, nicht die Folge, muß die Idee ergeben."

[1] Project member: Axel Pichler

Concepts and their Connections in Adorno[1]

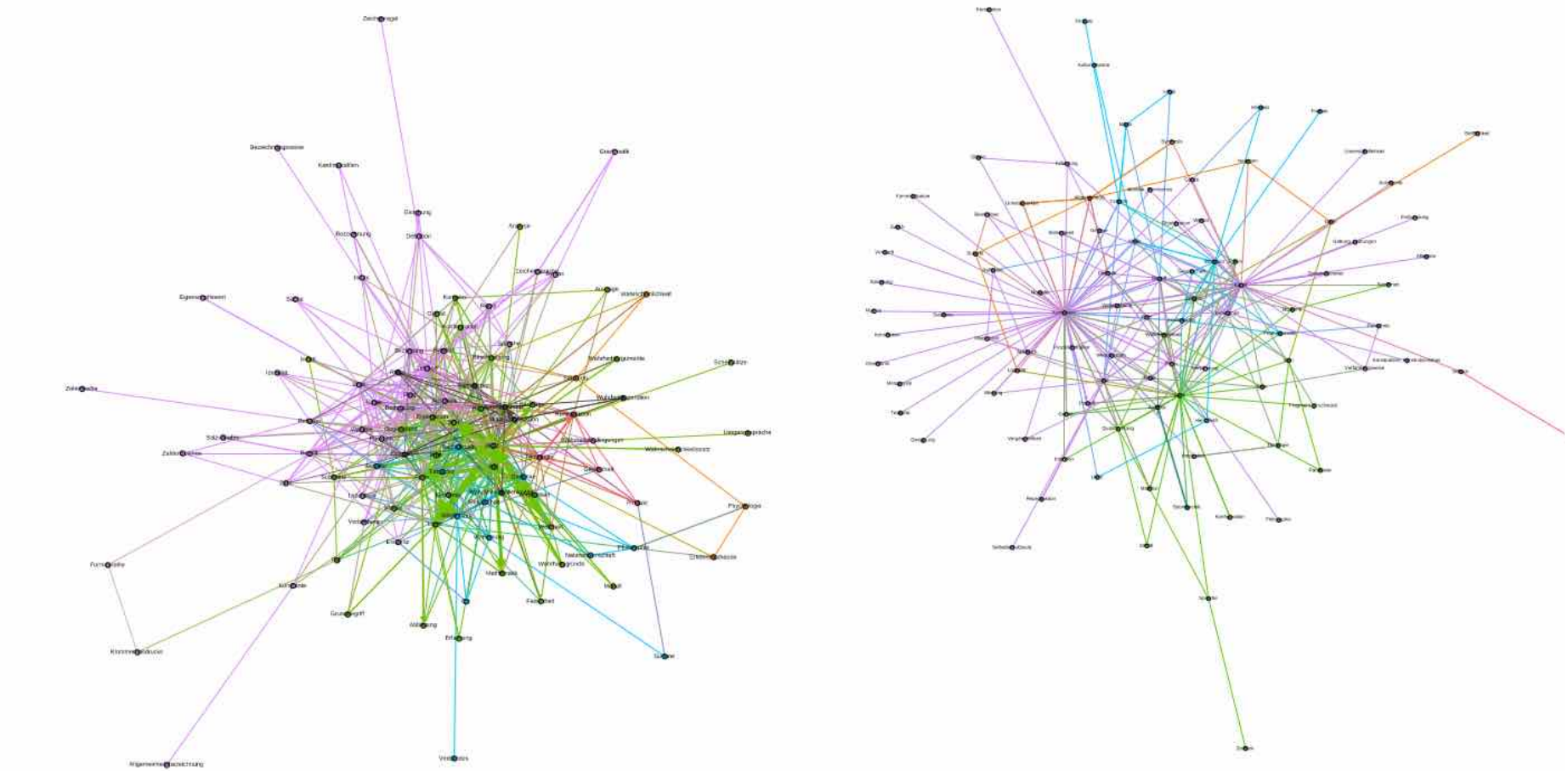


Figure: Term networks in Wittgenstein's Tractatus (left) and Adorno's "Ästhetische Theorie" (right)

[1] Project member: Axel Pichler

Common interest

- Individually distinguishable objects in the real or a fictional world
  - “Individually distinguishable” = by naming
  - “objects” does not imply physical objects
- References to such entities in texts
  - Proper names (“Angela Merkel”)
  - Descriptive noun phrases (“the chancellor”)
  - Pronouns (“she”)

Entities

- Parzival
  - Characters und groups of characters (“three knights”)
  - Locations
- Werther
  - Characters: Werther, Lotte, Albert
- Parliamentary debates
  - Political parties, international organisations (“EU commission”, “United Nations general assembly”)
- Aesthetic Theory
  - Philosophers, works, abstract concepts
- Dramatic Texts
  - Characters: Hamlet, Sara Sampson, the priest

Annotating Coreference

Coreference

Definition and Examples

- If two linguistic units co-refer, they refer to the same non-textual entity
  - [A woman] bought a house. [The woman] lived happily ever after. (prototypical)
  - [A woman] bought a house. [She] lived happily ever after. (anaphoric)
  - [Mary] bought a house. [The woman] enjoyed the garden. (coreferential)
  - [Mary likes to sing in the shower.] Her boyfriend hates [that]. (non-nominal antecedent)

Coreference

Definition and Examples

- If two linguistic units co-refer, they refer to the same non-textual entity
  - [A woman] bought a house. [The woman] lived happily ever after. (prototypical)
  - [A woman] bought a house. [She] lived happily ever after. (anaphoric)
  - [Mary] bought a house. [The woman] enjoyed the garden. (coreferential)
  - [Mary likes to sing in the shower.] Her boyfriend hates [that]. (non-nominal antecedent)
- Complications
  - Embedded mentions: [[Donald Trumps] Gouvernement]
  - Plural mentions: [John] was cooking for [Mary]. [They] enjoyed it.
  - Discontinuity/relative clauses: Ich habe [die Sendungen] zur Post gebracht, [[die] hier schon lange rumliegen].

Coreference

Two tasks

- Mention detection
  - Find all linguistic units that refer (to something)
  - Named entity detection + X
    - Noun phrases (the gardener)
      - Almost all NPs can be used both non-referring and referring
    - Pronouns (she)
      - Trivial, since it's a fixed list
- Coreference resolution
  - Group mentions that co-refer
  - Most simple approach: Classification of mention pairs

Coreference

Two tasks

- Mention detection
  - Find all linguistic units that refer (to something)
  - Named entity detection + X
    - Noun phrases (the gardener)
      - Almost all NPs can be used both non-referring and referring
    - Pronouns (she)
      - Trivial, since it’s a fixed list

Annotating Coreference

Goals

- Create a **gold standard** for a couple of relevant texts
- Allows for experimentation of simple hacks (dramas) as well as more complex systems
- Be *compatible* to standard coreference as much as possible
  - Reuse of annotation guidelines (Naumann, 2007)
  - Potential integration of tools
  - Potential combination of annotated data (→ domain adaptation)
- Apply the same/similar scheme to prose, drama, and non-literary texts

Annotating Coreference

Guidelines

- Singletons are not annotated
- Non-nominals are included (as antecedents)
- Deviations from standard coreference
  - Chains: Elements of a chain form an equivalence set
  - No annotation of link type (e.g., anaphoric vs. coreferential vs. cataphoric)
    - Can be detected automatically

Annotating Coreference

Corpus

- Prose
  - Goethe: *Die Leiden des Jungen Werther* (1787)
  - Perutz: *Zwischen neun und neun* (1918), *Nur ein Druck auf den Knopf* (1930), *Der Mond lacht* (1930)
  - Rowlandson: *Narrative of the Captivity and Restoration of Mrs. Mary Rowlandson* (1682, English)
  - Fairy tales
- Drama
  - Lessing: *Miß Sara Sampson* (Bürgerliches Trauerspiel)
  - Schiller: *Die Räuber* (Sturm und Drang)
  - Hofmannsthal: *Der Rosenkavalier* (?)
- Drama-like
  - Bundestag: Parliamentary debates (1996-2015)

Annotating Coreference

Annotation Issues

Annotation Issues

Use of generic noun phrases (1/2)

Der Wirt. […] Was liegt mir daran, ob ich es weiß, oder nicht, was Sie für eine Ursache hierher führt, und warum Sie bei mir im Verborgnen sein wollen? Ein Wirt nimmt sein Geld, und läßt seine Gäste machen, was ihnen gut dünkt. Waitwell hat mir zwar gesagt, […]

(Lessing, Miß Sara Sampson)

Annotation Issues

Use of generic noun phrases (1/2)

Der Wirt. […] Was liegt mir daran, ob ich es weiß, oder nicht, was Sie für eine Ursache hierher führt, und warum Sie bei mir im Verborgnen sein wollen? Ein Wirt nimmt sein Geld, und läßt seine Gäste machen, was ihnen gut dünkt. Waitwell hat mir zwar gesagt, […]

(Lessing, Miß Sara Sampson)

$[\text{Sie}]_1$ ließ $[[\text{ihren}]_2$ Regenschirm$]_3$ fallen. $[\text{Jeder junge Mann}]_4$ wird in einem solchen Fall blitzschnell nach $[\text{dem Schirm}]_5$ greifen und $[\text{ihn}]_6$ $[\text{der Dame}]_7$ überreichen. Und $[\text{die Dame}]_8$ bedankt sich vielmals. Aber diesmal geschah etwas Unerhörtes. $[\text{Stanislaus Demba}]_9$ ließ $[\text{den Schirm}]_{10}$ liegen.

(Perutz, Zwischen Neun und Neun)

Annotation Issues

Use of generic noun phrases (2/2)

- Generic expressions are relatively frequent
  - “[Ihr Mannspersonen] müßt doch selbst nicht wissen, was ihr wollt.”
  - “Mellefont besitzt alles, was uns [eine Mannsperson] gefährlich machen kann.”
  - Should generic expressions be co-referent (e.g., multiple references to the class `man`)?
    - Standard guidelines: No coreference between generic mentions
- Generic noun phrases not specific to literary texts
- Switching between generic and non-generic use is more frequent than in newspaper texts

Annotation Issues

Entity development and character perception

Minimalist fairy tale posters by Christian Jackson

- Characters change over the course of the text (frog prince)
- Characters in disguise (red riding hood: wolf poses as grandmother)

- Annotating reader knowledge possible in both cases
- What do we loose?
- What knowledge do we define?
  - Important for reaching inter-subjective annotations

Annotation Issues

World knowledge

Hermann. [...] Da Karl auf der Welt nichts mehr zu hoffen hatte, zog ihn der Hall von Friederichs siegreicher Trommel nach Böhmen. Erlaubt mir, sagte er [zum großen Schwerin], daß ich den Tod sterbe auf dem Bette der Helden [...].

(Schiller, Die Räuber)

- Friederich == großer Schwerin?

Annotation Issues

World knowledge

Hermann. [...] Da Karl auf der Welt nichts mehr zu hoffen hatte, zog ihn der Hall von Friederichs siegreicher Trommel nach Böhmen. Erlaubt mir, sagte er [zum großen Schwerin], daß ich den Tod sterbe auf dem Bette der Helden [...].

(Schiller, Die Räuber)

- Friederich == großer Schwerin?
  - No. Großer Schwerin = Kurt Christoph Graf von Schwerin
  - https://de.wikipedia.org/wiki/Kurt_Christoph_von_Schwerin

Annotation Issues

Lexical variation / coreference vs. bridging

Mellefont. Mit Unrecht tadelt sie die Verzögerung [einer Zeremonie] [...].

Sara. Neue Freunde sollen die Zeugen [unserer Verbindung] sein? [...]

Mellefont. Aber überlegen Sie denn nicht, Miss, dass [unserer Verbindung] hier diejenige Feier fehlen würde, die wir ihr zu geben schuldig sind?

(Lessing, Miß Sara Sampson)

Annotation Issues

Lexical variation / coreference vs. bridging

Mellefont. Mit Unrecht tadelt sie die Verzögerung [einer Zeremonie] [...].

Sara. Neue Freunde sollen die Zeugen [unserer Verbindung] sein? [...]

Mellefont. Aber überlegen Sie denn nicht, Miss, dass [unserer Verbindung] hier diejenige Feier fehlen würde, die wir ihr zu geben schuldig sind?

(Lessing, Miß Sara Sampson)

- "Verbindung" refers to both the wedding and the marriage
- Related to bridging: Reference to entities that are related to previously mentioned entities
  - Mary bought [a car]. [The engine] is strong.
    - (bridging relation between car and engine)

Annotation Issues

Colorful language

Mellefont. […] ich hatte noch keine verwahrlosete Tugend auf meiner Seele. Ich hatte noch keine Unschuld in ein unabsehliches Unglück gestürzt. Ich hatte noch keine Sara aus dem Hause eines geliebten Vaters entwendet, und sie gezwungen, einem Nichtswürdigen zu folgen, der auf keine Weise mehr sein eigen war.

- At utterance time, Mellefont has abducted Sara.
- “kein X”: Typically considered to be generic and therefore non-referring
- But clearly related to (and understood as) a non-generic entity (Sara)

Annotation Issues

Colorful language

Mellefont. […] ich hatte noch keine verwahrlosete Tugend auf meiner Seele. Ich hatte noch keine Unschuld in ein unabsehliches Unglück gestürzt. Ich hatte noch keine Sara aus dem Hause eines geliebten Vaters entwendet, und sie gezwungen, einem Nichtswürdigen zu folgen, der auf keine Weise mehr sein eigen war.

- At utterance time, Mellefont has abducted Sara.
- “kein X”: Typically considered to be generic and therefore non-referring
- But clearly related to (and understood as) a non-generic entity (Sara)

Sara. […] Wenn zum Exempel, [ein Mellefont] [eine Marwood] liebt, und sie endlich verläßt

- Proper nouns with indefinite article

Annotation Issues

Non-nominal antecedents

Mellefont. Eine schlechte Vorbereitung, eine [trost]suchende Betrübte zu empfangen. Warum sucht [ihn] auch bei mir? – Doch wo soll sie [ihn] sonst suchen?

- Only possible antecedent for „ihn“: trost
- Sub-token annotation!

Waitwell. Sara liebt ihren Vater noch. Wenn Sie nur [davon] überzeugt sein wollen

- „davon“ refers to the entire sentence

Annotation Issues

True ambiguities

"Vierzehn Personen, darunter zwei Professoren der Universität, […]
Der Universitätsprofessor versuchte es […]
[…] und einer von den Professoren versuchte […]
Einer von den Universitätsprofessoren trat hin […]"

(Perutz, Nur ein Druck auf den Knopf)

- Two references must be co-referential but we don't know which ones

Annotation Issues

True ambiguities

"Vierzehn Personen, darunter zwei Professoren der Universität, […]
Der Universitätsprofessor versuchte es […]
[…] und einer von den Professoren versuchte […]
Einer von den Universitätsprofessoren trat hin […]"

(Perutz, Nur ein Druck auf den Knopf)

- Two references must be co-referential but we don't know which ones
- Other examples:
  - Rhetorical "we" in parliamentary debates (we the opposition, we the CDU, we the Europeans, we the audience, we politicians, …)
  - Groupings in Schiller's *Räuber*: Who exactly belongs to the group at which point is impossible to determine

Annotating Coreference

Annotation Tool

Annotation Tool

Requirements

- Basics: Group mentions that refer to the same entity
- Handling of groups, discontinuous clauses
- Long texts
- A lot of entities
- Flexible Im- and Export
  - E.g., TEI/XML
- Handling of text structure (acts, scenes, who speaks, stage directions)

Annotation Tool

Failed but tried

- Annotators edit TEI/XML files directly

```
<sp who="#sir_william">
    <speaker>SIR WILLIAM.</speaker>
    <l> Hier <rs ref="#sara">meine Tochter</rs>? Hier in <rs xml:id="wirtshaus">
        diesem elenden Wirtshause</rs>?</l>
</sp>
```

  - Readily available
  - Well-formedness/validity can be ensured automatically
  - Slow, cumbersome, somewhat error-prone

Annotation Tool

Failed but tried

- Annotators edit TEI/XML files directly

```
<sp who="#sir_william">
    <speaker>SIR WILLIAM.</speaker>
    <l> Hier <rs ref="#sara">meine Tochter</rs>? Hier in <rs xml:id="wirtshaus">
        diesem elenden Wirtshause</rs>?</l>
</sp>
```

  - Readily available
  - Well-formedness/validity can be ensured automatically
  - Slow, cumbersome, somewhat error-prone
- WebAnno https://webanno.github.io
  - Conversion from/to TEI/XML
  - Web-based (distribution of documents, update of software)
  - Impossible to handle many long chains
  - UI unresponsive for long documents

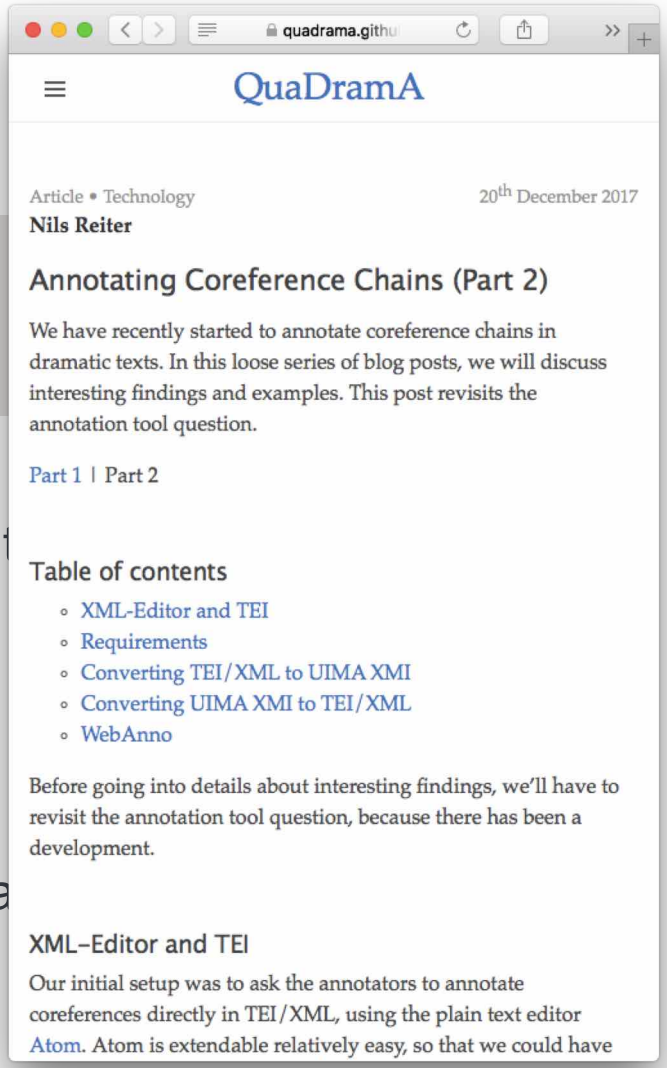
Annotation Tool

Failed but tried

- Annotators edit TEI/XML files directly

```
<sp who="#sir_william">
    <speaker>SIR WILLIAM.</speaker>
    <l> Hier <rs ref="#sara">meine Tochter</rs>? Hier
        diesem elenden Wirtshause</rs>?</l>
</sp>
```

  - Readily available
  - Well-formedness/validity can be ensured aut
  - Slow, cumbersome, somewhat error-prone
- WebAnno https://webanno.github.io
  - Conversion from/to TEI/XML
  - Web-based (distribution of documents, upda
  - Impossible to handle many long chains
  - UI unresponsive for long documents



Annotation Tool

CorefAnnotator

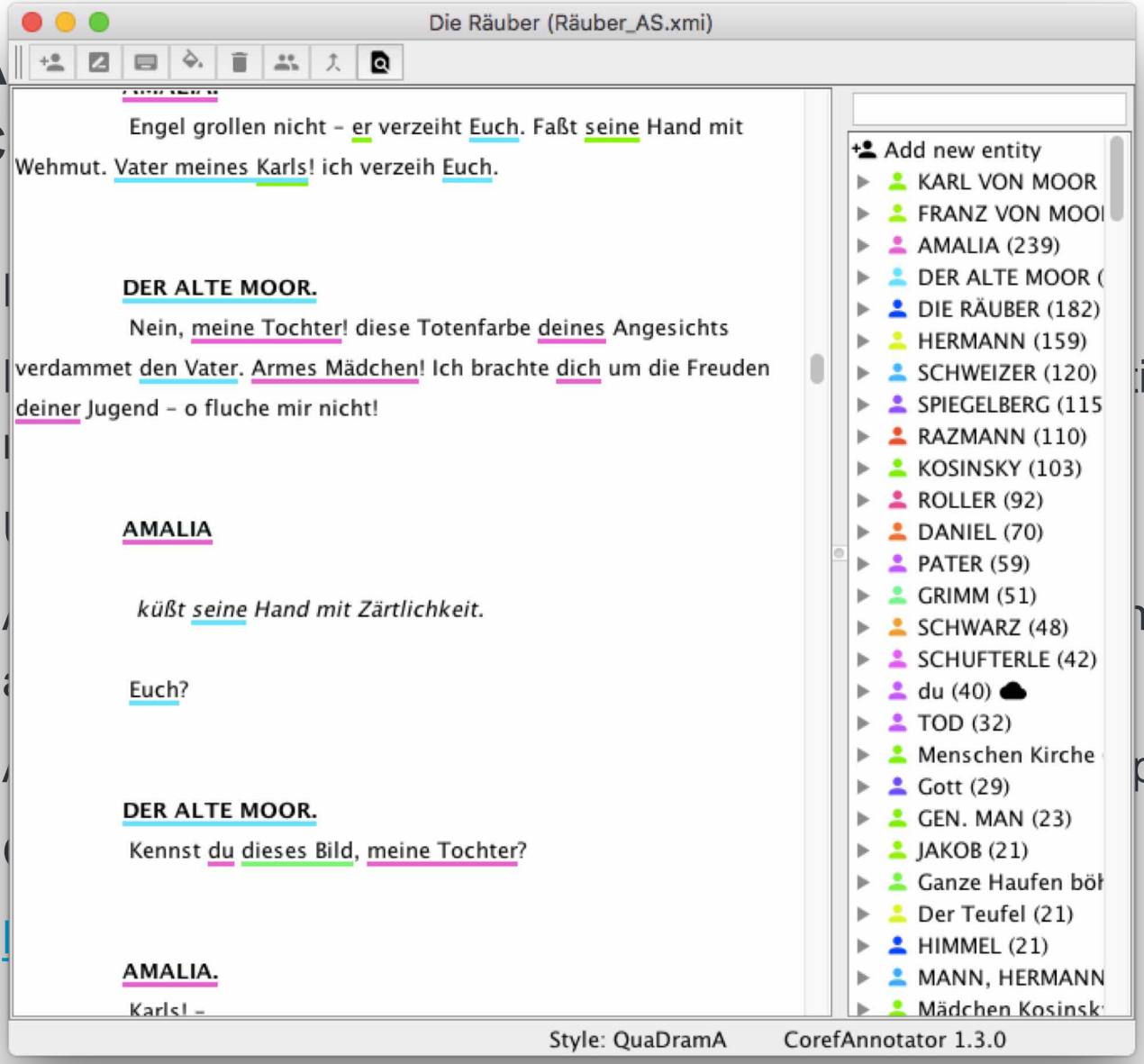
- Developed for long texts with many entities
- Entities represented as sets of mentions, no binary relations between mentions
- Useable with a keyboard for fast annotation
- Automatic generation of candidate entities based on context and previous annotations (semi-automatic annotation)
- Annotations stored in stand-off format using UIMA as a platform
  - Compatible to WebAnno
- Import/Export in various formats (CoNLL, TEI/XML*)
- Open Source: Apache License 2.0
- https://github.com/nilsreiter/CorefAnnotator

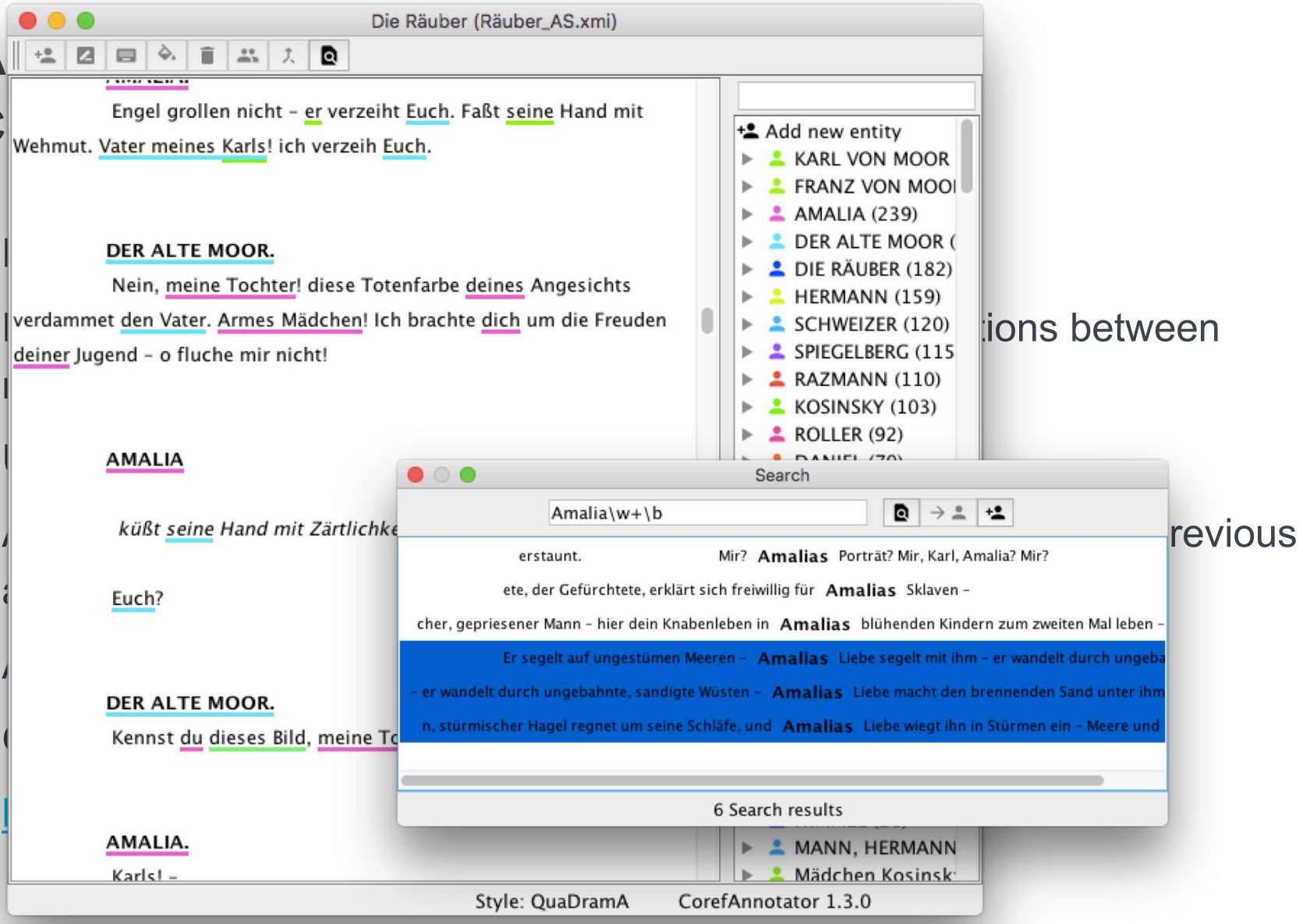
* Given some preconditions

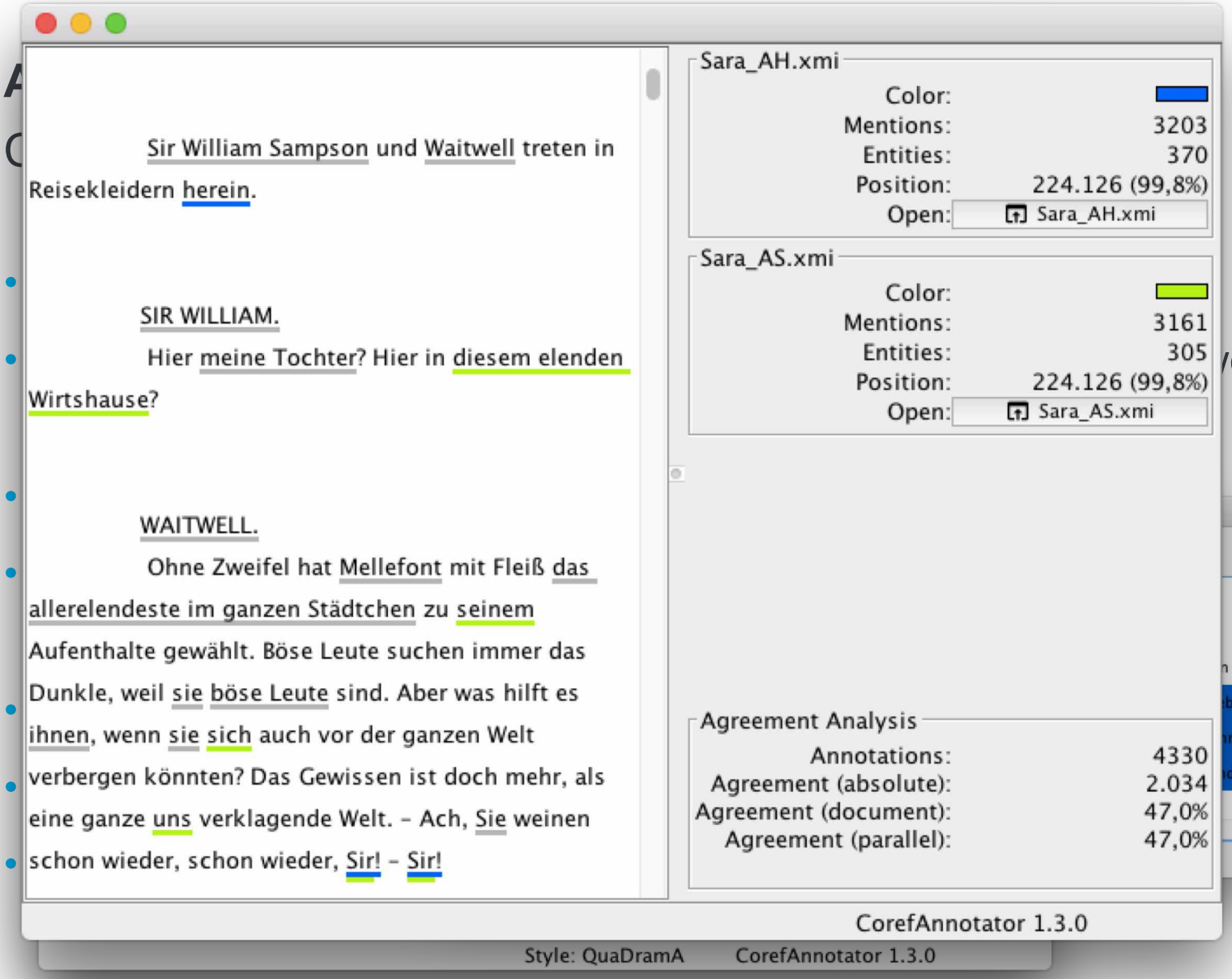
A

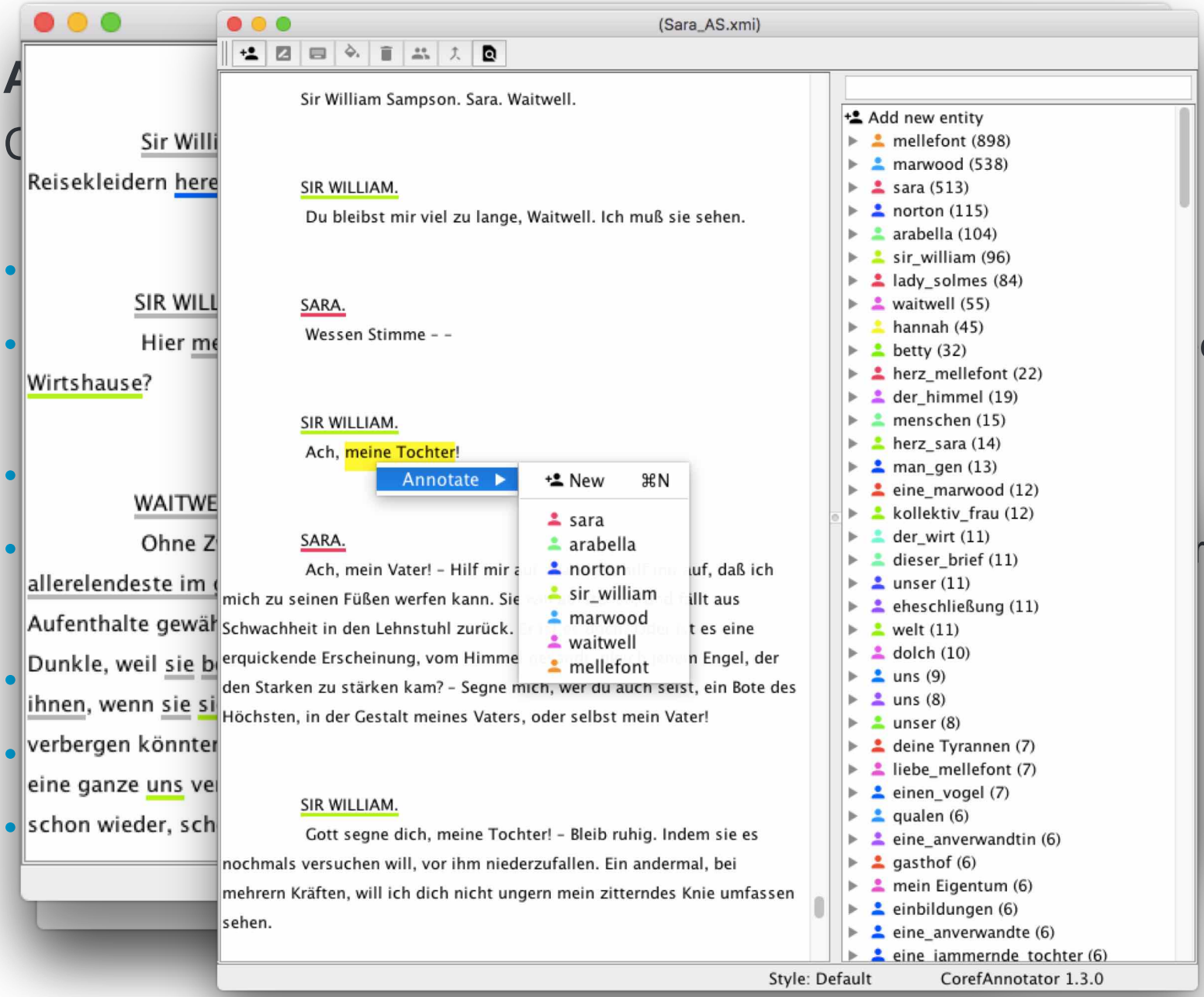
C

- [illegible]
- [illegible] tions between
- [illegible]
- [illegible] ntext and previous
- [illegible]
- [illegible] platform
- [illegible]
- [illegible]




Die Räuber (Räuber_AS.xmi)
Engel grollen nicht – er verzeiht Euch. Faßt seine Hand mit Wehmut. Vater meines Karls! ich verzeih Euch.
DER ALTE MOOR.
Nein, meine Tochter! diese Totenfarbe deines Angesichts verdammet den Vater. Armes Mädchen! Ich brachte dich um die Freuden deiner Jugend – o fluche mir nicht!
AMALIA
küßt seine Hand mit Zärtlichke
Euch?
DER ALTE MOOR.
Kennst du dieses Bild, meine To
AMALIA.
Karls! –
Add new entity
KARL VON MOOR
FRANZ VON MOOI
AMALIA (239)
DER ALTE MOOR (
DIE RÄUBER (182)
HERMANN (159)
SCHWEIZER (120)
SPIEGELBERG (115
RAZMANN (110)
KOSINSKY (103)
ROLLER (92)
MANN, HERMANN
Mädchen Kosinsk
Search
Amalia\w+\b
erstaunt. Mir? Amalias Porträt? Mir, Karl, Amalia? Mir?
ete, der Gefürchtete, erklärt sich freiwillig für Amalias Sklaven –
cher, gepriesener Mann – hier dein Knabenleben in Amalias blühenden Kindern zum zweiten Mal leben –
Er segelt auf ungestümen Meeren – Amalias Liebe segelt mit ihm – er wandelt durch ungeba
– er wandelt durch ungebahnte, sandigte Wüsten – Amalias Liebe macht den brennenden Sand unter ihm
n, stürmischer Hagel regnet um seine Schläfe, und Amalias Liebe wiegt ihn in Stürmen ein – Meere und
6 Search results
Style: QuaDramA
CorefAnnotator 1.3.0


Sir William Sampson und Waitwell treten in Reisekleidern herein.
SIR WILLIAM.
Hier meine Tochter? Hier in diesem elenden Wirtshause?
WAITWELL.
Ohne Zweifel hat Mellefont mit Fleiß das allerelendeste im ganzen Städtchen zu seinem Aufenthalte gewählt. Böse Leute suchen immer das Dunkle, weil sie böse Leute sind. Aber was hilft es ihnen, wenn sie sich auch vor der ganzen Welt verbergen könnten? Das Gewissen ist doch mehr, als eine ganze uns verklagende Welt. – Ach, Sie weinen schon wieder, schon wieder, Sir! – Sir!
Sara_AH.xmi
Color:
Mentions: 3203
Entities: 370
Position: 224.126 (99,8%)
Open: Sara_AH.xmi
Sara_AS.xmi
Color:
Mentions: 3161
Entities: 305
Position: 224.126 (99,8%)
Open: Sara_AS.xmi
Agreement Analysis
Annotations: 4330
Agreement (absolute): 2.034
Agreement (document): 47,0%
Agreement (parallel): 47,0%
CorefAnnotator 1.3.0
Style: QuaDramA
CorefAnnotator 1.3.0


(Sara_AS.xmi)
Sir William Sampson. Sara. Waitwell.
SIR WILLIAM.
Du bleibst mir viel zu lange, Waitwell. Ich muß sie sehen.
SARA.
Wessen Stimme – –
SIR WILLIAM.
Ach, meine Tochter!
Annotate
New ⌘N
sara
arabella
norton
sir_william
marwood
waitwell
mellefont
SARA.
Ach, mein Vater! – Hilf mir a
uf, daß ich
mich zu seinen Füßen werfen kann. Sie
llt aus
Schwachheit in den Lehnstuhl zurück.
t es eine
erquickende Erscheinung, vom Himme
n Engel, der
den Starken zu stärken kam? – Segne mich, wer du auch seist, ein Bote des
Höchsten, in der Gestalt meines Vaters, oder selbst mein Vater!
SIR WILLIAM.
Gott segne dich, meine Tochter! – Bleib ruhig. Indem sie es
nochmals versuchen will, vor ihm niederzufallen. Ein andermal, bei
mehrern Kräften, will ich dich nicht ungern mein zitterndes Knie umfassen
sehen.
Add new entity
mellefont (898)
marwood (538)
sara (513)
norton (115)
arabella (104)
sir_william (96)
lady_solmes (84)
waitwell (55)
hannah (45)
betty (32)
herz_mellefont (22)
der_himmel (19)
menschen (15)
herz_sara (14)
man_gen (13)
eine_marwood (12)
kollektiv_frau (12)
der_wirt (11)
dieser_brief (11)
unser (11)
eheschließung (11)
welt (11)
dolch (10)
uns (9)
uns (8)
unser (8)
deine Tyrannen (7)
liebe_mellefont (7)
einen_vogel (7)
qualen (6)
eine_anverwandtin (6)
gasthof (6)
mein Eigentum (6)
einbildungen (6)
eine_anverwandte (6)
eine_jammernde_tochter (6)
Style: Default
CorefAnnotator 1.3.0

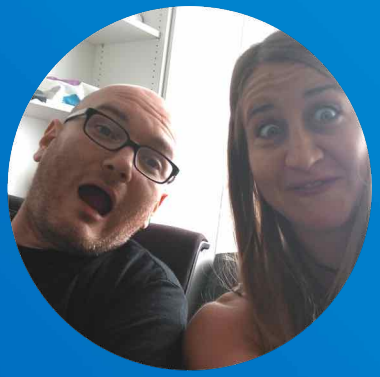
Summary

Summary

- Lessons learned
  - Clarify concepts through annotation and discussion
  - Annotation useful in DH, because close reading
  - Common sub tasks can be addressed across disciplines/research questions
  - Expert knowledge helpful for method development
  - Coreference fascinating task when applied to literary texts because it highlights the blind spots of linguistic theory / CL assumptions
- Collaboration
  - Two-way street
    - CL: New challenges for established workflows
    - DH: Quantification and inter-subjectivity

Universität Stuttgart

Vielen Dank!



Sarah Schulz, Nils Reiter

E-Mail firstname.lastname@ims.uni-stuttgart.de
Telefon +49 (0) 711 685-81354
Fax +49 (0) 711 685-81366

Universität Stuttgart
Institut für Maschinelle Sprachverarbeitung
Pfaffenwaldring 5b
70569 Stuttgart