

Melanie Andresen & Michael Vauth
melanie.andresen@uni-hamburg.de

**ADDED VALUE OF COREFERENCE
ANNOTATION FOR CHARACTER ANALYSIS
IN NARRATIVES**

Research Question

- What are the benefits of a time consuming **coreference annotation** for **character analysis**?
- Can we just base our analysis on proper nouns?

Character Analysis (in DH)

- Presence and copresence of characters
 - Where in the text does a character appear?
 - Which characters appear together frequently?
- Characterization
 - What are a character's properties?
 - Can we categorize the character (e. g. as the story's hero)?

(see Piper et al. 2017, Xanthos et al. 2016 for English,
Barth et al. 2018, Blessing et al. 2017, Krautter 2018 for German)

Coreference

[Sophies] Studentinnenzopf hüpfte fröhlich auf und ab, während [sie] beim Überfliegen des medizinischen Gutachtens vor sich hin nickte. [Sie] ist gut gelaunt, ohne besonderen Grund.

Case Study

Juli Zeh: Corpus Delicti (2009) about 46.000 tokens

Mitten am Tag, in der Mitte des Jahrhunderts

Rings um zusammengewachsene Städte bedeckt Wald die Hügelketten. Sendetürme zielen auf weiche Wolken, deren Bäuche schon lange nicht mehr grau sind vom schlechten Atem einer Zivilisation, die einst glaubte, ihre Anwesenheit auf diesem Planeten vor allem durch den Ausstoß gewaltiger Schmutzmengen beweisen zu müssen. Hier und da schaut das große Auge eines Sees, bewimpert von Schilfbewuchs, in den Himmel – stillgelegte Kies- und Kohlegruben, vor Jahrzehnten geflutet. Unweit der Seen beherbergen stillgelegte Fabriken Kulturzentren; ein Stück stillgelegter



picture: <https://www.amazon.de/Corpus-Delicti-Prozess-Juli-Zeh/dp/3442740665>

Coreference Annotation:

- CorefAnnotator by Nils Reiter
(<https://doi.org/10.5281/zenodo.1228105>)

Data Annotation

Dystopie Corvus Delicti (Lizli_Eph).txt (juli.zent.mergle.txt)

Datei Bearbeiten Ansicht Werkzeuge Einstellungen Hilfe

sie. »Scheint mir nicht recht zu Ihnen zu passen. Haben Sie keine Angst vor meinem Infektionspotential?« »Das Wichtigste im Leben ist Stil, Mia Holl. Und Hysterie ist die schlimmste Feindin des guten Stils.« »Ihr Gesicht«, sagt Mia nachdenklich, »ist wohl eine Art Etikett. Man kann es auf die unterschiedlichsten Ansichten kleben.« »Darf ich reinkommen?« »Sie verlangen, dass ich dem Mörder meines Bruders etwas zu trinken anbiete?« »Durchaus nicht. Für eine so plumpe Einschätzung sind Sie zu intelligent. Aber etwas zu trinken hätte ich tatsächlich gern. Eine Tasse heißes Wasser.« Kramer geht an Mia vorbei in die Wohnung und steuert das Sofa an, auf dem die ideale Geliebte schnell zur Seite rutscht. Kaum hat Kramer sich hingesetzt, wirkt das Sofa wie für ihn gemacht. Den angewiderten Blick der idealen Geliebten bemerkt er nicht, was ausnahmsweise weniger an seiner Selbstsicherheit liegt als an der Tatsache, dass er die ideale Geliebte nicht lehen kann. »Nur der Vollständigkeit halber: Ich habe Ihren Bruder nicht ermordet. Wir sollten vielleicht eher fragen, woher er im Gefängnis die Angelschnur hatte, um sich aufzuhängen.« Mia steht mitten im Raum, hat die Arme gekreuzt und die Finger ins Fleisch der Oberarme gekrallt, als

- Neue Entität anlegen
- m: Mia Holl (2467)
- h: Heinrich Kramer (896)
- Rosentreter (646)
- Murtz Holl (634)
- Sophie (279)
- Die ideale Geliebte (185)
- j: Hutschneider (129)
- Bell (123)
- Driss (112)
- Würmer (94)
- Mia Holl and Murtz Holl (61)
- Mia Holl and Rosentreter (51)
- (Lizli) (45)
- Mia Holl and Heinrich Kramer
- Die Poltsche and Driss and Lizli
- Sibylle (35)
- Die Poltsche (34)
- Erzähler-Wir (25)
- Die METHODE (20)

Stil: Default CorefAnnotator 1.4.4

Coreference Annotation:

- CorefAnnotator by Nils Reiter
(<https://doi.org/10.5281/zenodo.1228105>)
- guidelines for coreference annotation described in Rösiger et al. (2018)
- restricted to the annotation of characters, i. e. mentions of humans (roughly)
- four annotators (single annotation)
- discussion of difficult or ambiguous instances

Automatic Annotation:

- Part-of-speech
- Dependency syntax

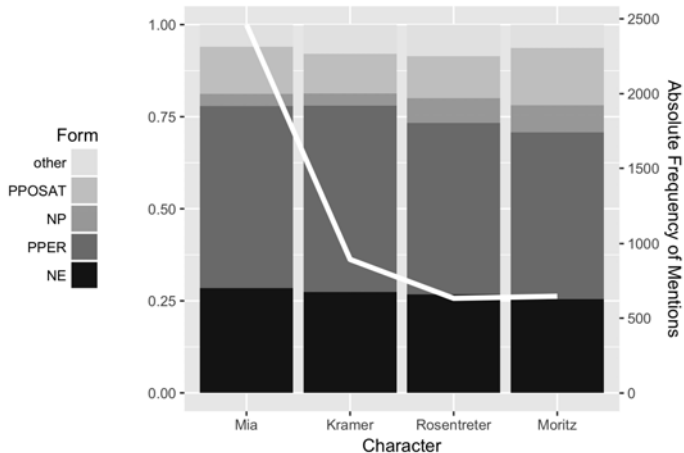
List of character mentions with information on

- the token span,
- the entity it refers to,
- the linguistic form (proper name, pronoun...),
- whether it occurs inside direct speech (detected by quotes) and
- the chapter in which it occurs.

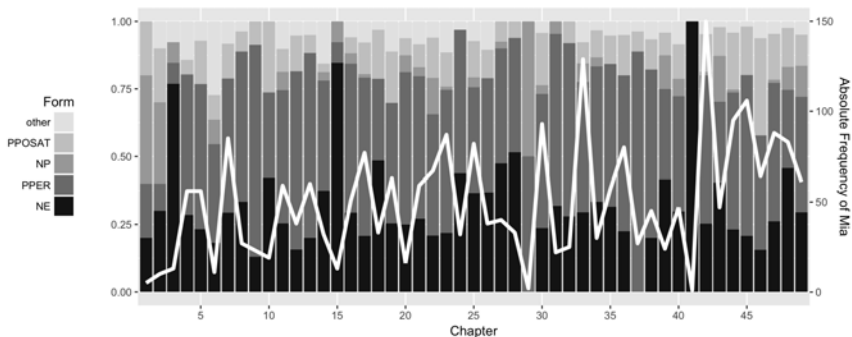
Download: <https://doi.org/10.5281/zenodo.1239701>.

Results

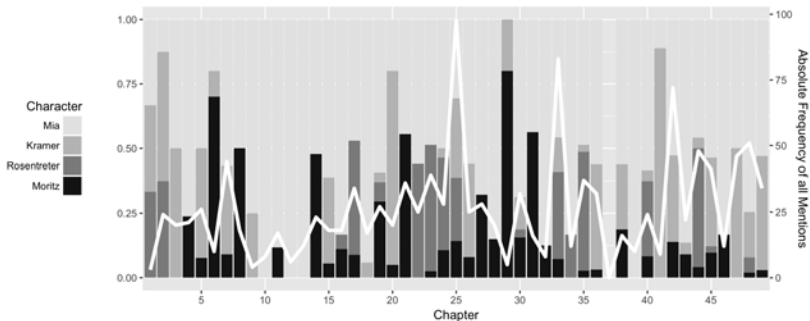
Form of Mentions



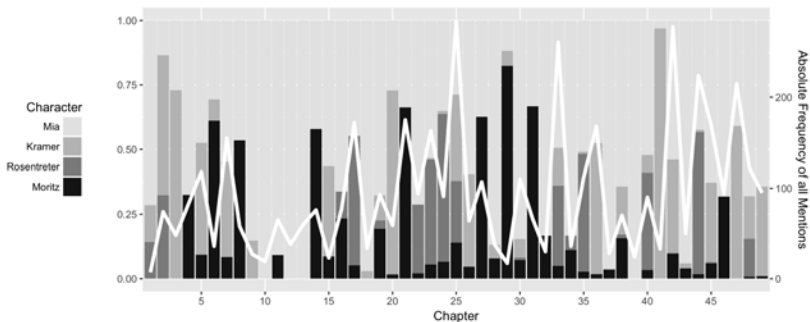
Mia across the Novel



Proper Names Only



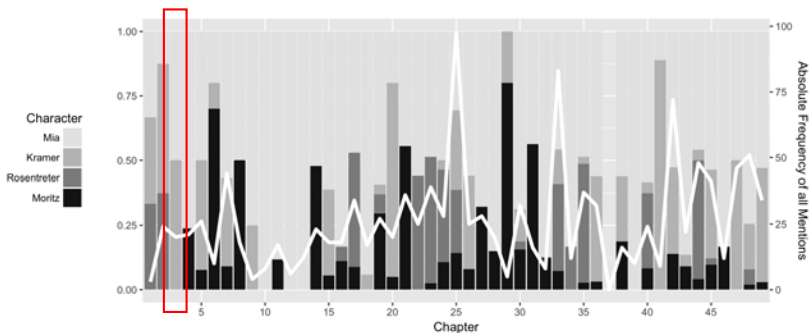
Coreference Annotation



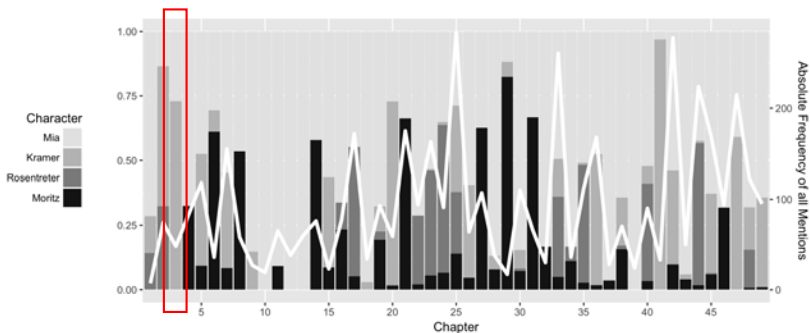
Correlation between the two conditions:

Mia: 0.87 – Kramer: 0.94 – Rosentreter: 0.94 – Moritz: 0.90

Proper Names Only



Coreference Annotation



Example (Chapter 3)

References to Mia Holl:

»Frau Holl wohnt ganz oben. Terrasse nach hinten.« »Tolle Wohnung«, sagt die Pollsche. »Mit der Biologie verdient man nicht schlecht.« »Zu Recht«, sagt Lizzie streng. »Schön«, sagt Kramer. »Und ist Frau Holl zu Hause?« »Immer!«, ruft Driss. »Zur Zeit, mein ich.« Sie beugt sich zu Kramer, als wolle sie ihm ein Geheimnis verraten. »Man sieht die Mia gar nicht mehr.« »Frau Mia Holl«, korrigiert Lizzie, »geht derzeit nicht arbeiten.« »Dann hat sie Urlaub?« »Ach was!«, platzt die Pollsche heraus. »So ein hübsches Kind und immer allein! Die guckt Angebote durch.« »Wir glauben«, sagt Lizzie vertraulich zu Kramer, »dass Frau Holl einen Partner sucht.« Kramer nickt. »Dann will ich mal.« »Die Mia ist eine Anständige.« »Das versteht sich doch von selbst, Driss.« »In einem Haus wie diesem.« »Danke.« Kramer nickt in die Runde, während er den Kreis der Nachbarinnen durchbricht. »Sie haben mir sehr geholfen. Und meinen Glückwunsch zu diesem schönen Haus.« Die Münder bleiben offen, aber stumm, während man Kramer und seinen Beinen und seiner ganzen elastischen Gestalt beim Treppensteigen zusieht.

Example (Chapter 3)

References to Kramer:

»Frau Holl wohnt ganz oben. Terrasse nach hinten.« »Tolle Wohnung«, sagt die Pollsche. »Mit der Biologie verdient man nicht schlecht.« »Zu Recht«, sagt Lizzie streng. »Schön«, sagt **Kramer**. »Und ist Frau Holl zu Hause?« »Immer!«, ruft Driss. »Zur Zeit, mein ich.« Sie beugt sich zu **Kramer**, als wolle sie **ihm** ein Geheimnis verraten. »Man sieht die Mia gar nicht mehr.« »Frau Mia Holl«, korrigiert Lizzie, »geht derzeit nicht arbeiten.« »Dann hat sie Urlaub?« »Ach was!«, platzt die Pollsche heraus. »So ein hübsches Kind und immer allein! Die guckt Angebote durch.« »Wir glauben«, sagt Lizzie vertraulich zu **Kramer**, »dass Frau Holl einen Partner sucht.« **Kramer** nickt. »Dann will **ich** mal.« »Die Mia ist eine Anständige.« »Das versteht sich doch von selbst, Driss.« »In einem Haus wie diesem.« »Danke.« **Kramer** nickt in die Runde, während **er** den Kreis der Nachbarinnen durchbricht. »Sie haben **mir** sehr geholfen. Und **meinen** Glückwunsch zu diesem schönen Haus.« Die Münder bleiben offen, aber stumm, während man **Kramer** und **seinen** Beinen und seiner ganzen elastischen Gestalt beim Treppensteigen zusieht.

Example (Chapter 3)

- Proper names partly cover **third person mentions** of a character
- Mentions in **first and second person** are not covered
- We might miss or underrepresent a direct conversation between two characters.
- However, this is a typical case of character interaction.

Characterization by Noun Phrases

Noun phrases referring to Mia:

| Noun Phrase | Translation | Frequency |
|--------------------|--------------------|------------------|
| Angeklagte | defendant | 32 |
| Schwester | sister | 7 |
| Beschuldigte | accused | 7 |
| Verurteilte | convicted | 6 |
| Mandantin | client | 4 |

Noun phrases referring to Moritz:

43 of 47 have the head Bruder ('brother')

Conclusions

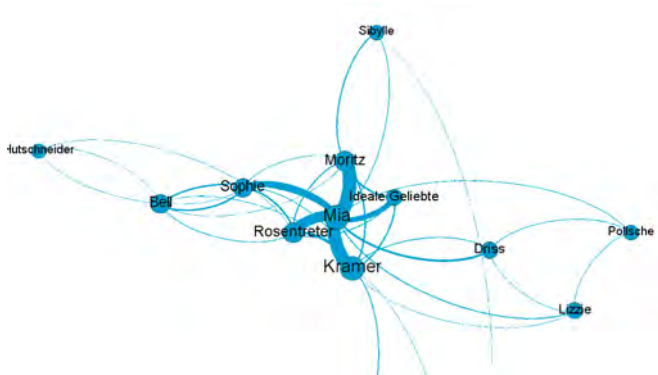
Conclusions

- Distribution of proper names (as a measure of character presence) is biased.
 - Mentions in first and second person are often not accompanied by proper names.
 - Coreference annotation greatly enhances possibilities of characterization.
 - more contexts → more context information
- Coreference annotation is highly beneficial,
- but not feasible for large corpora.

Future Work

- Multivariate model to further investigate interaction of variables
- Broaden dataset (four novels, two historic and two contemporary)
- Create character networks of the novel (Andresen and Vauth in preparation)

Future Work



Future Work

- Multivariate model to further investigate interaction of variables
- Broaden dataset (four novels, two historic and two contemporary)
- Create character networks of the novel (Andresen and Vauth in preparation)
- Characterization by non-verbal predicates (Andresen, Krüger, et al. submitted)

Explicit attributions by non-verbal predicates:

Mia is...

not a school girl

a scientist

a nihilist

a witness

a supporter of the METHOD

a saint

Kramer is...

a patient man

a machine

a fanatic

a media figure

a brilliant demagogue

a man of conviction

Acknowledgements

Thank you!

This work has been funded by the ‘Landesforschungsförderung Hamburg’ in the context of the *hermA* project (LFF-FV 35). We thank Lea Röseler and Daniel Fabian Klein for their help with the annotation and Piklu Gupta for checking our English. All remaining errors are our own.

References

- Andresen, Melanie, Katharina Krüger, Michael Vauth, and Heike Zinsmeister (submitted). *Can we describe a literary character by its explicit attributions based on syntactic annotation?*
- Andresen, Melanie and Michael Vauth (in preparation). *Figurenrelationen und Figurencharakterisierung. Interdisziplinarität zwischen Literaturwissenschaft und Computerlinguistik am Beispiel der Text- und Genreanalyse.*
- Barth, Florian, Evgeny Kim, Sandra Murr, and Roman Klinger (2018). "A Reporting Tool for Relational Visualization and Analysis of Character Mentions in Literature". In: *Book of Abstracts of DHD 2018*. Cologne, Germany, pp. 123–127.
- Blessing, Andre, Nora Echelmeyer, Markus John, and Nils Reiter (2017). "An End-to-End Environment for Research Question-Driven Entity Extraction and Network Analysis". In: *Proceedings of the Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*. Vancouver, Canada, pp. 57–67. DOI: 10.18653/v1/W17-2208.
- Krautter, Benjamin (2018). "Quantitatives „close Reading“? Vier Mikroanalytische Methoden Der Digitalen Dramenanalyse Im Vergleich". In: *Book of Abstracts of DHD 2018*. Cologne, Germany, pp. 295–300.
- Piper, Andrew, Mark Algee-Hewitt, Koustuv Sinha, Derek Ruths, and Hardik Vala (2017). "Studying Literary Characters and Character Networks". In: *Digital Humanities 2017, Conference Abstracts*. Montreal, Kanada, pp. 119–122.
- Rösiger, Ina, Sarah Schulz, and Nils Reiter (2018). "Towards Coreference for Literary Text: Analyzing Domain-Specific Phenomena". In: *Proceedings of LaTeCH-CLfL*.
- Xanthos, Aris, Isaac Pante, Yannick Rochat, and Martin Grandjean (2016). "Visualising the Dynamics of Character Networks". In: *Digital Humanities 2016: Conference Abstracts*. Kraków, pp. 417–419.