

Das Projekt CompAnno: Comparative Annotation to Explore and Explain Text Similarities

Flüh, Marie

marie.flueh@uni-hamburg.de
Universität Hamburg, Deutschland
ORCID: 0000-0002-1707-284X

Nantke, Julia

julia.nantke@uni-hamburg.de
Universität Hamburg, Deutschland

Pagel, Janis

jpagell@uni-koeln.de
Universität zu Köln, Deutschland

Reiter, Nils

nils.reiter@uni-koeln.de
Universität zu Köln, Deutschland
ORCID: 0000-0003-3193-6170

1. Comparative Annotation to Explore and Explain Text Similarities (CompAnno)

Das DFG-Projekt CompAnno entwickelt einen vergleichenden Annotationsworkflow zur computergestützten Detektion und Klassifizierung von literarischen Textähnlichkeiten am Beispiel von Figureneigenschaften als einer Kategorie, die für die Gestaltung literarischer Erzähltexte und für die Interpretation intertextueller Beziehungen zentral ist (Müller 1991:101). Der Workflow für eine computergestützte Untersuchung von Textähnlichkeit soll so gestaltet sein, dass er über die Erkennung von text-reuse hinausgeht und nicht auf ein festes Korpus bezogen ist. Gleichzeitig greifen wir mit der vergleichenden Annotation eine literaturwissenschaftliche Basismethode auf (Unsworth 2000, Epple et al. 2020:7) und entwickeln einen neuen Weg für die Arbeit mit interpretativen Kategorien.

2. Annotationsaufgaben

Wir arbeiten mit vier Annotatorinnen, wobei die Annotationsaufgaben zur Annotation von Figureneigenschaften ineinandergreifen. Die Annotatorinnen sind alle Germanistik-

und Linguistikstudentinnen. Da Figuren in der Regel zu Beginn eines Textes eingeführt werden, ist hier mit besonders zahlreichen Eigenschaften zu rechnen. Deshalb werden in unterschiedlichen Annotationsphasen jeweils die Anfangspassagen (circa 20.000 Tokens) aus verschiedenen Prosatexten aus *d-Prose* (1870–1920, Gius et al. 2021) annotiert.

2.1 Explorative Annotation mit CATMA

Um einen differenzierten und spezifischen Blick auf das Phänomen zu erlangen und eine möglichst große Bandbreite an Figureneigenschaften zu ermitteln, wird zunächst explorativ annotiert (vgl. Pagel et al. 2020: 127). Die qualitative Auswertung der Annotationsdaten bildet die Grundlage für Richtlinien für die manuelle Annotation von Figureneigenschaften.

2.2 Automatisierungsorientierte Annotation

Figureneigenschaften werden erst manuell ermittelt und kategorisiert. Auf Grundlage der Guidelines ist jede Annotatorin für die Analyse einer ausgewählten Kategorie zuständig. In Diskussionsrunden werden die Annotationsdaten besprochen und angepasst. Darauf aufbauend wird ein Ranking der Ähnlichkeiten erstellt, das zum Trainieren oder Prompten eines maschinellen Lernsystems bzw. großen Sprachmodells verwendet wird. Ziel dieses Modells ist die automatische Erkennung von Stellen, an denen eine Figureneigenschaft vorkommt, sowie die Kategorie der Figureneigenschaft, so dass eine vergleichende Annotation Sinn ergibt.

2.3 Vergleichende Annotation mit PhiTag

Im Gegensatz zu etablierten Annotationsansätzen beruht die vergleichende Annotation auf der gleichzeitigen Betrachtung mehrerer Textausschnitte: Auf Grundlage von Richtlinien für die vergleichende Annotation von Figureneigenschaften werden den Annotatorinnen jeweils zwei Textabschnitte vorgelegt, zu denen dann die ihnen enthaltenen Figureneigenschaften vergleichend zu annotieren sind. Benutzt wird hierzu die Webanwendung *PhiTag* (Schlechtweg, Kotchourko o.D.), die ursprünglich für die Annotation semantischer Ähnlichkeiten entwickelt wurde (Schlechtweg et al. 2020) und durch ihre quelloffene und modular erweiterbare Schnittstelle eine individuelle, projektspezifische Erweiterung für das vergleichende Annotieren unterstützt (s. Abb. 1).

Im weiteren Projektverlauf sollen die vergleichenden Annotationen dazu benutzt werden, die zuvor genannten intertextuellen Beziehungen zu beleuchten.



Abbildung 1: Screenshot eines zu annotierenden Textpaares in *PhiTag* mit Angabe von Unterkategorie und Referenzgröße; Textpaare stammen aus unterschiedlichen Texten, zum Vergleich vorgesehene Eigenschaften (rot markiert) stammen aus derselben Unterkategorie und werden auf eine Skala von "Gegensatz" zu "identisch" verortet

3. Erste Ergebnisse

Bisher entstandene Teilergebnisse sind Annotationsrichtlinien, Einblicke in die Annotationspraxis, qualitative und quantitative Einblicke in die Darstellung und Verteilung von Figureneigenschaften sowie erste Automatisierungsversuche.

3.1 Guidelines zur manuellen Annotation und zur vergleichenden Annotation von Figureneigenschaften

Aus der induktiven Auswertung der explorativen Annotationsphase, in der unter Einbezug etablierter Figurenkonzepte (Forster 1949, Hansen 2000, Jannidis 2004) vor allem konzeptuelle Fragestellungen im Fokus standen, ergeben sich fünf bzw. sechs Oberkategorien, die häufig für die Beschreibung literarischer Figuren verwendet werden und deshalb als Analysekategorien für die automatisierungsorientierte und vergleichende Annotation in Frage kommen (s. Abb. 2).

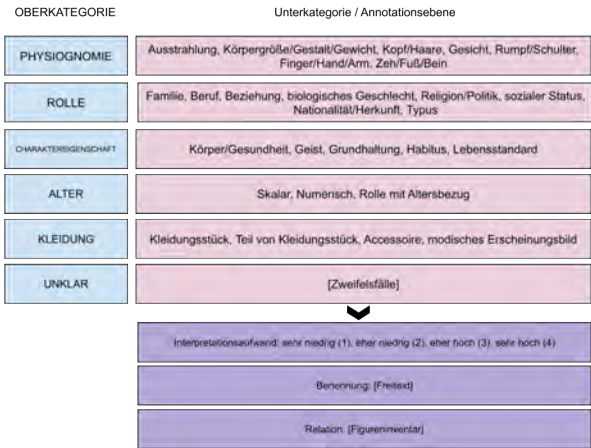


Abbildung 2: Schematischer Aufbau des Tagsets zur Annotation von Figureneigenschaften (mit Titeln der in den Guidelines detailliert definierten Ober- und Unterkategorien)

3.2 Quantitative und qualitative Auswertung der Annotationen und der Annotationsverfahren

Bisher zeigt sich, dass Figuren vor allem über Rollen und Charaktereigenschaften näher beschrieben werden (s. Tabelle 1).

Tabelle 1: Anzahl der Annotationen pro Kategorie in vier Beispieltexten

Kategorie	Annotationen
Rolle	1206
Charakter	328
Alter	289
Physiognomie	122
Kleidung	77

Für jede Annotation legen die Annotatorinnen den Interpretationsaufwand auf einer Skala von sehr niedrig bis sehr hoch fest. Ausgehend von der Annahme, dass ein geringer Interpretationsgrad auf explizit im Text thematisierte Figureneigenschaften und ein hoher Interpretationsgrad auf implizite Figureneigenschaften hindeutet, fungiert er als Marker für den Grad der Explizitheit.

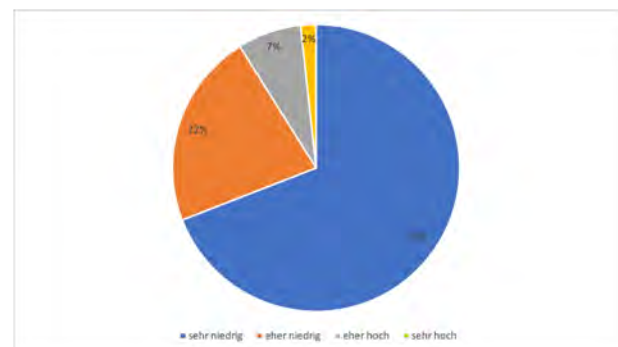


Abbildung 3: Interpretationsgrad für vier Beispieltexte im Überblick

Bei der Auswertung eines Teils der Annotationsdaten zeigt sich, dass der Interpretationsgrad meistens als gering eingeschätzt wird. Dieser Befund ist als individuelle Annotationsentscheidung aufzufassen. Außerdem lässt sich schlussfolgern, dass Eigenschaften vor allem explizit erwähnt werden (s. Abb. 3) und implizite Eigenschaften eher ein Randphänomen darstellen. Ein niedriger Interpretationsgrad findet sich vor allem in den Kategorien “Kleidung”, “Alter” und “Physiognomie” und ein höherer in den Kategorien “Rolle” und “Charakter”.

Eigenschaften, für die ein niedriger Interpretationsgrad angegeben wurde (explizite Eigenschaften), sind in allen Texten häufig. Sie sind ‘hochgradig intertextuell’ und relativ generisch (bspw. “jung”, “schön” oder “groß”). Eigenschaften, die mit einem hohen Interpretationsgrad ausgezeichnet wurden, kommen in einzelnen Texten und in geringerer Anzahl vor. Implizite Eigenschaften scheinen individuell zu sein, können aber gerade deshalb eine spezifischere Verbindung zwischen zwei Texten markieren als die explizit-generischen Eigenschaften.

4. Ansätze zur Automatisierung

Erste Pilotexperimente zur automatischen Klassifizierung von Figureneigenschaften zeigen, dass zwar schon moderat gute Ergebnisse mit relativ simplen Methoden zu erreichen sind, aber die Performanz noch ausbaufähig ist. Tabelle 2 zeigt die Durchschnittswerte für Precision, Recall und F1-Score für die Klassifikation von drei Modellen¹ (BERT (Devlin et al. 2019), ELECTRA (Clark et al. 2020) und RoBERTa (Liu et al. 2019)) für den Task in Sätzen mit Figureneigenschaft die korrekte Kategorie zu nennen. Benutzt wurden vier Texte, die Daten wurden in 80 % Trainingsdaten und 20 % Testdaten aufgeteilt. Alle Modelle wurden für 40 Epochen fine-tuned, mit einer Learning Rate von $4e-5$.

Die Ergebnisse zeigen, dass die Klassifikation von Charaktereigenschaft, Physiognomie und Rolle mit ca. 60 % F1-Score im ersten Anlauf akzeptabel funktioniert, die Modelle jedoch Probleme haben, “Alter” und “Kleidung” richtig zu klassifizieren (16 bzw. 36 % F1-Score), außerdem ist die Precision für Rolle mit 6 % deutlich niedriger als für die anderen Kategorien.

Tabelle 2: Klassifikationsergebnisse für das Erkennen von Figureneigenschaften

	Precision	Recall	F1-Score	Instanzen
Alter	0.25	0.12	0.16	11
Charaktereigenschaft	0.64	0.73	0.68	37
Kleidung	0.34	0.38	0.36	13
Physiognomie	0.63	0.74	0.68	26
Rolle	0.06	0.51	0.55	44

5. Ausblick

Geplante Arbeitspakete betreffen vor allem die Automatisierung der vergleichenden Annotation und die Verbesserung der automatischen Erkennung der Kategorien. Eine weitere offene Frage ist der ideale Kontext, den es braucht, um automatisiert Entscheidungen bezüglich der Figureneigenschaften zu treffen.

Bibliographie

Clark, Kevin, Minh-Thang Luong, Quoc V. Le und Christopher D. Manning. 2020. “ELECTRA: Pre-training Text Encoders as Discriminators Rather Than Generators”. In *International Conference on Learning Representations (ICLR) 2020*. <https://openreview.net/pdf?id=r1xMH1BtvB> (zugegriffen: 27. November 2024).

Devlin, Jacob, Ming-Wei Chang, Kenton Lee und Kristina Toutanova. 2019. “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding”. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, 4171–4186. <https://www.aclweb.org/anthology/N19-1423> (zugegriffen: 27. November 2024).

Epple, Angelika und Walter Erhart. 2020. “Practices of Comparing: A New Research Agenda Between Typological and Historical Approaches”. In *Practices of Comparing: Towards a new understanding of a fundamental human practice*, hg. von Angelika Epple, Walter Erhart und Johannes Grave, 11–38. Bielefeld: transcript.

Epple, Angelika, Walter Erhart und Johannes Grave. 2020. *Practices of comparing: Towards a new understanding of a fundamental human practice*. Bielefeld: transcript.

Forster, Edward Morgan. 1949. *Ansichten des Romans*. Berlin: Suhrkamp.

Gius, Evelyn, Svenja Guhr und Inna Uglanova. 2021. ““d-Prose 1870–1920” a Collection of German Prose Texts from 1870 to 1920”, In *Journal of Open Humanities Data*, 7(0), 11. <https://doi.org/10.5334/johd.30>.

Hansen, Per Krogh. 2000. *Die Rolle des Charakters. Aspekte einer literarischen Charakterologie*. PhD diss., Aalborg University.

Jannidis, Fotis. 2004. “Figur und Person. Beitrag zu einer historischen Narratologie”. In *Narratologia. Beiträge zur Erzähltheorie Band 3*, hg. von Fotis Jannidis, John Pier und Wolf Schmid. Berlin/New York: De Gruyter.

Liu, Yinhan, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer und Veselin Stoyanov. 2019. *RoBERTa: A Robustly Optimized BERT Pretraining Approach*. In *arXiv*. <https://doi.org/10.48550/arXiv.1907.11692>

Müller, Wolfgang G. 1991. “Interfiguralität. A Study on the Interdependence of Literary Figures”. In *Intertextuality*,

hg. von Heinrich. F. Plett, 101–122. Berlin/New York: De Gruyter.

Pagel, Janis, Nils Reiter, Ina Rösiger und Sarah Schulz. 2020. “Annotation als flexibel einsetzbare Methode.” In *Reflektierte algorithmische Textanalyse: Interdisziplinäre(s) Arbeiten in der Creta-Werkstatt*, hg. von Janis Pagel, Nils Reiter, Ina Rösiger und Sarah Schulz, 125–142. Berlin, Boston: De Gruyter. <https://doi.org/10.1515/9783110693973-006>.

Schlechtweg, Dominik, Barbara McGillivray, Simon Hengchen, Haim Dubossarsky und Nina Tahmasebi. 2020. SemEval-2020 Task 1: Unsupervised Lexical Semantic Change Detection . In *Proceedings of the Fourteenth Workshop on Semantic Evaluation*, pages 1–23, Barcelona (online). International Committee for Computational Linguistics.

Schlechtweg, Dominik und Serge Kotchourko (o.D.). PhiTag. Annotationsplattform . URL: <https://phitag.ims.uni-stuttgart.de/> (zugegriffen: 27. November 2024).

Unsworth, John. 2000. *Scholarly Primitives: What methods do humanities researchers have in common, and how might our tools reflect this?* <https://johnunsworth.name/Kings.5-00/primitives.html> (zugegriffen: 27. November 2024).