

Reflexive Passagen und ihre Attribution

Varachkina, Hanna

hanna.varachkina@stud.uni-goettingen.de
Georg-August-Universität Göttingen, Deutschland

Barth, Florian

barth@sub.uni-goettingen.de
Georg-August-Universität Göttingen, Deutschland

Gödeke, Luisa

luisa.goedeke@uni-goettingen.de
Georg-August-Universität Göttingen, Deutschland

Hofmann, Anna Mareike

annamareike.hofmann@uni-goettingen.de
Georg-August-Universität Göttingen, Deutschland

Dönicke, Tillmann

tillmann.doenicke@uni-goettingen.de
Georg-August-Universität Göttingen, Deutschland

Im Projekt MONA (Modes of Narration and Attribution) werden Phänomene in fiktionaler Literatur untersucht, die mit reflexiven Passagen assoziiert sind. Reflexive Passagen kommentieren die Handlung im Text oder den Schreibprozess oder generalisieren über die fiktive und / oder reale Welt. Da das Konzept der reflexiven Passagen in der Literaturwissenschaft bisher nicht formalisiert wurde, werden diese nicht direkt annotiert. Stattdessen annotieren wir drei Phänomene, die wir für starke Indikatoren reflexiver Passagen halten: Kommentar (Bonheim 1975; Chatman 1980), nicht-fiktionale Rede (Konrad 2017; Searle 1975) und Generalisierung (Leslie et al. 2016; Dönicke et al. 2021). Darüber hinaus beschäftigt sich das Projekt mit der Zuschreibung reflexiver Passagen zu Sprechinstanzen. Für die Identifikation und Klassifikation dieser Phänomene werden Modelle entwickelt. Dafür wird ein annotiertes Korpus deutschsprachiger fiktionaler Texte erstellt, das die Entwicklung dieser Phänomene über 350 Jahre der Literaturgeschichte abbildet.

Basierend auf den bisherigen Arbeiten haben wir Definitionen für die mit reflexiven Passagen assoziierten Phänomene formuliert bzw. weiterentwickelt. Unter Generalisierungen werden quantifizierte Aussagen über angenommene Instanzen einer Klasse oder Gruppe von Objekten, Individuen oder (Zeit-)Räumen verstanden, auf die nicht kontextuell referiert wird. Kommentare schließen Textstellen ein, in denen die erzählte Zeit unterbrochen und eine ergänzende Information zu Erzählung, Figuren, Handlung oder dem Akt des Erzählens eingefügt wird (Bonheim 1975). Nicht-fiktionale Rede bezeichnet Passagen in fiktionalen Texten, die Behauptungen bzw. Hypothesen über die reale Welt nahelegen (Konrad 2017). Generalisierung, Kommentar und nicht-fiktionale Rede können sich vollständig oder teilweise überlappen.

Und so begann der Hauptmann: »[An allen Naturwesen, die wir gewahr werden, bemerken wir zuerst, daß sie einen Bezug auf sich selbst haben. [...]]« (Goethe 2012)

In diesem Beispiel treten alle drei Phänomene auf. Die erzählte Zeit, die in der Erzählerrede fließt, wird unterbrochen und ein Kommentar über Naturwesen vorgenommen. Zugleich wird eine Aussage über angenommene Instanzen der Klasse der Naturwesen getroffen. Da auch in der realen Welt Naturwesen (jeglicher Art) vorkommen, ist die Proposition grundsätzlich auf die reale Welt übertragbar.

Für reflexive Passagen erstellen wir eine Goldannotation, auf der in einem nächsten Schritt eine Attributionsannotation vorgenommen wird. Die Attribution bestimmt, wem die in der Passage enthaltene Information zugeschrieben werden kann, wofür grundsätzlich Figuren, die Erzählinstanz oder die AutorIn in Frage kommen. Einige sprachliche Mittel im Text sind prädestiniert für bestimmte Attributionen, so markieren bestimmte Satzzeichen i. d. R. (in)direkte Rede und damit die Sprecher im Text. Dennoch gibt es Passagen, in denen sich die Sprechinstanz nicht eindeutig identifizieren lässt und sich unterschiedliche Interpretationen (Zuschreibungen) aufdecken lassen.

Zur automatischen Erkennung und quantitativen Analyse erstellen wir das Korpus MONACO (Modes of Narration and Attribution Corpus)¹, das aus deutschsprachigen fiktionalen Erzähltexten von 1600–1950 besteht. Jede AutorIn ist im Korpus nur einmal vertreten und es wird eine gleiche Verteilung der Texte über Jahrhunderte angestrebt. Die wichtigste Textquelle für MONACO ist Kolimo (Herrmann und Lauer 2017).

Die Annotation der Texte wird in CATMA 6.2² vorgenommen und basiert auf detaillierten Richtlinien, die iterativ entwickelt wurden. Dabei werden momentan lediglich die ersten 200 Sätze eines Textes annotiert, um mehr Texte zu annotieren und das Korpus möglichst divers gestalten zu können. Jedes der drei mit reflexiven Passagen assoziierten Phänomene wird von zwei studentischen AnnotatorInnen annotiert. Der Goldstandard für Generalisierung, Kommentar und nicht-fiktionale Rede wird in Kleingruppen von 2-3 DoktorandInnen erstellt. Zur Beschleunigung der Goldstandarterstellung haben wir einen „CATMA-Merger“ entwickelt, welcher eine neue Annotation als die Vereinigung mehrerer Annotationen erstellt, die dann von den DoktorandInnen überprüft und bestätigt, korrigiert oder gelöscht werden kann. Die Attribution wird im zweiten Schritt von allen (sechs) studentischen AnnotatorInnen auf dem Goldstandard für Generalisierung, Kommentar und nicht-fiktionale Rede annotiert. Für Attribution wird kein Goldstandard erstellt, um sämtliche mögliche Interpretationen zu erfassen.

Bisher wurden Goldstandards für achtzehn Texte erstellt. Der älteste Text stammt aus dem Jahr 1616, der jüngste aus dem Jahr 1930. Die annotierten Texte weisen im Durchschnitt ein moderates (> 0,4) oder gutes (> 0,6) Inter-Annotator Agreement mit κ -Werten (Fleiss et al. 1981) von 0,59 für Generalisierung, 0,44 für Kommentar und 0,66 für nicht-fiktionale Rede auf. Die Inter-Annotator Agreement-Werte für γ (Mathet et al. 2015) sind etwas höher: 0,66 für Generalisierung, 0,52 für Kommentar und 0,72 für nicht-fiktionale Rede.

Mit der zunehmenden Menge annotierter Texte werden schrittweise regelbasierte, statistische und neuronale Tagger für die einzelnen Phänomene entwickelt. Ihre Anwendbarkeit wird dabei auch für andere Textsorten wie Essays und enzyklopädische Texte erprobt. Letzten Endes soll eine ausreichend große Menge annotierter Daten nicht nur bessere Modelle ermöglichen, sondern auch diachrone oder genreübergreifende Perspektiven auf reflexive Passagen und ihre Attribution eröffnen.

Fußnoten

1. <https://gitlab.gwdg.de/mona/korpus-public>
2. <https://catma.de/>

Bibliographie

Bonheim, Helmut (1975): "Theory of narrative modes ", in: *Semiotica* 14/4: 329–344.

Chatman, Seymour Benjamin (1980): *Story and Discourse: Narrative Structure in Fiction and Film*. Ithaca / London: Cornell University Press.

Dönicke, Tillmann / Gödeke, Luisa / Varachkina, Hanna (2021): Annotating Quantified Phenomena in Complex Sentence Structures Using the Example of Generalising Statements in Literary Texts, in: *Proceedings of the 17th Joint ACL-ISO Workshop on Interoperable Semantic Annotation* 20-32.

Fleiss, Joseph L. (1971): "Measuring nominal scale agreement among many raters ", in: *Psychological bulletin* 76.5: 378.

Goethe, Johann Wolfgang (2012): "Die Wahlverwandtschaften ", in: *TextGrid Repository, Digitale Bibliothek*, <https://hdl.handle.net/11858/00-1734-0000-0006-6A93-D> [letzter Zugriff 26. November 2021]

Herrmann, Berenike / Lauer, Gerhard (2017): "KOLIMO. A corpus of Literary Modernism for comparative analysis ", <https://kolimo.uni-goettingen.de/about>

Konrad, Eva-Maria (2017): "Signposts of Factuality: On Genuine Assertions in Fictional Literature " in: Sullivan- Bissett, Ema / Bradley, Helen / Noordhof, Paul (eds.): *Art and Belief*. Oxford: Mind Association Occasional Series 42-62.

Leslie, Sarah-Jane / Lerner, Adam / Zalta, Edward Nouri (2016): "Generic Generalizations ", in: *Stanford Encyclopedia of Philosophy* <https://plato.stanford.edu/entries/generics/> [letzter Zugriff 26. November 2021]

Mathet, Yann / Widlöcher, Antoine / Métivier, Jean-Philippe (2015): "The unified and holistic method gamma (γ) for interannotator agreement measure and alignment ", in: *Computational Linguistics* 41.3: 437–479.

Searle, John Rogers (1975): "The logical status of fictional discourse ", in: *New literary history* 6(2): 319-332.