

PD Dr. Friedrich Michael Dimpel
Ausgezeichnete Mären analysieren – ein Werkstattbericht

1.

Digitale Korpora können dank der Fortschritte im Bereich der Digital Humanities mit zahlreichen quantitativen Analyseverfahren untersucht werden: Im Rahmen der Stilometrie beschäftigen sich Computerphilologen etwa mit Fragen der Autorschaftsattribuion oder der Werkchronologie sowie mit Fragen nach Spezifika von Gattungen oder Epochen.¹ Auch wenn digitale Texte mit zahlreichen quantitativen Analyseverfahren untersucht werden können: In der Regel sind vollständige Texte oder zusammenhängende Textabschnitte wie Buchkapitel der Gegenstand etwa von stilometrischen Studien. Die Möglichkeiten für automatische Analysen enden jedoch dort, wo solche spezifische Textebenen in den Blick zu nehmen wären, die repräsentieren, welche Eigenschaft auf welche Figur bezogen wird, und welche Figuren oder Erzähler diese Zuschreibung vornehmen: Wer spricht? Zudem: Über wen wird gesprochen – welche Terme werden auf welche Figuren bezogen? Gibt es dabei bspw. genderspezifische Distributionen? Auf welche Weise wird gesprochen? Es ist ein erheblicher Unterschied, ob eine Figur über ihre Liebe spricht und damit ihre Emotion in der erzählten Welt öffentlich macht, oder ob nur eine Gedankenrede einer Figur erzählt wird. Wenn man auf solche Informationen zugreifen möchte, dann ist es nötig, die Texte mit einer entsprechenden Annotation zu versehen.

Quantitative Analysen von Textsamples, die spezifische Textdaten enthalten wie Figurenrede bestimmter Aktanten, fokalisierte Textpassagen oder Textpassagen, die sich auf bestimmte Aktanten beziehen, sind derzeit nicht möglich ohne eine aufwendige manuelle Textaufbereitung für eine konkrete Fragestellung.

Wie wichtig jedoch eine narratologische Textauszeichnung in einem diachron angelegten Korpus wäre, unterstreichen Fotis Jannidis, Gerhard Lauer und Andrea Rapp: „Wer hat, wann und wo zum ersten Mal die Form der erlebten Rede eingesetzt, wer die episodische Reihung zu psychologischer Figurenzeichnung verdichtet? Wo verknüpfen sich populäre Novellenstoffe mit hochkulturellen Erzähltechniken [...]? Denn nur mit Hilfe des Computers lässt sich ein hinreichend großes Korpus über einen langen Zeitraum hinweg untersuchen und damit etablierte literaturhistorische Methoden um serielle, computerbasierte Verfahren so ergänzen, dass eine Geschichte des Erzählens überhaupt erst geschrieben werden kann.“²

¹ Vgl. exemplarisch JOHN F. BURROWS: ‚Delta‘. A Measure of Stylistic Difference and a Guide to Likely Authorship. In: *Literary and Linguistic Computing* 17, 2002, S. 267–287, FRIEDRICH MICHAEL DIMPEL: Computergestützte textstatistische Untersuchungen an mittelhochdeutschen Texten. Tübingen 2004, FOTIS JANNIDIS: Methoden der computergestützten Textanalyse. In: Vera Nünning / Ansgar Nünning (Hrsg.), *Methoden der literatur- und kulturwissenschaftlichen Textanalyse. Ansätze – Grundlagen – Modellanalysen*. Stuttgart 2010, S. 109–132, CHRISTOF SCHÖCH: Corneille, Molière et les autres. Stilometrische Analysen zu Autorschaft und Gattungszugehörigkeit im französischen Theater der Klassik. In: Christof Schöch / Lars Schneider (Hrsg.), *Literaturwissenschaft im digitalen Medienwandel*. 2014, S. 130–157.

² FOTIS JANNIDIS / GERHARD LAUER / ANDREA RAPP: Hohe Romane und blaue Bibliotheken. Zum Forschungsprogramm einer computergestützten Buch- und Narratologiegeschichte des

2.

Bei narratologisch annotierten Korpora zur deutschen Literatur handelt es sich um ein Desiderat.³ Daher soll ein Korpus von 100 Kurzerzählungen narratologisch annotiert werden. Das Korpus soll historisch relativ ausgewogen angelegt werden: 30 mittelhochdeutsche und frühneuhochdeutsche Mären, 10 Boccaccio-Novellen, 10 Chaucer-Tales, 30 neuhochdeutsche Novellen und 20 neuhochdeutsche Kurzgeschichten werden aufgenommen. Als Ebenen der Auszeichnung sind vorgesehen:

1. Welche Figuren befinden sich an dem Ort, von dem erzählt wird?
2. Welche Figurenbewegungen im Raum finden statt?
3. Welche temporalen Abweichungen ereignen sich (Prolepsen etc.)?
4. Welche Figur ist wie fokalisiert?
5. Um welche Art von Redewiedergabe (direkte Rede, Bewusstseinsdarstellung etc.) handelt es sich – und welche Figur denkt oder spricht?
6. Um welche Art von Erzählerrede (Descriptio, Bericht Figurenaktivität, Erzählerreflexion etc.) handelt es sich?
7. Auf welche Figur bezieht sich eine Figuren- oder Erzählerrede?
8. Auf welche Figur bezieht sich eine wertende Äußerung einer anderen Figur oder des Erzählers?
9. Steht eine Äußerung in Negation?
10. Liegt eine uneigentliche Rede vor (metaphorisch, ironisch etc.)?
11. Inwieweit besteht Unsicherheit hinsichtlich der Eindeutigkeit der Textdaten?

Die nötigen XML-Elemente sind noch nicht im TEI-Standard enthalten, daher muss ein geeignetes Tagset entwickelt werden.⁴ TEI-Kompatibilität wird jedoch angestrebt: Die Elemente können über das Roma-Tool in eine ODD-Datei integriert werden.⁵

Romans in Deutschland (1500-1900). In: Lucas Marco Gisi / Jan Loop / Michael Stolz (Hrsg.), *Literatur und Literaturwissenschaft auf dem Weg zu den neuen Medien*. germanistik.ch 2006. (= online unter http://www.germanistik.ch/publikation.php?id=Hohe_Romane_und_blaue_Bibliotheken).

³ Allerdings kann in vielfältiger Weise auf die wichtige Grundlagenstudie von ANNELEN BRUNNER: *Automatische Erkennung von Redewiedergabe in literarischen Texten*. Diss. masch. Würzburg 2012, aufgebaut werden. Brunner hat in ihrer Dissertation ein Annotationsverfahren für Redewiedergabe entwickelt und ein Korpus, das aus Texten von 1787 bis 1913 besteht, manuell annotiert. Auch wenn die automatische Erkennung der Redewiedergabe (regelbasiert und via Maschinelles Lernen) noch nicht eine Fehlerrate erreichen kann, die narratologische Auswertungen erlaubt, kann das vorliegende Projekt in konzeptioneller Hinsicht von Brunners Studie erheblich profitieren.

⁴ Zu narratologischen Desideraten von TEI-P5 vgl. FOTIS JANNIDIS: *TEI in a Crystal Ball*. In: *Literary and Linguistic Computing* 24, 2009, S. 253–265, hier S. 261f.

⁵ Vgl. <http://www.tei-c.org/Guidelines/Customization/odds.xml>.

Als XML-Elemente werden vorgestellt:

1. <FigurFokusort> (@Bezeichnung, @Figurengruppe)
2. <BewegungLokal> (@Typ)
3. <Chronologie> (@Typ)
4. <Fokalisiert> (@Bezeichnung, @Typ)
5. <Redewiedergabe> (@Typ, @Bezeichnung, @non-fact, @level)
6. <Erzählerrede> (@Typ, @Bezeichnung)
7. <Figurenbezug> (@Unmittelbar, @Mittelbar)
8. <Wertung> (@BezeichnungWertende, @BezeichnungGewertete, @Typ)
9. <Negation> (@Typ)
10. <UneigentlicheRede>
11. <certainty> sowie @cert als Attribut zu allen Elementen (> TEI P5)

3.

Zentral für den Workflow ist die Entwicklung von Annotationsrichtlinien. Es wird angestrebt, dass verschiedene Versuchspersonen beim gleichen Text zu homogenen Ergebnissen kommen. In ähnlicher Weise, wie man sich bei der Herstellung einer Edition über Kollationierungsregeln verständigen muss, sind hier Annotationsregeln zu erarbeiten. Solche Regeln sind nötig, weil selbst scheinbar eindeutig definierte narratologische Phänomene sich oft nicht eindeutig in Texten wiederfinden lassen. Zudem ist Ambiguität ein charakteristisches Merkmal von literarischen Texten.

Mit Blick auf das Homogenitätsziel werden Annotationsregel und Bearbeitungsdatum direkt im XML-Code dokumentiert, damit bei einer Weiterentwicklung der Annotationsrichtlinien einerseits rasch auf eine einschlägige Fallsammlung zugegriffen werden kann; andererseits können bei einer Regelrevision Entscheidungen gezielt aufgesucht und revidiert werden. Dabei hilft ein eigener Projekteditor, der in Perl/TK implementiert wurde, der neben dem Annotationsfenster in einem zweiten Fenster in Kurzform Informationen zu bereits ausgezeichneten Elementen einblendet.

4.

Annotiert wurden bislang sechs Texte. Vorgestellt werden einige Probleme, die sich bei der Auszeichnung von ‚Sperber‘ und das ‚Häslein‘⁶ etwa durch Segmentierung oder durch Ambiguitäten ergeben haben. Anhand von Auswertungsdaten soll exemplarisch aufgezeigt werden, in welcher vielfältiger Weise ein entsprechend annotiertes Korpus Analysen möglich macht; etwa in Bezug auf

a) multiple Methoden:

- i. Das Korpus kann wie andere Korpora auch mit einer Vielzahl an statistischen Methoden analysiert werden – etwa in Hinblick auf Heterogenität oder Homogenität.
- ii. Eine besondere Stellung nimmt das Korpus jedoch dadurch ein, dass auf Basis der Textauszeichnung eine Sample-Erstellung für spezifische Fragestellung möglich ist, die

⁶ Ausgabe: KLAUS GRUBMÜLLER: Novellistik des Mittelalters. Märendichtung. Frankfurt/Main 1996 (=Bibliothek des Mittelalters 23)

nicht nur chronologisch-lineare Zugriffe auf Korpussegmente erlaubt, sondern systematische Zugriffe auf gleichartig annotierte Korpussegmente. So lässt sich ein Korpussegment bspw. mit Bewusstseinsdarstellung von weiblichen Figuren mit einem Korpussegment vergleichen, das aus Erzählerrede besteht; die Figurenrede von Antagonisten lässt sich mit Figurenrede von Protagonisten vergleichen, u.v.m.

b) multiple Fragestellungen, beispielsweise:

- i. Wie steht es um die diachrone Entwicklung von Fokalisierung, um die Eigenschaften von Erzähler- und Figurenrede, um temporale Alternationen, wie verteilt sich der Redebezug auf verschiedene Figurentypen wie Protagonist oder Antagonist, wie steht es um quantitative Parameter bei uneigentlicher Rede?
- ii. Korrelation kulturwissenschaftlich relevanter Terme und aktantieller Rolle. Hier werden bspw. zahlreiche gender-bezogene Auswertungen möglich, indem eine Sample-Analyse mit Figuren- oder Erzählerrede möglich wird, die jeweils auf weibliche oder männliche Figuren bezogen ist oder durch eine Sample-Analyse mit Figurenrede, die jeweils von weiblichen oder männlichen Figuren stammt.
- iii. Studien zur Wertungstheorie: Wie sind evaluative Äußerungen auf Erzählerrede und Figurenrede verteilt? Welche Aktanten bewerten bevorzugt, welche werden bevorzugt bewertet?
- iv. Lassen sich für diese oder für andere Fragestellungen epochenspezifische Verteilungen ausmachen? Es werden Studien ermöglicht, die einen Beitrag zur Gattungsgeschichte leisten. So wären beispielsweise Theoriebildungen zu überprüfen, inwieweit und inwiefern das Verhältnis vom Märe zur Novelle in Anschluss an die Unterscheidungskriterien von Hans-Jörg Neuschäfer unter dem Gesichtspunkt eines „Noch-Nicht“ beschrieben werden kann.⁷

⁷ Vgl. HANS-JÖRG NEUSCHÄFER: *Boccaccio und der Beginn der Novelle. Strukturen der Kurz-erzählung auf der Schwelle zwischen Mittelalter und Neuzeit*. München 1969 (=Theorie und Geschichte der Literatur und der schönen Künste 8). Kritisch dazu FRIEDRICH MICHAEL DIMPEL: *Sprech- und Beißwerkzeuge, Kunsthandwerk und Kunst in Kaufringers ‚Rache des Ehemanns‘*. In: *Daphnis* 42, 2013, S. 1–27 (im Erscheinen).