

Workshop Computational Narratology

In den letzten fünf Jahren hat die computergestützte erzähltechnische Analyse von Einzeltexten wie von großen Textsammlungen solche Fortschritte gemacht, dass sich inzwischen eine ganze Reihe von Methoden und Werkzeugen etabliert haben. Durch eine Reihe von Einzeluntersuchungen [z.B. Brunner 2014, Mani 2013] ist dadurch ein neues Forschungsfeld sichtbar geworden, das man als *Computational Narratology* bezeichnen kann.¹ Dieser Workshop soll Interessierte an diesem innovativen Forschungsfeld, bei dem sich Erzählforschung, formale Modellierung und quantitative Textanalyse begegnen, zusammenbringen, um möglichst frühzeitig Felder zu identifizieren, in denen Kooperationen die unterschiedlichen Forschungsinteressen stärken und gegenseitig befruchten können.

Im Zentrum des Workshops stehen dabei folgende sechs Themenkomplexe:

I. Entwicklung von Standards und Datenstrukturen für gemeinsam nutzbare Textkorpora

Ein erstes Aufgabenfeld für eine Computational Narratology, das eine Grundlage für die koordinierte, verteilte Arbeit an einzelnen narratologischen Problemfeldern bildet, ist die Verabredung von Standards und Datenstrukturen für gemeinsam nutzbare Textkorpora. Die Aufgabenstellung umfasst dabei folgende Einzelaspekte:

- Entwicklung eines gemeinsamen Formats für Metadaten, die für narratologische Analysen relevante, die Texte beschreibende Informationen enthalten (einheitliche Feldnamen und kontrolliertes Vokabular, um den Anpassungsaufwand an neue Korpora zu minimieren; siehe auch unten: "Narratologische Basiskategorien").
- Ein gemeinsames Format für annotierte Trainingskorpora, sodass händisch bezüglich linguistischer und/oder narratologischer Phänomene annotierte Trainingskorpora eines Teams von anderen Teams für die Optimierung von automatischen Analyseverfahren nachgenutzt werden können.
- Ein gemeinsames Repository für solche Trainingskorpora, das die sichere Speicherung, eindeutige Identifizierung und Rechteverwaltung zulässt, sodass neue Ressourcen publiziert, ausgetauscht, gefunden und nachgenutzt werden können.
- Eine gemeinsame Strategien zum Einsatz von Crowdsourcing beim Erstellen von Ressourcen durch händische Annotationen sollte entwickelt werden, auch unter Berücksichtigung verschiedener institutioneller Akteure wie Forschungsgruppen oder Anbietern von Textarchiven.

II. Narratologische Basiskategorien für die Analyse von Textkorpora

¹ Wir fokussieren im Folgenden auf den dritten der in Mani 2013 angeführten Gegenstandsbereiche, nämlich die computergestützte narratologisch-literaturwissenschaftliche Textanalyse. Das umfassender Forschungsfeld umfasst nach Mani „ (...) approaches to storytelling in artificial intelligence systems and computer (and video) games, the automatic interpretation and generation of stories, and the exploration and testing of literary hypotheses through mining of narrative structure from corpora.“

Als zweites Aufgabenfeld soll auf dem Workshop die Definition eines generischen, projektübergreifend relevanten Sets von narratologischen Beschreibungskategorien in Form einer sog. Tag-Library diskutiert werden. Dabei stellen sich zwei Kernfragen:

- Theoretische Fundierung: Bereits seit den 1990er Jahren wird eine intensive fachwissenschaftliche Debatte zwischen Vertretern der "klassischen" (formalistisch-strukturalistisch orientierten) Narratologie und Vertretern der sog. "new narratologies" (d.h. der kognitivistische, thematische und ideologiekritische Forschungsfragen verfolgenden) geführt. Vor diesem Hintergrund ist zu erörtern, ob es möglich und forschungsmethodisch überhaupt sinnvoll ist, diesen narratologischen Methodenpluralismus in Form einer gemeinsamen Tag-Library kategorial abbilden zu wollen. Im Kontext der Computational Narratology wäre prinzipiell auch die Alternative denkbar: d.h. die Modellierung mehrerer alternativer, dafür jedoch in sich homogener Klassifikationssysteme.
- Wie kann man ausgehend von den 'klassischen' Beschreibungskategorien insbesondere Genette'scher Provenienz, die konzeptionell überwiegend auf eine 'dichte' narratologische Beschreibung von Einzeltexten etwa bis zur Satzebene hinauf abzielen, zu neuen synthetischen Kategorien fortschreiten, die auf einer deutlich höheren Komplexitäts- und Abstraktionsebene angesiedelt sind, wie sie für synchrone wie diachrone Korpusstudien erforderlich ist?

III. Automatisierung und Optimierung von Search&Retrieval-Operationen

Eine dritte Herausforderung an eine Computational Narratology stellt - im Anschluss an die formale Modellierung - das Retrieval von entsprechend beschreibbaren Textinstanzen dar. Hierzu sollen in dem Workshop zwei Operationalisierungsstrategien diskutiert werden, die auch komplementär eingesetzt werden können:

- die manuelle Auszeichnung aller Texte auf der Grundlage der formalen Modellierung durch idealerweise mindestens zwei Annotatoren;
- die manuelle Annotation eines definierten Teilkorpus, um auf dieser Basis entweder mit maschinellem Lernen oder regelbasiert den Rest der Textsammlung automatisch zu annotieren.

Beide Verfahren, die für ein Retrieval narratologischer Konzepte notwendig sind, werden aktuell bereits für Teilbereiche des narratologischen Kategoriensystems angewendet – dazu gehören beispielsweise Redewiedergabe (vgl. Brunner 2013) und Zeitgestaltung (vgl. Bögel et al. im Erscheinen). Dabei zeigt sich, dass einige „weiche“ Konzepte, wie beispielsweise die Kategorie *erlebte Rede*, deutlich schlechtere Ergebnisse beim Retrieval ergeben als die klarer definierbaren Begriffe. Ein Desiderat für die Computational Narratology ist entsprechend, *best practices* für das Retrieval von Konzepten unterschiedlichen Explikationsgrades zu etablieren. Auch dabei sind unterschiedliche Strategien möglich: so kann z.B. die Übereinstimmungsquote, die menschliche Annotatoren in der Anwendung der fraglichen Konzepte maximal erreichen (*interannotator agreement*), als Ziel des Retrievals festgelegt werden. Denkbar ist aber auch ein Ansatz, der weniger auf Normierung und stärker auf die Exploration und Abbildung von Deutungsvielfalt setzt, wie dies z.B. im Ansatz des sog. kollaborativen taggings (vgl. Meister 2012) versucht wird.

IV. Gattungsklassifikation auf Grundlage narratologischer Merkmale

Auf der Grundlage von detaillierten Informationen über die Präsenz bzw. Verteilung von kategorial definierten narrativen Phänomene in zahlreichen Einzeltexten ist heute erstmals eine narratologisch begründete, korpusbasierte und damit auch empirisch validierbare Textklassifikation möglich. Unter diesem Gesichtspunkt soll der Workshop zwei Teilfragen diskutieren:

- Inwieweit stehen die Mechanismen und Ergebnisse dieses Ansatzes in einem produktiven Spannungsverhältnis zu bestehenden (historischen oder fachwissenschaftlichen) Gattungskategorien?
- Wie ist es um die Anschlussfähigkeit der Computational Narratology an literaturwissenschaftliche Teilbereiche wie Literaturgeschichte oder Gattungstheorie bestellt?

V. Werkzeug-Entwicklung

Die Entwicklung und nachhaltige Bereitstellung von Werkzeugen und Plattformen für die Arbeit mit gemeinsam nutzbaren Textkorpora ist für die Computational Narratology von besonderer Relevanz. Aufgrund der Vielfältigkeit narratologischer Forschung stellen sich dabei besondere Herausforderungen an das Beschreibungsinventar, wie z. B. die Berücksichtigung dynamisch entwickelbarer Taxonomien, von Hierarchieüberlappungen und widersprüchlichen Annotationen. Analysewerkzeuge sollten immer auch Anpassungsmöglichkeiten für konkrete Forschungsfragen vorsehen. Vor dem Hintergrund dieser Problemstellung soll der Workshop zwei Fragen thematisieren:

- Textanalyse unter Einbeziehung von Textannotationen: Werkzeuge zur einfachen Analyse reiner Textdaten sind nicht zuletzt durch die Entwicklung im Rahmen von Suchmaschinen gut erforscht. Zur tiefergehenden Analyse von Textdaten ist allerdings die Einbeziehung von Kontextinformationen in Form von Annotationen nötig. Die Werkzeugentwicklung einer Computational Narratology sollte daher ein besonderes Augenmerk auf die Erstellung und Verarbeitung von Annotationen richten.
- Verknüpfung von Crowd-Sourcing und Machine Learning-Verfahren: Aufgrund der benötigten großen Menge von qualitativ hochwertigen manuellen Annotationen werden Crowd Sourcing fähige Werkzeuge benötigt. Angestrebt wird jedoch zugleich auch die automatische Generation von Annotationen. Gefordert sind daher Werkzeuge, die aus den manuell angefertigten Annotationen lernen und auf diesen aufbauen können.

VI. Tutorien zur Computational Narratology

Um die Nutzbarkeit der im Rahmen einer Computational Narratology gewonnenen Erkenntnisse und Ergebnisse zu gewährleisten und einer möglichst breiten Nutzergruppe zugänglich zu machen, sollen frei im Netz verfügbare Tutorien entstehen. Diese sollen die Verwendung neu entwickelter Methoden und Tools veranschaulichen, prototypische Arbeitsabläufe darlegen und die Einhaltung von best practices beschreiben. Dadurch soll im Bereich der computergestützten Textanalyse unerfahrenen Personen, weniger technisch versierten Narratologen oder auch Studenten der Digital Humanities oder Literaturwissenschaft der Einstieg in das Feld der Computational Narratology erleichtert werden.

Vor dem Hintergrund dieser Zielsetzung soll zum Abschluss des Workshops ein Gesamtkonzept für den Aufbau einer umfassenden Sammlung "Tutorien zur Computational Narratology" entworfen werden.

Der Workshop wird geleitet von:

Prof. Dr. Fotis Jannidis

Universität Würzburg, Institut für Deutsche Philologie, Philosophische Fakultät I, Am Hubland, D - 97074 Würzburg, fotis.jannidis@uni-wuerzburg.de

Prof. Dr. Jan Christoph Meister

Universität Hamburg, Institut für Germanistik, Fakultät für Sprache, Literatur und Medien, Von-Melle-Park 1, D - 20146 Hamburg, jan-c-meister@uni-hamburg.de

Dr. Christof Schöch

Universität Würzburg, Institut für Deutsche Philologie, Lehrstuhl für Computerphilologie, Am Hubland, D - 97074 Würzburg, christof.schoech@uni-wuerzburg.de

Organisatorisches

Der Workshop ist für bis zu 15 Teilnehmer ausgelegt. Als technische Grundausstattung wird ein Beamer und ausreichend Stromanschlüsse für die Teilnehmer sowie Zugang zum Internet benötigt. Ein workshopspezifischer Call wird nicht veröffentlicht.

Die oben umrissenen sechs Themenkomplexe sollen in einer Vormittags- und einer Nachmittagssession bearbeitet und die Zwischenergebnisse live in einem begleitenden Workshop-Blog dokumentiert werden.

Literatur

Bögel, Thomas/Gertz, Michael/Gius, Evelyn/Jacke, Janina/Meister, Jan Christoph/Petris, Marco/Strötgen, Jannik: Collaborative Text Annotation Meets Machine Learning. heureCLÉA, a Digital Heuristics of Narrative, in: DHCommons Journal (im Erscheinen).

Brunner, Annelen: Automatic recognition of speech, thought, and writing representation in German narrative texts. In Literary & Linguistic Computing 28: 4 (2013), S. 563-575.

Mani, Inderjeet: "Computational Narratology". In: Hühn, Peter et al. (eds.): *the living handbook of narratology*. Hamburg: Hamburg University. URL = <http://www.lhn.uni-hamburg.de/article/computational-narratology> [view date:10 Nov 2014]

Meister, Jan Christoph: „Crowd sourcing “true meaning”. A collaborative markup approach to textual interpretation.“ In: Willard McCarty, Marilyn Deegan (eds.), *Collaborative Research in the Digital Humanities*. Festschrift for Harold Short (Ashgate Publishers) 2012, 105-122.