

Informatik und Hermeneutik

Erste Erkenntnisse aus dem heureCLÉA-Projekt¹

In unserem aktuellen BMBF-eHumanities Projekt heureCLÉA arbeiten wir als Informatiker und Literaturwissenschaftlerinnen an einer digitalen Heuristik, die die Analyse von literarischen Texten unterstützen soll. Der Anwendungsfokus liegt dabei exemplarisch auf narratologischen Phänomenen der Zeit: d.h., das heuristische Modul von heureCLÉA soll die Funktionalität der Textanalyse und -annotationsumgebung CATMA² erweitern, indem es den Usern automatisch generierte Vorschläge zur Annotation narratologisch definierter Zeit-Phänomene in einem Text anbietet. Das Modul wird auf der Basis von drei Zugängen entwickelt: (1) Ausgangspunkt ist so genanntes "hermeneutisches Markup" (Piez 2010), das auf klassischen narratologischen Kategorien wie etwa Ordnung, Frequenz und Dauer beruht (vgl. Genette 1972, Lahn und Meister 2013) und von geschulten Annotatorinnen vergeben wird. Dieses Markup wird (2) mit regelbasierten Verfahren sowie (3) Machine-Learning-Ansätzen kombiniert.³

Aufgrund des Zusammenspiels von literaturwissenschaftlichen – und speziell: hermeneutischen – Verfahren und informatischen Verfahren der Information Extraction und der Statistik stehen sich non-deterministische Zugänge zu Texten und entscheidbare bzw. deterministische Verfahren gegenüber, die nicht ohne weiteres auf den jeweilig anderen Ansatz übertragen werden können. Deshalb ist die Reproduzierbarkeit von narratologischen Analysen für die Vereinbarkeit des literaturwissenschaftlichen und des informatischen Zugangs und damit für den Erfolg des heuristischen Moduls ausschlaggebend.

In unserem Beitrag präsentieren wir ein methodisches Desiderat im Bereich der Narratologie, das erst durch die interdisziplinäre Zusammenarbeit zwischen Geisteswissenschaftlern und Informatikern in den Fokus gerückt ist und aus unserer Sicht exemplarisch für eine solche Zusammenarbeit ist: die Notwendigkeit, narratologische Analysekatégorien eindeutiger zu konzeptionalisieren, um sie operationalisieren zu können.

¹ vgl. www.heureclea.de (gesehen am 10.12.2013)

² vgl. www.catma.de (gesehen am 10.12.2013)

³ Damit ist heureCLÉA ein Beitrag zur *computational narratology* im Sinne von Mani, da es zu "exploration and testing of literary hypotheses through mining of narrative structure from corpora" (Mani, 2013, para. 1) beiträgt.

³Zu den regelbasierten Verfahren vgl. Strötgen und Gertz (2010), die Gesamtarchitektur von heureCLÉA wird außerdem in einem weiteren eingereichten Beitrag vorgestellt.

Die Narratologie ist eine literaturwissenschaftliche Disziplin, die eine Reihe theoretischer Konzepte und Modelle für die Analyse und Interpretation erzählender Texte zur Verfügung stellt. Diese narratologischen Kategorien dienen normalerweise der Bezeichnung und Verortung textueller Eigenschaften, die (a) als typisch für narrative Texte angesehen werden und (b) für besonders interessant und geeignet befunden werden, um die speziellen Eigenschaften eines literarischen Einzelwerkes herauszustellen. Viele der Kategorien dienen der Bezeichnung struktureller Phänomene, die hauptsächlich an der Textoberfläche zugänglich sind. Das gilt insbesondere auch für die meisten Kategorien, die der Analyse explizit markierter Zeitphänomene dienen, wie sie im Rahmen von *heureCLÉA* untersucht werden.⁴ Obwohl narratologische Kategorien gemeinhin als theoretisch durchdacht und leicht operationalisierbar gelten, zeigten sich bei ihrer formalisierten Anwendung im Rahmen manueller, kollaborativer Annotation in *heureCLÉA* einige theoretische Unzulänglichkeiten. Typischerweise wurden solche Unzulänglichkeiten dann entdeckt, wenn sich die Annotatoren hinsichtlich der korrekten narratologischen Bestimmung konkreter Textstellen nicht einig waren. In Diskussionen über die Gründe für individuelle Annotations-Entscheidungen stellte sich dann oft die uneindeutige oder unvollständige Konzeption der jeweiligen Kategorie als Ursache uneinheitlichen Markups heraus. Die festgestellten theoretischen Versäumnisse lassen sich in zwei Gruppen einteilen, die je unterschiedliche Problemlösungsstrategien erfordern:

a) konzeptionelle Unvollständigkeit, die leicht durch eine Vervollständigung der Kategorie mittels funktionaler Entscheidungen behoben werden kann. Stellt sich bei der versuchten Anwendung einer Kategorie heraus, dass ihre Definition zu vage ist, um die Bestimmung einer fraglichen Textstelle vorzunehmen, müssen pragmatische Entscheidungen im Hinblick auf die Inklusion oder Exklusion bisher nicht bedachter textueller Oberflächenmerkmale getroffen werden.⁵

⁴ Der Klarheit wegen sollte angemerkt werden, dass keine der im Feld der Narratologie interessanten Phänomene rein formale Textmerkmale sind, da die *Bedeutung* von Wörtern und Sätzen stets ausschlaggebend für ihre Bestimmung ist. Das bedeutet, dass solche Phänomene zwar an der Textoberfläche zugänglich sind, ihre Bestimmung jedoch trotzdem in einem weiteren Sinne des Wortes interpretativ sein kann.

⁵ Ein derartiger Problemfall stellte ich an folgender Textstelle in Friedrich Hebbels Erzählung *Matteo* im Hinblick auf die Frage, ob es sich hier um eine Prolepse - bisher konzeptionalisiert als Vorgriff in der Zeit - handelt: "Sieh, morgen feire ich meine Hochzeit; zum Zeichen, daß du mir nicht mehr böse bist, kommst du auch, meine Mutter wird dich gern sehen." (Hebbel 1963: para. 4). Die Schwierigkeit ist hier dadurch gegeben, dass die angesprochene Figur am folgenden Tag nicht auf der Hochzeit erscheint. Um

Derartige Entscheidungen haben nur für die Anwendung der jeweiligen Kategorie Konsequenzen, nicht aber für weitere Konzepte. - Die zweite Kategorie betrifft dagegen

b) theoretische Unvollständigkeit, die ihrerseits auf die unzureichende Bestimmung fundamentaler narratologischer Konzepte zurückzuführen ist. Probleme dieses Typs können nicht einfach durch pragmatische Entscheidungen behoben werden, weil die für eine Problemlösung notwendigen Setzungen auf der Ebene grundlegender narratologischer Konzepte weitreichende Konsequenzen für viele erzähltheoretische Einzelkategorien nach sich zieht. Im Folgenden soll diese zweite Problemklasse anhand eines Beispiels erläutert werden.

In der Erzählung *Der Tod* von Thomas Mann ist bei dem Vergleich der Annotationsergebnisse im Hinblick auf die Geschwindigkeit der Erzählung⁶ folgende Passage in den Fokus der Aufmerksamkeit gerückt:

Ich habe die ganze Nacht hinausgeblickt, und mich dünkte, so müsse der Tod sein oder das Nach dem Tode: dort drüben und draußen ein unendliches, dumpf brausendes Dunkel. Wird dort ein Gedanke, eine Ahnung von mir fortleben und -weben und ewig auf das unbegreifliche Brausen horchen?

Mann 2004: 76

Diese Passage wurde von einigen Annotatoren ab dem ersten Komma als zeitraffend erzählt eingeordnet, von anderen dagegen als Erzählpause. Die Diskussion über die Gründe für die individuellen Entscheidungen hat gezeigt, dass die Annotatoren unterschiedliche Auffassungen darüber vertreten, was ein Ereignis ist. Betrachtet man mentale Vorgänge als Ereignisse, so muss man die zitierte Passage als zeitraffend klassifizieren, da die Gedanken des Erzählers in der fiktiven Welt vermutlich längere Zeit anhielten als die wenigen Sekunden, die in der Erzählung für ihre Wiedergabe eingeräumt werden. Ist man jedoch der Ansicht, dass es sich bei mentalen Prozessen nicht um Ereignisse handelt, so liegt in obiger Textstelle eine Pause vor: Der Bericht von Ereignissen wird unterbrochen durch die Darstellung nicht-ereignishafter Gegebenheiten. Die Frage danach, welche Konzeption von Ereignis korrekt oder sinnvoll ist, ist Gegenstand der Debatte um Narrativität: die für erzählende Texte konstitutive Eigenschaft, von Ereignissen zu berichten. Die unterschiedlichen Intuitionen der Annotatoren in Bezug auf die Definition von "Ereignis" korreliert hier mit Schmidts Konzeptionen von Ereignis I, das jegliche

entscheiden zu können, ob hier eine Prolepse vorliegt, muss entschieden werden, ob dieses Konzept auch antizipierte Ereignisse fassen soll, die im Verlauf der Erzählung nicht eintreten.

⁶ Unter "Erzählgeschwindigkeit" versteht man in der Narratologie das Verhältnis zwischen der Menge an Ereignissen, von denen berichtet wird, und der Zeit, die für diesen Bericht notwendig ist.

Form von Zustandsveränderung inkludiert, und Ereignis II, das zusätzliche Kriterien anführt, die Zustandsveränderungen aufweisen müssen, um als Ereignis zu gelten (Schmid 2003).⁷ Eine Entscheidung im Hinblick auf die richtige Narrativitätsdefinition, die für die Lösung von Annotationsproblemen im Bereich der Erzählgeschwindigkeit notwendig wäre, hätte nun nicht nur für die fraglichen Kategorien Konsequenzen, sondern beispielsweise auch für die Bestimmung des Gegenstandsbereichs der Narratologie und potenziell für eine Reihe weiterer Kategorien.⁸

Angesichts insbesondere dieser zweiten Sorte von Problem stellt sich die Frage, inwieweit solche grundlegenden Fragen im Rahmen von *heureCLÉA* geklärt werden können und sollten. Da die theoretische Arbeit an narratologischen Grundkonzepten nicht im Fokus des Projektes stehen sollte, war zunächst ein individueller Umgang der Annotatoren mit den anwendungsbezogenen Einzelproblemen vorgesehen. Es hat sich jedoch herausgestellt, dass diese Vorgehensweise weder aus narratologischer Sicht befriedigend ist, noch eine aus informationstheoretischer Perspektive verwertbare Datengrundlage liefert. Aus diesen Gründen haben wir uns dazu entschieden, den beiden geschilderten narratologischen Basisproblemen einige Aufmerksamkeit zu widmen: Für die Bestimmung von Ebenenwechsel und -zuordnung wird eine konsistente Lösung gefunden, so dass Ordnungsphänomene tatsächlich unterschiedlichen Erzählebenen zugeordnet werden können. Für die Bestimmung von „Ereignis“ streben wir eine plausible Konzeptionalisierung an, die ein möglichst wenig interpretatives Erkennen von Ereignissen erlaubt.

⁷ Zu diesen Kriterien zählt neben Resultativität, Relevanz, Unvorhersehbarkeit, Effekt, Irreversibilität und Nicht-Wiederholbarkeit auch das Kriterium der Faktizität, das die Eigenschaft von Zustandsveränderungen bezeichnet, tatsächlich in der fiktiven Außenwelt stattzufinden. Wertet man Faktizität als notwendige Eigenschaft von Ereignissen, muss die oben zitierte Passage aus *Der Tod* als Erzählpause klassifiziert werden.

⁸ Eine ähnliche Verknüpfung von Annotationsproblemen und ungeklärten narratologischen Basiskonzepten findet sich bei der Annotation von Phänomenen der zeitlichen Ordnung einer Erzählung einerseits und dem grundlegenden narratologischen Konzept der Erzählebenen. Es kann nur sinnvoll das zeitliche Verhältnis von solchen Ereignissen bestimmt werden, die sich auf derselben Erzählebene befinden. Anhand welcher Faktoren ein Ebenenwechsel festzumachen ist, wird in der narratologischen Forschung noch diskutiert (vgl. Ryan 1991, Coste/Pier 2011).

Die beschriebene Problematik ist ein spezifisch literaturwissenschaftliches Problem, die sie erzeugenden Rahmenbedingungen sind jedoch zugleich exemplarisch für das Zusammenspiel von Informatik und Geisteswissenschaften. Deshalb ist der entwickelte Lösungsansatz von entscheidender Bedeutung für das Gelingen des Projekts. Die Reproduzierbarkeit von Analyseergebnissen, die durch den Ansatz anvisiert wird, wird in den Geisteswissenschaften traditionell nicht thematisiert, da diese meist dem Konzept der intersubjektiven Übereinstimmung operieren, ohne diese weiter zu bestimmen. Die Reproduzierbarkeit von Analyseergebnissen ist jedoch zugleich auch eine von mehreren, bislang wenig erforschten Gelingensbedingungen für interdisziplinäre Projekte im Bereich der *Digital Humanities*. Diese disziplinäre Doppelperspektive auf ein methodisches Kriterium weist insofern auf die konzeptionellen Chancen, Probleme und Bedingungen einer Kooperation zwischen Geisteswissenschaftlern und Informatikern im Kontext von DH-Projekten.

Literatur

Coste, D. and Pier, J. (2013). Narrative Levels. *the living handbook of narratology*. <http://www.lhn.uni-hamburg.de/article/narrative-levels> (gesehen am 06.12.2013). First published 2011.

Genette, G. (1972). Discours du récit. In id., *Figures III*. Paris: Editions Du Seuil, pp. 67-282.

Lahn, S. and Meister, J. C. (2013). *Einführung in die Erzähltextanalyse: 2nd, updated edition*. Stuttgart: Metzler.

Mani, I. (2013). Computational Narratology. *the living handbook of narratology*. <http://www.lhn.uni-hamburg.de/article/computational-narratology> (gesehen am 06.12.2013).

Piez, W. (2010): Towards Hermeneutic Markup: an Architectural Outline. *Digital Humanities 2010. Conference Abstracts*. London: Office for Humanities Communication, Centre for Computing in the Humanities, King's College London, pp. 202-205.

Ryan, M.-L. (1991). *Possible Worlds, Artificial Intelligence, and Narrative Theory*. Bloomington: Indiana UP.

Schmid, W. (2003): Narrativity and Eventfulness. In T. Kindt & H.-H. Müller (eds.). *What Is Narratology? Questions and Answers Regarding the Status of a Theory*. Berlin: de Gruyter, 17–33.

Strötgen, J. and Gertz, M. (2010). HeidelTime: High Quality Rule-based Extraction and Normalization of Temporal Expressions. *Proceedings of the 5th International Workshop on Semantic Evaluation (ACL 2010)*. Uppsala, pp. 321-324.