

Aufbau eines Referenzkorpus “Erste Sätze in der deutschsprachigen Literatur”

Busch, Anna

annabussh@uni-potsdam.de

Theodor-Fontane-Archiv, Universität Potsdam

Roeder, Torsten

dh@torstenroeder.de

Bergische Universität Wuppertal, Germany

Stand der Forschung und Problem- aufriß

In literatur- und sprachwissenschaftlichen Untersuchungen ist der erste Satz eines narratologischen Zusammenhangs ein regelmäßig untersuchter Gegenstand (dazu u.a. Alt 2020, Haubrichs 1995, Hirdt 1974, Queng 2019, Miller 1965, Neuhaus 2019, Raulff 2019, Retsch 2000, Selbmann 2019). Das verwundert insofern wenig, gilt doch der erste Satz seit Wolfgang Isters Studie *Der Akt des Lesens* als Eingang in den Text durch die Lektüre, als Schlüsselstelle der Interaktion zwischen Text und Leser (1976: 38). Der erste Satz ist Verdichtungspunkt, sinnstiftender Ort für den Fortgang der Erzählung. Er unterliegt Moden, Bedingungen und Abhängigkeiten, bewegt sich in literarischen Traditionen, folgt Wirkungsabsichten, führt Reminiszenzen mit, steht für sich und erzeugt Kontext. Der erste Satz offenbart im Reichtum seiner unterschiedlichen Formen “die Schätze der Literatur in nuce” (Alt 2020: 18) und es ließe sich mit Alain Robbe-Grillet die These aufstellen, dass Literaturgeschichte aus der Untersuchung ihrer Anfangssätze zu schreiben ist (1992: 38).

Eine systematische, digital gestützte Untersuchung von “ersten Sätzen” steht bislang aus. Vereinzelt wurden händisch Korpora erster Sätze zusammengetragen (Beck 1992, Beck 1993, Wolkersdorf 1994) und Versuche unternommen, eine Typologie des ersten Satzes in der Literatur anhand ausgewählter Einzelanalysen zu entwerfen (zuletzt Alt 2020). Ergänzend dazu kann eine systematische Kategorisierung auf der Basis eines semiautomatisiert erstellten, größeren Untersuchungskorpus – wie sie hier projiziert wird – zielführend sein. Ähnlich gelagerte Untersuchungen, die nach der Quintessenz des Poetischen in der Literatur durch ihre Zählbarkeit fragen (vgl. beispielhaft Moretti 2009, auch Fischer/Strötgen 2015, Fischer/Jäschke 2018a/b), liegen vor, eine einzige sich dezidiert mit deutschsprachigen Erzählanfängen (nicht ersten Sätzen) beschäftigende quantifizierende Studie findet sich in der Arbeit von Herrmann 2018.

Der sämtlichen bisherigen Studien zu ersten Sätzen “mangelnden Gesamtsicht” (Alt 2020: 246) zu begegnen, ist Anliegen des Korpus “Erste Sätze in der deutschsprachigen Literatur”. Dazu wird ein Datenkorpus erstellt, publiziert und anschließend in einer Verzahnung von quantitativen und textanalytischen Herangehensweisen eine erste Auswertung unternommen.

Projekt, Vorgehen und Korpus

Als Ausgangsmaterial dienen mehrere Volltextkorpora (Deutsches Textarchiv, Zeno, u.a.), aus denen Texte nach Gattungen extrahiert wurden. Es ist deutlich, dass die vorhandenen Volltextangebote zwar unterschiedlich reichhaltige Strukturinformationen über das jeweilige Dokument bieten, aber die automatische Abgrenzung geschlossener Texteinheiten oft nicht trivial und ohne Einzelprüfung nicht zuverlässig möglich ist (z.B. bei Sammelbänden, Texten mit mehreren Kapiteln, Texte in mehreren Bänden). Dies bildet allerdings die Voraussetzung für das Extrahieren der ersten Sätze. Hinzu kommt, dass der Beginn des “poetischen Texts” durch z.B. vorangestellte Vorworte, Widmungstexte oder Einleitungen automatisiert nicht immer eindeutig zu lokalisieren ist.

Ferner ist die Abgrenzung von “ersten Sätzen” ein semantisches Problem. Sätze lassen sich als grammatisch-analytische Einheiten begreifen, die durch bestimmte Satzzeichen voneinander abgetrennt werden, was der maschinellen Verarbeitung entgegenkommt. Jedoch unterscheiden und verändern sich die zur Abgrenzung eines Satzes verwendeten Zeichen erheblich (man betrachte allein die Entwicklungen zwischen dem 17. und 18. Jahrhundert). Die absolute Trennschärfe mancher Satzzeichen steht zudem kontextabhängig infrage, weshalb Sätze teils auch als Sinneinheiten zu begreifen sind, in denen Satzzeichen eine strukturierende, aber nicht unterbrechende Funktion innewohnt (vgl. Abb. 2a/b). Sollte man also eher von einem fließenden “Beginn” oder “Anfang” sprechen? Bei der Bestimmung der “ersten Sätze” spielen somit Unschärfebereiche hinein, die sich wiederum auf Korpuskonsistenz und -vergleichbarkeit auswirken können.

Auswertung und Reflexion der Ergebnisse

Das derzeitig erstellte Korpus ist vollständig mitsamt Metadaten und Quellenangaben inkl. Positionsangaben in TEI codiert. Gattungsabhängig bewegt sich die Anzahl der Satzanfänge zwischen 100 und 1000 Einträgen. Mithilfe der manuell und automatisch erstellten Annotationen lässt sich das Korpus nach verschiedenen Parametern analysieren und visualisieren, beispielsweise nach Veröffentlichungsdatum, Textgattung, Geschlecht von Verfasserin oder Verfasser, Personen-, Orts- oder Zeitbezüge im Text (vgl. Abb. 1c) oder Länge des Gesamttexts. Außerdem wird dokumentiert, welchen Auswahlkriterien die jeweiligen Datenquellen unterlagen und wie dies im Hinblick auf die Ausgewogenheit des Korpus bei der Auswertung berücksichtigt werden sollte (vgl. Hug/Boenig 2021). Zur Dissemination des Korpus wurde 2021 das Twitter-Projekt “@satzomat” gelauncht, das täglich zwei erste Sätze sendet (vgl. Abbildungen 1–3).

Ziel ist es, eine “Typologie des ersten Satzes” mithilfe computerphilologischer Auswertungsverfahren zu erstellen sowie zu fragen, inwieweit Gattungen im Verlaufe der Geschichte bestimmte Typen von ersten Sätzen determinierten (z.B. Landschaftsbild, Rahmenhandlung) und ob sich weitere Korrelationen mithilfe der Metadaten und Annotationen feststellen lassen.

Abbildungen

”

Gewiß feid ihr alle voll Unruhe, daß ich so
lange — lange nicht geschrieben.

*erster Satz aus »Der Sandmann«
von E. T. A. Hoffmann*

”

Ach/ ich Unglückfeelig! was fange ich doch
nunmehr an?

*erster Satz aus »Der Academische Roman«
von Eberhard Werner Happel*

”

Es war Sommers-Frühe, die Nachtigallen fangen erft feit
einigen Tagen durch die Straßen, und verstumten heut in
einer kühlen Nacht, welche von fernen Gewittern zu uns
herwehte; der Nachtwächter rief die elfte Stunde an, da fah
ich, nach Haufe gehend, vor der Thür eines großen Gebäudes
einen Trupp von allerlei Gefellen, die vom Biere kamen, um
Jemand, der auf den Thürstufen faß, verfammelt.

*erster Satz aus »Geschichte vom braven Kasperl und dem schönen Annerl«
von Clemens Brentano*

”

Berlin schlief noch, aber es lag in jenem leifen
Schlummer, der dem Erwachen vorhergeht.

*erster Satz aus »Meister Timpe«
von Max Kretzer*

Abb. 2a/b/c: Twitter-Grafiken mit Roman-Anfängen.

”

Johann Heinrich Ludwig Hanemann wurde im Jahre
1803 in Hoya geboren, fiedelte in einem Alter von 5
Jahren nach dem hannöverfchen Städtchen Wunstorf im
Amt Blumenau, wo er bis zu feiner Konfirmation
verblieb, und begab sich dann, als er das Bäckergefchäft
erlernt, im Jahre 1819 nach Hamburg.

*erster Satz aus »Vom heimathlofen Vaterland«
von Ernst Dronke*

”

Weit hinaus im Meer ist das Wasser so blau, wie die
Blätter der schönsten Kornblume, und so klar, wie das
reinste Glas, aber es ist sehr tief, tiefer als irgend ein
Ankertau reicht; viele Kirchtürme müßten auf einander
gestellt werden, um vom Boden bis über das Wasser zu
reichen.

*erster Satz aus »Die kleine Seejungfrau«
von Hans Christian Andersen*

Abb. 1a/b/c: Twitter-Grafiken mit Novellen-Anfängen.

”

Wie –? Was –? rief man von allen Seiten.

*erster Satz aus »Die neuen Serapionsbrüder«
von Karl Gutzkow*

”

Louise naschte gern.

*erster Satz aus »Die Näscherinnen und das mäßige Kind«
von Karoline Stahl*



Es war einmal ein König, der hatte zwölf
Töchter, eine immer schöner als die andere.

erster Satz aus »Die zertanzten Schuhe«
von Jacob und Wilhelm Grimm

Abb. 3a/b/c: Twitter-Grafiken mit Märchen-Anfängen.

Bibliographie

- Alt, Peter-André** (2020): 'Jemand musste Josef K. verleumdet haben ...' Erste Sätze der Weltliteratur und was sie uns verraten. München: Beck.
- Beck, Harald** (1992): *Roman-Anfänge. Rund 500 erste Sätze*. Zürich: Haffmans.
- Beck, Harald** (1993): *Romanenden. Rund 500 letzte Sätze*. Zürich: Haffmans.
- Fischer, Frank / Strötgen, Jannik** (2015): "Wann findet die deutsche Literatur statt? – Zur Untersuchung von Zeitausdrücken in großen Korpora." Presented at the DHd2015 *Von Daten zu Erkenntnissen: Digitale Geisteswissenschaften als Mittler zwischen Information und Interpretation*. 2. Tagung des Verbands "Digital Humanities im deutschsprachigen Raum" (DHd2015), Graz: Zenodo. <http://doi.org/10.5281/zenodo.4623384> [letzter Zugriff: 6. Juli 2021]
- Fischer, Frank / Jäschke, Robert** (2018a): "Liebe und Tod in der Deutschen Nationalbibliothek. Der DNB-Katalog als Forschungsobjekt der digitalen Literaturwissenschaft." Presented at the DHd 2018 *Kritik der digitalen Vernunft*. 5. Tagung des Verbands "Digital Humanities im deutschsprachigen Raum" (DHd 2018), Köln: Zenodo. <http://doi.org/10.5281/zenodo.4622376> [letzter Zugriff: 9. Juli 2021]
- Fischer, Frank / Jäschke, Robert** (2018b): "Ein Quantum Literatur. Empirische Daten zu einer Theorie des literarischen Textumfangs." DFG-Symposium "Digitale Literaturwissenschaft". Villa Vigoni, 9.–13. Oktober 2017. [noch unveröffentlicht]
- Haubrichs, Wolfgang** (1995): "Kleine Bibliographie zu "Anfang" und "Ende" in narrativen Texten (seit 1965)", in: *Zeitschrift für Literaturwissenschaft und Linguistik* 25, 99: 36-50.
- Herrmann, Berenike** (2018): "Anschaulichkeit messen. Eine quantitative Metaphernanalyse an deutschsprachigen Erzählanfängen zwischen 1880 und 1926", in: Köppe, Tilmann / Singer, Rüdiger (eds.): *Show, don't tell: Konzepte und Strategien anschaulichen Erzählens*. Bielefeld: Aisthesis 167-212.
- Hirdt, Willi** (1974): "Incipit. Zu einer Poetik des Romananfanges", in: *Romanische Forschungen* LXXXVI: 419-436.
- Hug, Marius / Boenig, Matthias** (2021): *Die Geschichte der Digitalen Bibliothek, oder: Aller guten Kurationen sind drei+* <https://sprache.hypotheses.org/2436> [letzter Zugriff: 6. Juli 2021]
- Iser, Wolfgang** (1976): *Der Akt des Lesens. Theorie ästhetischer Wirkung*. München: Fink.
- Miller, Norbert** (1965): *Romananfänge. Versuch zu einer Poetik des Romans*. Berlin: Verl. Literarisches Colloquium.
- Moretti, Franco** (2009): "Style, Inc Reflections on Seven Thousand Titles (British Novels, 1740-1850)", in: *Critical Inquiry* 36, I: 134-158.
- Neuhaus, Stefan** (2019): "'Aber wehe, wehe, wehe! Wenn ich auf das Ende sehe!!' Wie in Romanen und Erzählungen durch Anfang und Ende ein Rahmen erzeugt wird", in: Neuhaus, Stefan / Weber, Petra (eds.): *Anfangen und Aufhören*. Paderborn: Wilhelm Fink 141-157.
- Queng, Jesse** (2019): "Syntaktische Strukturen als poetologisches Mittel des Anfangens in der Prosa: Der erste Satz von Heinrich Bölls Irischem Tagebuch", in: Neuhaus, Stefan / Weber, Petra (eds.): *Anfangen und Aufhören*. Paderborn: Wilhelm Fink 89-101.
- Raulff, Ulrich** (2019): "Letzte Sätze", in: *Zeitschrift für Ideengeschichte* 13: 129-142.
- Retsch, Annette** (2000): *Paratext und Textanfang*. Würzburg: Königshausen & Neumann.
- Richardson, Brian** (2008): *Narrative Beginnings: Theories and Practices*. University of Nebraska Press.
- Robbe-Grillet, Alain** (1992): "Warum und für wen schreibe ich", in: Bühler, Karl Alfred (ed.): *Robbe-Grillet zwischen Moderne und Postmoderne - "nouveau roman", "nouveau cinéma" und "nouvelle autobiographie"*. Tübingen: Narr.
- Selbmann, Rolf** (2019): "Lauter erste Sätze", in: Neuhaus, Stefan / Weber, Petra (eds.): *Anfangen und Aufhören*. Paderborn: Wilhelm Fink 67-87.
- Wolkersdorfer, Andreas** (1994): *Der erste Satz. Österreichische Romananfänge 1960-1980*. Wien: WUV Univ.-Verl.