

Digitale Modellierung von Figurenkomplexität am Beispiel des Parzival von Wolfram von Eschenbach

Braun, Manuel

manuel.braun@ilw.uni-stuttgart.de
Institut für Literaturwissenschaft, Universität Stuttgart

Klinger, Roman

roman.klinger@ims.uni-stuttgart.de
Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart

Padó, Sebastian

pado@ims.uni-stuttgart.de
Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart

Viehhauser, Gabriel

viehhauser@ilw.uni-stuttgart.de
Institut für Literaturwissenschaft, Universität Stuttgart

Einleitung

Figuren gehören zu den wichtigsten Bestandteilen literarischer Erzählungen. Narratologische Analysen haben sich bislang insbesondere mit zwei Aspekten von Figuren beschäftigt: ihrer strukturellen Bedeutung und ihrer Charakterisierung. Während der erste Aspekt in den Digital Humanities bereits modelliert worden ist (etwa im Rahmen von Netzwerkanalysen, vgl. Jannidis et. al. 2016, Piper et. al. 2017), steht die datengetriebene Untersuchung des zweiten noch ganz am Anfang, obwohl die hermeneutisch arbeitende Literaturwissenschaft wiederholt auf seine Bedeutung hingewiesen hat (etwa Jannidis 2004, 2009). Dieses Desiderat lässt sich darauf zurückführen, dass Figuren auf die unterschiedlichste Art gekennzeichnet werden – etwa durch ihr Handeln und Reden, aber auch durch Beschreibungen und Bewertungen des Erzählers – und dass neben expliziten Charakterisierungen schwerer zu greifende implizite stehen. Diese Informationen lassen sich nur schwer erheben und in einem umgreifenden Modell verrechnen.

Um dennoch einen Einstieg in die digitale Erfassung der Figurencharakteristik zu ermöglichen, nehmen wir eine vergleichsweise einfach zu modellierende Facette der Figurendarstellung in den Blick, und zwar das Konzept der Figurenkomplexität, das auch in der traditionellen Theoriebildung eine prominente Rolle spielt. Nach der

ebenso verbreiteten wie vielkritisierten Kategorisierung von Forster (1927) lassen sich Figuren grundsätzlich in *flat* und *round characters* einteilen. Trotz der zahlreichen Differenzierungsversuche späterer Forscher prägt dieses Modell noch heute Annahmen über die literaturgeschichtliche Entwicklung von Figurenentwürfen. So gelten die Protagonisten der mittelhochdeutschen Literatur klischeehaft als *flat characters*, die in erster Linie auf ihre Funktion für das Handlungsgefüge hin konstruiert werden und nicht die Tiefe moderner Individuen erreichen (vgl. zusammenfassend Schulz 2012). Im Widerspruch hierzu hat die mediävistische Literaturwissenschaft aber auch herausgestellt, dass die Konstruktion komplexer Figuren zu den grundlegenden Gestaltungsprinzipien mittelhochdeutscher Texte gehören kann.

Als Paradebeispiel für Figurenkomplexität kann der um 1200/1210 abgefasste Artusroman ‚Parzival‘ angesehen werden, dessen Autor Wolfram von Eschenbach im Prolog das Konzept eines ‚gemischten‘, nicht eindeutig als gut oder böse bewerteten Menschentypus entwirft, von dem er im Folgenden handeln will. Dementsprechend lassen sich im ‚Parzival‘ immer wieder Figuren finden, die nicht dem Ideal (nach ihm sind adelige Protagonisten nicht nur politisch mächtig, sondern zum Beispiel auch ethisch gut und körperlich schön) entsprechen, sondern in ihrer Darstellung widersprüchlich erscheinen und daher als komplex angesehen werden können.

Im Folgenden schlagen wir eine Methodik vor, mit der diese Einschätzungen der hermeneutischen Forschung digital modelliert und überprüft werden kann.

Methode

Wir untersuchen, ob man bereits mit extrem einfachen Methoden Vorhersagen zur Figurenkomplexität treffen kann, und verwenden dazu *distributionelle Analysen* (Harris 1954, Miller und Charles 1991). Distributionelle Analysen repräsentieren die Bedeutung eines Zielwortes durch Eigenschaften der Kontexte seines Vorkommens in Textkorpora – im einfachsten Fall durch die Häufigkeiten aller Kontextwörter innerhalb eines Fensters um die Vorkommen des Zielwortes („Kontextvektor“).

Spezifisch betrachten wir zwei Arten frequenzbasierter Kontextvektoren: (a) *lexikalische Kontextvektoren*, berechnet über lemmatisierte Kontexte; (b) *entitätsbasierte Kontextvektoren*, berechnet über Eigennamen in den Kontexten. Dabei misst (a) die *lexikalische* Vielfalt der Kontexte, in denen die jeweilige Figur vorkommt, und (b) entsprechend die *soziale* Vielfalt der Kontexte. Beide Arten von Repräsentationen verwenden als Kontexte die sogenannten *Dreißiger*-Abschnitte (die Unterteilung des ‚Parzival‘ in Abschnitte von 30 Versen nach Karl Lachmann) und greifen nicht auf sprachliche Strukturen innerhalb der Dreißiger wie Wortarten, Syntax, Satzgrenzen oder Anaphern zurück.¹

In der Computerlinguistik werden Kontextvektoren typischerweise miteinander verglichen, um die semantische

Ähnlichkeit der Zielwörter zu modellieren und finden breite Anwendung (Pantel und Turney 2010). In unserer Studie betrachten wir stattdessen den *Informationsgehalt* der Kontextvektoren als Prädiktor für die Figurenkomplexität und messen ihn per *Entropie*. Entropie eruiert den Informationsgehalt $H(p)$ einer Wahrscheinlichkeitsverteilung p , definiert als $H(p) = -\sum p(x) \log p(x)$ (Shannon 1948). Entropie kann verstanden werden als die Anzahl an Bits, die an Information in der Verteilung enthalten sind. Damit enthalten diejenigen Kontextvektoren eine höhere Entropie, die in gleichmäßiger verteilten Kontexten vorkommen, als die, die häufig in ähnlichen Kontexten auftreten.

Dies entspricht der (aus theoretischer Perspektive natürlich stark vereinfachenden) Annahme, wonach Figuren als komplexer wahrgenommen werden, wenn sie in reicheren und verschiedenere Kontexten vorkommen. Bei lexikalischen Kontexten (s.o.) sagen wir also voraus, dass höhere lexikalische Variabilität auf Komplexität hinweist – wobei Entropie freilich nicht zwischen lexikalisch reichen und semantisch widersprüchlichen Kontexten unterscheiden kann. Bei entitätsbasierten Kontexten übernimmt diese Rolle das gemeinsame Auftreten einer Figur mit mehreren anderen Figuren.

Methodisch ist anzumerken, dass Frequenz einen Störfaktor bei der Interpretation von Entropie darstellt: Häufigere Zielwörter kommen – *ceteris paribus* – häufiger in verschiedenen Kontexten vor und erhalten damit eine höhere Entropie. Da dieser Zusammenhang aber nicht linear ist, ist eine einfache Normalisierung nicht möglich. In einem ersten Schritt sind also nur Zielwörter mit ähnlicher Frequenz hinsichtlich ihrer Entropie gut vergleichbar.

Experiment

Als Textgrundlage verwenden wir den ‚Parzival‘ nach der 5. Auflage der Ausgabe Lachmanns (1891) in der digitalisierten, mit Lemmata und der Auszeichnung von Eigennamen versehenen Fassung von Yeandle (2014).

Für unser Hauptexperiment beschränken wir uns auf die sieben im Text am häufigsten namentlich genannten Frauenfiguren Cundrîe, Herzeloyde, Jeschûte, Cunnewâre, Arnîve, Bêne und Itonjê. Hinzu kommt Sigûne als ein in der Forschung oft genanntes Beispiel für eine weniger komplexe Figur im ‚Parzival‘. Sieben der acht Figuren liegen in einem engen Frequenzband zwischen 30 und 40 Nennungen, während Sigûne nur 14 Mal vorkommt. Wir präsentieren die Ergebnisse per Streudiagramm, mit Entropie und Frequenz als den beiden Achsen.

Abbildung 1 und 2 zeigen die Ergebnisse für die wichtigsten weiblichen Hauptfiguren. Diese lassen sich zwanglos mit den Einschätzungen der hermeneutischen Literaturwissenschaft in Einklang bringen. Die höchsten Werte weisen Cundrîe und Herzeloyde auf, auf deren Komplexität die ‚Parzival‘-Forschung immer wieder hingewiesen hat: Bei Cundrîe handelt es sich um die Gralsbotin, die zwar kultiviert und nach höfischen Sitten

gekleidet auftritt, zugleich aber monströs hässlich ist. Sie klagt Parzival zwar berechtigterweise an, fällt dabei aber aus dem höfischen Rahmen. Herzeloyde zieht ihren Sohn Parzival aus (übermäßiger?) Trauer um ihren im Kampf gefallenen Ehemann fern von der höfischen Welt in einer Waldeinöde auf, um zu verhindern, dass auch er einst Ritter wird – ein Verhalten, das die Forschung unterschiedlich bewertet. Mittelwerte erreichen Cunnewâre und Jeschute, die in der Forschung durchaus kontrovers diskutiert werden. Bêne, Arnîve und Itonjê lassen sich hingegen eher als Nebenfiguren bezeichnen, für die keine hohe Komplexität zu erwarten war.

Die in der Trauer um ihren Geliebten Schionatulander verharrende Sigûne, die gerade in ihrer Geradlinigkeit einen Gegenentwurf zu den auf Ruhm und Ehre versessenen Artusrittern darstellt, zeigt erwartungsgemäß keine hohen Entropiewerte. Allerdings erlaubt dieser Befund keine starke Interpretation, da für Sigûne aufgrund der niedrigeren Frequenz von vorneherein niedrigere Entropiewerte zu erwartet sind. Um dennoch eine Vorhersage zur Komplexität Sigûnes zu erhalten, vergleichen wir in den Abbildungen 3 und 4 ein breiteres Spektrum an Figuren im Frequenzbereich zwischen 10 und 20 Nennungen bezüglich ihrer Entropie. Dies entspricht einer erweiterten Version unserer Hypothese: Figuren, die *verglichen mit anderen Figuren des gleichen Frequenzbereiches* besonders hohe bzw. niedrige Entropien aufweisen, sind *relativ zu diesen* mehr bzw. weniger komplex.

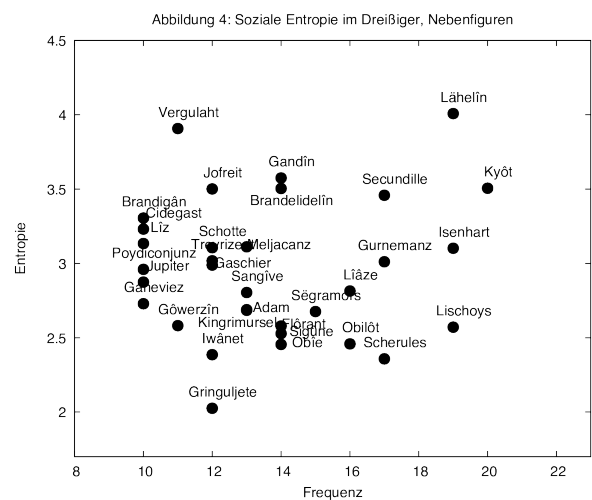
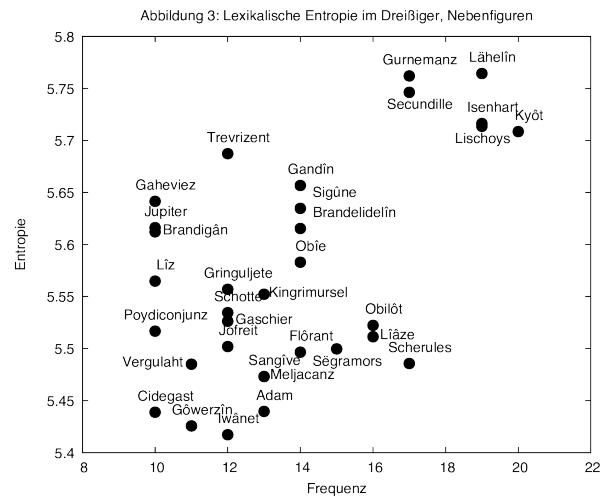
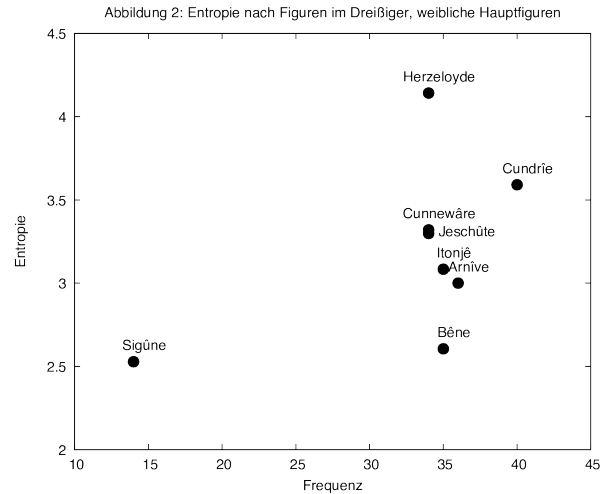
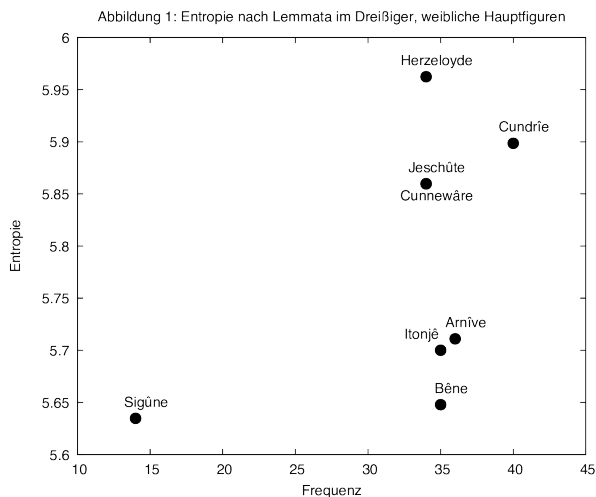
Interessanterweise liefern die beiden Kontextdefinitionen für Sigûne abweichende Vorhersagen: Die unterdurchschnittliche soziale Kontextvielfalt entspricht der Isoliertheit der Figur, die sich aus Trauer von der Welt zurückzieht und wenn, dann fast nur noch mit Parzival interagiert. Die vergleichsweise hohe lexikalische Vielfalt könnte demgegenüber darauf zurückzuführen sein, dass Sigûne neben ihrer Trauer auch die weitere Funktion hat, Parzival über seine Herkunft aufzuklären: Während Sigûne als Figur eher statisch erscheint, sind ihre Erzählungen über die Gralswelt informationshaltig. Für eingehendere Untersuchungen wären daher weitere Faktoren der Figurendarstellung wie Figurenbeschreibung und Figurenrede mit einzubeziehen sowie das Verhältnis von lexikalischer und sozialer Kontextvielfalt näher zu bestimmen.

Diskussion

Unsere Ergebnisse legen nahe, dass sich die Analyse der lexikalischen und ‚sozialen‘ Vielfalt von Figurenkontexten durchaus als Maß für eine erste Annäherung an das Konzept der Figurenkomplexität eignet. Damit konnte mit erstaunlich einfachen Mitteln ein Zugang zum komplexen narratologischen Themenfeld der Figurencharakterisierung gewonnen werden. Auf der technischen Ebene bleiben zwei Fragen offen, zum einen die nach den sprachlichen Mechanismen, die zur Korrelation von Entropie und

Figurenkomplexität führen, zum anderen die nach dem tatsächlichen Zusammenhang von Frequenz und Entropie: Gibt es also auch selten auftretende *round characters* oder häufig vorkommende flache Figuren?

Auf der konzeptuellen Ebene bieten unsere Ergebnisse den Einstieg in eine differenziertere Modellierung von Figurenkomplexität, deren Erarbeitung auf die traditionelle narratologische Theoriebildung zurückwirken kann: Die digitale Modellierung macht es notwendig, die Faktoren, aus denen sich das Konzept der Figurenkomplexität zusammensetzt (etwa Figurenhandeln, -rede, -beschreibung und -bewertung), genauer und expliziter zu bestimmen sowie über ihre Gewichtung nachzudenken. Außerdem wäre zu klären, was das Konzept der Komplexität genau meint und wie es sich zu verwandten Konzepten wie denen der Hybridität, der Dynamik oder der Aktantenhaftigkeit von Figuren verhält. Im Sinne einer kritisch reflektierten digitalen Methodik werden somit vielschichtige literaturwissenschaftliche Phänomene wie die Charakterisierung von Figuren durch den formalisierenden Zugang nicht nivelliert, vielmehr führt die Verbindung von quantitativen und qualitativen Methoden zur wechselseitigen Erhellung.



Fußnoten

1. Experimente mit kleineren Kontexten lieferten schlechtere Ergebnisse.

Bibliographie

Forster, Edward M. (1927): *Aspects of the Novel*. New York: Harcourt.

Harris, Zellig S. (1954): "Distributional Structure", in: *Word* 10(2-3): 146–162.

Jannidis, Fotis (2004): *Figur und Person*. Beitrag zu einer historischen Narratologie. Berlin: de Gruyter.

Jannidis, Fotis (2009): „Character“, in: Hühn, Peter / Pier, John / Schmid, Wolf / Schönert, Jörg (eds.) *Handbook of Narratology*. Berlin: de Gruyter 14-29

Jannidis, Fotis / Reger, Isabella / Krug, Manfred / Weimer, Lukas / Macharowsky, Luisa / Puppe, Frank (2016): „Comparison of Methods for the Identification of Main Characters in German Novels“, in: *Digital Humanities 2016: Conference Abstracts. Jagiellonian University & Pedagogical University, Kraków* 578-582.

Manning, Christopher D. / Schütze, Hinrich (1999): *Foundations of Statistical Natural Language Processing*. Cambridge, MA: MIT Press.

Miller, George A. / Charles, Walter G. (1991): "Contextual Correlates of Semantic Similarity", in: *Language and Cognitive Processes* 6(1): 1–28.

Piper, Andrew / Algee-Hewitt, Mark / Sinha, Koustuv / Ruths, Derek / Vala, Hardik (2017): "Studying Literary Characters and Character Networks", in: *Digital Humanities 2017: Conference Abstracts. Montreal, Canada, August 8-11, 2017*.

Schulz, Armin (2012): *Erzähltheorie in mediävistischer Perspektive*. Berlin: De Gruyter.

Shannon, Claude E. (1948): "A Mathematical Theory of Communication", in: *Bell System Technical Journal* 27(3): 379–423.

Turney, Peter D. / Pantel, Patrick (2010): "From Frequency to Meaning: Vector Space Models of Semantics", in: *Journal of Artificial Intelligence Research* 37(1): 141–188.

Yeandle, David (o. J.): *Stellenbibliographie zum 'Parzival' Wolframs von Eschenbach für die Jahre 1753–2004*. <http://wolfram.lexcoll.net/homeidx.htm> [letzter Zugriff 9. Januar 2017].