

Digitale Modellierung literarischen Raums

Barth, Florian

florianbarth@ilw.uni-stuttgart.de
Universität Stuttgart, Deutschland

Viehhauser, Gabriel

viehhauser@ilw.uni-stuttgart.de
Universität Stuttgart, Deutschland

Problemstellung

Im Anschluss an den 1990 durch den Humangeographen Edward Soja ausgerufenen ‚Spatial Turn‘ (Soja 1990) haben sich zahlreiche kulturwissenschaftliche Forschungsarbeiten mit einer Beschreibung des Raums beschäftigt. In der Literaturwissenschaft fanden dabei u.a. kartografische Darstellungen große Resonanz: Franco Moretti etwa untersuchte in seinem „Atlas of the European Novel“ Orte der literarischen Produktion und Rezeption (Moretti 1998), Barbara Piattis Studie „Die Geographie der Literatur“ richtete den Fokus auf die Illustration einer konkreten literarisch thematisierten Gegend (die Zentralschweiz, vgl. Piatti 2008). Besondere Aufmerksamkeit wurde literarischen Karten auch im Kontext der Digital Humanities zu Teil, in denen geografische Informationssysteme (GIS) zum Einsatz kommen (typische Workflows beschreiben Gregory et. al. 2015)

Den meisten dieser Ansätze ist dabei gemein, dass sie für ihre Datengrundlage in erster Linie auf konkrete Nennungen von Ortsnamen (Toponymen) rekurrieren und weitere Ortsmarker weniger stark berücksichtigen. An der Konstitution literarischer Räume sind jedoch in der Regel auch komplexere Faktoren beteiligt, zu deren Beschreibung bereits erste narratologische Ansätze vorliegen (etwa von Kathrin Dennerlein [2009 und 2011] oder Gabriel Zoran [1984], vgl. auch die Überlegungen bei Piatti [2008]), die jedoch im Kontext der Digital Humanities bislang noch zu wenig Beachtung gefunden haben.

In unserem Beitrag möchten wir diese Ansätze aufgreifen, um das Instrumentarium der digitalen Textanalyse hinsichtlich der Kategorie des Raums zu schärfen und zu erweitern. Dazu scheinen uns insbesondere zwei Aspekte von Bedeutung: Zum einem die Unterscheidung von Raummarkierungen hinsichtlich ihrer Handlungsrelevanz (I), zum anderen die Ausweitung der Analyse auf räumliche Begriffe, die über bloße Namensnennungen hinausgehen (II). Für beide Problemfelder präsentieren wir erste Verfahren zur automatischen Auswertung und geben Ausblicke auf die Möglichkeiten einer vergleichenden Analyse.

I. Differenzierung von Räumen nach Handlungsrelevanz

Im Anschluss an Dennerleins Narratologie des Raumes hat sich insbesondere der Terminus der *räumlichen Gegebenheit* als Grundeinheit zur Bezeichnung von Ort und Raum durchgesetzt. Sind räumliche Gegebenheiten der Schauplatz eines konkreten Ereignisses, werden sie als *Ereignisregionen* spezifiziert. Diese haben als die zentralen handlungsrelevanten Räume besondere Bedeutung gegenüber *erwähnten räumlichen Gegebenheiten*, die bei nicht-situationsbezogener Thematisierung von Raum entstehen. In ähnlicher Weise unterscheiden Piatti (2008) sowie Piatti et. al. (2011) zwischen *Schauplatz* und *projizierten Orten*.

Am Beispiel von Jules Vernes „Reise um die Erde in 80 Tagen“ lässt sich die Wichtigkeit dieser Unterscheidung aufzeigen: So bietet z.B. Kapitel 14 eine Zugfahrt durch das Gangestal mit Aufenthalten in Allahabad und Benares sowie der Ankunft in Calcutta. Genannt werden im Text jedoch auch weitere Städte Indiens und das nächste Ziel Hongkong; eine Vielzahl der extrahierten Ortsnamen bezieht sich somit nicht auf den Handlungsort des Kapitels.

Erst die kategoriale Trennung der Raummarkierungen in Ereignisregionen und erwähnte räumlichen Gegebenheiten ermöglicht die valide Rekonstruktion einer Reiseroute, die dann in einer GIS-Darstellung visualisiert werden kann (Abbildung 1).



Abbildung 1: Wegstrecke anhand zentraler Ereignisregionen in *Reise um die Erde in 80 Tagen* (QGIS-Grafik)

Eine automatische Unterscheidung dieser Kategorien muss daher das Fernziel computerunterstützter Raumuntersuchungen sein. Ansätze zu einer solchen Differenzierung suchen wir in der Anwendung von computerlinguistischen Methoden der Relationsextraktion, bei denen sich aus strukturellen Auffälligkeiten Regeln zur Klassifikation von Ereignisregionen und erwähnten räumlichen Gegebenheiten ableiten lassen. Hierzu zwei Beispiele:

1. Das 14. Kapitel der „Reise um die Erde in 80 Tagen“ beginnt mit dem in Abbildung 2 dargestellten Satz:

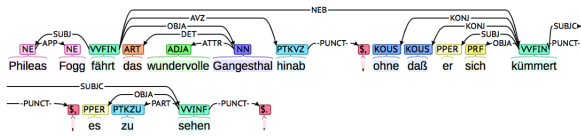


Abbildung 2: Dependenz-Pfad einer Ereignisregion (Malt Parser, Uni Tübingen)

Vor allem der Hauptsatz mit dem Pfad [Figur – SUBJ – Bewegungsverb – OBJ – Raumnomen] zwischen "Phileas Fogg" und "Gangesthal" kennzeichnet letzteres als Ereignisregion. Dabei stellt insbesondere die Verbkategorie ein Indiz für die Klassifikationsentscheidung dar: Statische Verben ("stehen", "sitzen") oder Verben der Bewegung ("gehen", "fahren") zeugen in spezifischen Satzstrukturen häufig von einer Ereignishaftigkeit im Gegensatz zu Verben der Kognition ("denken").

Zur Einteilung der Verben nutzen wir das von der Universität Tübingen entwickelte lexikalisch-semantische Netz *GermaNet* (Hamp/Feldweg 1997, Henrich/Hinrichs 2010), in dem eine Systematisierung vorliegt, die auf den Verbkategorien von Levin und dem Valenzwörterbuch von Schumacher basiert (Levin 1993, Schumacher 1986).

2. Findet ein erzähltes Ereignis innerhalb einer Bewegung im Raum statt, können mehrere Räume zu einem *Bewegungsbereich* zusammengefasst werden. Diese können bei Strukturen wie der folgenden leicht maschinell erkannt werden:

"Ich fuhr [...] von Königstadt, wo unser Hauptbüro war, nach Gründerheim, wo wir eine Nebenstelle hatten. Dort holte ich dringende Korrespondenz, Gelder und schwebende Fälle." (Böll, Heinrich: Über die Brücke)

Im zweiten Satz findet sich zudem eine Referenz auf das Antezedens "Gründerheim". Eigentlich werden derartige deiktische Adverbialausdrücke ("hier", "da" und "dort") bei der Koreferenz-Resolution nicht berücksichtigt. Deshalb planen wir eine Erweiterung der Trainingssets bestehender Koreferenz-Systeme hinsichtlich dieser Termini.

Die hier an zwei Beispielfällen dargestellten Regeln zur Unterscheidung narratologischer Raumeinheiten sollen in Zukunft kontinuierlich erweitert und zunächst so gestaltet werden, dass sie eine hohe Präzision erzielen. Anschließend können sie als Features für spätere maschinelle Lernverfahren verwendet werden.

II. Anreicherung der Raumbeschreibung

Netzwerkvisualisierung von Raumnomen

Eine kartografische Darstellung, wie sie in Abbildung 1 ersichtlich ist, bleibt auf realweltliche Ortsnamen beschränkt und lässt wesentliche Aspekte der Raumdarstellung außer Acht. Einen ersten Ansatz, die Vielfältigkeit der tatsächlichen Handlungsräume und ihrer Zusammenhänge abzubilden, bietet das Netzwerk in Abbildung 3. Hier werden für das angesprochene Kapitel 14 neben den Toponymen auch unspezifische Raumnomen als Knoten berücksichtigt und Verbindungen immer dann

etabliert, wenn zwei Raumbegriffe gemeinsam in einem Satz auftreten.

Insbesondere landschaftliche (grün) und architektonische Raumnomen (grau) stellen relevante Klassen von Raummarkern dar, die neben den konkreten Ortsnamen zentrale Komponenten der literarischen Raumbeschreibung bilden. Als räumliche Gegebenheiten kommen aber auch bewegliche Objekte, in denen sich Figuren aufhalten können, in Frage, wie die in der Grafik blau markierten Fahrzeuge.

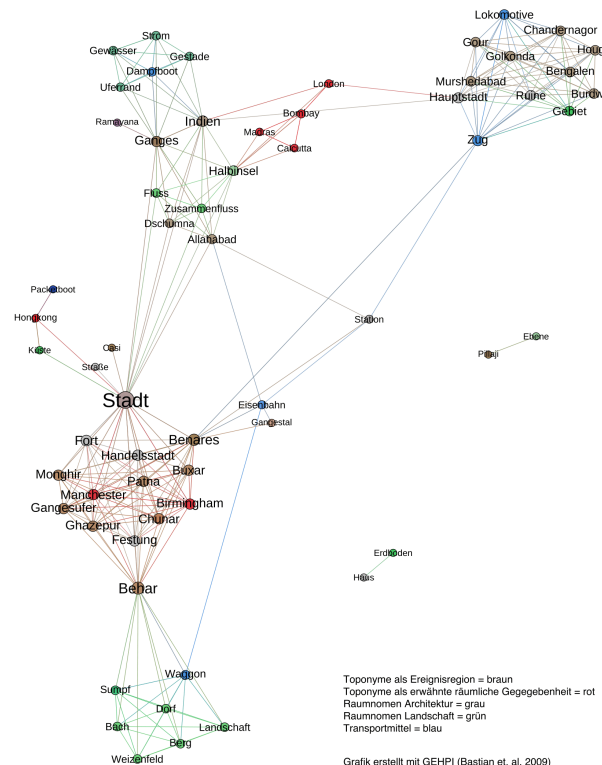


Abbildung 3: Netzwerk

Lexikon und Taxonomien für Raumbegriffe

Zur lexikalischen Erfassung von realweltlichen Toponymen greifen wir auf die Named-Entity-Recognition von *Weblicht* (2012) zurück, deren Ergebnisse wir mit dem Rückgriff auf die frei zugänglichen Datenbanken *GeoNames* (www.geonames.org) und *OpenStreetMap* (www.openstreetmap.org, Datendownload über www.geofabrik.de) zu verfeinern trachten. Aufgrund des Problems der möglichen Ambiguität von Ortsnamen (Leidner / Liebermann 2011, Gregory / Hardie 2011) ist jedoch eine manuelle Nachbearbeitung nötig. Listen von unspezifischen Raumnomen („Berg“, „Bach“, etc.) erstellen wir (ebenfalls semi-automatisch) auf der Basis von *GermaNet*.

Innerhalb dieses Lexikons planen wir zudem eine Einordnung der Raumbegriffe in spezifische Taxonomien:

i) Vertikale Raum-Ort-Hierarchie

Narratologisch wird unter dem Begriff *Raum* ein umfassendes Gebiet in der erzählten Welt verstanden, welches ein Innen und Außen besitzt und wiederum lokalisierbare, punktuelle *Orte* beinhaltet (Dennerlein 2009). Diese Zuordnung erfolgt jedoch meist relational zur Erzählsituation: In Alfred Döblins Roman *Berlin Alexanderplatz* bildet die Stadt den Raum mit einzelnen Plätzen und Straßen als Orten. Unter einer geringfügigen Erweiterung des Erzählspektrums wäre Berlin aber potentiell nur ein Ort unter vielen im übergeordneten Raum Deutschland.

Statt einer festen Zuschreibung nähern wir uns dem Verhältnis von Ort und Raum über eine vertikale Taxonomie von räumlichen Gegebenheiten an, die von der Planetenebene bis zu jenen Objekten reicht, in denen sich unter Annahme faktualer Gesetzmäßigkeiten keine Figur mehr aufhalten kann. Im Sinne des *principles of minimum departure* (Ryan 1980) kann dabei so lange von einer nach realweltlichen Gesetzen eingerichteten Erzählwelt ausgegangen werden, bis deren bewusste Aufhebung innerhalb fiktionaler Texte in spezifischen Fällen eine gezielte Anpassung der Taxonomie erfordert (z.B. bei Aladins Flaschengeist in der Wunderlampe).

Abbildung 4 zeigt basierend auf den kapitelweise extrahierten Ereignisregionen in „Reise um die Erde in 80 Tagen“ die obersten taxonomischen Stufen: 1. Kontinent, 2. Land, 3. Stadt bzw. landschaftliche Region (inklusive Transportmittel, markiert in blau).

Europa					Afrika			Asien						
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Großbritannien					Ägypten			Indien						
					Ägypten (SK)									
L	L	L	L	L	S	S	DM	DM,	B,	Zug,	Latenenwald	P,	A,	Z,
				(N)	(N)			Aden,	Zug,	Callyan	bei	BE,	BE,	C
								B			Allahabad	C	Behar,	C

Abbildung 4: Oberste Raumebenen der ersten 15 Kapitel in „Reise um die Erde in 80 Tagen“

Abkürzungen: (N) - Nebenfiguren

Europa: L - London

Afrika: S - Suez, SK - Sueskanal, DM - Dampfer Mongolia

Asien: RM - Rotes Meer, SA - Saudi Arabien, IO - Indischer Ozean, B - Bombay, A - Allahabad, P - Pillaji, BE - Benares, C - Calcutta

Während in den ersten beiden Ebenen dieses Beispiels ausschließlich Toponyme vorkommen, beinhaltet zumindest die dritte Stufe in der Nominalphrase „Latenenwald bei Allahabad“ ein unspezifisches Raumnomen. Weitere allgemeine Begriffe wären vor allem auf einer hier nicht dargestellten vierten Ebene zu finden (z.B. „Sumpf“, „Bach“, „Weizenfeld“ innerhalb der Landschaft „Behar“, vgl. Abb. 3).

Zur automatischen Erstellung einer solchen Hierarchie bieten sich bei Toponymen die in *GeoNames* vorhandenen Metadaten an, in denen bei jedem Stadt-Eintrag Informationen zu Land und Kontinent vorhanden sind. Bei unspezifischen Raumnomen eignet sich hingegen die hierarchische Struktur von *GermaNet*. So stellt etwa der Begriff „Bach“ ein Hyponym zum übergeordneten Synset

„Wasser/Gewässer“ dar, letzteres besitzt wiederum die Hyperonyme „Land/Gegend/Gefilde“.

ii) Wortfelder

Wie in Abb. 3 ersichtlich, speisen sich Raumnomen zu großen Teilen aus den Wortfeldern Architektur und Landschaft. Die folgende Analyse basiert auf semiautomatisch erstellten Wortlisten, die auf der Basis von *GermaNet* durch die Auswertung der entsprechenden Synsets und Implikationen (Hyperonymie und Hyponymie) von zentralen Begriffen aus beiden Wortfeldern gewonnen wurden.

Makroperspektive

Das Potential digitaler korpusgestützter Raum-Analysen soll anhand des Vergleichs dreier ‚Berlin-Romane‘ exemplarisch aufgezeigt werden. Dazu werden die Texte in jeweils zehn Segmente aufgeteilt und hinsichtlich der Frequenz spezifischer Raumbegriffe aus den Wortfeldern Architektur und Landschaft untersucht:

Dabei lassen sich deutlich höhere Anteile des architektonischen Vokabulars gegenüber dem Segmentmittelwerten eines Vergleichskorpus erkennen, das aus 451 im Textgrid-Repository enthaltenen Romanen besteht (Abbildung 5, oben).

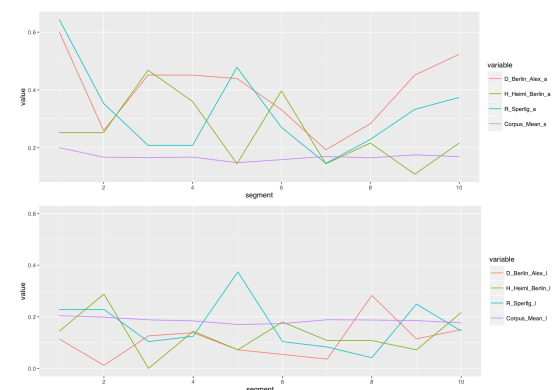


Abbildung 5: Architektonisches (oben) und landschaftliches (unten) Vokabular in *Berlin Alexanderplatz* (Alfred Döblin), *Heimliches Berlin* (Franz Hessel) und *Die Chronik der Sperlingsgasse* (Wilhelm Raabe) gegenüber dem mittleren Segmentverlauf im Korpus

Während die Verteilung des architektonischen Wortschatzes in Hesses „Heimliches Berlin“ nur temporäre Spitzen zeigt, sind die Segmentverteilungen von „Berlin Alexanderplatz“ und Wilhelm Raabes „Die Chronik der Sperlingsgasse“ gegenüber der mittleren Verteilung des Korpus signifikant verschieden. Dies wurde sowohl mit dem Wilcoxon-Rangsummentest (Annahme der Varianzhomogenität und Gleichverteilung zwischen den Sampleverteilungen) sowie dem Mood's Median-Test (keine Verteilungsannahme) überprüft (Abbildung 6).

Das landschaftliche Vokabular hingegen liegt bei den Berlin-Romanen tendenziell etwas unter dem Mittel des Korpus, allerdings sind die Abweichungen nur im Fall von „Berlin Alexanderplatz“ eindeutig signifikant (Abb. 5 unten, Abb. 6)

	Wilcoxon Architektur	Mood's Median Architektur	Wilcoxon Landschaft	Mood's Median Landschaft
Berlin Alexanderplatz	0.0001224	0.001093334	0.001408	0.001093334
Heimliches Berlin	0.07008	0.1788954	0.0225	0.1788954
Die Chronik der Sperlingsgasse	0.0007523	0.001093334	0.2406	0.6562818

Abbildung 6: p-Werte aus den Signifikanztests der Architektur- und Landschaftsverteilungen gegenüber den mittleren Segmenten im Korpus

Ungeachtet dieser Unterschiede sind die Zusammenhänge zwischen beiden Wortfeldern auffällig (Abbildung 7). Die Spearman-Korrelation zwischen architektonischen und landschaftlichen Begriffen bei „Berlin Alexanderplatz“ beträgt 0.5030488 und bei „Die Chronik der Sperlingsgasse“ sogar 0.7454545. So kann trotz abweichender Anteile der Wortfelder hinsichtlich ihrer Frequenz eine starke Verflechtung spezifischer Klassen von Räumen angenommen werden.

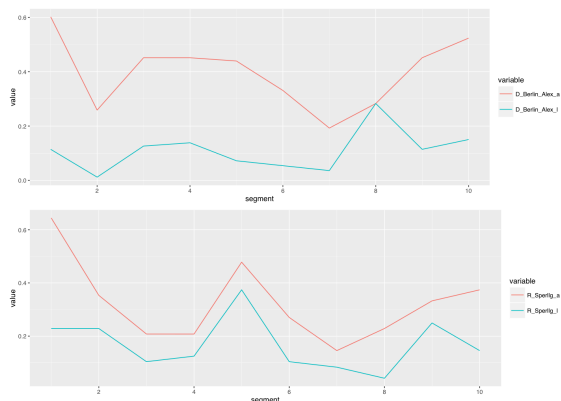


Abbildung 7: Korrelation zwischen den Wortfeldern Architektur und Landschaft in Berlin Alexanderplatz (oben) und Die Chronik der Sperlingsgasse (unten)

Ausblick

Die vorgestellten Ansätze verstehen sich als Anregung für die Entwicklung eines differenzierten Instrumentariums der digitalen Raumanalyse, das in Zukunft weiter ausgebaut werden soll und die Grundlage für die Behandlung weiterführender literaturwissenschaftlicher Fragestellungen bildet, die etwa Aspekte der Semantisierung von Räumen (Lotman 1972), des raumzeitlichen Entwurfs von Erzählwelten (Bachtin 1989) und der Bedeutung von Raumkonstellationen für die Gattungspoetik beinhalten (vgl. zusammenfassend Nünning 2009).

Bibliographie

Bachtin, Michail Michailowitsch (1989): *Formen der Zeit im Roman. Untersuchungen zur historischen Poetik*. Ed. von Kowalski, Edward / Wegner, Michael. Frankfurt am Main: Fischer.

Bastian Mathieu / Heymann Sebastian / Jacomy Mathieu (2009): „Gephi. An open source software for exploring and manipulating networks“, in: *International AAAI Conference on Weblogs and Social Media*.

Dennerlein, Katrin (2009): *Narratologie des Raumes*. Berlin: de Gruyter.

Dennerlein, Katrin (2011): „Raum“, in: Matías Martínez (ed.): *Handbuch Erzählliteratur: Theorie, Analyse, Geschichte*. Stuttgart / Weimar: Metzler 158–165.

Gregory, Ian / Hardie, Andrew (2011): „Visual GISTing: bringing together corpus linguistics and Geographical Information Systems“, in: *LLC* 26: 297–314.

Gregory, Ian / Cooper, David / Hardie, Andrew / Rayson, Paul (2015): „Spatializing and Analyzing Digital Texts. Corpora, GIS, and Places“, in: David Bodenhamer / John Corrigan / Trevor Harris: *Deep Maps and Spatial Narratives*. Bloomington: Indiana University Press 150–178.

Hamp, Birgit / Feldweg, Helmut (1997): „GermaNet - a Lexical-Semantic Net for German“, in: *Proceedings of the ACL workshop Automatic Information Extraction and Building of Lexical Semantic Resources for NLP Applications*. Madrid.

Henrich, Verena / Hinrichs Erhard (2010): „GernEdiT - The GermaNet Editing Tool“, in: *Proceedings of LREC 2010* 2228–2235.

Leidner, Jochen / Lieberman, Michael (2011): „Detecting Geographical References in the Form of Place Names and Associated Spatial Natural Language“, in: *SIGSPATIAL Special* 3: 5–11.

Levin, Beth (1993): *English Verb Classes and Alternations*. University of Chicago Press.

Lotman, Juri (1972): *Die Struktur literarischer Texte*. München: Fink.

Moretti, Franco (1998): *Atlas of the European novel. 1800-1900*. London / New York: Verso.

Nünning, Ansgar (2009): „Formen und Funktionen literarischer Raumdarstellung: Grundlagen, Ansätze, narratologische Kategorien und neue Perspektiven“, in: Wolfgang Hallet / Birgit Neumann (eds.): *Raum und Bewegung in der Literatur: Die Literaturwissenschaften und der Spatial Turn*. Bielefeld: Transcript 33–52.

Piatti, Barbara (2008): *Die Geographie der Literatur. Schauplätze, Handlungsräume, Raumphantasien*. Göttingen: Wallstein.

Piatti, Barbara / Reuschel, Anne-Kathrin / Hurni, Lorenz (2011): „A Literary Atlas of Europe – Analysing the Geography of Fiction with an Interactive Mapping and Visualisation System“, in: *Proceedings of the 25th International Cartographic Conference*. Paris.

Ryan, Marie Laure (1980): „Fiction, Non-Factuals, and Minimal Departure“, in: *Poetics* 8: 403–422.

Schumacher, Helmut (1986): *Verben in Feldern: Valenzwörterbuch zur Syntax und Semantik deutscher Verben*. Berlin / New York: de Gruyter Verlag,

Soja, Edward (1990): *Postmodern Geographies: The Reassertion of Space in Critical Social Theory*. London / New York: Verso.

WebLicht (2012): CLARIN-D/SfS-Uni. Tübingen 2012. *WebLicht: Web-Based Linguistic Chaining Tool*. <https://weblicht.sfs.uni-tuebingen.de/> [letzter Zugriff 1. Dezember 2016]

Zoran, Gabriel (1984): „Towards a theory of space in narrative“, in: *Poetics Today* 5: 309–335.