

Auf der Suche nach dem erfüllten Raum:
Digitale Korpusanalyse in der
Literaturwissenschaft
am Beispiel von Ilse Aichinger

Christine Ivanovic (Universität Wien)
Andrew U. Frank (Technische Universität Wien)

Universität Wien, Abt. für Vergleichende Literaturwissenschaft
Sensengasse 3A A-1090 Wien
christine.ivanovic@univie.ac.at

TU Wien, Geodäsie und Geoinformation
Gusshausstrasse 27-29/E120.2
A-1040 Wien
frank@geoinfo.tuwien.ac.at

1 Einleitung

Literarische Topographien sind mehr als geographisch verifizierbare Ortsangaben. In Literatur kann auf komplexe Weise das Profil historisch-kulturell identifizierbarer Räume musterhaft gestaltet, zu „Topoi“ verdichtet und diese als Kommunikationsformen sui generis eingesetzt werden. Ihre Analyse ermöglicht Aussagen über den Funktionszusammenhang von Gesellschaften im historischen Wandel. Sie bedingt notwendigerweise eine genaue Bestimmung der sprachlich-literarischen Formalisierung, welche die dichte Raumgestaltung im Rahmen literarischer Topographien erst ermöglicht.

2 Literarische Topographien in Digitaler Analyse

Um die Anlage und Wirkungsweise literarischer Topographien zuverlässig und umfangreich zu erfassen, versucht man seit einiger Zeit digitale Methoden für die Analyse genau begrenzter Korpora einzusetzen. Dabei werden verschiedene Ansätze verfolgt: Der historiographisch-diskursanalytisch orientierte Ansatz von

Franco Moretti beruhte auf statistischen Methoden, wie sie in der Soziologie Anwendung finden [1999], während eine Forschergruppe der ETH Zürich um Barbara Piatti eng in Verbindung mit der Kartographie arbeitete, um präzise Profile für die literarische Gestaltung konkreter geographischer Räume in eng umgrenzten Zeiträumen zu erstellen [2008, 2009]. Todd Presner begründete mit dem Projekt „HyperCities“ [2009] ein *Thick Mapping* insbesondere im Bereich der Großstadtliteratur.

Diese Ansätze entfernen sich dezidiert von traditionellen Methoden philologischer Textanalyse und deren Darstellungsmodi: Moretti vernachlässigte vor allem in seinen frühen Arbeiten vorsätzlich die Analyse der Textstrukturen zugunsten einer Erhebung all jener Daten, die sich auf die Schauplätze von Literatur beziehen (intra- wie extratextuell); Piatti und Presner konzentrieren sich auf digitale Visualisierungen zur Veranschaulichung der von ihnen aus den Texten gewonnenen topographischen Daten, um damit neue Analyseergebnisse zu generieren.

In dem hier vorzustellenden Projekt hingegen wird ein geschlossenes Œuvre untersucht im Hinblick auf die Generierung literarischer Topographien, auf die den Text strukturierenden Verfahren und auf die allgemeine Funktion der Raumdarstellung.

3 Raumreferenzen im Werk von Ilse Aichinger

Das Gesamtwerk der österreichischen Autorin Ilse Aichinger (geb. 1921 in Wien) bietet aus mehreren Gründen ein dafür geeignetes Untersuchungsbeispiel:

3.1 Günstige Korpusgröße

Aichingers Werk entstand innerhalb von sechs Jahrzehnten (1946-2006) und umfasst insgesamt 2100 Seiten. Die Texte liegen in einer leicht greifbaren Werkausgabe in 8 Bänden [1991] sowie in vier danach erschienenen Einzelbänden vor. Alle Bücher wurden gescannt und in eine OCR-Textdatei umgewandelt.

3.2 Klare Profilierung literarischer Topographien

Aichinger bezieht sich in einem großen Teil ihrer Werke explizit oder implizit auf den Raum der Stadt *Wien*, in der sie geboren wurde, wo sie die Judenverfolgungen miterlebte und in die sie nach jahrzehntelanger Abwesenheit 1988 zurückkehrte [Fässler, 2014].

Des Weiteren profiliert sie den Raum *England* als Sehnsuchtsort [Ivanovic, 2011a] sowie weitere Räume, an denen sie sich kürzere Zeit aufhielt (Frankreich, USA) oder wo sie länger lebte (österreichisch-bayerisches Grenzland).

Schließlich elaboriert sie *Heterotopien* wie das Kino [Ivanovic, 2011b], den Friedhof, das Meer sowie *typische/topische Räume*.

4 Arbeitshypothese

Aichinger „füllt“ einen (geographisch identifizierbaren, realen) Erfahrungsraum mit diversen, oftmals heterogenen Erinnerungsstücken unterschiedlicher geographischer, historischer und medialer Provenienz und verkettet diese mittels gleichbleibender sprachlicher Verfahren. Sie generieren keine neue Bedeutung im Sinne syntagmatischer Verknüpfung, stellen aber auf paradigmatischer Ebene neue Zusammenhänge her.

Dieser Vorgang entspricht grundsätzlich der Verknüpfung von Daten im www: die Anhäufung von Daten wird durch Verweise neu gruppiert und generiert dadurch neue Datensätze, die von sich aus aber noch keine in sich abgeschlossene, diskursiv organisierte Aussage bilden.

Bei Aichinger entsteht dadurch im Text - und nur hier - ein „erfüllter Raum“. Dieser bildet weder einen realen Raum noch ein einzelnes (historisches) Ereignis, das hier stattfand (oder stattfindet), ab. Der Text stellt vielmehr dar, wie das Bewusstsein an diesem konkreten Ort im Durchgang durch verschiedene Zeitschichten Geschichte bearbeitet.

Die digitale Analyse setzt bei der systematischen Erfassung der Raumreferenzen an. Sie zielt auf die Erarbeitung der strukturellen Merkmale und textuellen Strategien zur Generierung literarischer Topographien.

5 Erfassung

Die digitale Erfassung der Texte soll dementsprechend Analysen in zwei Untersuchungsbereichen ermöglichen:

1. Welche Orte werden mit welchen „Erinnerungsstücken“ verknüpft?
2. Welche wiederkehrenden sprachlich-literarischen Konstruktionen lassen sich dabei feststellen?

Mit dem gewählten Verfahren lassen sich die in Frage 1 fokussierten Informationen - Nennung von Orten, Zeiten, Personen, Medien - systematisch und vollständig erfassen. Das Erfassungssystem ist so aufgebaut, dass damit auf einfache Weise Abfragen auch in Bezug auf die in Frage 2 fokussierten sprachlichen Konstruktionen - Wiederholungen, Paarungen, Kombinationen, Kontextbildung etc. - möglich werden.

6 RDF als Hilfsmittel zur Codierung: Aus Informationen werden Daten

Die in Aichingers Texten explizit genannten Personen sowie die Orts- und Zeitangaben werden systematisch, einheitlich und vollständig erfasst. Dazu wird das gesamte publizierte Werk von Ilse Aichinger mit Hilfe einer speziellen Markup-Sprache kodiert und danach in RDF (Resource Description Framework)[Manola

et al., 2004, Hitzler et al., 2008] codiert: Aus Informationen werden digital analysierbare Daten.

Die Erfassung dieser Informationen im Gesamtkorpus ermöglicht eine Auswertung, die zuverlässige Aussagen über das Gesamtwerk der Autorin erlaubt. Die computergestützte Auswertung ersetzt die literaturwissenschaftliche Analyse nicht: RDF und SPARQL werden lediglich als zeitgemäße Werkzeuge zur Sichtung und Aufbereitung der in den Texten vorhandenen Informationen eingesetzt. Und von allen Informationen, die ein literarischer Text enthält, wird in unserem Ansatz lediglich eine einzige Kategorie in drei unterschiedlichen Bereichen berücksichtigt: zunächst geht es nur und allein um die Erfassung von Personennamen sowie Orts- und Zeitangaben.

7 Digitale Analyse

Der annotierte Text wird in RDF (im Turtle Format) formatiert und in einen SPARQL endpoint eingebracht (wir verwenden 4store). Einfache Abfragen zur Kontrolle der Erfassung und zum Finden von Text-Abschnitten nach den üblichen Kriterien (Seitenzahl, Titel, Textstellen) werden von einem derartigen System erwartet.

Beispiele für quantitative Abfragen:

- Wie viele (und welche) Texte enthalten Bezugnahmen auf geographisch referenzierbare Orte, wie viele (und welche) Texte enthalten nur Bezugnahmen auf typische Orte? Welche Text enthalten keine Bezüge zu Orten?
- Welche sind die am häufigsten von Aichinger in Wien genannten Orte, welche kommen nicht vor?
- In welchem Zeitraum ihres Schreibens spricht sie von Wien, in welchem Zeitraum gibt es keine Bezugnahmen auf die Stadt?
- Wie verhalten sich Nennungen öffentlicher Räume zu Nennungen von privaten Lokalitäten?
- Wie verhalten sich Bezugnahmen auf Aussenräume zu Bezugnahmen auf Innenräume?

Die Darstellung von Orten in einer Karte erlaubt den Überblick über den durch den Text beschriebenen Raum und zeigt allenfalls auch Räume, die ausgespart werden („Negative Räume“) und Ziel einer möglicherweise lohnenden Analyse sein könnten. Es kann untersucht werden, ob sich der örtliche Schwerpunkt der Bezüge im Laufe der Schreibzeit oder im Laufe der Erzählzeit verändert. All dies kann in Diagrammen und Chronologien dargestellt werden.

```

»Haben schon gewählt?« heißt es bei Demel. Aber welche Wahl
im Leben ist offen?
.propText
    lit:ort      :Demel.
:Demel lit:name  "K. & K. Hofzuckerbäcker Demel";
    lit:google  "Kohlmarkt 14 , 1010 Wien";
    lit:wiki    "Demel".

```

Abbildung 1: Ausschnitt aus codiertem Text; ein Ort muss nur einmal im Detail (*name*, *google*) beschrieben werden, später genügt der eingeführte Bezeichner (*Demel*).

8 Wahl der technischen Lösung

Unsere Ziele für die technische Realisierung waren: (1) Langfristige Nutzung: Die erzeugten Daten können in spätere Anschlussysteme migrieren; (2) Quelloffen; (3) Standardisierung.

Die von der W3C, dem Steuerungsgremium des World Wide Web, in den Richtlinien "Semantic Web" [Berners-Lee et al., 2001] vorgeschlagenen Methoden zur Erfassung von Bedeutung in Form von strukturierten Daten und einer zugehörigen Abfragesprache, scheinen für die vorliegende Aufgabe passend. Die Trennung in eine Markup Sprache, die Kodierungen im Text erlaubt, und die Speicherung und Verarbeitung mit RDF hat sich soweit bewährt.

Die technische Lösung und insbesondere die Codierung muss ausbaufähig bleiben; die gesuchten Eigenschaften können möglichst objektiv bestimmt werden. Das erlaubt es, die Codierungs-Arbeit z.B. in einem Seminar auf alle Teilnehmer aufzuteilen (crowdsourcing) und dennoch einen einheitlichen Datenbestand zu generieren.

Die Codierung muss kontextrelevant angelegt sein, wobei der relevante Kontext der Abschnitt (nicht ein Wort, nicht ein Satz) ist, in dem ein Ort erwähnt wird. Bei Abfragen werden die angesprochenen Abschnitte ausgegeben um die Interpretation zu gewährleisten.

8.1 Methode der Annotation

In die mittels OCR erzeugten Textdateien werden nach jedem Abschnitt in RDF Turtle Syntax die erwähnten Orte, Personen und Zeiten codiert und mit anderen Datenbanken (z.B. Google Maps oder Wikipedia) verknüpft (Fig. 1). Dieser annotierte Text wird dann von einem eigenen Programm in RDF Syntax umgewandelt und diese in einen SPARQL endpoint geladen. Damit kann aus dem Web mit einem üblichen Browser gesucht und die Ergebnisse angezeigt werden.



Abbildung 2: Ein Ausschnitt mit einigen Textstellen zu Wien

8.2 Kartographische Ausgabe

Karten, die die gewonnene Information räumlich darstellen, sind ein wichtiger Teil des Projektes und lassen sich leicht mit bereits bestehenden Werkzeugen erstellen:

1. Textstellen, mit bestimmten Eigenschaften und die darin erwähnten Orte werden mittels einer SPARQL Abfrage herausgesucht und als CSV Datei dargestellt (CSV - Comma Separated Values, ein übliches Format für die Übermittlung von Tabellen). Zum Beispiel, um alle Abschnitte, die einen konkreten Ort beschreiben, herauszusuchen und mit der Ortsbeschreibung auszugeben genügt:

```
Select ?absatz ?titel ?googleName where
{ ?absatz lit:ort ?ort .
  ?ort lit:google ?googleName .
  ?absatz lit:titel ?titel .
}
```

2. Die Datei mit den drei Kolonnen *Absatz*, *Titel* und *Location* wird in Google Docs hochgeladen und dort mittels Fusion Tables angezeigt. Dabei werden die im Feld *Location* (= *googleName*) verwendeten Adressen in Koordinaten umgesetzt und als rote Punkte angezeigt; die Karte kann im Massstab verändert und der Ausschnitt frei gewählt werden; beim Klicken auf einen Punkt wird die Textstelle geöffnet (Fig. 2):

Auf ein Beispiel kann unter

<https://www.google.com/fusiontables/DataSource?docid=1BCoyc2Pho9GkVTTc9gfKbfQJ4Mf4nHiy6-33rQww#map:id=3>

zugegriffen werden.

9 Zusammenfassung und Ausblick

Ziel ist die Erarbeitung einer zuverlässigen, anwendungsneutralen und anpassungsfähigen digitalen Analysemethodik, die auch bei anderen Textcorpora und veränderten Fragestellungen eingesetzt werden kann. Besonders interessant ist die Möglichkeit der Verknüpfung mit bestehenden Datenbanken; bei Aichinger z.B. die Verbindung zwischen den erwähnten Filmtiteln und der Filmdatenbank, die Darsteller etc. bereitstellt und in SPARQL Abfragen nutzbar macht. Es wird also möglich, alle Abschnitte, in denen ein Film mit einem bestimmten Schauspieler erwähnt wird, zu finden.

Literatur

- Ilse Aichinger. *Werke*. Fischer-Taschenbuch-Verl., 1991.
- Tim Berners-Lee, James Hendler, Ora Lassila, et al. The semantic web. *Scientific american*, 284(5):28–37, 2001.
- Simone Fässler. *Von Wien her, auf Wien hin. Ilse Aichingers "Geographie der eigenen Existenz"*. Böhlau, 2014.
- Pascal Hitzler, Markus Krötzsch, Sebastian Rudolph, and York Sure. Semantic web. *Berlin, Heidelberg*, 2008.
- Christine Ivanovic. Nach England! Zur Geschichte einer Sehnsucht. In Rüdiger Görner, Christine Ivanovic, and Sugi Shindo, editors, *Wort-Anker Werfen. Aichinger und England*. Königshausen & Neumann, 2011a.
- Christine Ivanovic. Masse. Medien. Mensch. Ilse Aichingers bioskopisches Schreiben. In Christine Ivanovic and Sugi Shindo, editors, *Absprung zur Weiterbildung. Geschichte und Medien bei Ilse Aichinger*. Stauffenburg, 2011b.
- Frank Manola, Eric Miller, Brian McBride, et al. RDF primer. *W3C recommendation*, 10(1-107):6, 2004.
- Franco Moretti. *Atlas of the European novel, 1800-1900*. Verso, 1999.
- Barbara Piatti. *Die Geographie der Literatur: Schauplätze, Handlungsräume, Raumphantasien*. Wallstein, 2008.
- Barbara Piatti, Hans Rudolf Bär, Anne-Kathrin Reuschel, Lorenz Hurni, and William Cartwright. *Mapping literature: Towards a geography of fiction*. Springer, 2009.
- Todd Samuel Presner. Hypercities: Building a web 2.0 learning platform. *Teaching Literature at a Distance*, page 171, 2009.