

Wann findet die deutsche Literatur statt?

Zur Untersuchung von Zeitausdrücken in großen Korpora

Frank Fischer¹ und Jannik Strötgen²

¹ Göttingen Centre for Digital Humanities, Universität Göttingen, Deutschland

² Institut für Informatik, Universität Heidelberg, Deutschland

1 Einleitung

Exakte Datumsangaben sind ein Merkmal vieler Prosatextsorten. In der Literatur finden sich dagegen bevorzugt ungefähre Datumsangaben, die Interpretationsräume öffnen. So sind etwa alle 19 Monatsnennungen in Theodor Storms *Schimmelreiter* ungefährender Natur („an einem Octobernachmittag“, „Ende März“ usw.). Ausnahmen von dieser Regel bilden Tagebuch-, Brief-, Abenteuer- und historische Romane: In Goethes *Werther* oder Jules Vernes *Tour du monde en quatre-vingts jours* wimmelt es genretypisch von exakten Datumsangaben, im letztgenannten Roman sind sie in Form eines Countdowns gar unentbehrlicher Teil der Handlung. Ansonsten lässt sich als Hypothese formulieren: Kommt in der erzählenden Literatur ein exaktes Datum vor, ist das eine narrative Setzung, die der näheren Analyse lohnt. Davon ausgehend lassen sich dezidiert literaturwissenschaftliche Fragen ins Blickfeld nehmen:

- Ist die Vermeidung exakter Datumsnennungen tatsächlich ein durchgehendes Merkmal bestimmter literarischer Genres? In welchem Verhältnis stehen exakte zu ungefähren Datumsangaben?
- Kann die Frequenzanalyse von Datumsangaben zu Genreuntersuchungen genutzt werden?
- Welche andere Bedeutung haben Datumsangaben neben der zeitlichen Verortung der Handlung? (‘Semiotisierung’ eines Datums)
- Gibt es zeitliche Konjunkturen für bestimmte Datumsnennungen? Wenn ja, warum?
- Welche Rolle spielen fiktive Daten? (vgl. etwa Erich Kästners *35. Mai* und Shakespeares 80. April in *The Winter’s Tale*, Autolycus’ Ballade im 4. Akt)

In unserem Vortrag widmen wir uns den Datumsangaben als einem isolierbaren *feature* literarischer Korpora im Sinne Matthew Jockers: „Indeed, the very object of analysis shifts from looking at the individual occurrences of a feature in context to looking at the trends and patterns of that feature aggregated over an entire corpus“ (Jockers 2013, S. 24). Anhand dieses Features (der expliziten Datumsnennung) als einzeln betrachtbare Einheit soll die Praxis der literaturwissenschaftlichen Makroanalyse methodisch bereichert werden. Die Extraktion der Zeitangaben erfolgt automatisiert mithilfe eines Temporal Taggers. Indem es dabei um die Untersuchung der Repräsentanz eines außerliterarischen Phänomens (Zeit, Datumsangaben) in literarischen Texten geht, wird auch ein Beitrag zur Erzählforschung geleistet. Die mit den Methoden der Digital Humanities erzielten Erkenntnisse werden dadurch in die Fachdisziplin (in diesem Fall die Literaturwissenschaften) zurückgespielt.

2 Vorgehen

Unser Vorgehen bestand aus vier meist parallel ablaufenden Schritten: 1. Zusammenstellung geeigneter (deutschsprachiger) Korpora. 2. Erhebung der Daten durch Einsatz des Temporal Taggers HeidelbergTime (Strötgen & Gertz 2012) zur automatischen Extraktion zeitlicher Ausdrücke im Sinne der Temporal Markup Language TimeML (Pustejovsky et al. 2003). 3. Analyse der Daten (von Heatmaps zu Einzelfällen). 4. Entwicklung einer Android-App zur explorativen Analyse des „literarischen Jahres“.

Zunächst haben wir mit dem TextGrid Repository¹ und Gutenberg-DE² zwei große literarische Korpora zusammengebracht, mit HeidelbergTime auf Datumsstrukturen untersucht und anhand der expliziten (und damit sehr sicher richtigen) Ausdrücke eine kalendarische Heatmap erstellt (‘1’ bedeutet 0–9 Vorkommen, ‘2’ bedeutet 10–19 Vorkommen usw., ‘+’ bedeutet 90 oder mehr Vorkommen). Dabei zeigten sich erwartete und unerwartete Konjunkturen:

```
JAN: +566554435455576554574445555455
FEB: 65655546454635554446554666762
MAR: 84553334565364646+4644435444465
APR: +55555344664466554465555646447
MAY: +777765557566565674836454565466
JUN: 957486574646656657586656444576
JUL: 9479486554468975676654555565465
AUG: +6+565555+5665+6974755775555556
SEP: 9745735555446575+7695554546457
OCT: +375564536445665+76734555645555
NOV: +68557546656549665554645545456
DEC: 77455755464455547644554+5533457
```

Wie man sieht, kommen Monatserste und fixe Feiertage (Neujahr, Weihnachten) besonders häufig vor. Ansonsten fiel die Konjunktur weiterer Tage auf, etwa des ‘18. März’. Die Vermutung lag nahe, dass die Nennung dieses Datums in Werken nach 1848 und damit nach der Märzrevolution ansteigt, da dieser Tag wegen der blutigen Ereignisse in Berlin eine eigene Semantik annahm. Die Erhebung und Analyse solcher Datumskonjunkturen soll systematisch ausgebaut werden.

3 Untersuchung des DTA-Korpus

Dieser ersten quantitativen Analyse lagen die beiden erwähnten Korpora zugrunde, in denen aber auch nichtliterarische Werke und (v. a. bei Gutenberg-DE) auch Übersetzungen fremdsprachiger Literatur vorkommen. Um mit dieser Methode belastbare Ergebnisse zu gewinnen, bedurfte es eines qualifizierteren Textkorpus. Das Deutsche Textarchiv (DTA) ist zwar weit weniger umfangreich, aber es beinhaltet (nomen est omen) nur original deutschsprachige Texte und ist sowohl grob zeitlich – nach Jahrhunderten – als auch thematisch – nach Genres – sortiert, sodass dort Hinweise auf Konjunkturen in literarischen Texten zu erwarten waren. In Tabelle 1 sind Informationen zu den DTA-Teilkorpora dargestellt.

Die Anzahl der enthaltenen Texte stellt zwar letztlich keine kritische Masse für large-scale Untersuchungen wie die unsere dar, wir versprachen uns aber weitere Hinweise auf Chancen und Grenzen der Methode. Eine Analyse des einzeln abrufbaren Belletristik-Korpus ergab eine weit weniger stabile Heatmap als oben und als häufigste Daten (zwischen ca. 20 und 70 Nennungen) die folgenden: 1. 1., 1. 4., 20. 4., 1. 5., 10. 8., 3. 11.

Die DTA-Stichprobe hebt wieder die Bedeutung des 1. Mai heraus, unter den unerwarteten Daten sei der ‘10. August’ herausgegriffen, der neben der zeitlichen Verortung fiktionaler Handlungen wiederum auch auf ein historisches Datum zurückzuführen ist, den Tuileriensturm am 10. August 1792 (vgl. in Büchners *Dantons Tod*: „ERSTER BÜRGER: Danton war unter uns am 10. August,

¹<http://www.textgridrep.de/>

²<http://gutenberg.spiegel.de/>

	Dokumente	Sätze	Tokens	TimeML Zeitausdrücke	explizite Tagausdrücke
1600–1699	124	957,950	15,779,536	34,425	635
1700–1799	341	1,866,106	30,077,557	97,219	1,531
1800–1899	559	3,565,758	60,750,795	289,697	18,644
Belletristik	401	1,774,209	29,303,975	106,988	3,366
Gebrauchsliteratur	103	624,829	11,041,708	35,079	787
Wissenschaft	532	4,103,679	68,268,553	298,195	17,170

Tabelle 1: Informationen zu den Teilkorpora des DTA-Korpus.

Danton war unter uns im September.“). Das DTA-Korpus ist aber, wie gesagt, relativ klein (nur 401 belletristische Werke, wobei Belletristik hier neben Fiktionalem auch Reiseberichte und Lebenserinnerungen einschließt) und nicht sehr belastbar für Korpusanalysen.

4 Bedeutung für die Literaturwissenschaft

Es wäre analog zu Hans Ulrich Gumbrechts Untersuchung *1926* denkbar und wünschenswert, dass man die Literaturgeschichte einzelner Tage schreibe. Dass jedes Datum seine eigene Literaturgeschichte hat, dies also in Ansätzen schon untersucht wird (allerdings noch ohne Korpusanalysen o. Ä.), zeigt das Beispiel Paul Celan und der ‘20. Jänner’.

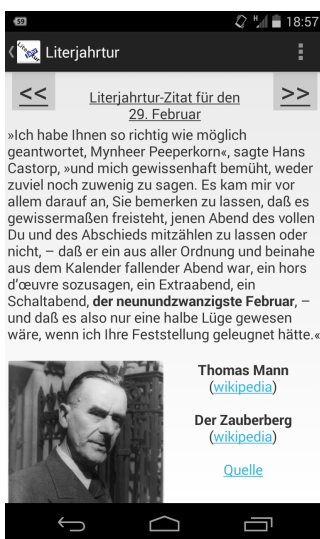
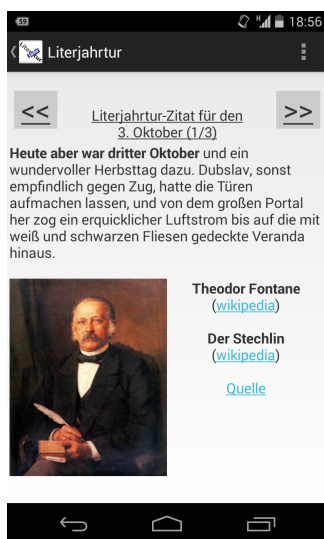
In Celans Prosagedicht *Gespräch im Gebirg*, 1960 in der *Neuen Rundschau* (Heft 2) erschienen, wird auf Georg Büchners Erzählung *Lenz* angespielt, in der ebenfalls ein Gang durchs Gebirge geschildert wird. Büchners Text beginnt mit dem Satz: „Den 20. [Jänner] ging Lenz durchs Gebirg.“ Dass Lenz’ Wanderung durchs Gebirge am 20. Jänner stattfindet, darauf weist auch Celan in seiner Büchner-Preis-Rede *Der Meridian* hin. Er erweitert den Anspielungsraum noch, indem er auf einen weiteren 20. Januar verweist, nämlich den des Jahres 1942, als in Berlin die Wannseekonferenz stattfand. Und er folgert: „Vielleicht darf man sagen, daß jedem Gedicht sein ‘20. Jänner’ eingeschrieben bleibt?“ (vgl. Sieber 2007, S. 114ff).

Die computergestützte Erhebung von Zeitangaben aus großen Korpora macht solche Gleichzeitigkeiten sichtbarer und ermöglicht so auch deren systematische Erforschung.

5 Android-App zur Exploration expliziter Zeitausdrücke in der Weltliteratur

Um eine Idee für die Jahreszeitlichkeiten der Literatur zu entwickeln, haben wir parallel zum Projekt eine Android-App entwickelt, die als Kalender funktioniert und für jeden Tag des Jahres Passagen aus der Weltliteratur anzeigt, die an diesem Tag spielen. Beispiele sind in Abbildung 1 dargestellt.

Die von uns entwickelte und mit HeidelbergTime belieferte App soll die einfache Exploration expliziter Zeitausdrücke in der Weltliteratur ermöglichen. Dass James Joyce’ *Ulysses* etwa am 16. Juni 1904 stattfindet (‘Bloomsday’), ist allgemein bekannt, allerdings im Text nicht sofort ersichtlich. In der App wird die einzige Stelle des überlangen Romans, an dem das Datum erwähnt wird, zitiert (die Sekretärin Miss Dunne tippt es auf ihrer Schreibmaschine ein). Weitere Beispiele für Passagen sind der 12. Juni in der *Blechtrommel* (Oskar Matzeraths erklärter Sohn Kurt wird geboren), der 29. Februar in Thomas Manns *Zauberberg* (der als spezielle Variante der Walpurgisnacht eine zentrale Rolle spielt, siehe Abbildung 1) oder der 27. Juli in Stefan Zweigs *Schachnovelle* (der Protagonist Dr. B. eignet sich an diesem Tag das Schachbuch an). Die App stellt somit ein erweiterbares Korpus mit Datumsnennungen in der Weltliteratur dar, das der weiteren Forschung zur Verfügung steht. Um allerdings das gesamte „literarische Jahr“ abzubilden, müssen auch die *ungefähren* zeitlichen Verortungen in den Blick genommen werden, was im nächsten Abschnitt versucht werden soll.



	Fontane	Storm
JAN	30	5
FEB	13	3
MAR	18	7
APR	13	7
MAY	28	11
JUN	17	9
JUL	16	6
AUG	17	10
SEP	36	10
OCT	41	11
NOV	27	13
DEC	25	1

Abbildung 1: Screenshots unserer Android-App *Literjahrtr*.

Tabelle 2: Fontane vs. Storm.

6 Die Jahreszeiten der Literatur

Von dem für literarische Texte sehr spezifischen Verhältnis zwischen exakten und ungefähren Datumsangaben war schon die Rede. Die Analyse des DTA-Korpus nur mit (auch nicht-literarischen) Texten aus dem 19. Jahrhundert hat eine besondere Konjunktur für die Monate März bis Juli erbracht:

JAN: +575544555454453554443444444667
 FEB: 53553443533424444343445645352
 MAR: +6667676789669+49+8+6877888699+
 APR: +78878876+76597+75699+89668869
 MAY: ++6+968758+899+8886+9+98987+9++
 JUN: +8489768+++++978697+976++9988+
 JUL: +789987+758978+7765887565767755
 AUG: 655654445845555554454455445345
 SEP: 644333434433345345334322333334
 OCT: 722222222233332231242323233324
 NOV: 845553454434436445534344443434
 DEC: 2423222216222222112222242141123

Bei der Suche nach Abweichungen im Werk einzelner Autoren des 19. Jahrhunderts stießen wir etwa auf Theodor Fontane und Theodor Storm. Eine Erhebung nur der Monatsnennungen in ausgewählten fiktionalen Texten beider Autoren ergab das in Tabelle 2 verzeichnete Bild.

Analog zur Popularität des 1. Mai ist auch der Gesamtmonat bei beiden stark repräsentiert. Doch für die Sommermonate gilt das nicht. Fontanes Romane und Erzählungen scheinen vor allem von September–Januar stattzufinden, Storms Texte von August–November. Auch unter der Maßgabe, dass der Monatsname als sprachlich-klangliches Zeichen einen stilistischen Effekt hat, scheinen beide Autoren herbstlich-winterliche Settings und Stimmungen zu bevorzugen.

Mit den vorgestellten Methoden zur Ermittlung von Datumskonjunkturen, zur Beschreibung des Verhältnisses zwischen ungefähren und exakten Datumsangaben, zum Aufbau eines Korpus mit exakten Datumsnennungen und zur Jahreszeitlichkeit der Literatur und bestimmter Autoren kann die im Titel gestellte Frage „Wann findet die deutsche Literatur statt?“ tatsächlich makroanalytisch beantwortet werden. Damit die getroffenen Aussagen tatsächlich literaturhistorisch belastbar sind, soll für die weitere Forschung ein größeres und besser (v. a. mit Entstehungs-/Veröffentlichungsdaten) ausgezeichnetes Korpus zusammengestellt werden.

Literatur

- Matthew Jockers. *Macroanalysis. Digital Methods and Literary History*. Chicago: University of Illinois Press, 2013.
- James Pustejovsky, José M. Castaño, Robert Ingria, Roser Sauri, Robert J. Gaizauskas, Andrea Setzer, Graham Katz and Dragomir R. Radev. TimeML: Robust Specification of Event and Temporal Expressions in Text. In *New Directions in Question Answering*, S. 28–34, 2003.
- Mirjam Sieber. *Paul Celans "Gespräch im Gebirg". Erinnerung an eine versäumte Begegnung*. Tübingen: Niemeyer, 2007.
- Jannik Strötgen, Michael Gertz. Temporal Tagging on Different Domains: Challenges, Strategies, and Gold Standards. In *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC 2012)*, S. 3746–3753, 2012.