



FORSCHUNGSSTELLE
DIGITAL HUMANITIES CENTER
UNIVERSITÄT KONSTANZ

Scalable Reading

Paul Heyses *Deutscher Novellenschatz*
zwischen Einzeltext und Makroanalyse

Kulturwissenschaftliches Kolleg Konstanz
Schloss Seeburg, Kreuzlingen
4. - 6. Juni 2015

Organisation und Kontakt
Prof. Dr. Thomas Weitin
thomas.weitin@uni-konstanz.de

Wissenschaftliche Zielsetzung

1. Textgegenstand zwischen Literaturgeschichte und Digital Humanities

Gegenstand der Tagung ist der von Paul Heyse und Hermann Kurz 1871-1876 herausgegebene *Deutsche Novellenschatz*, eine 24-bändige Anthologie mit insgesamt 86 Novellen. Die Literaturgeschichte hat nicht wenige Einzeltexte davon (z.B. Hoffmanns *Fräulein von Scuderi*, Drostes *Judenbuche*, Kellers *Romeo und Julia auf dem Dorfe*) bis in die gymnasiale Mittelstufe unserer Tage hinein kanonisiert, diese Novellen liegen in historisch-kritischen Ausgaben vor, sind breit kommentiert und haben ein entsprechendes Volumen an Sekundärliteratur hervorgebracht. Andere sind hauptsächlich Experten innerhalb der Germanistik ein Begriff, wieder andere außerhalb der Sammlung kaum noch gedruckt zu beschaffen. Was die Herausgeber „des Aufhebens wert“ (Einleitung Bd. 1) befanden, versammelt mithin beides: den kleinen Ausschnitt kanonischer Texte, die hermeneutisch intensiv durchdrungen und dadurch repräsentativ geworden sind innerhalb ‚ihrer‘ literaturgeschichtlichen Epoche und jene heute mehr oder weniger vergessenen Arbeiten von Autoren, die als Vertreter des tatsächlichen Buchangebots und Zeitschriftenwesens ihrer Zeit zum *big unread* der Literaturgeschichte zählen.

Hier setzt die wissenschaftliche Zielsetzung der Tagung an. Retrodigitalisierungen auf allen Ebenen (von EU-geförderten Großinitiativen und Bibliotheksverbänden bis zu einzelnen Forschungsvorhaben und zahllosen Laienprojekten) haben die Ausschnitthaftigkeit der Literaturgeschichtsschreibung als systematisches Problem auf den Schirm der Philologien zurück gebracht und die Frage, wie sich hermeneutische Einzeltextlektüren zur digitalen Analyse vor allem größerer und großer Textmengen verhalten, in den Fokus gerückt. Zugleich haben die technischen Möglichkeiten zur digitalen Textanalyse und zur statistischen Makroanalyse von Texten einen Entwicklungsstand erreicht, der neue Erkenntnismöglichkeiten für die Literaturgeschichte eröffnet (Jockers 2013).

Entgegen mancher Rhetorik der *Digital Humanities* sind die Ergebnisse wirklicher big data-Forschung in der Literaturwissenschaft und in den Geisteswissenschaften überhaupt noch überschaubar. Die Tagung geht davon aus, dass es gerade die oft unterschätzte Arbeit mit mittelgroßen Korpora ist, die es erlaubt, aus konkreter Forschung heraus Antworten auf die methodische Grundfrage nach dem Verhältnis hermeneutischer und statistischer Methoden, qualitativer und quantitativer Forschung zu formulieren. Dafür braucht es ein integratives Konzept, das die vieldiskutierte Frontstellung von *close* und *distant reading* (Moretti 2013) überwindet und die Erkenntnisse des verstehenden Lesens mit den Daten computergestützter Analysen

zusammenführt. Das soll auf der Tagung zu Heyses *Novellenschatz* unter dem Stichwort des „Scalable Reading“ (Mueller 2013) geschehen.

2. Methodisches Konzept

Das Konzept des *Scalable Reading* schließt prinzipiell alle Akte des Lesens und Analysierens von Texten ein, wobei davon ausgegangen wird, dass wir es in der Regel schon beim verstehenden Lesen mit Text-„Surrogaten“ zu tun haben. Wer die Oxford-Klassiker-Ausgabe der *Odyssee* liest (so das Beispiel bei Mueller 2013), ist im Grunde bereits ein ‚distant reader‘ der Gesänge, die Homer zugeschrieben worden sind. Und gerade der *distant reader* im herkömmlichen Verständnis, der mit digitalen Analysen Daten und Visualisierungen erzeugt, muss diese verstehen und interpretieren. „Scalable Reading“ bedeutet indes nicht nur, dass sich close und distant reading methodisch durchdringen, es steht für ein integriertes Verständnis aller Akte des Lesens und Untersuchens auf einer weiten „scale“ medialer Formen und analytischer Aufbereitungen. Das reicht vom analogen Textdruck über digitalisierte Textkorpora bis hin zu digitalen Surrogaten, die auf der Grundlage von Worthäufigkeitslisten Delta Scores ermitteln (Burrows 2002, Jannidis/Lauer 2014) oder den Fließtext in ein *bag of words* umwandeln, um ihn modellieren zu können.

3. Umsetzung des Konzepts – Vorbereitung der Tagung

Auf der Tagung soll untersucht werden, ob und wie sich dieses Konzept in der gemeinsamen Arbeit an einem konkreten Textgegenstand bewährt. Auf literaturwissenschaftlicher Seite werden dazu Expertinnen und Experten der Älteren deutschen Literatur (Märe und Novelle), der Literaturtheorie des Realismus, für bestimmte Autoren des Heyseschen *Novellenschatzes* sowie für die Wissensgeschichte der Literatur des 19. Jahrhunderts mit Kolleginnen und Kollegen zusammenarbeiten, die in den letzten Jahren auf dem Gebiet der Computerphilologie und der Digital Humanities im deutschsprachigen Raum besonders hervorgetreten sind. Auf der gemeinsamen *scale* von einzeltextorientierten Untersuchungen bis zu quantitativen Analysen, die mit Textgruppen oder dem Gesamtkorpus der Sammlung arbeiten, bewegen sich auch die Ansätze der beteiligten Informatiker und SprachwissenschaftlerInnen.

Um das gemeinsame *Scalable Reading* durchführen zu können, ist eine intensive Vorbereitung notwendig, mit der bereits im Dezember 2014 begonnen worden ist. Die 24 Bände des *Deutschen Novellenschatzes* wurden vollständig als pdf-Dateien gescannt, gleichzeitig wurden im OCR-Verfahren txt- und xml-Korpora erstellt, das xml-Korpus soll nach der Korrekturphase im TEI-Standard mit den Metadaten der Einzeltexte aufbereitet werden. Alle Texte und Daten werden den TeilnehmerInnen rechtzeitig vor der Tagung (Zieldatum: 1.5.2015) über eine Online-Plattform zur

Verfügung gestellt, so dass schon im Vorfeld neben den Lektüren auch Analysen mit einschlägigen *tools* und Programmen zur digitalen und statistischen Textanalyse, zur Netzwerk- und Sentimentanalyse oder zum Topic Modeling möglich sind. So können die einzelnen Fragestellungen und Themen der Tagung (s.u. Sektionen und Programm) von allen Beteiligten von vornherein im Modus experimenteller Bewährung begriffen werden.

4. Der *Novellenschatz* als Textsammlung: Ausgangspunkte und Leitfragen

Mit der Methode eines integrativen mixed methods-Ansatzes betritt die interdisziplinäre Tagung Neuland im Hinblick auf einen literaturgeschichtlichen und philologischen Textgegenstand, der dadurch — so die Erwartung — gleichfalls auf bislang nicht gekannte Weise erschlossen werden kann. So gut erforscht bestimmte kanonische Einzeltexte des Deutschen *Novellenschatzes* sind, so vollständig stellt er als Textsammlung ein Desiderat der Forschung dar. Um Textsammlungen und mithin Mengen von Text, die nicht ohne weiteres durchgelesen werden können, umfassend zu analysieren, fehlten bis zur digitalen Wende in der Philologie die technischen und methodischen Mittel (wenngleich die Geschichte quantitativer Verfahren in der Literaturwissenschaft bis ins 19. Jahrhundert zurückreicht; vgl. Kelih 2008 zur Slawistik). Ein mittelgroßes Korpus, das noch in der potentiellen Reichweite individueller Lektüre liegt und zugleich schon eine für statistische Analysen repräsentative Größe hat, bietet sich in dieser für die weitere Entwicklung der philologischen Fächer entscheidenden Umbruchphase als Untersuchungsgegenstand an.

Den Herausgebern des *Novellenschatzes*, Paul Heyse und Hermann Kurz, ging es darum, einen „Sammelpunkt zu gründen, wo aufbewahrt wird, was unter der Masse des Eintags lebendigen [sic] Dauer verspricht“. Das normative Sammlerinteresse und der Wille zur Kanonisierung von Autoren, die heute selbst als „Paradebeispiel für Dekanonisierung“ (Grube 2014: 15) gelten, geben für die Analysen der Textsammlung wichtige Hinweise. Die Einleitung zum *Novellenschatz*, die über die so genannte ‚Falkentheorie‘ der Novelle das Einzige darstellt, was von Heyse als erstem deutschen Schriftsteller mit Literaturnobelpreis im aktiven kulturellen Gedächtnis bewahrt worden ist, bezieht ihre Koordinaten aus einer Transformationszeit, in der qualitative und quantitative Literaturerfahrung in einem produktiv erlebten Konkurrenzverhältnis stehen. Zwar grenzt sich die „Mustersammlung“ gegen die Massenware des „Journalismus“ ab, dessen Publikationsorgane der Novelle zugleich „Hausrecht“ gewähren, die Herausgeber sind sich aber im Klaren darüber, dass der „Zeitgeschmack“ eine „Macht“ ist, „der ein doktrinärer Eigensinn sich nicht in den Weg stellen darf“.

Der *Novellenschatz* orientiert sich zwischen literaturgeschichtlichen und gattungspoetischen Ambitionen, die angestrebte chronologische Ordnung wird rasch wieder aufgegeben, nur die

ersten drei Bände geben zu jeder Novelle einen kritischen Autorenkommentar, im Folgenden begnügt man sich mit einer Kurzbiographie. Die „Unabsehbarkeit des Stoffes“ lässt befürchten, dass „eine strenge Gruppierung nach geschichtlichen und ästhetischen Gesichtspunkten“ ein Werk für „Jahrzehnte“ wäre. Umso deutlicher steht die realistische Gattungspoetik der Novelle unmittelbar im Zeichen programmatischer Literaturgeschichtsschreibung. Goethe, Kleist und die Romantiker bilden den Anfang der Sammlung, die jedoch keinen Zweifel daran lässt, dass erst die „Wendung zum Realismus“ der deutschsprachigen Novelle zum Durchbruch habe verhelfen können (mindestens 51 der 86 Novellen stammen aus der Zeit ab 1848). Die Herausgeber sehen die Rückkehr von Politik und Philosophie „auf den Boden des Tatsächlichen“ sowie die methodische Orientierung der Geschichtswissenschaft an der „Quellenforschung“ und der Naturwissenschaften am „Experiment“ als begünstigende Faktoren an. Dementsprechend formuliert der *Novellenschatz* die heute mit Heyses ‚Falkentheorie‘ assoziierte Poetologie der Verdichtung: Die große Wirkung der Novelle liege darin, dass sie analog zur „Isolierung des Experiments“ in der Konzentration auf ein Ereignis, eine Situation, einen Fall „den Eindruck [...] verdichtet, auf einen Punkt sammelt und dadurch zur höchsten Gewalt zu steigern vermag“. Formal gehöre dazu eine „starke Silhouette“, ein charakteristisches Motiv, das die Verdichtungsqualität eines Dingsymbols besitzt wie der aus Liebe verzehrte Falke bei Boccaccio; „dramatische Unmittelbarkeit“ und „ein gewisser nervöser, herzklopfender Stil“ sind stilistische Merkmale, die für die Fähigkeit der Novelle, „mit einem raschen Schlage uns das innerste Herz zu treffen“, verantwortlich sind und sie den Herausgebern zufolge „höchst modernen Aufgaben“ gewachsen zeigen.

Für ein *Scalable Reading*, das qualitative literaturgeschichtliche Diskussionen mit quantitativen Analysen verbinden will, wirft der *Deutsche Novellenschatz* eine Reihe grundlegender Fragen auf, die auf der Tagung sicher nicht einfach beantwortet, aber doch so weit konsolidiert werden können, dass klar wird, wie und womit sich weiter arbeiten lässt. Bei weitem nicht alle Fragen lassen sich anhand des zugrundeliegenden Korpus im Verhältnis zu den Einzeltexten überhaupt ausreichend untersuchen. So bedarf etwa die Frage, ob und inwiefern die realistische Novellen-Programmatisierung der Herausgeber eine Entsprechung in statistisch feststellbaren Textsignalen besitzt, eines Abgleichs mit Referenzkorpora sowohl anderer Gattungen (etwa des realistischen Romans) als auch anderer Epochen (vor allem der fortwährend als „falsche Kunst“ geschmähten Romantik). Angesichts der offenkundigen Schwierigkeiten, eine chronologische Ordnung in die „Mustersammlung“ zu bringen oder diese nach historischen oder ästhetischen Kriterien zu gruppieren, erscheint es gleichwohl vielversprechend, auch innerhalb der Sammlung mit Mitteln der quantitativen Makroanalyse (z.B. PCA, Clustering) zu untersuchen, wie die Texte sich auf der Basis ihres stilometrischen Worthäufigkeitsprofils tatsächlich gruppieren und ob darin ‚Signale‘ der Gattung, der Epoche oder aber — hierzu bedarf es freilich wiederum eines Referenzkorpus — von

Autorschaft identifiziert werden können. Anhand der zur Verfügung stehenden Metadaten lassen sich makroanalytische Untersuchungen auch hinsichtlich der Dekaden von 1800-1870 (jeweils nach Datum Erstdruck), in Bezug auf Genderfragen (AutorInnengeschlecht) oder mit Blick auf die Geografie der literarischen Kommunikation (Erscheinungsort des Textes, Wirkungsort des Autors) durchführen.

5. Sektionen / Beiträge der TeilnehmerInnen

Die Arbeitstagung ist so zusammengesetzt, dass über die einzelnen Teilnehmerinnen und Teilnehmer an zentrale Probleme der literaturwissenschaftlichen Novellen- und Realismusforschung angeknüpft werden kann, die sich für die gemeinsame *mixed methods*-Analyse anbieten. Dazu gehören das Verhältnis der modellhaften Schreibweisen des Poetischen Realismus zu den „Routines“ (Moritz Baßler) der frühen Moderne und die besondere Bedeutung der semantischen Topics ‚Bildlichkeit‘ und ‚Bildbeschreibung‘ für die Dramaturgie der realistischen Novelle (Ralf Simon). Hannelore Schläffer hat für die Figurenkonstellationen der Novelle den Begriff der „Novellenmathematik“ (Schläffer: 1993, 36) fortgeprägt, er stammt vermutlich aus den Arbeiten der russischen Formalisten zu dieser Gattung. Mit den Mitteln der quantitativen Netzwerkanalyse wird der Informatiker Ulrik Brandes mathematischen Mustern in den Figurenkonstellationen des *Novellenschatzes* auf der *scale* vom Einzeltext bis zum Gesamtkorpus nachgehen. Daran anknüpfend stellt Fotis Jannidis heuristische Überlegungen zu literaturwissenschaftlichen Erkenntnismöglichkeiten zwischen *close* und *distant reading* zur Diskussion (Sektion 1).

Die zweite Sektion widmet sich einer wichtigen offenen Flanke in der literaturgeschichtlichen Programmatik des *Deutschen Novellenschatzes*. Während Heyse und Kurz ihre Sammlung mit dem Ziel gründen, alle aufhebenswerten Novellen normativ gegen den Verfall und das Vergessen im Zeitalter der literarischen Massenproduktion zu sichern, übersehen oder tilgen sie ihrerseits die Vorgeschichte, Rezeption und Pflege deutschsprachiger mittelalterlicher Kurzerzählungen, die die zeitgenössische Diskussion unter den Begriff der Novelle stellt, wobei die Bezeichnung ‚Novellenschatz‘ explizit geprägt wird (z.B. in Felix Liebrechts „Beiträgen zur Novellenkunde“ in der *Germania* 1856). Bent Gebert rekonstruiert diese verdrängte Vorgeschichte anhand der vielbeachteten Editionen Friedrich Heinrich von der Hagens und Adelbert von Kellers. Von letzterem stammt auch der *Italiänische Novellenschatz*, der bereits 1851 erschien (bis 1908 folgten unter der Herausgeberschaft Heyses der *Neue Deutsche Novellenschatz* und der *Novellenschatz des Auslandes* sowie der *Jüdische Novellenschatz* von Julius Moses). Friedrich Michael Dimpel schlägt mit einem Digital Humanities-Ansatz gleichfalls einen Bogen von der älteren zur neueren deutschen Literatur. Vor dem Hintergrund seines langfristigen Projekts zur narratologischen

Annotation und statistischen Auswertung eines Korpus von Kurzerzählungen, das mittel- und frühneuhochdeutsche Mären und neuhochdeutsche Novellen enthält, wird er Beispiele aus Heyses *Novellenschatz* annotieren, um dabei die methodischen Probleme narratologischer Auszeichnung zur Diskussion zu stellen. Der Historiker Stefan Nellen vom Schweizerischen Bundesarchiv in Bern kommentiert beide Beiträge aus der Perspektive editionsphilologischer Archivpraxis und im Hinblick auf eine mögliche Archivtheorie der Digital Humanities.

Auch in der dritten Sektion wird mit der statistischen Auswertung annotierter Novellen gearbeitet. Während die Sprachwissenschaftlerin Miriam Butt gemeinsam mit Michael Pollich und Christin Schätzle den Kasus als Stilmittelgebrauch in der Novelle untersucht, geht es bei Frank Fischer und Katja Mellmann in unterschiedlicher Perspektive um Zeitstrukturen. Fischer analysiert das Gesamtkorpus des *Novellenschatzes* auf temporale Typologien der erzählten Zeit, Mellmann befasst sich mit dem Aufbau der Zeitstruktur im Erzähleingang von Novellen. Wie schnell dort über welche Textmerkmale ein Zeitrahmen definiert wird und wie diese Merkmale koordiniert sind, soll auf der Basis lektürebasierter Annotationen quantitativ nachvollzogen werden. Im Hintergrund steht die Forschungsfrage, welche textuellen Zeitstrukturen wie ein kognitiver Schalter Fiktionalitätssignale erzeugen. Die kognitive Wirkung der Novellen, die Heyse und Kurz als Gattungsmerkmal herausstellen, interessiert auch Thomas Weitin. Sein Beitrag fragt, ob sich die harte Epochengrenze, die die Herausgeber zwischen Romantik und Realismus aufrichten, hinsichtlich dieses Merkmals aufrechterhalten lässt. Als Referenz dienen dabei neben der Lektüre auch die Ergebnisse einer Sentiment-Analyse und des statistischen Clusterings des Gesamtkorpus.

In der vierten Sektion sollen der Methode des *Scalable Reading* entsprechend noch einmal Textlektüren mit quantitativen Analysen zusammengeführt werden. Ausgehend von der Einleitung der Herausgeber beschäftigen sich Nicolas Pethes und Daniela Gretz in ihrem Beitrag mit der literarischen Selbstreflexion massenhafter Textproduktion und der Verarbeitung von „Papierfluten“ (K. Gutzkow) im *Novellenschatz*. Dominik Müller beleuchtet anhand der Schweizer Autoren des *Novellenschatzes* das ethnographisch dichte Schreiben als wissenschaftsgeschichtliche Kehrseite des Siegeszugs der Statistik im 19. Jahrhundert. Der Informatiker Sven Kosub beschließt die Tagung mit einer Analyse der Wortdynamiken in ausgewählten Texten des *Novellenschatzes* und einem Vorschlag für eine quantitative Klassifizierung.

6. Zitierte Literatur

Moritz Baßler: *Entsagung und Routines. Aporien des Spätrealismus und Verfahren der frühen Moderne*. Berlin/Boston 2013.

J.F. Burrows: "'Delta': a measure of stylistic difference and a guide to likely authorship", In: *Literary and Linguistic Computing* 17 (2002), S. 267-287.

Christoph Grube: *Warum werden Autoren vergessen? Mechanismen literarischer Kanonisierung am Beispiel von Paul Heyse und Wilhelm Raabe*. Bielefeld 2014.

Fotis Jannidis und Gerhard Lauer: „Burrow’s Delta and Its Use in German Literary History.“ In: Matt Erlin, Lynne Tatlock (Hrsg.): *Distant Reading. Topologies of German Culture in the Long Nineteenth Century*, Rochester 2014, S. 29-54.

Matthew L. Jockers: *Macroanalysis. Digital Methods and Literary History*. Urbana/Chicago/Springfield 2013.

Emmerich Kelih: *Geschichte der Anwendung quantitativer Verfahren in der russischen Sprach- und Literaturwissenschaft*. Hamburg 2008.

Felix Liebrecht: „Beiträge zur Novellenkunde mit besonderem Bezug auf die Ältere deutsche Literatur.“ In: *Germania* 1 (1856), S. 257-272.

Franco Moretti: *Distant Reading*. London/New York 2013.

Martin Mueller: „Morgenstern’s Spectacles or the Importance of Not-Reading.“ In: *Scalable Reading* (blog), 21.01.2013, <https://scalablereading.northwestern.edu/2013/01/21/morgensterns-spectacles-or-the-importance-of-not-reading/>.

Hannelore Schläffer: *Poetik der Novelle*. Stuttgart/Weimar 1993.

Ralf Simon: *Die Bildlichkeit des lyrischen Textes. Studien zu Hölderlin, Brentano, Eichendorff, Heine, Mörike, George und Rilke*. München/Paderborn 2011.

7. Vorläufiges Programm und Zeitplan

Donnerstag, 4. Juni — Sektion 1

Anreise

15.00 Begrüßung (Thomas Weitin)

15.15—16.00

Moritz Baßler (Literaturwissenschaft, Münster):
Modellierungen realistischer Verfahrenspoetologie

16.00—16.45

Ralf Simon (Literaturwissenschaft, Basel):
Bild-Topics in der szenischen Dramaturgie bei Raabe und Hackländer

16.45—17.00 Kaffeepause

17.00—17.45

Ulrik Brandes (Informatik, Konstanz)
Muster in den Figurenkonstellationen des *Deutschen Novellenschatzes*

17.45—18.30

Fotis Jannidis (Literaturwissenschaft, Würzburg):
Literaturwissenschaftliche Erkenntnismöglichkeiten zwischen Close und Distant Reading

18.30—19.00 Diskussion Ergebnisse Sektion 1

20.00 Gemeinsames Abendessen

Freitag, 5. Juni 2015

Sektion 2

10.00—10.45

Bent Gebert (Literaturwissenschaft, Konstanz):
Versunken „in den Stürmen der Zeit“? Märe und Novelle 1850—1871: Koordinaten eines Verdrängungssystems

10.45—11.30

Friedrich Michael Dimpel (Literaturwissenschaft, Erlangen):
Novellenschätze narratologisch auszeichnen

11.30—11.45 Kaffeepause

11.45—12.30

Dr. Steffen Nellen (Schweizer Bundesarchiv, Bern):
Archivpraxis und Archivtheorie der Digital Humanities

12.30—14.30 Gemeinsames Mittagessen

Sektion 3

14.30—15.15

Miriam Butt, Michael Pollich, Christian Schätzle (Sprachwissenschaft, Konstanz):
Kasus als Stilgebrauch in der Novelle

15.15—16.00

Frank Fischer (Digital Humanities, Göttingen):
Temporale Typologien im *Deutschen Novellenschatz*

16.00—16.15 Kaffeepause

16.15—17.00

Katja Mellmann (Literaturwissenschaft, Göttingen):
Der Aufbau der Zeitstruktur im Erzähleingang von Novellen

17.00—17.45

Thomas Weitin (Literaturwissenschaft, Konstanz):
Gattungskognition und Epochengrenzen im *Deutschen Novellenschatz*

20.00 Gemeinsames Abendessen

Samstag, 6. Juni 2015 — Sektion 4

9.30—10.15

Nicolas Pethes, Daniela Gretz (Literaturwissenschaft, Köln):
„mitten unter den Papierfluten“ (Karl Gutzkow). Thematisierung und Reflexion des Umgangs mit Textsammlungen und großen Datenmengen im *Deutschen Novellenschatz*

10.15—11.00

Dominik Müller (Literaturwissenschaft, Genf):

Keller, Gotthelf und Frey zwischen Ethnographie und Exotismus

11.15—11.30 Kaffeepause

Abreise