

Ein digitales Korpus der Novellen und Journalprosa des 19. Jahrhunderts

Herausforderungen der Metadatenerschließung

Julian Schröter¹ Johannes Leitgeb² Theresa Valta³

¹ Digitale Literaturwissenschaften, Ludwig-Maximilians-Universität München

² Institut für Deutsche Philologie, Julius-Maximilians-Universität Würzburg

³ Institut für Geschichte, Julius-Maximilians-Universität Würzburg



Julian Schröter, Johannes Leitgeb, Theresa Valta: „Ein digitales Korpus der Novellen und Journalprosa des 19. Jahrhunderts“. *DARIAH-DE Working Papers* No. 46. Göttingen: DARIAH-DE, 2023.

DOI: [10.47952/gro-publ-131](https://doi.org/10.47952/gro-publ-131).

This article is published under the
licence [Creative-Commons Attribution 4.0](https://creativecommons.org/licenses/by/4.0/) (CC-BY).

The *DARIAH-DE Working Papers* are published by Mirjam
Blümm, Thomas Kollatz, Stefan Schmunk and Christof Schöch.



Dieser Beitrag entstand im Rahmen eines Walter-Benjamin-Stipendiums, DFG-Projektnummer 449668519.

Zusammenfassung

Maßgeblich für die Konstruktion literarischer Korpora ist neben der Aufbereitung der digitalisierten Texte die Erschließung relevanter Metadaten. Zunehmend anspruchsvollere literaturgeschichtliche Frage- und Problemstellungen im Feld der *Computational Literary Studies* (CLS) erfordern die Erschließung und Modellierung komplexerer Metadatenstrukturen als die lange Zeit scheinbar hinreichende Metadatentrias von Autor, Epoche und Gattung. Der Aufsatz hat zwei Aufgaben: Im Zentrum steht zunächst die Beschreibung eines zwischen 2018 und 2022 aufgebauten Korpus aus Novellen und Erzählungen im mediengeschichtlichen Kontext der Journale und Taschenbücher des 19. Jahrhunderts. Auf Grundlage dieser Beschreibung diskutieren wir anschließend die für künftige Forschung leitende Frage, welche Metadaten in welcher Weise erfasst werden sollten, um kultur-, medien- und sozialgeschichtliche Fragen im Rahmen computergestützter Analyseverfahren beantworten zu können. Unsere Antwort wird durch drei Thesen strukturiert: In Abschnitt 3.1 entfalten wir an einer Reihe von exemplarischen Befunden zu den historischen Inhaltsverzeichnissen der Medienformate, die üblicherweise als Archiv zur Gewinnung von Metadaten etwa zur Gattungszuordnung der einzelnen Texte dienen, die These der vielschichtigen Abhängigkeit der Gattung vom Medium. Abschnitt 3.2 entwickelt die These, dass es – neben den Texten selbst – auch die Medienformate mit ihren Publikationsereignissen – den täglich, wöchentlich, monatlich, viertel-, halb- oder jährlich erscheinenden Drucken sowie gegebenenfalls den jährlich zusammengebundenen Jahrgängen – als Einheiten für die Metadatenerschließung zu betrachten gilt. Nur so lassen sich Fragen der Gattungspolitik und einer sozialgeschichtlichen Prägung der Gattungen beantworten. In Abschnitt 3.3 schließlich entfalten wir die These, dass es weitere Metadaten zum medialen Publikationskontext zu erschließen gilt, um literaturgeschichtlichen Wandel kausal erklären zu können. Auf diese Weise beschreibt der Aufsatz nicht nur das Korpus eines konkreten Projekts zur Novellengeschichte des 19. Jahrhunderts, sondern leistet mit einer exemplarischen Beschreibung von Herausforderungen und Chancen für eine mediengeschichtlich orientierte breite Metadatenerschließung neue Hinweise zu einer künftigen Methodologie der Korpuskonstruktion für die in den CLS wichtiger werdende quantitative Sozial- und Mediengeschichte der Literatur. In einem Ausblick werden wir daher als Agenda für künftige Forschung anregen, nicht ausschließlich Textobjekte als die basalen Entitäten literarischer Korpora zu modellieren, sondern auch die medialen Publikationsereignisse als Elemente der Korpora zu begreifen.

Schlagwörter

Korpus, Metadaten, Novelle, Mediengeschichte, Literaturgeschichte, 19. Jahrhundert, Gattungsgeschichte, Computational Literary Studies

Keywords

corpus, metadata, novella, media history, literary history, 19th century, history of genre, Computational Literary Studies

Inhaltsverzeichnis

1	Einleitung	4
2	Beschreibung des Korpus und Systematik der Metadatenerfassung	6
3	Explorative Beschreibung der medialen Situation	12
3.1	Inhaltsverzeichnisse und die Abhängigkeit zwischen Gattung und Medium	12
3.2	Praxeologie und Gattungspolitik der Inhaltsverzeichnisse: Das Medium als eigenständige Entität im Korpus	16
3.3	Mediale Tiefen-Metadaten zur Erklärung kulturellen Wandels	19
4	Zusammenfassung und Ausblick	20
5	Literaturverzeichnis	23
6	Anhang	25

1 Einleitung

Die Fähigkeit, große Textmengen in kurzer Zeit zu verarbeiten und mit Verfahren aller technologischen Komplexitätsgrade zu erschließen, macht computergestützte und quantitative Analysen zu einem immer wichtiger werdenden Baustein literaturgeschichtlicher Forschung. Entscheidend für die Konstruktion digitaler Korpora ist neben der Aufbereitung der Texte deren Ergänzung durch Metadaten zu materialen, sozialen, medialen und kulturellen Kontexten. Zunehmend anspruchsvollere literaturgeschichtliche Fragestellungen und reichhaltigere Kontextualisierungen erfordern die Erschließung und Modellierung komplexerer Metadatenstrukturen.¹ Computergestützte literaturgeschichtliche Forschung sollte daher in Zukunft nicht nur »*data-rich*«² sein, sondern auch *metadata-rich*.

Dieser Aufsatz präsentiert das Korpus für ein Projekt zum Verhältnis von Gattungs- und Mediengeschichte in der Novellistik des 19. Jahrhunderts.³ Die eine Zeit lang scheinbar hinreichende Metadatentrias von Autor, Epoche und Gattung, mit der noch Jockers (2013) auskam, genügt für medien- und sozialgeschichtlich ausgerichtete Projekte nicht. Doch die Frage, welche Kontexte in welcher Form und in welcher Struktur für ein Untersuchungskorpus zu erschließen sind, ist alles andere als trivial und bislang nicht systematisch geklärt.⁴ Unser Interesse am Einfluss von Medienformaten wie Zeitung, Taschenbuch, Anthologie, Familienblatt oder Rundschau auf die Gattung lässt sich konkret über eine von Meyer (1987) vertretene These und über eine Anschlussfrage fassen: Laut Meyer ist der in der literarischen Kultur des 19. Jahrhunderts etablierte Novellenbegriff ein Effekt von medial hervorgebrachten Schreibverfahren. Diese These eröffnet eine Untersuchungsperspektive, wonach die Gattungszugehörigkeit von Texten in erster Linie eine Gattungsetikettenpolitik im medialen Kontext zu sein scheint. Meyers These lässt sich abstrahieren und als die generelle literaturgeschichtliche Frage reformulieren, welche ökonomischen, kultur- und literaturpolitischen Gründe maßgeblich dafür sind, dass einzelne Medienformate die Gattungsbezeichnungen in je spezifischer und idiosynkratischer Weise gebrauchen und so indirekt Einfluss auf die weitere Entwicklung der Gattungsemantiken nehmen.

In Kapitel 2 beschreiben wir im Folgenden die Struktur und Eigenschaften unseres zwischen 2018 und 2022 aufgebauten literarischen Untersuchungskorpus zu Novellen- und Journalprosa des 19. Jahrhunderts. Dieses Korpus galt es dabei so zu konstruieren, dass es die Prüfung der von Meyer vertretenen These sowie die Beantwortung der anschließenden Forschungsfrage erlaubt. Im Zentrum steht dabei eine historisierende Erfassung von Gattungszuschreibungen und medialem Publikationskontext. Da der Aufbau von Korpora nach Kriterien der Repräsentativität und Balanciertheit von einer angemessenen Berücksichtigung von Texten entsprechend den zu untersuchenden Gattungen und Medienformaten

¹Dies gilt, wenn man auf die jüngsten Publikationen blickt, für sozialgeschichtliche Modellierungen (vgl. Underwood u. a. 2022) sowie für statistische Modellierungen relevanter Faktoren, für die zuletzt auch Methoden der Bayesschen Statistik wichtig geworden sind (vgl. Kestemont und Karsdorp 2021).

²So der Untertitel in Bode (2018).

³Textdaten, abgeleitete Textformate, Metadaten und Skripte zur Exploration des Korpus sind im GitHub-Repository <https://github.com/julianschroeter/19CproseCorpus> verfügbar. Umfangreicher und modular aufgebauter Programmcode zu den Textanalysen im Rahmen des übergreifenden Projekts befindet sich im Repository <https://github.com/julianschroeter/PyNovellaHistory>. Der Aufbau des Korpus wurde durch Mittel aus dem Forschungsfonds der Philosophischen Fakultät der Universität Würzburg gefördert.

⁴Entscheidend vorangetrieben wurde die Entwicklung um neue Metadaten ergänzter literarischer Korpora im Rahmen des Projekts »Creating the European Literary Text Collection« (ELTeC), (vgl. Schöch u. a. 2021).

ausgehen muss, führt die hier geleistete Beschreibung des Korpus Metadatenreflexion und systematische Korpusbeschreibung zusammen. Das Kapitel hat das Ziel, die künftige Nachnutzung des Novellenkorpus in den Computational Literary Studies (CLS) zu ermöglichen.

Daraus, dass das Erkenntnisinteresse des Projekts eine historisierende Erfassung von Gattungszuordnungen erfordert, ergibt sich jedoch eine interessante neue Problemlage und Herausforderung. Ein philologisch genauer Blick in die Journale und andere Publikationsorgane des 19. Jahrhunderts zeigt, dass die Zuweisung des Gattungslabels zu Texten alles andere als eindeutig und trivial ist. Während computergestützte Analysen grundsätzlich zunächst von der Unabhängigkeit der durch Metadatenkategorien erfassten Produktionsfaktoren – Epoche, Autor, Gattung und im Folgenden: Medium – ausgehen, um untersuchen zu können, wie stark der Einfluss der einzelnen Faktoren auf die Textgestalt ist, zeigt sich bei genauem Hinsehen eine beträchtliche Abhängigkeit der Gattung vom Medium. Kapitel 3 beginnt in Abschnitt 3.1 mit einer exemplarischen Beschreibung solcher Abhängigkeiten und einer exemplarischen Diskussion der Probleme systematischer Metadatenerfassung. Wir können keine systematische Lösung für diese Probleme vorlegen. Unsere exemplarischen Beschreibungen sollen der Theoriebildung und Methodologie der Digital Humanities jedoch künftig dazu dienen, die Abhängigkeit zwischen Metadatenkategorien besser abzubilden.

Der Blick auf die Verschränkung und Abhängigkeit von Medium und Gattung führt uns schließlich zu zwei weiteren Beobachtungen, die sich in künftiger Forschung durch entsprechende Metadatenerfassung systematisch abbilden lassen. In Abschnitt 3.2 plädieren wir für eine Blickwendung vom Text hin zum Medium. Wenn man nicht nur fragt, mit welcher Gattungsbezeichnung und in welchem Medium ein Text erschienen ist, sondern auch auf der Ebene des Mediums erfasst, wie die Texte dort disponiert sowie in welcher Länge und Folgenanzahl und in welcher Relationierung zu anderen Textsorten sie dort präsentiert werden, eröffnet sich eine praxeologisch interessante Perspektive auf Gattungspolitik. Es kann dann sichtbar gemacht werden, wie Gattungsgeschichte beispielsweise durch Praktiken der Verknappung oder der Marktüberschwemmung mitgeprägt wird. Anhand einzelner Beispiele werden wir illustrieren, dass dieser Perspektivwechsel eine Chance ist, mediale Gattungspolitik besser zu verstehen.

Diese Chance wird uns in Abschnitt 3.3 am Beispiel der Familienblätter aus der zweiten Hälfte des 19. Jahrhunderts zu einer abschließenden Beobachtung führen. Gattungs- und mediengeschichtlicher Forschung geht es um Kausalrelationen; etwa um die Frage, ob die Gattung das Medium prägt oder umgekehrt das Medium die Gattung. Wenn das Medium als eine prägende Instanz angenommen wird, stellt sich bei näherer Betrachtung die Frage, ob es materiale Bedingungen oder einzelne personale Instanzen wie Verleger und Herausgeber gibt, die systematisch auf Textproduktion und Gattungszuordnung Einfluss nehmen.

Der Aufsatz hat damit im Kern zwei Aufgaben: Zum einen die Präsentation eines neuen digitalen Korpus zur Journalprosa des 19. Jahrhunderts einschließlich der systematischen Klärung der Metadatenerfassung und seiner Komposition unter den Gesichtspunkten von Balanciertheit und Repräsentativität. Zum anderen möchten wir einer künftigen Modellierung und theoretischen Konfiguration von sozial- und mediengeschichtlich orientierten digitalen Korpora den Boden bereiten, indem wir auf konkrete Herausforderungen und Chancen sowie auf den Bedarf nach einer neuen Ontologie für medienkommunikations-orientierte anstelle von textobjekt-orientierten Korpora aufmerksam machen.

2 Beschreibung des Korpus und Systematik der Metadatenerfassung

Wir beginnen mit einer Beschreibung des Korpus und seiner wichtigsten Merkmale mit Blick auf die allgemeinen Kriterien der Korpusbildung und unter dem Blickwinkel eines einfachen Metadatenschemas, das neben Autor, Publikationsjahr und Gattung noch das Medium als vierte Kategorie hinzunimmt. Korpora können nach den Kriterien der Repräsentativität oder der Balanciertheit gebildet oder nach opportunistischen Gesichtspunkten gewonnen werden (vgl. Schöch 2017, Biber, Conrad, und Reppen 1998; Wynne 2004; Percillier 2017). Die opportunistisch gewonnene Datensammlung, bei der Texte nach dem Kriterium der einfachsten und schnellsten Verfügbarkeit (z.B. mittels direktem Import über Plattformen wie Gutenberg, Zeno, HathiTrust, DTA oder Textgrid) dem Korpus einverleibt werden, bringt – obwohl sie etabliert und in vielen Fällen ein brauchbarer Erstzugang ist – die größten Schwächen mit sich, auf die wir vereinzelt zu sprechen kommen werden. Da für die Novellengeschichte die Grundgesamtheit weder bekannt noch verfügbar ist und das Korpus daher nicht als repräsentative Zufallsstichprobe aus der Grundgesamtheit gebildet werden kann, war der Aufbau des Korpus durch das Kriterium der Balanciertheit bestimmt. Balanciertheit haben wir für die vier Kategorien des Jahres der Erstpublikation, der Zuordnung des Gattungslabels der Erstpublikation, des Medienformats und des Kanonisierungsgrads angestrebt. Aufgrund der ungleichen Verteilungen der Merkmale dieser Kategorien in der Grundgesamtheit lässt sich Balanciertheit jedoch nicht auf allen Ebenen erreichen. Tabelle 1 zeigt die Verteilung nach Gattungslabels. Balanciert ist das Korpus dabei nach den drei im Zentrum der Untersuchung stehenden Untergattungen, die hier als »Novelle«, »Erzählung« und »sonstige Erzählprosa«, das heißt also mittellange Erzählprosa ohne Gattungsbezeichnung oder mit sonstigen Gattungsbezeichnungen (wie »Geschichte«, »Begebenheit« etc.), gelabelt sind.

Die Auswahl der Gattungen für die Korpusbildung führt zu einem der meistdiskutierten Probleme der Gattungstheorie. Ausgehend von der Annahme, dass die Gattungsnamen »Novelle« und »Erzählung« im Untersuchungszeitraum höchst uneinheitlich gebraucht wurden, scheint es nahezuliegen, auf die Berücksichtigung der historischen Gattungsbezeichnungen ganz zu verzichten und stattdessen mit rein systematischen Kategorien für Textklassifikationen bzw. mit systematisch vorab definierten Gattungsbegriffen etwa der Novelle und der Erzählung zu operieren (vgl. Lukas 1998). Dieses Verfahren wurde in der Gattungstheorie häufig als deduktives Verfahren bezeichnet und dem induktiven Verfahren der Korpuskonstruktion nach den historischen Gattungsemantiken gegenübergestellt (vgl. Mueller 2010). In Schröter (2019) wurde eine Schwäche des deduktiven Verfahrens unter historiographischer Perspektive herausgearbeitet. Diese Schwäche besteht darin, dass man sich den potenziellen Zugang zu historischen Gattungsemantiken verbaut, wenn man die Gattungszuschreibungen ignoriert.⁵ Dennoch ist eine systematische Konstruktion des Korpus unerlässlich. Für eine Novellengeschichte, die an einer Erschließung der historischen Gattungsemantik interessiert ist, ist daher ein induktives Verfahren erforderlich. Anders als die Kategorien des Erscheinungsjahrs, des Autors und des Medienformats ist die Gattung keine systematische beschreibungssprachliche Kategorie, sondern eine Information darüber, als was der Text in einer bestimmten historischen Situation kommuniziert wurde.⁶ Mit Underwood (2019) kann man von einer perspektivischen Gattungszuordnung sprechen. Die Gattungszuordnung kommt

⁵Für eine historisierende Perspektive auf Genres wurde häufig argumentiert und plädiert (vgl. Fishelov 1993).

⁶In seiner gattungsgeschichtlichen Studie sieht Calvo Tello (2021) dieses Problem für den Gattungsbegriff, löst das Problem jedoch etwas unglücklich über die Suche nach »durchschnittlichen« Gattungszuordnungen auf, um sie damit wieder in den Bereich einer systematischen Beschreibungskategorie zurückzuholen; dies allerdings zu dem Preis, dass die Kategorie ihren Bezug zum historischen Kommunikationskontext verliert.

einem Text also nicht absolut zu, sondern als Zuschreibungskategorie relativ zu einer historischen Kommunikationssituation. Dies hat zur Folge, dass einige Texte, wie z.B. Goethes *Unterhaltungen deutscher Ausgewanderten*, Kleists Erzählungen oder Droste-Hülshoffs *Judenbuche*, bei ihrer Erstpublikation nicht als Novellen veröffentlicht, später aber als solche kanonisiert wurden und somit in unterschiedlichen historischen Kommunikationssituationen unterschiedliche Gattungszuordnungen haben.

Daher wurde folgender systematischer Umgang mit den Gattungslabels gewählt. Zum einen werden die in historischen Situationen als Novellen und als Erzählungen wahrgenommenen Texte, die sich zu einer Textsorte der mittellangen, meist realistisch erzählten Erzählprosa zusammenfassen lassen, auf zwei angrenzende Textsorten bezogen: Märchen und Romane. Zum anderen werden Erzählungen und Novellen innerhalb ihres Erscheinungszusammenhangs der Journale und Taschenbücher im Verhältnis zu den nicht nur fiktionalen, sondern auch faktualen Nachbartexten untersucht (auf diese, auch faktualen Journalgenres werden wir in Kapitel 3 zurückkommen). Das Korpus umfasst also drei Gruppen: (a) Fiktionale mittellange Erzählprosa unter der Perspektive historischer Gattungszuordnungen, (b) in exemplarischer Auswahl und im Untersuchungszeitraum veröffentlichte Romane und Märchen sowie (c) weitere Journaltexte, innerhalb der Medienformate, in denen Erzählungen und Novellen erschienen sind. Mit einem Gesamtanteil der mittellangen Erzählprosa von 70% am Gesamtkorpus (siehe rechte Spalte in Tabelle 1) ist gewährleistet, dass aussagekräftige Analysen der Erzählprosa im Vergleich zu den Nachbargattungen des Romans und des Märchens möglich sind.

Tabelle 1: Balanciertheit des Korpus nach Gattungen.

	Absolute Zahl	Relativer Anteil der Gattungen	Anteil der Hauptgruppen nach Anzahl der Texte
Novelle	227	0.32	Gruppe (a) 0.7
Erzählung	107	0.15	
Sonstige Prosaerzählung	157	0.22	
Roman	138	0.20	Gruppe (b) 0.25
Märchen	39	0.05	
Sonstige Journalinhalte	30	~ 0.05	Gruppe (c) 0.05
Summe	698	(1.00)	

Die in Tabelle 1 aufgeschlüsselte Verteilung nach Genres bzw. Textsorten wird in Abbildung 1 nach der Verteilung über 30-Jahres-Zeiträume von 1790 bis 1940 (*Terminus umfasst nur die Zeit zwischen 1789 bis 1914*) aufgelöst.

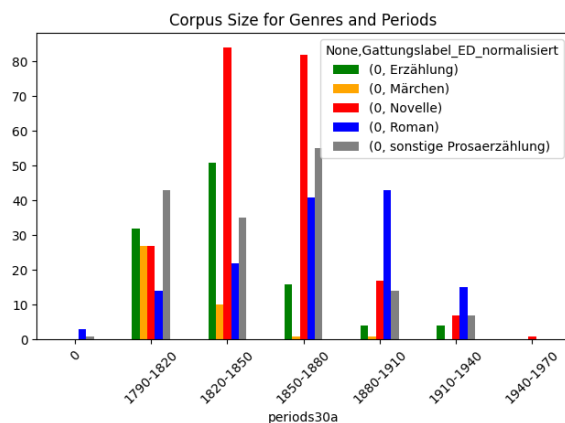


Abbildung 1: Korpusgröße nach Gattungen und Perioden.

Die hier gewählten 30-Jahres-Zeiträume mit den Periodengrenzen auf 1790, 1820, 1850 und 1880 sind zunächst heuristischer Natur.⁷ Dennoch lässt sich vorläufig mit Blick auf literaturgeschichtliches Wissen und die zentralen historiographischen Hypothesen dieser Arbeit davon ausgehen, dass diese Periodengrenzen den Wandel der Gattungsgeschichte recht gut fassen. Zudem decken sie grob die literaturgeschichtlichen Perioden – Goethezeit, Restaurationszeit, Realismus, Naturalismus/frühe Moderne, klassische Moderne – ab. Für die Perioden der Erstpublikationen strebt das Korpus Balanciertheit insbesondere hinsichtlich der für das Projekt zentralen Zeit zwischen 1790 und 1880 an. Texte aus der Zeit nach 1880 wurden opportunistisch ins Korpus aufgenommen, um Analysen für das sogenannte lange 19. Jahrhundert zu ermöglichen. Der Fokus der Analysen liegt auf der Restaurationszeit und der Zeit des Realismus sowie auf dem diachronen Vergleich beider Perioden.

Tabelle 2, die sich auf den Kernzeitraum zwischen 1790 und 1880 beschränkt, zeigt die Verteilung der Texte im Korpus nach ihren Medienformaten und relativ zum Zeitraum der Erstpublikation.⁸

Tabelle 2: Sub-Samples der Medienformate nach Perioden.

	Taschenbuch	Familienblatt	Anthologie	Sonst. Journale	Werke/Bücher	gesamt
1790–1820	26	0	79	19	20	144
1820–1850	120	0	58	10	13	201
1850–1880	4	91	21	21	67	204
gesamt	150	91	158	50	100	549

⁷Dass in den CLS, insbesondere infolge der Arbeit von Jockers (2013), häufig mit Generationen von 30 Jahren für den Wandel literarischer Gattungen gearbeitet wird, spielt für den Begründungs- oder Rechtfertigungszusammenhang an dieser Stelle keine Rolle. Zu einer Kritik dieses Generationenverständnisses vgl. Underwood (2019).

⁸Aufgrund der Beschränkung auf den Kernzeitraum und die zentralen Medienformate fehlen in dieser nur 549 Texte erfassenden Darstellung 138 Texte aus dem 687 Texte umfassenden Gesamtkorpus.

Ein nicht-triviales Unterfangen stellt eine literaturgeschichtlich und medientheoretisch angemessene Definition von Medienformaten dar. Für Tabelle 2 wurden neben der exakten Erfassung des Titels der Publikation als mögliche Medientypen ›Taschenbuch‹, ›Familienblatt‹, ›Anthologie‹, ›Rundschau‹, ›eigenständige Buchpublikation‹ und – als Residualkategorie – ›sonstige Journale‹ gewählt. Was in dieser Residualkategorie zusammengefasst ist, ist in der Metadatenerschließung des Korpus in Tageszeitung, Wochenzeitung, Monatsschrift/Monatsheft und Rundschau unterteilt. Hierbei handelt es sich um in literatur- und mediengeschichtlicher Forschung etablierte Kategorien.⁹ Man sieht jedoch schnell, dass sich hier systematische Kriterien (Periodizität) mit historischen Benennungskategorien »Rundschau« und »Familienblatt« als etablierte Bezeichnungen aus der Kultur des 19. Jahrhunderts kreuzen. Die in der Novellenforschung bislang anspruchsvollste mediengeschichtliche Untersuchung von Meyer (1987) hat von Tageszeitung bis Taschenbuch einen sehr weiten Begriff des Journals. Alles, was kein eigenständig gedrucktes Buch ist, ist für Meyer ein Journal. Dank der Studie von Spoerhase (2018), in der die Nähe des Taschenbuchs zu anderen Buchformaten und die feinen Unterschiede zwischen den Formaten nachgewiesen wurden, ist eine solche Verallgemeinerung insbesondere unter materialgeschichtlichen Gesichtspunkten nicht länger haltbar. Die Herausforderungen einer angemessenen Systematik und Taxonomie der Medienformate können wir hier jedoch lediglich andeuten.

Eine der wichtigsten Leistungen der Metadatenkonfiguration unseres Novellenkorpus liegt darin, nicht nur die Medientypen der Erstpublikationen zu erfassen, sondern neben Zweit- und Drittpublikation insbesondere auch die Folgepublikation in einer der Novellensammlungen des 19. Jahrhunderts. Dies sind die 1828 bis 1831 erschienene Sammlung *Pantheon* und der zwischen 1871 und 1874 herausgegebene *Deutsche Novellenschatz*.¹⁰ Die Folgepublikationen in einer der Novellensammlungen, in Anthologien oder Werkausgaben kommen in unserer Metadatenstruktur additiv dazu und verdrängen nicht die Verweise auf die zumeist in Taschenbüchern erfolgten Erstpublikationen. Voraussetzung hierfür ist die oben eingeführte induktive Struktur der Gattungszuweisungen nach historischen Situationen. Diese Metadatenstruktur ist wertvoll für die Untersuchung der Kanonisierungsgeschichte der Gattung; denn es zeigt sich, dass zahlreiche Texte insbesondere durch den von Heyse und Kurz herausgegebenen *Deutschen Novellenschatz* nicht nur kanonisiert wurden, sondern auch erstmals als Novellen kategorisiert wurden. Die Prozesse des Re-Labelings der Gattungszuweisungen lassen sich durch diese Metadatenstruktur systematisch und mit quantitativen Verfahren analysieren. Die in diesem Korpus verfolgte medienrelative Erschließung unterschiedlicher Gattungszuweisungen für jeden Text ist zugleich ein Novum in der Konstruktion literarischer Textkorpora. Auf einige der empirischen Folgeprobleme dieser medienrelativen Erschließung kommen wir in Kapitel 3 zurück.

Insgesamt ist das Korpus, wie aus der unteren Randzeile und der rechten Randspalte in Tabelle 2 hervorgeht, für die Kernzeiträume und für die Verteilung nach Medienformaten balanciert, sofern man das Familienblatt als eine Form des Journals auffasst, so dass Taschenbuch, Anthologien und Journale auf jeweils ca. 150 Texte kommen und die beiden Kernperioden 1820 bis 1850 und 1850 bis 1880 auf jeweils ca. 200 Texte. Balance zwischen Medienformaten wurde dadurch erreicht, dass

⁹Ausgewertet haben wir hierfür neben den Arbeiten von Meyer (1987; 1998) die Studien von Zuber (1955); Schröder (1970); Jäger (2003).

¹⁰In einer künftigen Erweiterung des Korpus wird sich auch noch der von Pustkuchen zwischen 1822 und 1824 herausgegebene *Novellenschatz des deutschen Volkes* aufnehmen lassen. Für diese Sammlung liegen jedoch im Moment noch keine Scans vor.

neben der Integration bestehender Textsammlungen¹¹ die Novellensammlung *Pantheon* sowie die Jahrgänge aus zahlreichen Taschenbüchern, für die über die Bayerische Staatsbibliothek, HathiTrust und andere digitale Bibliotheken einzelne Bilddateien vorliegen, einer optischen Schrifterkennung (OCR) unterzogen, manuell korrigiert und als *plain-text* Dateien ins Korpus aufgenommen wurden.¹² Texte aus dem Familienblatt *Die Gartenlaube* konnten aus bestehenden Quellen extrahiert werden. Für Texte aus dem Familienblatt *Daheim* lagen bereits Digitalisierungen vom Lehrstuhl für Computerphilologie der Universität Würzburg vor.

Da die Erzählprosaerzählungen des 19. Jahrhunderts für die drei Gruppen der Novellen, der Erzählung und der sonstigen Prosaerzählungen balanciert erfasst wurden, ergibt sich – da keine gegenüber der Grundgesamtheit bewussten oder indirekten Verzerrungen erfolgen – eine Ungleichverteilung bei den Medienformaten, und zwar sowohl für sich genommen als auch im Verhältnis zur Entstehungszeit. Dies ergibt sich aus zwei zentralen Charakteristika der Grundgesamtheit für den Journalmarkt des 19. Jahrhunderts, nämlich erstens daraus, dass mit Meyer (1998) zu erwarten ist, dass 97% der Prosaerzählungen der Restaurationszeit in Journalen (inklusive Taschenbücher) erschienen sind, und zweitens daraus, dass es Taschenbücher de facto nur bis ca. 1850 gab, während die Familienblätter sowie die Rundschau-Formate die Taschenbücher ab 1850 vom Markt verdrängt haben. Dies wiederum hat zur Konsequenz, dass im Korpus keine Taschenbücher nach 1850 und keine Familienblätter vor 1850 enthalten sein können und aufgrund der historischen Realität keine Balanciertheit für alle Teilgruppen hergestellt werden kann.

Die vierte und letzte Dimension, für die wir ein Mischverhältnis aus Balanciertheit und Repräsentativität angestrebt haben, ist das Verhältnis zwischen kanonischen und nicht-kanonischen Texten. Diese Dimension ist wichtig, um die Frage adressieren zu können, ob bestimmte Entwicklungen gattungstypischer Schreibverfahren möglicherweise nur innerhalb eines Hochliteratursegments oder im ganzen Feld des Literaturmarkts stattfinden. Würde man eine repräsentative Verteilung hinsichtlich des Kanonisierungsgrads anstreben, müsste man – indem man die Schätzung von Meyer zur Menge der heute vergessenen, weil nicht nachgedruckten Texte als unteren Grenzwert zugrunde legt – davon ausgehen, dass mindestens 85% des historischen Textbestands vergessen sind und höchstens 15% der Grundgesamtheit schwach bis stark kanonisierte Texte bilden. Da man jedoch davon ausgehen kann, dass die am besten kanonisierten Texte in den meisten Fällen auch die in den digitalen Repositorien und Archiven verfügbaren Texte sind, wird ein opportunistisch nach einfacher Verfügbarkeit gebildetes Korpus zu einem Übergewicht kanonischer Texte führen. In allen bestehenden Textsammlungen, etwa bei Gutenberg, Zeno.org und folglich auch im TextGrid-Repository, zeichnet sich eine derartige Verzerrung der Verhältnisse in der Grundgesamtheit ab.

¹¹Von besonderem Wert war neben dem TextGrid-Repository die Digitalisierung des *Deutschen Novellenschatzes* im Rahmen des von Thomas Weitin und Ulrich Brandes geleiteten Projekts »Reading at Scale«.

¹²Die Kodierung im TEI-Format ist für das künftige Projekt einer digitalen Edition geplant.

Tabelle 3: Komposition des Korpus nach Kanonisierungsgrad.

	absolut	relativ	Population (geschätzt)
0 (nicht-kanonisiert)	452	0.66	>0.85
1 (schwach)	71	0.10	
2 (mittel)	83	0.12	<0.15
3 (stark kanonisiert)	81	0.12	

Auch unser Novellenkorpus weist gegenüber der mutmaßlichen Grundgesamtheit ein deutliches Untergewicht der nicht-kanonischen Texten auf, hat aber auf etwa gleichgroße Untergruppen schwach, mittel und stark kanonisierter Texte hingearbeitet und erstmals eine absolute Mehrheit in der Grundgesamtheit dominierenden Gruppe der heute dekanonisierten Texte erreicht (vgl. Tabelle 3).

Zudem wurden zu jedem Text folgende Informationen erfasst: (i.) Pseudonym der Autorin / des Autors und (ii.) Gender, im Fall eines erst posthum erschienenen Werks (iii.) das mutmaßliche Entstehungsjahr für entstehungsgeschichtliche Untersuchungen, (iv.) die Serialität (für die Texte der Familienblätter *Die Gartenlaube* und *Daheim* in der exakten Kodierung der Einzelfolgen, für die übrigen Texte mit der Booleschen Information, ob der Text in mehreren Folgen erstpubliziert wurde), (v.) der Status als Überarbeitungsfassung (d.h., ob es sich bei einem Text um eine überarbeitete Fassung eines früher gedruckten und ggf. in das Korpus aufgenommenen älteren Texts handelt), (vi.) der Verleger und Herausgeber in einem gemeinsamen Metadatenfeld. Als weitere Metadaten auf der Ebene des Digitalisierungsprozesses wurden der urheberrechtliche Status sowie die Bearbeitungsqualität des OCR der Texte erfasst.¹³ Bei diesen Metadaten handelt es sich nicht um aus den Texten analytisch extrahierten Metadaten wie etwa bei Paige (2020), also um Informationen, die häufig auch als automatische Annotationen bezeichnet werden. Stattdessen handelt es sich um situativ-kontextbasierte Informationen, die sich – epistemisch gesehen – aus Paratexten, den Kommentaren zu Werkausgaben, den Medien des Erstdrucks, oder Normdatenabgleichen extrahieren lassen.

Gegenüber der in älteren Arbeiten meistverwendeten Trias von Autor, Epoche, Gattung bietet das hier beschriebene Basismodell mit fünf Haupt- und den genannten Nebenkategorien ein Gerüst, mit dem sich eine ganze Reihe von Fragen nach dem medialen Einfluss auf die Gattungssemantik mit quantitativen Analyseverfahren adressieren lassen, indem beispielsweise mit Verfahren überwachten maschinellen Lernens nicht nur auf die Gattung, sondern auch auf den medialen Kontext trainiert wird. Zudem lassen sich Fragen nach dem medialen Einfluss auf Gattungen innerhalb einzelner Medienformate und innerhalb kanonischer oder nicht kanonischer Textgruppen beantworten. Üblicherweise werden in den CLS die Metadatentypen als voneinander unabhängig modelliert. Die Annahme der Unabhängigkeit erleichtert es zu untersuchen, welche der Kategorien im Sinn textprägender Faktoren mehr oder weniger starken Einfluss auf die Textgestalt hatten. Im folgenden Kapitel werden wir an konkreten Beispielen zeigen, dass diese Annahme der Realität nicht gerecht wird und dass sich vielfältige Formen der Abhängigkeit der Gattung vom Medium beobachten lassen. Künftige Forschung sollte

¹³Im gegenwärtigen Projektstand sind die Metadaten vollständig und tabellarisch in der Datei metadata.csv im GitHub-Repository enthalten (s. Anm. 4).

diese Abhängigkeiten auch auf der Ebene der Korpus- und Metadatenerschließung berücksichtigen, um literaturgeschichtliche Fragestellungen fundierter und mit höherem Grad der Berücksichtigung historischer Komplexität beantworten zu können. Quantitative Analysen können über die Konzepte der Variableninteraktion und konditionalen Abhängigkeit auch ein komplexeres Zusammenspiel von Kausalfaktoren modellieren und einer empirischen Analyse zuführen. Aufgabe dieses Beitrags ist aber nicht die statistische Analyse, sondern die Beschreibung der Abhängigkeiten der Gattung vom medialen Kontext.

3 Explorative Beschreibung der medialen Situation

Am Beispiel der Familienblätter *Daheim* und *Die Gartenlaube* zeigen wir für ihre erste Blütephase von den 1850er bis in die 1880er Jahre im vorliegenden Kapitel zunächst die komplexen Formen der Abhängigkeit zwischen Medium und Gattungszuweisung auf (Abschnitt 3.1). Aus dieser dichten und philologisch »nahen« Beschreibung werden wir im nächsten Schritt (Abschnitt 3.2) eine Chance für künftige korpusbasierte Forschung skizzieren und abschließend (Abschnitt 3.3) einen theoretischen Auftrag an die Theoriebildung und Methodologie der Korpuskonstruktion formulieren. Für die Herausforderungen haben wir noch keine systematischen Lösungen. Indem wir aber die Problematik exemplarisch und dabei möglichst konkret sichtbar machen, möchten wir auf Anforderungen an und Möglichkeiten für künftige Lösung des Problems hinweisen.

3.1 Inhaltsverzeichnisse und die Abhängigkeit zwischen Gattung und Medium

Für die in Taschenbüchern, Familienblättern und Journalen veröffentlichten Novellen und Erzählungen gibt es hauptsächlich zwei Stellen, an denen die Gattung einem Text mittels Paratextlabel zugeordnet werden kann: der Untertitel und das Inhaltsverzeichnis. Bei genauerem Hinsehen ergeben sich bereits hier die ersten Probleme. Wir unterscheiden im Folgenden drei Problemebenen, die wir exemplarisch diskutieren, indem wir gängige Annahmen (a) zur Eindeutigkeit oder Verbindlichkeit des Labels, (b) zur Autorität des ursprünglichen Labels und (c) zu semantischen Kriterien der Vergabe von Gattungslabels infrage stellen. Auf allen drei Ebenen stoßen wir auf das Problem, dass die Voraussetzung der Unabhängigkeit von Gattungszuordnung und medialem Publikationskontext nicht gegeben ist.

(Zu a) In den Inhaltsverzeichnissen der publizierten Jahrgänge haben die historischen Zeitschriften und Familienblätter die Texte selbst in eine übergeordnete Systematik von Rubriken eingeordnet und so ihren zeitgenössischen Leserinnen und Lesern sowie der Nachwelt systematisch erschließbar gemacht. Die Inhaltsverzeichnisse sind mithin die zentrale Quelle für die Erschließung der Gattung als Metadatenkategorie der Texte. In einigen Fällen weichen jedoch die Gattungszuweisungen zwischen Untertitel, Nennung und Rubrikenüberschrift im Inhaltsverzeichnis voneinander ab. Zum Beispiel wird der Text »Aus eigener Kraft« Wilhelmine von Hillerns, welcher 1870 in der *Gartenlaube* erschien, in einer Fußnote nicht nur als Roman, sondern auch als Erzählung bezeichnet, obgleich er im Titel keine Gattungszuschreibung aufweist.¹⁴ Das gleiche Phänomen findet sich in Bezug auf den Text »Der Dorfcaplan: Erzählung aus Oberbaiern nach einer wahren Begebenheit«, über den in einer Anmerkung als Novelle gesprochen

¹⁴Anmerkung des Herausgebers. In: *Gartenlaube*, 23 (1870): 368.

wird.¹⁵ Diese gar nicht seltenen Fälle haben wir erfasst, indem wir für unser Metdatenmodell zwischen Gattungslabel im Untertitel und Inhaltsverzeichnis unterschieden haben. Beide Rubrikenlabel wurden in unterschiedlichen Spalten in die Metadatentabelle aufgenommen.

Die Gattungszuordnung ist also nicht nur, wie in Kapitel 2 ausgeführt, situationsrelativ, sondern überdies innerhalb einzelner medialer Situationen uneindeutig. Für die Metadatenerschließung haben wir mit der vorläufigen Hypothese operiert, dass die Rubrizierungen der Inhaltsverzeichnisse eine systematische Kategorisierung nach Gattungen leistet und somit eine pragmatische Zuordnung erlaubt. Es bleibt jedoch noch philologisch und für einzelne Fälle zu prüfen, in welchem Grad sich kulturelle Diskurse am Inhaltsverzeichnis abzeichnen und in welchem Maß in den Änderungen der Rubriken mediale Spielregeln der Familienblätter deutlich werden, die darauf schließen lassen, dass die Rubriken nicht nur klassifikatorische, sondern auch politische und marktstrategische Funktionen, etwa der Rezeptionsbeeinflussung, erfüllen. Hinsichtlich der methodologischen Frage, wie sich eindeutige und situationsrelative Gattungslabels als Metadaten zu den Texten aus den Paratexten und historischen Kontexten extrahieren lassen, zeigt sich mithin ein komplexes Verhältnis zwischen historischen Daten und systematischer Metadatenstruktur. Es ist nicht so, dass Gattungslabels aus einem historischen Archiv einfach in eine Metadatentabelle übertragen werden können. Vielmehr überlagern sich unterschiedliche Archive, die es in einer komplexen Metadatenstruktur dem Forschungsinteresse entsprechend abzubilden gilt. Auf mögliche Modellierungen einer solchen Metadatenstruktur werden wir im Fazit noch einmal zu sprechen kommen.

(Zu b) Die insbesondere in radikal historisierenden Ansätzen gängige Annahme, die Gattungszuweisungen im Erstdruck seien so etwas wie die ›ursprünglichen‹ Textsortenklassifikationen mit einer besonderen Autorität für so etwas wie eine ›eigentliche‹ Gattungszuordnung, wird im Fall der Familienblätter problematisch.¹⁶ Die Inhaltsverzeichnisse mitsamt der Einteilung in Rubriken wurden am Ende eines Jahrgangs für die Publikation der wöchentlichen Ausgaben als Sammelband erstellt. Somit gilt grundsätzlich, dass die Rubriken nicht unmittelbar für die (Erst-)Publikation der Einzeltexte und -folgen in Erscheinung treten. In unserer Metadatenontologie konnten wir daher noch nicht abbilden, dass die Erstellung der Verzeichnisse zwei Funktionen erfüllen konnte: Als Archivierungspraxis können die Rubriken induktiv motiviert sein und eine erst im Nachhinein erzeugte Ordnung der vielfältigen Texte erzeugen, um diese retrospektiv besser zu erschließen. Als Projektion einer medieninternen Struktur, die sich an etwaigen, deduktiv im Voraus festgelegten Richtlinien der Zeitschriften und Familienblätter ausrichtet, käme den Inhaltsverzeichnissen dagegen die Funktion der direkten Auskunft über die Gattungs- und Genrepolitik der jeweiligen Redaktionen zu.¹⁷ Weiterhin lassen sich in den Rubrizierungen potenzielle rezeptions- wie produktionslenkende Funktionen erkennen, deren genauere Analyse auf Basis von biographischen Zeugnissen als Desiderat anzumerken ist. Zu klären, welche Funktionen in den Mediensegmenten des 19. Jahrhunderts dominiert hat, gehört demnach zu den künftigen Aufgaben mediengeschichtlicher Forschung. Die in Kapitel 2 beschriebene Erfassung einer einfachen Gattungszuordnung und eines medialen Erscheinungskontexts erlauben für sich genommen noch nicht die Beantwortung dieser Frage.

¹⁵Anmerkung des Herausgebers. In: *Gartenlaube*, 49 (1865): 771.

¹⁶Explizit vertritt Schröder (1970) diese Position; operativ und heuristisch arbeitet auch Meyer (1987 und 1998) mit den Gattungsbezeichnungen der Erstpublikation.

¹⁷Nur für Einzelfälle liegen dabei bereits medienhistorische Studien vor: So basiert die Rubrikensystematik auf einem Entwurf, den der Herausgeber Ernst Keil bereits 1852 vorab der Publikation entworfen und für die subsequenten Jahrgänge mit einigen Änderungen beibehalten hat (das sog. ›Hubertusburger Konzept‹) (vgl. Stockinger 2018, 83ff.).

Wir werden daher im Folgenden exemplarisch umreißen, wie eine Metadatenstruktur der nächsten Generation aussehen könnte, um diese Frage zu klären.

Einen ersten Anhaltspunkt für die Beobachtung medienpezifischer Funktionen der Rubriken und damit der Gattungszuordnungstechniken einzelner Medienformate bietet ein Vergleich der Rubriken der Familienblätter *Die Gartenlaube* und *Daheim*. (vgl. Tabelle 4).

Tabelle 4: Äquivalente Rubriken der Familienblätter *Daheim* und *Gartenlaube*.

Daheim	Gartenlaube
Erzählungen und Novellen	Erzählungen und Novellen
Gedichte	Gedichte
Geschichts- und Zeitbilder	Beschreibende und geschichtliche Aufsätze
Literatur- und Lebensbilder	Biographien und Charakteristiken
Skizzen aus der Heimat und Fremde [Unterrubrik <i>Elsaß und Lothringen</i> nur 1871]	[fehlt]
Naturwissenschaftliches und Medicinisches	Naturwissenschaftliches und Medicinisches [vor 1875 getrennt]
Sociales und Volkswirtschaftliches [ab 1870]	[fehlt]
Vermischtes	Vermischtes
Am Familientische	Blätter und Blüten
Der deutsch-französische Krieg [nur 1871]	Der deutsch-französische Krieg [nur 1870 und 1871]

Die Ähnlichkeit der Systematiken weist zunächst auf eine weitere Abhängigkeit hin: die der gegenseitigen Beeinflussung einzelner Medien, die an dieser Stelle zur medienübergreifenden Etablierung von Gattungen und Labels führt. Besonders deutlich fällt in beiden Systematiken der scheinbar klar gezogene Unterschied zwischen Texten mit faktuellem Anspruch, fiktionaler Erzählprosa («Erzählungen und Novellen») und Poesie («Gedichte») auf. Diese Einteilung muss jedoch vor dem Hintergrund der zugehörigen Texte relativiert werden, da sich einerseits auch in Texten anderer Rubriken narrative und fiktionale Texte finden lassen sowie andererseits, weil auch Texte der Mischgattungen in den Familienblättern anzutreffen sind. Einige tragen beispielsweise das Label »novellistisches Kulturbild«¹⁸ oder »kulturhistorische Novellette«¹⁹ und sind in entsprechenden faktualen Rubriken verzeichnet. Der überaus

¹⁸B. L. »Der erste deutsche Kalendermann«. In *Daheim*, 1 (1876): 14-16.

¹⁹Jonathan. »Bischof und Burgermeister. Kulturhistorische Novellette aus Danzigs Vergangenheit«. In *Daheim*, 32 (1876): 393-399.

heterogener Umgang mit Gattungen, Rubriken und Labels stellt daher eine besondere Herausforderung für die Erstellung eines entsprechenden Korpus dar, da zwar der historisch-mediale Kontext und Sprachgebrauch für eine Analyse erhalten bleiben, die Erfassung im Korpus aber eine pragmatisch sinnvolle Kategorisierung der Texte vornehmen muss, um geeignete Perspektiven auf die Texte zu gewinnen. Für eine philologische Auswertung der Inhaltsverzeichnisse gilt es daher zu reflektieren, dass die oben angesprochene Unschärfe in der Funktion der Verzeichnisse beachtet und eine entsprechende Unterscheidung zwischen systematischer Textklassifikation und deren Zuordnung zu einer Rubrik kontinuierlich mitvollzogen werden sollte.

(zu c) Eine grundlegende Annahme für die Zuordnung von Texten zu Gattungen besteht darin, dass jeder Text der Kategorie zugeordnet wird, unter die er am ehesten fällt.²⁰ Dagegen spricht bereits, dass einem Text im selben medialen Kontext verschiedene Gattungen zugeordnet werden können, wie oben gezeigt wurde. Wir werden im Folgenden, ganz unabhängig vom Aspekt intensionaler Passung zwischen Einzeltext und Gattungszuordnung, auf einer rein äußerlich medialen Ebene ein Beispiel dafür liefern, dass die generelle Annahme der ›Passung‹ in historischen medialen Kontexten problematisch ist. Die Anzahl der Erzähltexte, die in der *Gartenlaube* mit einer Gattungszuschreibung versehen sind, nimmt Ende der 1870er Jahre im Verhältnis zu den ungelabelten Texten deutlich zu. Unabhängig davon, welche Rolle textuelle Merkmale für die Gattungszuordnung gespielt haben und welche weiteren Faktoren relevant waren, legt diese Entwicklung nahe, dass mediale und situative Faktoren auch eine Rolle in der Wahl der Gattungszuschreibung spielen. Die folgende Tabelle gibt eine Übersicht über die Verteilung der Label innerhalb der Stichprobe (vgl. Tabelle 5).

Tabelle 5: Gattungsetiketten der Erzähltexte für ausgewählte Jahrgänge der *Gartenlaube*.

	1860	1865	1870	1871	1875	1876	1879	1881	1882	1883	1889
Erzählun- gen	3	4	0	0	2	1	0	6	1	2	5
Novellen	2	0	2	0	0	0	4	1	8	1	3
anderes Label (z.B. Roman, Kalenderge- schichte)	0	1	2	1	4	1	5	2	5	3	6
ohne Label	11	10	5	5	5	4	8	2	2	5	5

Im Anhang findet sich eine Übersicht über die Häufigkeit der Genrelabels im konkurrierenden Familienblatt *Daheim*. Ohne an dieser Stelle weiter ins Detail zu gehen, lässt sich beobachten, dass zwischen den Familienblättern und insbesondere im Verhältnis zu anderen Medienformaten wie etwa der *Deutschen*

²⁰Diese Annahme prägt sowohl die einschlägigen Auffassungen von Gattungen als Klassifikationsinstrumenten (Fricke 2010) als auch die Auffassung von Gattungen als ästhetischen Kategorien (Walton 1970).

Rundschau die Wahrscheinlichkeiten für Texte, kein Gattungslabel oder ein ganz bestimmtes Gattungslabel zu erhalten, stark auseinandergehen. Im Folgenden beschäftigen wir uns nicht mit den textuellen Merkmalen und Strukturen der Gattungsgruppen. Wir verweisen nur darauf, dass es hinreichende Evidenz für die bereits von Meyer (1987 und 1998) vertretene These gibt, dass die Unterschiede in der Zuteilung der Textmerkmale nicht allein die strukturellen Unterschiede in der Zuordnung von Gattungslabels erklären können, sondern dass handfeste pragmatische, marktstrategische, literaturpolitische Gründe und manchmal bloße Willkür hinter den Gattungszuordnungen stecken. Das wiederum bedeutet: Ob ein Text überhaupt eine Gattungszuschreibung erhalten hat, scheint mithin sowohl vom Medium als auch vom Publikationszeitraum abhängig zu sein.

Die erläuterten Herausforderungen lassen sich aus datenanalytischer Perspektive als ein Problem der Abhängigkeit fassen. Ein Text besitzt nicht etwa eine Gattungszuordnung und eine Medienzugehörigkeit als zwei voneinander unabhängige Eigenschaften. Die Abhängigkeit der Auswahl des Labels im Fall von Alternativen (z.B. ›Novelle‹ oder ›Erzählung‹) ist dabei nur ein Problem. Diese Form der Abhängigkeit wird von unserer Systematik der Metadatenerschließung bereits berücksichtigt und analysierbar gemacht. Wie dieser Abschnitt zeigen sollte, reichen die Probleme jedoch noch tiefer. Je nach Medium gibt es unterschiedliche Grade der Eindeutigkeit der Zuordnung; zudem kann, je nach Medium, die Entscheidung für eine Gattung ganz unterschiedlich motiviert sein: Sie muss nicht immer auf den Versuch zurückzuführen sein, eine sinnvolle Kategorisierung zu finden oder die Intention des Autors oder der Autorin abzubilden.

3.2 Praxeologie und Gattungspolitik der Inhaltsverzeichnisse: Das Medium als eigenständige Entität im Korpus

Im folgenden Abschnitt werden wir unsere Perspektive von einer Orientierung am Text und dessen Medien- und Gattungszugehörigkeit ab-, und stattdessen dem Medienformat selbst zuwenden. Dabei wollen wir zeigen, dass mit einer solchen Blickwendung bislang unbeachtete Aspekte von Gattungspraktiken jenseits unmittelbar textueller Merkmale wahrgenommen werden können. Für die Untersuchung der Dominanz der Rubriken haben wir die Inhaltsverzeichnisse ausgewählter Jahrgänge stichprobenartig betrachtet und die Textanzahl pro Rubrik erfasst (vgl. Tabelle 6 sowie Tabelle 7 für das Familienblatt *Daheim* im Anhang). Gezählt wurden dabei ganze Texte, auch wenn diese in Einzelfolgen über mehrere Ausgaben hinweg publiziert wurden (s. u.).

Tabelle 6: Textanzahl nach Rubriken für ausgewählte Jahrgänge der Gartenlaube. Das Fehlen einer Rubrik, die nur für abgrenzbare Zeiträume eingeführt wurde, wurde mit / gekennzeichnet, das okkasionelle Fehlen einer Rubrik mit 0. Für die kürzeren Texte der Rubrik »Blätter und Blüten« wurden nicht alle Jahrgänge ausgewertet.

Rubrik im Inhaltsver- zeichnis	1865	1866	1867	1868	1869	1870	1871	1872	1873	1874	1875	1876
Erzählungen und Novellen	15	13	7	13	7	9	6	8	8	7	11	6
Gedichte	11	10	9	10	16	12	18	24	16	23	13	14
Beschreibende und geschichtliche Aufsätze	147	168	140	122	121	44	40	128	113	126	142	102
Biographien und Charakte- ristiken	28	16	41	47	30	30	18	29	22	27	20	29
Der deutsch- französische Krieg	/	/	/	/	/	344	343	/	/	/	/	/
Naturwissen- schaftliches und Medicinisches	/	/	/	/	/	/	/	/	/	22	24	24
Naturwissen- schaftliche Aufsätze / Naturwissen- schaft	13	14	16	22	16	10	9	22	28	/	/	/
Medicinisches / Medicin	6	10	7	6	0	6	5	4	/	/	/	/
Vermischtes	0	0	0	6	24	10	8	13	11	14	12	18
Blätter und Blüten	119	142	n.e.	n.e.	115	35	69	141	n.e.	n.e.	155	160

Die besondere Stellung der Rubrik *Der deutsch-französische Krieg* fällt bereits bei Betrachtung der bloßen Zahl publizierter Texte auf. In beiden Familienblättern wird die Rubrik infolge der zeitgenössischen Kriegsereignisse eingeführt, im *Daheim* für das Jahr 1871, in der *Gartenlaube* sogar für beide Kriegsjahre 1870 und 1871. Darüber hinaus weist *Daheim* im Jahr 1871 für die Rubrik *Skizzen aus der Heimat und Fremde* die Unterrubrik *Elsaß und Lothringen* auf – eben jene Region, die als ein Streitpunkt des Krieges mit politischer Bedeutung aufgeladen ist.²¹

Die Kriegs-Rubrik nimmt einen übermäßig großen Anteil der Jahrgänge ein, in *Daheim* 1871 rund ein Drittel der publizierten Texte, in der *Gartenlaube* 1870 und 1871 sogar je rund drei Viertel. Um der aktuellen politischen Berichterstattung Raum zu bieten, wurden in den Familienblättern andere Rubriken gekürzt: So ist für das *Daheim* 1871 der vollständige Wegfall der Rubriken *Geschichts- und Zeitbilder*, *Naturwissenschaftliches und Medicinisches* sowie *Soziales und Volkswirtschaftliches* zu beobachten, während in der *Gartenlaube* die entsprechende Rubrik *Beschreibende und geschichtliche Aufsätze* von 168 Texten vor dem Krieg auf 40 bzw. 44 Texte in den Kriegsjahren reduziert wird. Diese Dynamiken sind umso bemerkenswerter als sie die fiktionalen Texte nicht betreffen. In beiden Familienblättern ist zwar eine Verminderung der absoluten Anzahl zu bemerken, diese erzeugt in beiden Fällen jedoch nicht die charakteristische Zäsur wie sie in den anderen Rubriken zu beobachten ist; die Werte aus den Kriegsjahren werden auch in den folgenden Jahren erneut erreicht. Diese Beobachtung wird auf der Mikroebene einzelner Texte teilweise relativiert, denn es ist durchaus ein Anstieg militärischer Sujets in den fiktionalen Texten zu beobachten. Als Beispiel lässt sich die mit der ersten Ausgabe des *Gartenlaube*-Jahrgangs 1871 einsetzende Erzählung »Pulver und Gold« von Levin Schücking nennen, die im zeitgenössischen Krieg angesiedelt ist und das Schicksal eines deutschen Offiziers schildert. Damit wird die politisch-kulturelle Anpassung der Texte angesprochen, die vor dem Hintergrund einer analog entstehenden medienpolitischen Entwicklung abläuft: Als wöchentliche Publikation reagieren die Journale auf das Zeitgeschehen und ändern damit auch die eigenen medialen Regeln. Als Marker für einen außerliterarischen und -medialen Diskurs kann die Aufnahme der zusätzlichen Rubrik neue Vergleichsperspektiven ergeben. Dabei können die Dominanz- und Gewichtungsverhältnisse der Gattungen im Verhältnis zueinander Auswirkung auf das Wertesystem der historischen Gattungssemantik haben.

Um aus praxeologischer Perspektive die politische Dimension von axiologischen, pragmatischen und semantischen Verschiebungen der Gattungsbegriffe untersuchen zu können, bietet es sich mit Blick auf die Korpuskonstruktion an, Metadaten zum medialen Kontext nicht nur als Informationen zu den einzelnen Texten zu erfassen, sondern auf der Ebene der Medienformate für die periodisch erscheinenden Formate. Auf diese Weise sind Kontextualisierungen etwa des Gebrauchs von Gattungsbezeichnungen im Verhältnis zu anderen Textsorten möglich, die dann Rückschlüsse auf die Gattungspolitik erlauben. Solche Rückschlüsse wären indes nicht möglich, wenn man zu den einzelnen Texten lediglich die Medien- und die Gattungszugehörigkeit erfasst. Wir können im Rahmen dieses Beitrags kein umfassendes Metadatenmodell für eine solche praxeologische Perspektivierung vorstellen, werden aber im Fazit einige Ideen für eine entsprechende Forschungsagenda diskutieren.

²¹Stockinger (2018) belegt diesen Umstand auch mit soziokulturellen Aspekten, so konnte ein erheblicher Anstieg der Abonnements in den Kriegsjahren festgestellt werden, die neue »Produktions- wie Rezeptionsformen« (Stockinger 2018, 313) seitens des Familienblattes zur Folge hatten.

3.3 Mediale Tiefen-Metadaten zur Erklärung kulturellen Wandels

Das übergreifende Projekt, in dessen Rahmen das hier vorgestellte Korpus aufgebaut wurde, fragt nach dem Einfluss des Mediums auf Gattungszuschreibungen. Medium ist hierbei zunächst eine recht abstrakte Kategorie. Wenn man vom Medium als Kausalfaktor spricht, kann dies ganz unterschiedliche Aspekte meinen: Entscheidungen der exklusiv für ein Medium arbeitenden Autorinnen und Autoren, autoritative Überstimmungen der Herausgeber oder Verleger, materiale Bedingungen und Einschränkungen, oder Marktlogiken, die – beispielsweise aus feldanalytisch-soziologischer Sicht – die Entscheidungen von Medien als Positionierungen auf einem konkurrierenden Medienmarkt beschreiben. Literatur-, Medien- oder sozialgeschichtliche Forschung wird daher, je nach konkretem Erkenntnisinteresse, zumeist an einzelnen der potenziell relevanten ›Tiefenfaktoren‹ innerhalb der abstrakten Kategorie ›Medium‹ interessiert sein. Unser Novellenkorpus erfasst insbesondere die Verleger als potenziell relevante Faktoren, weil zahlreiche Studien von einem starken Einfluss der Verleger und Herausgeber auf die Autorinnen und Autoren der Taschenbücher der Restaurationszeit sowie der Familienblätter ausgegangen sind (vgl. Scherer und Stockinger 2016, S. 277). Wir werden abschließend und wiederum exemplarisch vorführen, wie sich durch die im zurückliegenden Abschnitt vorgestellte Blickwendung vom Text auf das Medium als Grundbaustein des Korpus die kausalen Kräfte zwischen Medium und Gattung sichtbar machen lassen. Hierfür untersuchen wir die Korrelation zwischen Herausgeberwechsel in den Familienblättern (*Die Gartenlaube* sowie *Daheim*) und der Veränderung der Dominanz einzelner Rubriken.

Die Gartenlaube wurde im Jahr 1853 erstmals publiziert; und zwar mit 31 Erzähltexten. Im Laufe der 1850er Jahre ging deren Anzahl deutlich zurück. Abbildung 2 macht die Entwicklungen der Erzähltextanzahl in *Gartenlaube* und *Daheim* bis in die 1890er Jahre sichtbar:²²

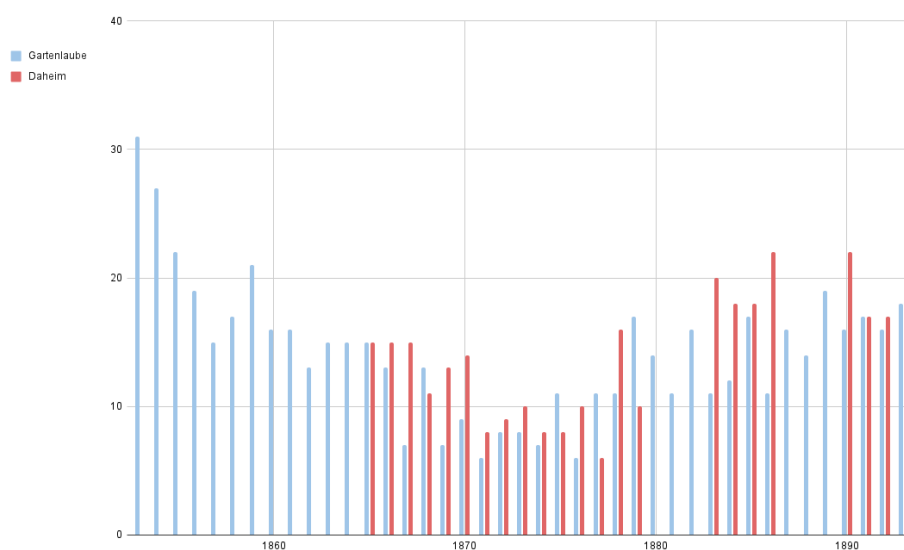


Abbildung 2: Anzahl der Erzähltexte im Vergleich zwischen *Gartenlaube* und *Daheim*.

²²Die Daten der Jahre 1880 bis 1882 sowie der Jahre 1887 bis 1889 konnten für das Familienblatt *Daheim* nicht erhoben werden, da hier kein Material vorlag. Die Datenerhebung setzt im Jahr 1865 ein, da das Blatt erst im Oktober 1864 gegründet wurde.

Zwischen 1867 und 1876 lag die Zahl der veröffentlichten Erzähltexte in der *Gartenlaube* meist im einstelligen Bereich. Ab 1879 stieg die Gesamtmenge wieder deutlich an. Für den Teilzeitraum zwischen 1865 und 1876 konnten wir eine Abnahme des relativen Anteils an der Gesamtzahl der im Jahrgang publizierten Texte nachweisen. Daraus leiten wir vorläufig ab, dass die in Abbildung 2 sichtbare Zunahme von Erzähltexten ab 1879 in absoluten Zahlen auch einer Zunahme ihres relativen Gewichts im Jahrgang entspricht. Stichprobenartige Untersuchungen für die Jahrgänge 1885 und 1889 bestätigen diese Annahme. Eine systematische Prüfung steht noch aus. Eine mögliche Begründung für die angenommene Zunahme der Dominanz der Erzähltexte können Präferenzen der wechselnden Herausgeber sein. Zwischen 1862 und 1878 wurde die *Gartenlaube* von Ernst Keil (1816–1878) verlegt, der das Familienblatt im Jahr 1853 begründet hatte. Nach seinem Tod übernahm zunächst Ernst Ziel (1841–1921) bis 1883 die Herausgeberschaft, woraufhin dieser 1884 von Adolf Kröner (1836–1911) ersetzt wurde. Im Laufe der Zeit verloren politische Themen, die vor allem im Rahmen der Reichseinigung zum Tragen kamen, an Bedeutung. Stattdessen rückte die Unterhaltung der Leserinnen und Leser in den Vordergrund (vgl. Stockinger 2018, 61 und 66). Geht man davon aus, dass die Anzahl der Erzähltexte auch anteilig anstieg, spiegelt sich diese Entwicklung in den angeführten Zahlen wider, da fiktionale Prosa ab 1879, also kurz nach dem Tod Keils, wieder zunahm.

Im Familienblatt *Daheim* kann man hinsichtlich der Erzähltextanzahl eine äquivalente Entwicklung verzeichnen. So sank diese in den 1870er Jahren in den einstelligen Bereich ab und stieg in den 1880er Jahren wieder deutlich an. Auch in diesem Fall kam es im Jahr 1878 zu einem sprunghaften Anstieg der Erzähltexte, was ebenfalls die Folge eines Herausgeberwechsels sein könnte, da ab August 1877 Theodor Hermann Pantenius (1843–1915) Mitverleger war. Zu berücksichtigen ist allerdings, dass die Gestaltung des Familienblattes *Daheim* insgesamt an die *Gartenlaube* angelehnt war. Die Orientierung am Vorbild könnte eine alternative Begründung für die gezeigte Entwicklung liefern. Zu erklären wäre dann jedoch, warum *Daheim*, vor allem in den 1880er Jahren in der Summe mehr Erzähltexte veröffentlicht hat als die *Gartenlaube*. In jedem Fall zeigt sich, dass Herausgeberschaft einen wichtigen ›Tiefenfaktor‹ innerhalb des Mediums darstellt. Gleichzeitig kann deren Berücksichtigung nicht immer eine letztgültige Begründung für die untersuchten Entwicklungen liefern. Für die Frage nach einer geeigneten Metadaten-systematik wird es deshalb künftig entscheidend sein, nicht einfach nur das Medienformat als abstrakte Kategorie zu erfassen, sondern auch die für die mediale Gestalt relevanten Faktoren und Instanzen.

4 Zusammenfassung und Ausblick

Dieser Aufsatz hat zwei Ziele verfolgt: Zunächst ging es (in Kapitel 2) darum, ein neu erstelltes Novellenkorpus vorzustellen und hinsichtlich neuer Metadaten-dimensionen zum medialen und sozialgeschichtlichen Kontext zu beschreiben. Das Korpus ist hinsichtlich der Kernperioden der Restaurationszeit (etwa 1820 bis 1850) und des Realismus (1850 ist etwa 1880) und hinsichtlich der Verteilung der Gattungsetiketten der Erstpublikation und der Medienformate balanciert und es soll mit Blick auf den Kanonisierungsgrad der Texte zwischen Repräsentativität und Balanciertheit vermitteln. Wesentlich für die Modellierung der Metadaten war hierbei, dass Gattung nicht als Beschreibungskategorie, sondern als historische kommunikative und situative Zuschreibung konfiguriert wurde. Diese Konfiguration ist erforderlich, um die historisierende Analyse von Gattungsetiketten zu ermöglichen. Die Metadaten-ontologie unseres Korpus bietet zugleich eine Erweiterung digitaler Literaturkorpora, insofern neben Informationen zum

Kanonisierungsstand weitere mediale Metadaten zur Serialität, zu Nachfolgepublikationen und deren Gattungslabels sowie zu den verantwortlichen Herausgebern und Verlegern ebenso erfasst wurden wie die Bearbeitungsqualität der Textdateien.

Den damit erreichten Stand einer Metadatenontologie haben wir im dritten Kapitel einer kritischen Reflexion unter drei Perspektiven unterzogen. In Abschnitt 3.1 haben wir mit dem exemplarischen Blick auf die Inhaltsverzeichnisse und deren Rubrizierungen in den Familienblättern das Verhältnis zwischen Gattungszuweisung und Medium problematisiert. Dabei hat sich gezeigt, dass die Gattungszuweisung häufig nicht eindeutig ist, dass das Gattungslabel der Erstpublikation durchaus nicht immer besondere Autorität genießt, und dass die Gattungszuweisung auf allen relevanten Ebenen vom medialen Publikationskontext und dem Zeitpunkt der Publikation abhängt. Dies hat zur Konsequenz, dass wir nicht immer davon ausgehen können, dass die historischen Gattungslabels intensional sinnvolle Kategorisierungen nach textmerkmalsbasierten Klassifikationsregeln darstellen.

Davon ausgehend haben wir in Abschnitt 3.2 einen Blickwechsel vorgeschlagen, weg vom Einzeltext und hin zum Medium als Publikationseinheit. Wir haben dies am Beispiel von Praktiken der Akzentuierung, Stabilität und Verdrängung von Rubriken zum Zeitpunkt des deutsch-französischen Kriegs 1870/71 vorgeführt. Damit ist nur ein Aspekt praxeologisch relevanter Handlungstypen aufgerufen. Zu denken ist überdies an bewusste Verknappungen zur Wertsteigerung oder der Marktüberflutung (vgl. Schröder 1970 für die Restaurationszeit), die eine Marginalisierung einer Gattung zur Folge haben können. Derartige Prozesse können (manchmal) beabsichtigte und (zumeist) unbeabsichtigte Effekte auf das kulturelle Wertesystem literarischer Gattungen innerhalb historischer Kulturen haben. Dies ist aus unserer Sicht ein wichtiger Grund dafür, Daten zu Textlänge, Anzahl oder Folgenanzahl systematisch und für quantitative Analysen auf Medienebene zu erfassen.

Der von uns vorgeschlagene Blickwechsel hat zur Konsequenz, dass im Metadatenmodell das Medium zur Entität wird und nicht länger nur als Wert in der Metadatenkategorie namens ›Medium‹ für die Einzeltexte erscheint. Eine solche Ontologie mit Medienpublikationsereignissen als basalen Entitäten wurde für das hier beschriebene Novellenprojekt lediglich exemplarisch vorangetrieben. Als systematische Korpusstruktur ist dies eine Aufgabe für digitale Korpora der nächsten Generation. Aufgabe dieses Aufsatzes war, eine solche Entwicklung vorzudenken. Nachzudenken wäre auch über eine Integration eines solchen Metadatenmodells in eine übergreifende *linked-open-data*-Struktur (LOD), wobei Texte, Publikationseinheiten, historische Kontexte und situationsbedingte Gattungszuschreibungen flexibel miteinander verknüpft werden können. Erst wenn man die medialen Publikationsereignisse, wie z.B. das wöchentliche Erscheinen eines Hefts der *Gartenlaube* und dann nochmals die zusammengefasste und mit einem Inhaltsverzeichnis ausgestattete Jahresausgabe als je eigene Entitäten auffasst, lassen sich praxeologische Fragen wie die nach der Gattungspolitik durch Verdrängungen, Verknappungen oder Marktüberflutungen einzelner Gattungen adressieren.

In Abschnitt 3.3 haben wir schließlich die Anforderungen an Metadatenerschließung von einer Perspektive historischer Erklärungen aus in den Blick genommen. Am Beispiel veränderter Gewichtungen einzelner Rubriken und Gattungen im Familienblatt *Die Gartenlaube* haben wir exemplarisch die kausale Erklärung eines derartigen Wandels durchgespielt, indem wir eine auffällige Änderung mit Herausgeberwechseln korreliert haben. Um in der hier exemplarisch vorgeführten Weise quantitative und korpusbasierte Untersuchungen für sozial- und mediengeschichtliche Fragestellungen in Angriff nehmen zu können, genügt es nicht, mediale Metadaten lediglich hinsichtlich der Publikationsnamen (wie etwa *Gartenlaube*

und *Daheim*) und der Formate (z.B. Familienblatt, Rundschau, Taschenbuch) zu erfassen. Zu ergänzen sind die tiefer liegenden, potenziell relevanten Kausalfaktoren. Die Rollen des Verlegers und Herausgebers sind nur ein Typ relevanter Faktoren. So unterschiedliche Aspekte wie Erscheinungsort, Reichweite, Region, die Frage nach serieller oder nicht serieller Publikation kommen ebenfalls infrage.

Unsere in Abschnitt 3.2 entfaltete These, dass es Metadaten auf der Ebene der Medienformate zu erschließen gilt, sowie das in Abschnitt 3.3 vorgetragene Plädoyer für die Extraktion möglichst vieler und kausal potenziell relevanter medialer sowie materialer Aspekte führt uns abschließend zu einem abstrakten Gedanken für eine generelle Korpus- und Metadatenmodellierung: Digitale Korpora beruhen bislang zumeist auf einer Ontologie, in der einzelne Texte die Grundbausteine darstellen. Die basalen Entitäten und Instanzen jeder Stichprobe sind damit Textobjekte. Dies bedeutet, dass man auch die unbekannte und nur abstrakt verfügbare Grundgesamtheit des zu untersuchenden literarischen Raums als einen Raum aus Textobjekten begreift. Der von uns vorgeschlagene Blickwechsel auf publizierte Medien-einheiten legt indes nahe, künftig mehr darüber nachzudenken, literarische Kommunikationsereignisse als Grundbausteine des Korpus zu betrachten. Für eine Historiographie literarischer Kommunikation mit quantitativen und statistischen Verfahren wird es darauf ankommen, Kommunikationsereignisse als basale Entitäten zu betrachten. Das kann zur Folge haben, dass die Instanzen eines Werks in vielen unterschiedlichen, aber für die jeweilige Forschung relevanten, historischen Situationen als ein jeweiliges Kommunikationsereignis aufgefasst werden und somit mehrfach als Instanz im Korpus auftauchen.

Es könnte sich lohnen, auch Novellengeschichte in dieser Weise zu modellieren. Bislang war es in der Novellenforschung – mit der Ausnahme von Schröder (1970) und Meyer (1987, 1998) – üblich, den Aspekt der kulturellen Relevanz der einzelnen Novellen in historischen Kontexten literarischer Kommunikation zu ignorieren. Wenn man jedoch in Rechnung stellt, dass einzelne Medienformate, wie etwa das Familienblatt *Die Gartenlaube*, mit einer Auflage von mehreren hunderttausend Exemplaren in den 1870er Jahren möglicherweise mehr als hundertmal so viele Leser erreicht hat wie die wichtigsten der Taschenbücher der Restaurationszeit (wie etwa die *Urania* mit einer Auflage von maximal 5000 Exemplaren, vgl. Stockinger 2018 und Schröder 1970), dann kann man auch die Frage des Einflusses des Mediums auf den Gattungsbegriff nicht länger unabhängig von der Relevanz der Medien stellen; denn die Wahrscheinlichkeit, dass ein in einem bestimmten Medienformat entstandenes Gattungsmerkmal oder ein Schreibverfahren einen Effekt auf die kulturelle Semantik hat, etwa auf die Ausbildung eines Gattungsbegriffs, hängt davon ab, welche Akteure zu welchem Zeitpunkt und mit welchen Medien in welcher Intensität in Kontakt kamen. Ausgehend von der Annahme, dass Medienformate auch Gattungsemantiken prägen können, ist es nicht mehr weit zu der Annahme, dass die in Gattungsbegriffen und -poetiken sedimentierten ästhetischen Kategorien in viel höherem Maß von der Popularität der Verbreitung und den konkreten Verlagspolitiken der einzelnen Medienformate abhängen, in denen die Gattungen groß wurden. Daher möchten wir die Diskussion anregen, ob und in welchen Kontexten die Modellierung eines digitalen Literaturkorpus nicht nur als Textkorpus, sondern als Korpus von Kommunikationsereignissen sinnvoll und produktiv sein könnte.

5 Literaturverzeichnis

- Biber, Douglas, Susan Conrad, und Randi Reppen. 1998. »Introduction: Goals and Methods of the Corpus-based Approach (Chapter 1)«. In *Corpus Linguistics. Investigating Language Structure and Use*, herausgegeben von Douglas Biber, Susan Conrad, und Randi Reppen, 1–18. Cambridge: Cambridge University Press.
- Bode, Katherine. 2018. *A World of Fiction. Digital Collections and the Future of Literary History*. Ann Arbor: University of Michigan Press.
- Calvo Tello, José. 2021. *The Novel in the Spanish Silver Age. A Digital Analysis of Genre Using Machine Learning*. Bielefeld: Bielefeld University Press.
- Fishelov, David. 1993. *Metaphors of Genre: The Role of Analogies in Genre Theory*. University Park (PA): Pennsylvania State UP.
- Fricke, Harald. 2010. »(A) Aspekte Der Literaturwissenschaftlichen Gattungsbestimmung. 1 Methodische Aspekte. 1.1 Definitionen Und Begriffsformen.« In *Handbuch Gattungstheorie*, herausgegeben von Rüdiger Zymner, 7–10. Stuttgart, Weimar: Metzler.
- Jäger, Georg. 2003. *Geschichte des deutschen Buchhandels im 19. und 20. Jahrhundert*. Frankfurt am Main: Buchhändler-Vereinigung GmbH.
- Jockers, Matthew L. 2013. *Macroanalysis: Digital Methods and Literary History*. 1st edition. Urbana: University of Illinois Press.
- Kestemont, Mike, und Folgert Karsdorp. 2021. *Humanities data analysis*. Princeton, Oxford: Princeton University Press.
- Lukas, Wolfgang. 1998. »Novellistik«. In *Zwischen Restauration und Revolution 1815–1848*, herausgegeben von Gert Sautermeister, 251–80. Hansers Sozialgeschichte der deutschen Literatur 5. München: Hanser.
- Meyer, Reinhart. 1987. *Novelle und Journal, I: Titel und Normen: Untersuchungen zur Terminologie der Journalprosa, zu ihren Tendenzen, Verhältnissen und Bedingungen*. Stuttgart: Steiner.
- . 1998. »Novelle und Journal«. In *Zwischen Restauration und Revolution 1815–1848*, herausgegeben von Gerd Sautermeister, 234–50. Hansers Sozialgeschichte der deutschen Literatur 5. München: Hanser.
- Mueller, Ralph. 2010. »1.7 Korpusbildung«. In *Handbuch Gattungstheorie*, herausgegeben von Rüdiger Zymner, 23–25. Stuttgart, Weimar: Metzler.
- Paige, Nicholas D. 2020. *Technologies of the Novel: Quantitative Data and the Evolution of Literary Systems*. Cambridge: Cambridge University Press, <https://doi.org/10.1017/9781108890861>.
- Percillier, Michael. 2017. »Creating and Analyzing Literary Corpora«. In *Data Analytics in Digital Humanities*, herausgegeben von Shalin Hai-Jew, 91–118. Cham (Schweiz): Springer.
- Scherer, Stefan, und Claudia Stockinger. 2016. »Archive in Serie. Literatur- und Kulturzeitschriften des 19. Jahrhunderts.« In *Archiv/Fiktionen. Verfahren des Archivierens in Kultur und Literatur des langen 19.*

- Jahrhunderts.*, herausgegeben von Daniela Gretz und Nicolas Pethes, 255–78. Freiburg i. Br., Berlin, Wien: Rombach.
- Schöch, Christof. 2017. »Aufbau von Datensammlungen«. In *Digital Humanities: Eine Einführung*, herausgegeben von Fotis Jannidis, Hubertus Kohle, und Malte Rehbein, 223–233. Stuttgart: Metzler.
- Schöch, Christof, Roxana Patras, Tomáš Erjavec, und Diana Santos. 2021. »Creating the European Literary Text Collection (ELTeC): Challenges and Perspectives«. *Modern Languages Open*, Nr. 1 (Dezember): 25. <https://doi.org/10.3828/mlo.v0i0.364>.
- Schröder, Rolf. 1970. *Novelle und Novellentheorie in der frühen Biedermeierzeit*. Tübingen: Niemeyer.
- Schröter, Julian. 2019. »Gattungsgeschichte und ihr Gattungsbegriff am Beispiel der Novellen«. *Journal of Literary Theory* 13 (2): 227–57. <https://doi.org/10.1515/jlt-2019-0009>.
- Spoerhase, Carlos. 2018. *Das Format der Literatur: Praktiken materieller Textualität zwischen 1740 und 1830*. Göttingen: Wallstein Verlag.
- Stockinger, Claudia. 2018. *An den Ursprüngen populärer Serialität. Das Familienblatt »Die Gartenlaube«*. Göttingen: Wallstein Verlag.
- Underwood, Ted. 2019. *Distant Horizons. Digital Evidence and Literary Change*. Chicago, London: The University of Chicago Press.
- Underwood, Ted, Kevin Kiley, Wenyi Shang, und Stephen Vaisey. 2022. »Cohort Succession Explains Most Change in Literary Culture«. *Sociological Science* 9 (Mai): 184–205. <https://doi.org/10.15195/v9.a8>.
- Walton, Kendall L. 1970. »Categories of Art.« *Philosophical Review* 79: 334–67.
- Wynne, Martin, Hrsg. 2004. *Developing Linguistic Corpora: a Guide to Good Practice*. Oxford: AHDS Literature, Languages and Linguistics.
- Zuber, Margarete. 1955. *Die deutschen Musenalmanache und schöngeistigen Taschenbücher des Biedermeier. 1815 – 1848*. München: Univ.Diss.

6 Anhang

Rubrik im Inhaltsverzeichnis	1865	1866	1870	1871	1875	1876
Erzählungen und Novellen	15	15	14	8	8	10
Gedichte	10	7	11	24	8	3
Deutsch-Französischer Krieg	/	/	/	77	/	/
Geschichts- und Zeitbilder	27	36	33	0	29	24
Literatur- und Lebensbilder	23	25	38	34	21	23
Skizzen aus der Heimat und Fremde	60	38	28	20	22	32
(davon Unterkategorie <i>Elsaß und Lothringen</i>)	/	/	/	11	/	/
Naturwissenschaftliches und Medicinisches	32	24	24	0	11	20
Soziales und Volkswirtschaftliches	/	/	12	0	8	11
Verschiedenes	16	18	46	42	11	12
(Am Familientisch)	111	n.e.	89	n.e.	91	n.e.

Tabelle 7: Textanzahl nach Rubriken für ausgewählte Jahrgänge des Familienblatts *Daheim*.²³

²³Das Fehlen einer Rubrik, die nur für abgrenzbare Zeiträume eingeführt wurde, wurde mit / gekennzeichnet, das okkasionelle Fehlen einer Rubrik mit 0. Für die kürzeren Texte der Rubrik »Am Familientische« wurden nicht alle Jahrgänge ausgewertet.

	1865	1870	1871	1875	1876	1879	1883	1885	1890	1892
Erzählun- gen	2	0	1	2	2	1	3	2	8	5
Novellen	4	5	2	4	5	4	4	3	1	2
anderes Label (z.B. Roman, Humores- ke)	6	8	3	1	3	3	9	11	3	9
ohne Label	3	1	2	1	0	2	4	2	9	1

Tabelle 8: Gattungslabel der Erzähltexte für ausgewählte Jahrgänge des Familienblatts *Daheim*.