

# Historische Korrespondenzen und Social Media Analytics

Eine experimentelle Analyse der Briefe aus Jean Pauls Umfeld

Frederike Neuber

## Einleitung

Wenn die Briefkultur um 1800 Parallelen zu den Social Media der Gegenwart aufweist, sollten ihre Kommunikationsstrukturen und -praktiken auch mit ähnlichen Methoden analysierbar sein, so lautet die Eingangsthese dieses Beitrags. Während wissenschaftliche Editionen traditionell Ergebnis und Gegenstand qualitativer Forschung sind, in denen der Einzeltext oder wenige Texte im Zentrum der Beobachtung und Argumentation stehen, werden die umfangreichen Datenkorpora der Social Media vorrangig aus quantitativer Perspektive untersucht.<sup>1</sup> Durch den digitalen Wandel der Gesellschaft ist das Sammeln, Überwachen, Analysieren und Visualisieren von Informationen aus den sozialen Medien im letzten Jahrzehnt für die verschiedensten Disziplinen und Arbeitsfelder immer relevanter geworden, darunter Kommunikationswissenschaft, Wirtschaft, Informatik, Politik und öffentliche Verwaltung.<sup>2</sup> Diesem Kontext entstammen die Social Media Analytics, die als interdisziplinäres Forschungsfeld und Querschnittsmethode wissenschaftliche Verfahren zur Auswertung von sozialen Interaktionen und Inhalten entwickeln.<sup>3</sup> Das dabei entstehende ‚distant reading‘<sup>4</sup> der Kommunikation, das sich von der Detaillektüre einzelner Texte löst, schafft eine abstrakte und übergeordnete Perspektive auf einen größeren Textbestand und macht allgemeine Merkmale

---

<sup>1</sup> Neben quantitativen Analysen von Daten der sozialen Medien, die numerisch-statistische Ergebnisse liefern, kommen qualitative Verfahren vorrangig in Userstudien wie Interviews oder Fokusgruppen zum Einsatz. Qualitative Verfahren können tiefe Einblicke in Verhaltensweisen der User geben, sind aber meist wesentlich aufwendiger durchzuführen und ihre Ergebnisse nicht generalisierbar bzw. repräsentativ. Vgl. Karen E. Sutherland, *Strategic Social Media Management: Theory and Practice*, Singapur 2020, hier S.14–16.

<sup>2</sup> Für einige Beiträge aus den verschiedenen Feldern siehe Klaus Bruhn Jensen, *A Handbook of Media and Communication Research: Qualitative and Quantitative Methodologies*, 3. Aufl., New York / London 2020; Markus Strohmaier, Maria Zens, *Analyse Sozialer Medien an der Schnittstelle zwischen Informatik und Sozialwissenschaften*, in: *Soziale Medien: Gegenstand und Instrument der Forschung*, Schriftenreihe der Arbeitsgemeinschaft Sozialwissenschaftlicher Institute (2014), S. 73-95; Stefan Stieglitz, Tobias Brockmann, Linh Dang Xuan, *Usage of social media for political communication*, in: *Proceedings of 16th Pacific Asia conference on information systems*, Ho Chi Minh City 2012; Daniel Zeng, Hsinchun Chen, Robert Lusch, Shu-Hsing Li, *Social media analytics and intelligence*, in: *IEEE Intelligent Systems* 25.6, 2010, S. 13–16.

<sup>3</sup> Stefan Stieglitz, Linh Dang-Xuan, *Social media and political communication: a social media analytics framework*, in: *Social Network Analysis and Mining*, Jg. 3, H. 4 (2013), S. 1277–1291, hier S. 1290.

<sup>4</sup> Unter dem maßgeblich von Franco Moretti geprägten Schlagwort ‚close reading‘ versteht man Methoden und Verfahren aus den digitalen Literaturwissenschaften zur computationalen Analyse von großen Mengen an Textdaten. Das Lesen einzelner Texte wird demgegenüber als ‚close reading‘ bezeichnet. Vgl. Franco Moretti, *Conjectures on World Literature*, in: *New Left Review* 1 (2000).

und Muster der Social Media-Kommunikation identifizierbar. Eine Übertragung der Verfahren und Methoden der Social Media Analytics auf historische Korrespondenzen kann nicht nur einen neuen Blick auf die epistolare Kommunikation generieren, sondern eröffnet auch einen Spielraum, um die ‚Social-Media-haftigkeit‘ der Briefkultur um 1800 zu erproben.

Anhand der Korrespondenz im Umfeld Jean Pauls, einer ‚Social Media-Community‘ um 1800 bestehend aus Familie, Freundinnen und Kolleginnen des Schriftstellers, erprobt der Beitrag die Potenziale der Social Media Analytics für die Untersuchung historischer Briefkommunikation. Seit 2019 wird ein ausgewählter Briefbestand aus Jean Pauls Umfeld an der Berlin-Brandenburgischen Akademie der Wissenschaften ediert und erscheint seit 2020 als Teil der Editionsplattform *Jean Paul – Sämtliche Briefe digital*.<sup>5</sup> Ausgehend von den Gemeinsamkeiten zwischen epistolarer und digitaler Kommunikation zieht der Beitrag einen Vergleich zwischen Social Media-Daten und Editionsdaten und illustriert den Analyseworkflow. Die Auswertungen der Umfeldbriefe nehmen drei Metriken bzw. Kennzahlen der Social Media Analytics in den Blick, welche Auskunft über die Rolle der Korrespondentinnen im Korpus, die Gewichtung der verhandelten Inhalte sowie die Tonalität der Kommunikation geben.

### Eine Social Media-Community um 1800

Soziale Netzwerke sind kein reines Internetphänomen. Konstitution und Dynamik digitaler Netzwerke wie Facebook und Instagram weisen Ähnlichkeiten zu den Briefnetzwerken des 18. und 19. Jahrhunderts auf. Auf beiden ‚Plattformen‘ formieren sich Communities, d. h. Gruppen von Individuen mit gemeinsamen Interessen und Bekanntschaften. Das Umfeld Jean Pauls ergibt sich durch die gemeinsame Bekanntschaft zum Schriftsteller, direkt oder über Dritte. Zu dem derzeit aus 1156 Briefen bestehenden Editions-korpus (Stand Juni 2022, v.5.0<sup>6</sup>) zählen u. a. Caroline Richters Korrespondenz jenseits der Briefe mit ihrem Ehemann Jean Paul, die Briefe ihrer Schwester und Berufsschriftstellerin Minna Spazier sowie die Korrespondenzen von Jean Pauls Freunden, darunter der vielvernetzte jüdische Kaufmann Emanuel (Osmund).

---

<sup>5</sup> Digitale Edition der Briefe aus Jean Pauls Umfeld, bearbeitet von Selma Jahnke und Michael Rölcke (2020–), in: *Jean Paul – Sämtliche Briefe digital*, herausgegeben im Auftrag der Berlin-Brandenburgischen Akademie der Wissenschaften von Markus Bernauer, Norbert Miller und Frederike Neuber (2018–) <<http://jeanpaul-edition.de>> [29.6.2022].

<sup>6</sup> Siehe Daten der Edition *Jean Paul – Sämtliche Briefe digital* (Version 5.0), hrsg. im Auftrag der Berlin-Brandenburgischen Akademie der Wissenschaften von Markus Bernauer, Norbert Miller und Frederike Neuber, 2018–2022, Versionspaket <[https://github.com/telota/jean\\_paul\\_briefe/releases/tag/v.5.0](https://github.com/telota/jean_paul_briefe/releases/tag/v.5.0)>, Daten <<https://zenodo.org/record/4109518>> [29.6.2022].

Innerhalb des Umfeldkorpus lassen sich verschiedene Kommunikationszirkel ausmachen, wie die Kommunikation des Ehepaars Richter mit den Kindern Emma, Max und Odilie oder das Korrespondenznetz, das sich während Jean Pauls Zeit in Weimar bildete, dem u. a. Caroline Richter und Johann Gottfried Herder angehören. Wie in den Social Media von heute (Stichwort Influencerinnen) gibt es auch in der Umfeldcommunity einige Personen, darunter Caroline Richter, bei denen die Kommunikation besonders stark gebündelt wird.

Gemeinsame Themen und Interessen sind der Kitt der epistolaren Kommunikation. Ähnlich wie Userinnen auf Instagram unter den Hashtags #travelgram oder #instatravel von ihren Reisen berichten, ergeben sich im Umfeldnetzwerk thematische Bündelungen der Briefe zum Thema Reisen, wenn auch ohne Markierung durch ein entsprechendes Hashtag. Sowohl in den Social Media des World Wide Web als auch im sozialen Medium Brief gestalten die User die Inhalte selbst (i. e. ‚user generated content‘), weshalb diese „Informationen zu allen möglichen Lebensbereichen enthalten [können], die kaum professionell oder institutionell gefiltert würden oder den Anspruch hätten, Teil einer soziopolitischen Öffentlichkeit zu sein.“<sup>7</sup> Dies gilt auch für die Edition der Umfeldbriefe, die durch die Verhandlung von Themen wie Kindererziehung, Feierlichkeiten, Finanzen und Krankheiten Einblicke in das Leben bürgerlicher Kleinfamilien um 1800 gibt. Die Kommunikationsformen sind damals wie heute teilweise konventionalisiert, um Beziehungen und Haltungen zu bestimmten Korrespondenzpartnerinnen oder Themen auszudrücken. Während in den Social Media rhetorische Marker wie Emojis die Tonalität des Geschriebenen explizit machen können, ist die ‚Stimmung‘ in historischen Briefen dem Text und seiner sprachlichen Gestaltung inhärent.<sup>8</sup>

#### Editionsdaten als ‚Smart Data‘

Sowohl bei Social Media-Inhalten als auch bei digitalen Editionstexten handelt es sich im Kern um Datensätze,<sup>9</sup> die sich allerdings in einigen Aspekte grundlegend unterscheiden. Datensätze sozialer Netzwerke kann man meist über technische Schnittstellen, Screen Scraping oder über

---

<sup>7</sup> Strohmaier, Zens 2014, hier S. 73.

<sup>8</sup> Freilich gibt es weitere Parallelen zwischen digitalen und epistolaren Netzwerken, die jedoch im Kontext dieses Beitrags nur marginal relevant und daher nicht weiter ausgeführt sind, darunter multipolare Korrespondenzstrukturen, d. h. Nachrichten von und an mehrere Personen sowie die Entgrenzung von Privatheit und Öffentlichkeit.

<sup>9</sup> Zum Verständnis von Editionen als Daten siehe Patrick Sahle, Zwischen Mediengründenheit und Transmedialisierung. Anmerkungen zum Verhältnis von Edition und Medien, in: editio, Bd. 24, Berlin 2010, S. 23–36.

spezielle Monitoringtools beziehen,<sup>10</sup> oft in Echtzeit. Auch wenn die Daten der Social Media für Auswertungen meist vorverarbeitet werden, beispielsweise durch die Extraktion von Metadaten, ist die Überlieferung relativ unmittelbar auch Analysedatensatz. Demgegenüber wird die Überlieferung historischer Korrespondenzen durch die digitale Edition bzw. ihre Kodierung repräsentiert, in welcher editorische Entscheidungen über die Erfassung und Erschließung des Textes zum Tragen kommen.<sup>11</sup> Die Distanz zwischen originärer Kommunikation und Daten ist bei Briefkorpora damit grundsätzlich größer als bei Social Media-Datensets.

Neben der Distanz zwischen Daten und Überlieferung unterscheiden sich Social Media-Korpora und Editionsdaten vor allem in Qualität und Umfang. Die „Sozialen Medien [bieten] zwar eine Überfülle von Informationen (Big Data), aber zunächst einmal keine hochqualitativen Daten im herkömmlichen Sinn“<sup>12</sup>, wie sie in digitalen Editionen vorliegen und welche in die Kategorie ‚Smart Data‘ fallen, die Christoph Schöch wie folgt definiert:

Smart data is data that is structured or semi-structured; it is explicit and enriched, because in addition to the raw data, it contains markup, annotations and metadata. And smart data is ‚clean‘, in the sense that imperfections of the process of capture or creation have been reduced as much as possible, within the limits of the specific aspect of the original object being represented. This also means that smart data tends to be ‚small‘ in volume, because its creation involves human agency and demands time.<sup>13</sup>

Mit 1156 Dokumenten, die als Briefe eine Textlänge von einigen Seiten nicht überschreiten, ist das Korpus der Umfeldbriefe sowohl aus Sicht der Social Media-Analyse als auch aus Perspektive der quantitativen Textanalyse ein eher kleines, aber ein ‚intelligentes‘ Datenset. Die Briefe liegen im Standardformat XML vor und sind nach den Richtlinien der *Text Encoding Initiative* (TEI),<sup>14</sup> dem de facto Standard zur Kodierung wissenschaftlicher digitaler Editionen

---

<sup>10</sup> Vgl. Jensen 2020, S. 319. Bei technischen Schnittstellen bzw. Application Programming Interfaces (APIs) obliegt es den Anbieterinnen der Daten, welche Informationen sie herausgeben, u. a. Twitter stellt seine Daten über APIs zur Verfügung. (Screen) Scraping ist bei der Datenaggregation wesentlich aufwendiger und bezeichnet die Sammlung von Daten über Website-Frontends, d. h. über die graphische Benutzeroberfläche. Kommerzielle Monitoringtools integrieren meist Netzwerke wie TikTok oder Facebook und bieten über Dashboards verschiedene Analyseoptionen an, die Unternehmen v. a. dazu nutzen, um zukünftige Entwicklungen wie Trends und Kundeninteressen zu prognostizieren.

<sup>11</sup> Die editorische Bearbeitungsschicht kann den Informationsgehalt der Überlieferung einerseits reduzieren, wenn beispielsweise textkritische Phänomene stillschweigend normalisiert werden, sie andererseits aber auch mit neuem Wissen anreichern, beispielsweise durch die Referenzierung von Personennamen auf Normdatensätze.

<sup>12</sup> Strohmaier, Zens 2014, hier S. 74.

<sup>13</sup> Christof Schöch, Big? Smart? Clean? Messy? Data in the Humanities, in: *Journal of the Digital Humanities* 2.3 (2013), <<http://journalofdigitalhumanities.org/2-3/big-smart-clean-messy-data-in-the-humanities>

<sup>14</sup> World Wide Web Consortium w3c: Extensible Markup Language 1.0, 2008, <<https://www.w3.org/TR/xml/>>; TEI Consortium, TEI P5: Guidelines for Electronic Text Encoding and Interchange (Version 4.4.0), 2022<<http://www.tei-c.org/P5/>> [29.6.2022].

erfasst.<sup>15</sup> Mit der TEI werden sowohl strukturelle Informationen zur Textgestalt (z. B. Überschrift, Absatz, Zeile etc.) als auch inhaltlich angereicherte Informationen (z. B. zu textkritischen Phänomenen, zu Personen- oder Ortsnamen) explizit kodiert. Damit kann man diese Informationen nicht nur in einer digitalen Edition visualisieren, sondern sie auch computergestützt auswerten, um Muster, Beziehungen und Anomalien in Bezug auf das Gesamtkorpus festzustellen. Entscheidend für ‚smarte Daten‘ ist außerdem das Vorhandensein und die Qualität von Metadaten, d. h. von Informationen, mit denen das eigentliche Datenset beschrieben wird. Für die im Beitrag anvisierten Analysen sind vor allem zwei Informationstypen relevant: Briefmetadaten und (thematische) Verschlagwortung.

Die strukturierte Erfassung der Briefmetadaten umfasst mit dem TEI-Element <correspSearch> Informationen zu Senderinnen, Sendedatum und -ort sowie Empfängerinnen. Diese sind, wenn vorhanden, mit Normdatensätzen verlinkt, darunter die Identifikatoren der *Gemeinsamen Normdatei* und der geografischen Datenbank *GeoNames*.<sup>16</sup> Die Verwendung von Normdaten und Standards erfüllt verschiedene Funktionen<sup>17</sup> und stellt die Weichen für kontextualisierbare Analysen, in denen man bestimmte Fragen an die Brieftexte in Bezug zu Personen oder Zeiträumen setzen kann. Im Gegensatz zu den Sende- und Empfangsinformationen zählt die Verschlagwortung der Dokumente nicht zu den gängigen Metadatenkategorien in Korrespondenzeditionen. Um die Vielstimmigkeit der Umfeldcommunity in den Daten abzubilden, wurde ein zweigliedriges Schlagwortsystem aus Korrespondenzkreisen und Themen implementiert. Die Verschlagwortung erfolgt briefweise innerhalb des TEI-Abschnitts <textClass> mit Referenz auf zwei Registertaxonomien.<sup>18</sup> Zu den thematischen Schlagworten zählen u. a. ‚Reisen‘ (147), ‚Berufliches‘ (17 Briefe) und ‚Bittschreiben‘ (15 Briefe).<sup>19</sup>

---

<sup>15</sup> Das Kodierungsschema der Umfeldbriefe orientiert sich an zwei TEI-Subsets: dem Basisformat des Deutschen Textarchivs und ediarum.BASE, welches der Software ediarum, mit dem im Rahmen der Umfeldbriefe ediert wird, zugrunde liegt. Siehe Berlin-Brandenburgischen Akademie der Wissenschaften (Hrsg.): DTABf Deutsches Textarchiv – Basisformat, 2011–2020, <<http://deutschestextarchiv.de/doku/basisformat>>; Ebd. (Hrsg.): ediarum/ediarum.BASE.edit (Version 2.0), bearbeitet von Stefan Dumont, Nadine Arndt, Sascha Grabsch und Lou Klappenbach, 2011–2022, <<https://github.com/ediarum/ediarum.BASE.edit>> [29.6.2022].

<sup>16</sup> Gemeinsame Normdatei, <[https://www.dnb.de/DE/Professionell/Standardisierung/GND/gnd\\_node.html](https://www.dnb.de/DE/Professionell/Standardisierung/GND/gnd_node.html)>; GeoNames, <<https://www.geonames.org/>> [29.6.2022].

<sup>17</sup> U. a. ermöglicht die Verwendung von Normdaten die Einbindung der Metadaten in andere Kontexte, wie beispielsweise in die den Webservice *correspSearch*, der Briefmetadaten verschiedener Editionen aggregiert. Siehe Stefan Dumont, Sascha Grabsch und Jonas Müller-Laackman (Hrsg.), *correspSearch – Briefeditionen vernetzen* (Version 2.0.0), Berlin 2021, <<https://correspSearch.net>> [29.6.2022].

<sup>18</sup> Beide Categoriesysteme, Korrespondenzkreise und Themen, fungieren in der digitalen Edition als Register, über die man auf die Briefe zugreifen kann und die Gemeinsamkeiten zwischen Briefen, auch wenn sie unterschiedlichen Teilkorrespondenzen entstammen, sichtbar machen. Die Verschlagwortung entsteht parallel zur Erschließung der Edition, d. h. sie kann sich bei wachsender Briefmasse verändern oder erweitern.

<sup>19</sup> Siehe Themen in Briefen aus dem Umfeld Jean Pauls, Edition der Umfeldbriefe (Anm. 5), <<https://www.jeanpaul-edition.de/themen.html>> [29.6.2022]. Die Korrespondenzkreise bündeln Briefe nach Bekanntschaftsverhältnissen. Zwei größere Korrespondenzkreise bzw. Teilcommunities im Umfeld sind der

Allgemeine Themen sind in der zweistufigen Taxonomie weiter ausdifferenziert, darunter ‚Reisen‘ in u. a. ‚Ausflüge‘ (7), ‚Besuche‘ (117) und ‚Jean Pauls Besuche in Heidelberg‘ (7).

### Workflow und Vorbereitung der Analysen

Zur Auswertung von Social Media-Datensätzen schlagen Stieglitz und Dang-Xuan ein fünfstufiges Social Media Analytics Framework<sup>20</sup> vor, das sich wie folgt zusammensetzt: i.) die Definition einer Zielstellung der Analyse, ii.) die Sammlung von Daten, iii.) die Vorverarbeitung der gesammelten Daten für die Analyse (engl. preprocessing), bei der beispielsweise irrelevante Informationen aus den Daten entfernt werden, iv.) die eigentliche Datenanalyse und v.) der Ergebnisbericht.<sup>21</sup> Der Workflow bei der Analyse der Umfeldbriefe orientiert sich im Wesentlichen an diesem Modell (Abb. 1).

#### **Abb\_Neuber\_1 / Abb. 1: Umfeldbriefe im Analyseworkflow.**

Die drei Fragenbereiche nach der Rolle von Korrespondentinnen, der Gewichtung der verhandelten Inhalte sowie der Tonalität der Kommunikation bilden den Ausgangspunkt der Analysen. Ihnen wird mittels drei Metriken bzw. Kennzahlen der Social Media Analytics nachgegangen, die im Zuge der Analysen ausführlicher definiert und diskutiert werden: Reach, Share Of Voice und Sentiment Analysis. Die Editionsdaten wurden über das Forschungsdatenrepositorium *Zenodo* bezogen<sup>22</sup> und zu zwei Analysedatensets vorverarbeitet (engl. preprocessing), welche auf die Aspekte reduziert sind, die für die geplanten Metriken relevant sind:<sup>23</sup> Erstens, ein Metadatenkorpus in XML aus editorisch angereicherten Informationen wie Titeldaten, Korrespondenzmetadaten und Themenschlagworten.<sup>24</sup> Zweitens,

---

‚Caroline Richter-Kreis‘ (189 Briefe), der die Korrespondenz um Jean Pauls Frau aggregiert, oder der ‚Zeitung für die elegante Welt-Kreis‘ (113 Briefe), in dem sich die Korrespondenz um die von Jean Pauls Schwager Karl Spazier 1800 gegründete gleichnamige Zeitschrift sammelt. Siehe Korrespondenzkreise im Umfeld Jean Pauls, Ebd., < <https://www.jeanpaul-edition.de/korrespondenzkreise.html> > [29.6.2022].

<sup>20</sup> Stieglitz, Dang-Xuan 2013. Die Autoren entwickelten das Framework für Social Media Analytics im Kontext von politischer Kommunikation, wobei das Modell durch das konzeptionelle Abstraktionslevel generalisierbar für alle Formen der Datenanalyse gelten kann.

<sup>21</sup> Bei der Datenanalyse (iv.) wird zwischen drei Forschungsfeldern differenziert: Themen der Kommunikation, die vorrangig durch Text Mining-Verfahren ermittelt werden, Stimmung der Kommunikation, identifiziert im Verfahren der Sentiment Analyse, und Kommunikationsstrukturen, ausgewertet in Netzwerkanalysen. Bedingt durch den geringen Strukturierungsgrad von Social Media-Daten schlägt das Modell von Stieglitz und Dang-Xuan vorrangig Verfahren des maschinellen Lernens zur Analyse vor, während die Analysen in diesem Beitrag auf den ‚smarten‘ Editionsdaten basieren (abgesehen von der Sentimentanalyse).

<sup>22</sup> Siehe Anm. 6.

<sup>23</sup> Die Analysedatensets sind samt Ergebnisdaten und Visualisierungen auf GitHub verfügbar; siehe Frederike Neuber, *jeanpaulanalytics* (GitHub-Repositorium), 2022, <<https://github.com/FrederikeNeuber/jeanpaulanalytics>> [29.6.2022].

<sup>24</sup> Die Themenschlagwörter aus der zweistufigen Taxonomie (siehe Abschnitt zu den Editionsdaten) wurden auf die Ebene der Oberbegriffe (59) zurückgeführt, insofern es sich nicht ohnehin schon um solche handelte. Für die

ein Textkorpus (plain text) bestehend aus den Brieftexten, deren Orthografie und historische Varianz mit der Software CAB<sup>25</sup> normalisiert wurde.<sup>26</sup> Die Umfeldbriefe wurden einerseits statistisch, andererseits mit dem Text Mining-Verfahren der Sentimentanalyse<sup>27</sup> ausgewertet. Die Eingabe und Verarbeitung der Daten wurde im Verlauf des Experiments in einer Art ‚Analysespirale‘ mehrfach angepasst, parallel zum wachsendem Verständnis über die Datengrundlage und die Parameter der Analyse sowie über die gewonnenen Ergebnisse.

### **Abb\_Neuber\_2 / Abb. 2: Jahrweise Verteilung der Briefe im Korpus.**

Im Vorfeld der Analysen wurde eine ‚Bestandsaufnahme‘ der Daten und der Informationen, die für die Auswertungen relevant sind, vorgenommen: Insgesamt korrespondieren 189 Personen in verschiedenen bzw. mehreren Rollen in der Umfeldcommunity (Verfasserin, Empfängerin und Mitleserin). Die 1156 Briefe im Umfeldkorpus haben 109 Verfasserinnen, wobei 85 Briefe von mehr als einer Person erstellt wurden. 129 Personen haben Briefe explizit empfangen, wobei 15 Briefe mehr als eine Empfängerin haben. 26 Personen haben 259 Briefe anderer Empfängerinnen mitgelesen (ohne selbst explizit Empfängerin zu sein). Eine jährweise Zählung der Briefe zeigt ein starkes Ungleichgewicht der Briefmenge zwischen den Jahren (Abb. 2). Die relativ geringe Anzahl an edierten Briefen aus den Jahren 1812–1815 hängt vermutlich u. a. mit der napoleonischen Besatzung und den Befreiungskriegen zusammen, die das Postwesen beeinträchtigt haben. 1823/24 sind viele der Hauptprotagonisten des Korpus bereits verstorben, z. B. Jean Pauls Sohn Max und der Romanschriftsteller Johann Ernst Wagner. Sowohl die jährweise Verteilung der Dokumente im Korpus als auch die multipolaren Kommunikationsstrukturen müssen bei den Analysen stets berücksichtigt werden.

Reach

---

ersten beiden Analysen wurde das Metadaten-set mit XSLT prozessiert, die Ergebnisse als Tabellen in CSV gespeichert und mit Microsoft Excel in verschiedenen Diagrammen visualisiert.

<sup>25</sup> Siehe Bryan Jurish, *Finite-state Canonicalization Techniques for Historical German*, Potsdam 2012; Deutsches Textarchiv, DTA::CAB Web Service (v1.115), Berlin, <<https://www.deutschestextarchiv.de/demo/cab/>> [29.6.2022]. Die sprachliche Normalisierung wurde vorgenommen, um lexikonbasierte Analysen zu ermöglichen (siehe Abschnitt zu Sentiment Analysis).

<sup>26</sup> Das Textdatenset wurde mit dem Tool SentText für die Ermittlung der ‚Stimmung‘ bzw. der Tonalität der Kommunikation ausgewertet, die Ergebnisse mit Microsoft Excel visualisiert. Zu SentText siehe Thomas Schmidt, Johanna Dangel, Christian Wolff, *SentText: A Tool for Lexicon-based Sentiment Analysis in Digital Humanities*, in: Thomas Schmidt, Christian Wolff (Hrsg.), *Information between Data and Knowledge. Information Science and its Neighbors from Data Science to Digital Humanities (Proceedings of the 16th International Symposium of Information Science)*, Glückstadt 2021, S. 156 –172; SentText (Tool), <<https://thomasschmidtur.pythonanywhere.com/>> [29.6.2022].

<sup>27</sup> Für weitere Ausführungen siehe den Abschnitt zu Sentiment Analysis.

Um die Rolle von Personen für das Kommunikationsnetz im Umfeld Jean Pauls zu bestimmen, kann man die Reichweite (engl. Reach) als Kennzahl aus Online Marketing und Social Media Analytics heranziehen. Je nach Kontext kann die Definition der Kennzahl variieren, wobei über die Reichweite allgemein ausgedrückt wird, wie viele Personen erreicht werden können, und zwar je nach Kontext durch verschiedene Kommunikationsträger wie Personen, Social Media Plattformen, Websites im Allgemeinen sowie Werbeträger und Marken, die über Websites präsentiert werden.<sup>28</sup>

Für das Korpus der Umfeldbriefe wird die Reichweite der Verfasserinnen in den Blick genommen. Um den multipolaren Kommunikationsstrukturen des Umfelds Rechnung zu tragen, wird nicht die Anzahl von Briefen, die eine Person allein oder gemeinschaftlich verfasst hat, sondern die Anzahl der damit erreichten Empfängerinnen, im Folgenden als ‚Empfangskontakt‘ bezeichnet, als Bezugsgröße der Rechnung genommen. Bedingt durch die multipolaren Kommunikationsstrukturen liegt die Summe aller Empfangskontakte mit 1435 höher als die Summe der Briefe (1156), wobei hier direkte Empfängerinnen und Mitleserinnen von Briefen gleichermaßen mitgezählt werden. Im Schnitt hat jede Verfasserin rund 13 Empfangskontakte generiert, allerdings haben von den 109 Verfasserinnen lediglich 22 Personen 13 oder mehr Empfänger erreicht, d. h. nur rund 20%. Allein die 5 aktivsten Verfasserinnen machen mit 792 Empfangskontakten rund 64% der Kommunikation im Korpus aus: Caroline Richter (289 Empfangskontakte), ihr Vater Johann Siegfried Wilhelm Mayer (192), ihr Sohn Max Richter (109), der Romanschriftsteller Johann Ernst Wagner (103) und Jean Pauls Freund und Kaufmann Emanuel Osmund (99). Im derzeitigen Datenbestand liegt also eine enorme Ballung der Kommunikation auf allein fünf Verfasserinnen, wobei Caroline Richter 23% aller Empfangskontakte generiert.

### **Abb. Neuber\_3 / Abb. 3: Brutto- und Nettoreichweite der 10 Verfasserinnen mit der höchsten Nettoreichweite.**

Die Zählung der Empfangskontakte legt die Gewichtung auf einzelne Korrespondentinnen im Korpus offen, sagt aber nur bedingt etwas über deren Reichweite im Umfeld Jean Pauls aus. Es wird daher zwischen der Summe aller Empfangskontakte einer Person, bei der Überschneidung von gleichen Empfängerinnen nicht berücksichtigt werden, und der Zahl der verschiedenen Empfängerinnen, die erreicht wurden, unterschieden. Im Online-Marketing differenziert man

---

<sup>28</sup> Die Anzahl der erreichten Personen bemisst sich bei Webseiten beispielsweise an der Nummer der Zugriffe. Zur Reichweite siehe bspw. Manfred Bruhn, Unternehmens- und Marketingkommunikation: Handbuch für ein integriertes Kommunikationsmanagement, 2. Aufl., München 2012, S. 1153; Franz-Rudolf Esch, Andreas Herrmann, Henrik Sattler, Marketing: Eine managementorientierte Einführung, 5. Aufl., München 2017, S. 310.



diese beiden Größen als Bruttoreichweite und Nettoreichweite.<sup>29</sup> Eine Auswertung und Visualisierung der 10 Senderinnen mit der höchsten Nettoreichweite ergibt folgendes Bild zur Rolle der Korrespondentinnen im Korpus (Abb. 3): Caroline Richter, deren hohe Bruttoreichweite bereits angesprochen wurde, hat gleichzeitig die höchste Nettoreichweite (Brutto 298 / Netto 48). Ihr Vater, Johann Siegfried Wilhelm Mayer, der anteilig die zweithöchste Summe an Ebdn aufweist, hat eine verhältnismäßig geringe Nettoreichweite (103 / 29), da ein Großteil seiner Korrespondenz die Briefe mit seiner Tochter ausmachen. Jean Paul, der als ‚eigenständiger‘ Korrespondent nicht im Umfeldkorpus vertreten ist, ist durch gemeinschaftliche verfasste Briefe mit Caroline und seiner Tochter Emma in der Aufstellung vertreten (60 / 50). Der Korrespondent mit der vierthöchsten Bruttoreichweite von 103 Empfangskontakten, Johann Ernst Wagner, hat die zweithöchste Nettoreichweite im Umfeldkorpus (103 / 29). Der Romanschriftsteller kontaktierte ab 1802 viele Personen in ganz Deutschland, um Werbung für eine von ihm geplante Kunstschule zu machen.<sup>30</sup> Daneben sticht Heinrichs Voß in der Abbildung ins Auge, da er eine geringe Bruttoreichweite hat, aber 50 Prozent seiner Empfangskontakte an verschiedene Personen gerichtet hat (18 / 9). Seine Briefe wurden von den Editorinnen gezielt danach ausgewählt, ob sie Schilderungen über Jean Pauls Besuche in Heidelberg beinhalten.<sup>31</sup> Emanuel Osmund, fester Protagonist des Umfelds, hat die gleiche Nettoreichweite wie Voß, aber wesentlich mehr Empfangskontakte generiert (99 / 9).<sup>32</sup>

Die Berechnung von Brutto- und Nettoreichweite gibt Aufschluss über die Rolle der Senderinnen im Umfeld hinsichtlich der Zahl der Empfangskontakte. In ihrer Einfachheit verdeutlichen die Kennzahlen, dass Menge an geschriebenen Briefen nicht zwangsläufig etwas über den Vernetzungsgrad einer Korrespondentin innerhalb des Korpus aussagt. In einem nächsten Analyseschritt könnte man ergänzende Parameter wie z. B. die Brieflänge oder die Menge an erwähnten Personen in die Berechnung aufnehmen, um die Intensität der Kommunikation weiter zu messen. Ebenfalls denkbar wäre die Anwendung der Metrik auf

---

<sup>29</sup> Vgl. Bruhn 2017, S. 1153; Anne Marx, *Media für Manager: Was Sie über Medien und Media-Agenturen wissen müssen*, 2012, hier S. 95–96. Die Nettoreichweite kann man in absoluten Zahlen angeben oder in Bezug auf eine Zielgruppengröße prozentual bestimmen. Vorliegende Berechnung erfolgt in absoluten Zahlen.

<sup>30</sup> Auch im Umfeld sucht Johann Ernst Wagner nach Unterstützern, siehe u. a. seine Briefe an Georg Joachim Göschel, 6. Februar 1805, Edition der Umfeldbriefe (Anm. 5), <<http://jeanpaul-edition.de/brief.html?num=JP-UB1035>>; An Christian Freiherr Truchseß von Wetzhausen, Ebd., <<http://jeanpaul-edition.de/brief.html?num=JP-UB0870>>; An Friedrich von Müller, 4. März 1808, Ebd., <<http://jeanpaul-edition.de/brief.html?num=JP-UB0867>> [29.6.2022].

<sup>31</sup> Siehe dazu den Beitrag „Sei vorsichtig mit diesem Briefe [...] Es ist ein Privatbrief.“ Copy & paste in Heinrich Voß' Berichten über Jean Pauls Besuche in Heidelberg von Michael Rölcke in diesem Band.

<sup>32</sup> Andere Korrespondentinnen wie Ernestine Mahmann (Bruttoreichweite 47 / Nettoreichweite 3) und Charlotte von Kalb (47 / 2), die eine relativ hohe Bruttoreichweite haben, sind aufgrund ihrer geringen Nettoreichweite nicht in der Aufstellung vertreten.

bestimmte Korrespondenzkreise, um innerhalb dieser reichweitenstarke Akteurinnen auszumachen.

### Share of Voice

Von den Personen zu den Inhalten der Kommunikation. Im digitalen Marketing gibt der Share of Voice Auskunft über die Sichtbarkeit und Relevanz einer Marke oder eines Thema in den sozialen Medien.<sup>33</sup> Übertragen auf die Umfeldbriefe kann man mit der Kennzahl ermitteln, welche Rolle verschiedene Themen in der Gesamtkommunikation spielen.

Als erster Schritt der Analyse wurden die manuell durch die Bearbeiterinnen der Editionen vergebenen Themenschlagworte ausgewertet.<sup>34</sup> Berücksichtigt man in der Analyse nur die erste Ebene der Thementaxonomie, dann gibt es 59 verschiedene Schlagworte, die bis dato insgesamt 2880-mal vergeben wurden, womit jedem Brief im Schnitt rund 2,5 Themen zugewiesen sind. Die fünf am häufigsten vergebenen Schlagworte sind ‚Reisen‘ (247), ‚Krankheit bzw. Gesundheitszustand‘ (199), ‚Verlage / Verlegerisches‘ (169), ‚Familie/n‘ (152), ‚Finanzen‘ (144).

#### **Abb\_Neuber\_4 / Abb. 4: Share of Voice der drei am häufigsten vergebenen Themenschlagworte.**

Setzt man die Anzahl der erwähnten Themen in Bezug zu den Jahren, um das Auf- bzw. Abflammen bestimmter Kommunikationsinhalte zu untersuchen, sind absolute Zahlen nicht aussagekräftig. In Zeiträumen, für die besonders viele Briefe überliefert und in die Edition aufgenommen sind, wie beispielsweise die Jahre 1808–1811 (siehe Abb. 2),<sup>35</sup> sind folglich auch mehr Themen verzeichnet. Als relative Metrik eignet sich daher der Share of Voice, mit dem man die anteilige Sichtbarkeit und Relevanz einer Marke oder eines Themas in den sozialen

---

<sup>33</sup> Vgl. Christoph Burmann, Tilo Halaszovich, Michael Schade, Rico Piehler: Identitätsbasierte Markenführung. Wiesbaden 2018, hier S. 266. Im Bereich der Social Media Analytics werden heutzutage oftmals Verfahren des maschinellen Lernens wie Topic Modeling eingesetzt; alternativ wird das Vorkommen von Themen oder Marken auf Basis von Erwähnungen, Hashtags oder speziell definierten Keywords untersucht. Vgl. Dimitrios Milioris, Topic Detection and Classification in Social Networks: The Twitter Case, Cham 2017, u. a. S. 13; zu den Potentialen von Topic Modeling für digitale Editionen am Beispiel der Briefe aus Jean Pauls Umfeld siehe Ulrike Henny-Krahmer, Frederike Neuber, Topic Modeling in Digital Scholarly Editions, in: Bernhard Geiger Ulrike Henny-Krahmer, Fabian Kaßner, Marc Lemke, Gerlinde Schneider, Martina Scholger (Hrsg.): Machine Learning and Data Mining for Digital Scholarly Editions (Schriften des Instituts für Dokumentologie und Editorik 18), Norderstedt 2023.

<sup>34</sup> Siehe dazu den Abschnitt zu Editionsdaten als ‚Smart Data‘ sowie Anm. 18 und 19.

<sup>35</sup> Siehe zum Thema ‚Krankheit bzw. Gesundheitszustand‘ den Beitrag Der Körper im sozialen Medium Brief, Arten und Funktionen der Thematisierung von Körper und Krankheit in Briefen aus dem Umfeld des Dichters Jean Paul von Selma Jahnke in diesem Beitrag. Die darin verfügbare Grafik operiert mit absoluten Zahlen.

Medien in Bezug auf alle erwähnten Marken bzw. Themen beziffern kann.<sup>36</sup> Um die Ermittlung des Share of Voice in einer Analyse übersichtlich zu gestalten, umfasst eine exemplarische Auswertung die drei am häufigsten vertretenen Themen im Umfeldkorpus: ‚Reisen‘, ‚Krankheit bzw. Gesundheitszustand‘ und ‚Verlage / Verlegerisches‘ (Abb. 4). Bei der Ergebnisbetrachtung fällt auf, dass das Schlagwort ‚Verlage / Verlegerisches‘<sup>37</sup> für Briefe der Jahre 1812/13 und 1822–1824 nicht vergeben wurde. Im ersten Zeitraum hängt dies teilweise damit zusammen, dass es um die zwei ‚Protagonistinnen‘ zu diesem Thema, Johann Ernst Wagner und Minna Spazier, in diesen Jahren still wurde. Johann Ernst Wagner war noch im Frühjahr 1812 nach Krankheit verstorben; Spazier erholte sich in diesem Zeitraum von einer Krankheit und Trennung. Die Berufsschriftstellerin nahm die Korrespondenz zum Thema nach einigen Jahren wieder auf (im Korpus ab 1816) und führte sie bis 1821 fort. Die relative Intensivierung des Themas ‚Verlage / Verlegerisches‘ um 1825/26 erklärt sich wiederum dadurch, dass Caroline Richter nach Jean Pauls Tod am 25. November 1825 mit der Organisation und den Verhandlungen zur Herausgabe der Gesamtausgabe befasst war und darüber u. a. mit dem Verlegern Eduard Vieweg und Johann Leonhard Schrag korrespondierte.<sup>38</sup> Das Thema ‚Reisen‘ ist in allen Jahren des Korpus präsent, allerdings kann man dabei ebenfalls An- und Abstiege des Share of Voice ausmachen. Im Jahre 1812 könnte die vergleichsweise geringe Thematisierung mit der napoleonischen Besatzung bzw. den Befreiungskriegen zusammenhängen, da die Reisetätigkeiten der Gesellschaft dadurch grundsätzlich etwas eingeschränkt waren. Vergleichsweise stark fällt der Share of Voice 1812 für ‚Krankheiten bzw. Gesundheitszustand‘ aus, was u. a. damit zusammenhängt, dass Johann Ernst Wagner zu dieser Zeit an einem „fakeligem Nerfenfieber“<sup>39</sup> litt, was in verschiedensten Briefwechseln<sup>40</sup> thematisiert wurde und an dem er kurze Zeit später verstarb,

Die Inhalte der Briefe und die Ermittlung des Share of Voice lassen sich schlüssig zueinander in Bezug setzen. Abgesehen von der Ermittlung der Kennzahl für Themen, könnte man den

---

<sup>36</sup> Vgl. Roland Fiege, *Social Media Balanced Scorecard: Erfolgreiche Social Media-Strategien in der Praxis*, Wiesbaden 2012, hier S. 102.

<sup>37</sup> Siehe die Übersicht zum Schlagwort ‚Verlage / Verlegerisches‘, Edition der Umfeldbriefe (Anm. 5), <<https://www.jeanpaul-edition.de/thema.html?id=JP-012572>> [29.6.2022].

<sup>38</sup> Die Gesamtausgabe der Jean Paulschen Werke sind ein eigenes Unterthema von Verlag / Verlegerisches, dem ausschließlich Briefe ab Oktober 1825 zugeordnet sind: Thema ‚Gesamtausgabe (Jean Pauls sämtliche Werke)‘, Edition der Umfeldbriefe (Anm. 5), <<https://www.jeanpaul-edition.de/thema.html?id=JP-011966>>.

<sup>39</sup> Von Antonie von Mützscheffel an Emanuel. Meinigen, 31. Dezember 1811 bis 1. Januar 1812, Ebd., <<http://jeanpaul-edition.de/brief.html?num=JP-UB0354>> [29.6.2022].

<sup>40</sup> U.a. tauschen sich Johann Ernst Wagners Söhne Carl und Anton mit Christian Freiherr Truchseß über die Krankheit des Vaters aus; siehe Christian Freiherr Truchseß von Wetzhausen an Carl und Anton Wagner, 2. Januar 1812, Ebd., <<https://www.jeanpaul-edition.de/umfeldbrief.html?num=JP-UB1117>>; Carl und Anton Wagner an Christian Freiherr Truchseß von Wetzhausen, 6. Januar 1812, Ebd., <<https://www.jeanpaul-edition.de/umfeldbrief.html?num=JP-UB0976>> [29.6.2022].

Share of Voice beispielsweise auch personenbezogen berechnen. Mit den in den Briefen vorgenommenen Personenindizierungen als Grundlage wäre auswertbar, welche Personen zu welchem Zeitpunkt anteilig besonders relevant in der Kommunikation des Umfelds sind.

## Sentiment

Neben dem ‚wer‘ und dem ‚worüber‘ der Kommunikation, ist das ‚wie‘ in Social Media Analysen hoch relevant, d. h. die Tonalität des Geschriebenen. Eine Methode, um Emotionen, Stimmungen, Bewertungen und Einstellungen in Texten auszuwerten, ist Sentiment Analysis (auch Opinion Mining, Sentimentanalyse). Sie trägt dazu bei, die Gefühle bzw. die Tonalität von Userinnen gegenüber einem Thema oder einer Marke zu identifizieren.<sup>41</sup> Sentimentanalysen können auf Verfahren des maschinellen Lernens basieren und bzw. oder lexikonbasiert unter Verwendung von Techniken des Natural Language Processing und des Text Minings erfolgen.<sup>42</sup> In einer lexikonbasierten Analyse, wie sie für die Umfeldbriefe vorgenommen wurde, wird – kurz gefasst – jedem Wort bzw. Satz in einem Dokument ein Wert zugewiesen wird, der auf einer positiven oder negativen Gewichtung in einem Wörterbuch basiert. Die Werte reichen von +1 bis -1, wobei 0 neutral, 1 stark positiv und -1 stark negativ ist.<sup>43</sup> Die Kombination aller Sentimentwerte der positiven und negativen Wörter im Text ergibt den Sentiment Score für einen Gesamttext, d. h. im Fall der vorliegenden Analyse für einen Brief. Da Gefühle oft an Ereignisse geknüpft sind,<sup>44</sup> wird die Methode für die Umfeldbriefe dazu eingesetzt, besonders einschneidende Ereignisse – ob positiv oder negativ – in den Briefen einzelner Senderinnen auszumachen.

Zunächst wurde die ‚Stimmung‘ aller Senderinnen, die mehr als 20 Briefe im Korpus verfasst haben, mit dem Tool SentText<sup>45</sup> konvolutvergleichend ausgewertet. Von den 14 Konvoluten

---

<sup>41</sup> Carlos A. Iglesias, Antonio Moreno, Editorial, in: dies. (Hrsg.), *Sentiment Analysis for Social Media*, Basel 2020, S. 1–4. Häufig kommt die Methode zur Analyse der politischen Stimmungslage zum Einsatz; siehe Melanie Siegel, Jennifer Deuschle, Barbara Lenze, Marina Petrovic, Sascha Starker, *Automatische Erkennung von politischen Trends mit Twitter – brauchen wir Meinungsumfragen noch?*, in: *Information - Wissenschaft & Praxis*, 68,1, 2017, S. 67–74.

<sup>42</sup> Bing Liu, *Sentiment Analysis: Mining Opinions, Sentiments and Emotions*, Cambridge 2015, 121f.

<sup>43</sup> Die Lexika werden meist manuell erstellt und verwenden unterschiedliche Messungen von Emotionen als Grundlage. Da orthografische Varianz dazu führen kann, dass der Abgleich zwischen Text und Wörterbuch nicht funktioniert, wurden die Brieftexte im Vorfeld der Analyse mit der Software CAB normalisiert (siehe Anm. 25); für vorliegende Analyse wurde das Wörterbuch SentiWS verwendet (siehe Robert Remus, Uwe Quasthoff, Gerhard Heyer, *SentiWS – a Publicly Available German-language Resource for Sentiment Analysis*, in: *Proceedings of the 7th International Language Resources and Evaluation*, S. 1168–1171, 2010).

<sup>44</sup> Federico Alberto Pozzi, Elisabetta Fersini, Enza Messina, Bing Liu (Hrsg.), *Sentiment analysis in social networks*, Amsterdam 2016, hier S. 178.

<sup>45</sup> Siehe Anm. 26.

weisen die die Texte von 13 Senderinnen einen positiven Sentiment Score auf, wobei die Brieftexte von Jean Paul und Caroline Richters Freundin Henriette Freifrau von Ende und die des Romanschriftstellers Johann Ernst Wagner die positivsten Scorewerte aufweisen. Von Ende war eine alleinstehende, unabhängige und wohlhabende Frau, deren Briefe u. a. von den Italienreisen mit ihrem Sohn Leopold handeln und einen überschwänglichen und positiven Duktus haben (Sentiment Score +0.00944). Johann Ernst Wagner war Anfang des 19. Jahrhunderts, wie bereits im Zusammenhang mit der Berechnung zur Reichweite erwähnt, in Begriff, eine deutschlandweite Kunstschule zu gründen und suchte in seinen Briefen mit motivierter Tonalität Mitstreiterinnen für dieses Vorhaben (+0.008797). Auffällig an der Auswertung der 14 Briefkonvolute ist das untere Ende der Sentimentskala, an dem die 21 Briefe des Verlegers Friedrich Arnold Brockhaus stehen, deren Sentiment Score den einzigen kumulativen Negativwert hat (-0.000518).<sup>46</sup>

#### **Abb\_Neuber\_5 / Abb. 5: Sentimentanalyse der Briefe von Friedrich Arnold Brockhaus.**

Eine chronologische und briefweise Visualisierung des Sentiment Score der Brockhaus-Briefe zeigt (Abb. 5), dass 4 Scheitelwerte, d. h. vier Brieftexte, signifikant hervorstechen, da sie im Positiv- bzw. Negativbereich den Wert 0,01 bzw. -0,01 überschreiten. An den Scheitelwerten (referenziert über Kleinbuchstaben) kann man unmittelbar den Verlauf von Brockhaus Liebesbeziehung mit der Berufsschriftstellerin und Caroline Richters Schwester Minna Spazier ablesen:

a) An Friedrich Bornträger, 28. August 1810: Seit Anfang August ist Brockhaus mit Spazier verlobt, welche offenbar im Zusammenhang mit dem von ihr bei Brockhaus herausgegebenen Taschenbuch *Urania* Zwist mit Rahel Varnhagen hatte,<sup>47</sup> wofür Bornträger vorhergehenden Brief wohl Verständnis geäußert hatte.

In dem, was Sie mir über Minna sagen, erkenne ich Ihr gefühlvolles theilnehmendes Freundesgemüth. Ich danke Ihnen dafür. Ich vertraue und glaube, Alles wird wohl werden. Nur Muth, Thätigkeit und festes Wollen, moralisch gut zu handeln! Ich und Minna vertrauen für dort auf Sie. Vertrauen Sie auf uns!<sup>48</sup>

---

<sup>46</sup> Die Briefe richten sich an vier verschiedene Empfängerinnen: Buchhändler und Verleger Friedrich Bornträger (12 Briefe), der in Altenburg ansässige Kammervorwalter und Publizist Ernst Karl Friedrich Ludwig (7), Caroline Richter (1) und Jurist und Schriftsteller Friedrich Ferdinand Hempel (1). Siehe Briefe von Friedrich Arnold Brockhaus, Edition der Umfeldbriefe (Anm. 5), <<https://www.jeanpaul-edition.de/briefe.html?sort=date&corpus=context&sender=JP-000501>> [29.6.2022].

<sup>47</sup> Varnhagen hatte in dem von Spazier herausgegebenen *Urania*. Taschenbuch für das Jahr 1810 den Beitrag „Die Strafe im Voraus“ (S. 180–210) beigegeben; über einen Streit sind keine Details bekannt. Siehe Edition der Umfeldbriefe (Anm. 5), <<https://www.jeanpaul-edition.de/umfeldbrief.html?num=JP-UB0926>>.

<sup>48</sup> Ebd., <<https://www.jeanpaul-edition.de/umfeldbrief.html?num=JP-UB0929>> [29.6.2022].

b) An Friedrich Bornträger, 21. November 1810: In der Zwischenzeit ist Minna an einem Nervenfieber erkrankt und ist laut Brockhaus nicht wiederzuerkennen. Unter Fieber wird Spazier redselig über vergangene Liebschaften und ein uneheliches und früh verstorbenes Kind.

Wo soll ich Worte hernehmen, um Ihnen den namenlosen Jammer auszudrücken, worin ich gestürzt bin! [...] Schon in meinem letzten Briefe muß ich Ihnen gesagt haben, daß Minna krank sei. Sie ist es geblieben – sie ist es noch – sie ist – entsetzen Sie sich nicht – sie ist – wahnsinnig! [...] In einer Stunde, die sie glaubte ihre Todesstunde werden zu sollen, hat sie mir über alle ihre seitherigen Verhältnisse die vollständigsten Aufschlüsse gegeben und mir die schriftlichen Belege darüber zu Händen gestellt! Diese Aufschlüsse machen es mir unmöglich – ihr je meine Hand zu geben!<sup>49</sup>

c) An Friedrich Bornträger, 15. Januar 1811: Es deutet sich an, dass die Beziehung zwischen Brockhaus und Spazier auseinandergehen wird, nicht nur, weil Brockhaus sich durch Spaziers krankheitsbedingten Charakterwandel zunehmend von ihr distanziert, sondern auch, weil die Familie Spaziers die Trennung der beiden forciert.

Von meiner Reise nach Berlin mit der armen Minna in der furchtbarsten Kälte, von unsern Beschwerden auf derselben, meinen Sorgen und meinem Jammer, von unserer Ankunft im Hause des Vaters, von der Scene der Zusammenkunft mit diesem und Julius, von dem allgemeinen und besondern Benehmen des Vaters und der (Stief-)Mutter, endlich von der herzerreißenden Stunde des Abschieds und der Trennung.<sup>50</sup>

d) An Ernst Karl Friedrich Ludwig, 26. März 1811: Brockhaus und Spazier haben sich getrennt. Der Verleger wirkt einerseits betroffen, andererseits erleichtert:

Heute etwas über der armen Minna Schicksal. Gestern erhielt ich von Karolinen Briefe. Auch sie betrachtet unsere Trennung – Minna's und meine – als entschieden durch den Willen des Vaters. Mein Herz zuckt krampfhaft bei dieser Entscheidung, denn Minna war mir unendlich und ist mir noch sehr teuer. Mein Verstand tritt aber der Entscheidung des Vaters mit Beifall bei.<sup>51</sup>

Die Sentimentanalyse macht das Auf und Ab der Gefühle Brockhaus', wie sie in den Briefen ausgedrückt werden, geradezu minutiös nachverfolgbar. Exemplarisch wird damit an einer überschaubaren Textmenge deutlich, dass die Methode, angewandt auf ein größeres Textkorpus, Stimmungen und Ereignisse identifizierbar machen kann. Ein anderes

---

<sup>49</sup> Ebd., <<https://www.jeanpaul-edition.de/umfelddbrief.html?num=JP-UB0938>> [29.6.2022].

<sup>50</sup> Ebd., <<https://www.jeanpaul-edition.de/umfelddbrief.html?num=JP-UB0945>> [29.6.2022].

<sup>51</sup> Ebd., <<https://www.jeanpaul-edition.de/umfelddbrief.html?num=JP-UB0958>> [29.6.2022].

Anwendungsszenario von Sentiment Analysis wäre beispielsweise einen Zusammenhang zwischen behandelten Themen und Stimmung herzustellen, um herauszufinden, welche Gesprächsinhalte positiv bzw. negativ konnotiert sind.<sup>52</sup>

### Fazit

Auch wenn es sich bei der Edition der Briefe aus Jean Pauls Umfeld nicht um ‚Big Data‘ handelt, zeigen die Analysen, dass sich aus den hoch strukturierten und informationsreichen Daten Erkenntnisse über die im Korpus abgebildete Community ableiten lassen.<sup>53</sup> Durch die drei experimentell auf die Umfeldbriefe übertragenen Konzepte Reichweite, Share of Voice und Sentiment wurde gezeigt, dass man historische Briefkorpora nicht nur mit den Social Media der Gegenwart vergleichen kann, sondern ihre Kommunikationsstrukturen durch den Methodentransfer der Social Media Analytics untersuchen kann. Zwei Aspekte der Datengrundlage sind dabei für die Kommunikation und Deutung der Analyseergebnisse relevant: Zum einen liegt es in der Natur des Umfelds, keine klar definierbaren Ränder zu haben, weshalb es an den Editorinnen ist, diese durch die Selektion bzw. den Ausschluss von Quellen zu bestimmen; darüber hinaus kann das Auswahlkorpus je nach Überlieferungslage Lücken aufweisen. Zum anderen ist das Briefkorpus der Edition ein wachsender Datensatz, weshalb sich die Ergebnisse durch die Integration weiterer Korrespondenzen zukünftig verändern können. Für die Kommunikation und Deutung der Analysen aufgrund beider Aspekte unabdingbar, die Ergebnisse stets auf die Datengrundlage zu beziehen und nicht auf die tatsächliche Korrespondenzrealität im Umfeld. Überdies müssen die Analyseergebnisse, meist Zahlenwerte, in eine für die Leserinnen verständliche Form gebracht werden, d. h. visualisiert werden. Wie die Analyse selbst ist die Generierung von Visualisierungen ab einer bestimmten Datenmenge meist mit einer Reduktion von Komplexität verbunden, damit Übersichtlichkeit und Lesbarkeit gewahrt bleiben.<sup>54</sup> Damit ist eine gewisse Datenmenge zwar

---

<sup>52</sup> Jedoch ist ein grundsätzliches Problem bei der Anwendung der Methode auf historische Texte, dass Sentimentlexika vorrangig Gegenwartssprache abbilden. Die Normalisierung des historischen Sprachstandes fängt diese Problematik nur teilweise ab, denn einige Ausdrücke und Formulierungen der Zeit um 1800 finden sich nicht in den Sentimentwörterbüchern. Dadurch wird ein Großteil der Wörter im Dokument als ‚neutral‘ klassifiziert und hat keinen Einfluss auf den Sentiment Score. Vgl. Thomas Schmidt, Manuel Burghardt, Christian Wolff, Herausforderungen für Sentiment Analysis bei literarischen Texten, in: Manuel Burghardt, Claudia Müller-Birn (Hrsg.), INF-DH 2018, Bonn 2018, S. 5.

<sup>53</sup> Allgemeiner formuliert in danah boyd, Kate Crawford, Critical Questions for Big Data, in: Information, Communication & Society 15,5, 2012, S. 662–679, hier S. 670: „The size of data should fit the research question being asked; in some cases, small is best“.

<sup>54</sup> Im Fall der Reichweitenberechnung umfasst die Visualisierung nur die 10 Korrespondentinnen mit der höchsten Nettoreichweite (anstatt alle 109 Senderinnen), für die Ermittlung des Share of Voice wurden lediglich die drei am häufigsten vergebenen Schlagworte anteilig in Beziehung gesetzt (und nicht alle 59!) und bei der

einerseits die Voraussetzung, damit Analysen sinnhaft eingesetzt werden können (nämlich dann, wenn man nicht mehr alle Texte lesen kann), andererseits aber auch eine Herausforderung bei der Kommunikation von Analyseergebnissen.<sup>55</sup>

Insgesamt stellt der Blick auf die Daten aus der ‚Vogelperspektive‘ Sachverhalte und Zusammenhänge des Korpus heraus, die den Leserinnen für gewöhnlich nicht explizit in einer Edition veranschaulicht werden. Selbst bei einem noch überschaubaren Korpus von 1156 Briefen muss man davon ausgehen, dass sich die Leserinnen nicht mehr mit jedem Brief im Detail beschäftigen können. Somit können selbst einfache Metriken wie die Reichweite das Verständnis der Leserinnen für Gewichtungen und Beziehungen im Korpus schärfen, was wiederum auch die Einschätzung der Kommunikationsstrukturen im Umfeld befördert. Ungleichgewichte im Korpus, beispielsweise in der Zahl der Dokumente pro Jahr, können durch anteilig konzipierte Metriken wie den Share of Voice ausgeglichen werden. Starke Gefühle und einschneidende Ereignisse, die in den Brieftexten – positiv oder negativ – beschrieben sind, werden durch Sentimentanalysen identifizierbar. In allen drei Analysenbeispielen gibt die quantitative Perspektive den Impuls, in das Korpus und auf einzelne Brieftexte zu ‚zoomen‘, also ‚distant reading‘ und ‚close reading‘ zu verbinden. Die wissenschaftliche Detailerschließung der Briefe bleibt damit weiterhin zentrales Instrument der Forschung, kann aber durch quantitativen Analysen um ein überblick- und erkenntnisgenerierendes Instrument ergänzt werden. Lässt man der Fantasie freien Lauf und stellt sich das Dashboard eines Social Media Monitoringtools als Teil einer digitalen Edition der Umfeldbriefe vor, auf dem man als Leserin verschiedene Analyseübersichten und -optionen hat, dann wäre dies nicht nur eine neue, durch die Social Media geprägte Perspektive auf die Briefkultur um 1800. Vor allem wäre ein solches Editions-Dashboard ein spielerischer und zeitgemäßer Zugang zur Kommunikation in Jean Pauls Umfeld, der gerade die junge Leserschaft für wissenschaftlichen Editionen begeistern könnte.

---

Sentimentanalyse nach Senderinnenkonvoluten wurden nur die Verfasserinnen von mindestens 20 Briefen berücksichtigt.

<sup>55</sup> Vollständige Analysen auf GH