universität wien

MASTERARBEIT | MASTER'S THESIS

Titel | Title

Möglichkeiten und Methoden zur Untersuchung von Textgenese mithilfe quantitativer Textanalyse am Beispiel von Parts-of-Speech-Tagging ausgewählter Gedichte von Friederike Mayröcker

verfasst von | submitted by

Johanna Maria Unterholzner BA

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of

Master of Arts (MA)

Wien | Vienna, 2024

Studienkennzahl lt. Studienblatt | Degree programme code as it appears on the student record sheet:	UA 066 647
Studienrichtung lt. Studienblatt | Degree programme as it appears on the student record sheet:	Masterstudium Digital Humanities
Betreut von | Supervisor:	Ass.-Prof. Mag. Mag. Dr. Andreas Baumann

Ich züchte mir mein Gedicht. […]
Jetzt schießen die Fäden ans dichterwerdende Gewebe.
(Mayröcker 1983a, S. 106)

Herzlichen Dank gilt meinem Betreuer Ass.-Prof. Mag. Mag. Dr. Andreas Baumann sowie Mag. Dr. Daniel Syrovy für die literaturwissenschaftliche Beratung.

Ein großes Dankeschön geht an Susanne Rettenwander, MA vom Literaturarchiv der Österreichischen Nationalbibliothek, Wien, für die zuvorkommende Bereitstellung des Archivbestands sowie an die Rechteinhaberin Edith Schreiber für die Erlaubnis, mit den Textzeugen aus Friederike Mayröckers Nachlass so eingehend arbeiten zu dürfen.

Abstract (Deutsch)

Diese Masterarbeit geht dem Versuch nach, die Genese von Lyrik mithilfe quantitativer Textanalyse zu beschreiben und führt dies am Beispiel von Parts-of-Speech-Tagging ausgewählter Gedichte Friederike Mayröckers durch. Somit wird der Fragestellung nachgegangen, welche Arbeitsschritte und Methoden helfen, um eine Textgenese mit mehreren Textzeugen und Revisionsstufen zu modellieren und für eine quantitative Analyse vorzubereiten. Zudem wird ermittelt, inwiefern sich über die quantitative Textanalyse mit dem Fokus auf Parts of Speech (POS) im Rahmen des Fallbeispiels Vermutungen über mögliche Bearbeitungstendenzen anstellen lassen. Mit der Kombination aus editions- sowie literaturwissenschaftlichen Perspektive und linguistisch quantitativer Textanalyse liegt der Fokus auf der notwendigen Aufbereitung durch Transkription, automatischem POS-Tagging, Modellierung in XML nach TEI-Standards, Extrahierung der Textstufen und den Analysemöglichkeiten des Materials. Als Beispielkorpus werden fünf Gedichte von Friederike Mayröcker herangezogen, welche mit insgesamt 29 Textzeugen in Literaturarchiv der Österreichischen Nationalbibliothek, Wien, vorliegen, 1981 verfasst und in den Gedichtbänden *Guten Morgen, guten Abend 1978-1981* (1982) und *Winterglück. Gedichte 1981-1985* (1986) erstmal erschienen sind. Die primär deskriptive Analyse der POS-Häufigkeitsverteilungen innerhalb der Gedichtgenese wird ergänzt um den Blick auf die Gesamtzahl der Tokens, die in XML/TEI codierten, handschriftlichen Eingriffe, die in Revisionen involvierten Parts of Speech und häufig auftretende Sequenzen der Wortarten. Während die gelungene Vorgehensweise für weitere Analysekategorien erweiterbar scheint, wäre ein umfangreicherer Textkorpus notwendig, um die nuancierten Tendenzen zu Mayröckers Arbeits- und Schreibweise fundierter bestätigen zu können.

Abstract (English)

This thesis attempts to describe the genesis of poetry through the lens of quantitative text analysis, using parts-of-speech (POS) tagging of selected poems by Friederike Mayröcker. The main research question focuses on identifying the steps and methods that aid in modeling the text genesis with several text witnesses and revision stages and preparing them for a quantitative analysis. In addition, it investigates the extent to which quantitative text analysis focused on parts of speech can provide insights into possible editing tendencies in the case study. Combining perspectives from editorial studies, literary studies, and linguistically-oriented quantitative text analysis, the focus is placed on the necessary preparation through transcription, automated POS tagging, modeling in XML according to TEI standards, extraction of textual stages, and the analytical potential of the material The sample corpus includes five poems by Friederike Mayröcker, supported by a total of 29 textual witnesses archived in the Literaturarchiv der Österreichischen Nationalbibliothek, Wien, written in 1981 and first published in the poetry collections *Guten Morgen, guten Abend 1978-1981* (1982) and *Winterglück. Poems 1981-1985* (1986). The primarily descriptive analysis of the POS frequency distributions within the poem genesis is supplemented by a look at the total number of tokens, the handwritten interventions encoded in XML/TEI, the parts of speech involved in revisions and recurring word sequences. While the approach can be potentially be extended for further categories of analysis, a more extensive text corpus would be necessary in order to confirm the nuanced tendencies of Mayröcker's working and writing methods in a more substantiated way.

Inhaltsverzeichnis

1 Einleitung

Die Entstehung eines literarischen Textes ist in jedem Fall ein individueller Prozess, der im Nachhinein nur durch erhalten gebliebene Textzeugen nachvollzogen werden kann. Trotz mancher Lücken ist über das eingehende Studieren und Vergleichen von Entwürfen, Fassungen und Überarbeitungen bis hin zum veröffentlichten Text die Schreibarbeit zu erschließen. Mit dem Fokus auf Veränderungen zwischen den Textzeugen oder aufgrund handschriftlicher Revisionen wie Streichungen und Einfügungen sind spezifische Aspekte genauer in der Genese nachzuverfolgen. Damit sind Entscheidungen während des Verfassens zu erahnen, die in weiterer Folge die Schreib- und Arbeitsweise einer* Autorin*, eines Autors mitunter charakterisieren können.

Fokus dieser Masterarbeit ist es, sich der Textgenese anhand von quantitativer Textanalyse anzunähern. Dabei steht die Aufbereitung der Textzeugen und die Weiterverarbeitung bis zur Auswertung im Vordergrund. Als grundlegende Analysekategorie werden die Wortarten herangezogen, da diese mit dem Tagging von Parts of Speech (POS, d.h. morphosyntaktische Wortklasse) automatisch ermittelt werden können. Zudem werden Sequenzen in Betracht gezogen, die häufig aufeinanderfolgende Wortarten miteinander bilden, sowie die Art der handschriftlichen Eingriffe in die Textzeugen.

Für eine solche Analyse ist eine umfangreiche, aber dennoch überschaubare Materialgrundlage von Vorteil. Wie Brüning festhält, bieten sich „kleinere Textmengen, v. a. Gedichte“ (2022, S. 317) für die Untersuchung von variantenreicher Textgenese an. Im Rahmen dieser Arbeit ist die Wahl auf Gedichte von Friederike Mayröcker gefallen. Diese sind mit mehreren Textzeugen in Mayröckers umfangreichen Nachlass des Literaturarchivs der Österreichischen Nationalbibliothek, Wien, untergebracht. Wenngleich der Fokus auf den Möglichkeiten und Methoden einer textgenetischen Analyse mit quantitativer Ausrichtung liegt, so ist es mit einem geringen Korpus von Gedichten dennoch möglich, bestimmte Annahmen über die Schreib- und Arbeitsweise der Dichterin im Zuge der Analyse zu verfolgen und zu prüfen.

Konkret werden zwei Forschungsfragen formuliert:

- Welche Arbeitsschritte und Methoden helfen, um eine Textgenese mit mehreren Textzeugen und handschriftlichen Eingriffen für eine quantitative Analyse vorzubereiten und auszuwerten?

- Inwiefern lassen sich über die quantitative Textanalyse mit dem Fokus auf Parts of Speech im Rahmen des Beispielkorpus Vermutungen über Friederike Mayröckers Schreibweise und mögliche Bearbeitungstendenzen anstellen?

Insgesamt ist eine breitgefächerte Herangehensweise an die Thematik gefordert, angefangen bei quantitativer sowie digitaler Literaturwissenschaft und Editionswissenschaft hin zu linguistischen Analysekriterien und computergestützten Methoden. Die ersten Kapitel der Arbeit widmen sich diesen grundlegenden Zugängen.

Zunächst wird im Kapitel 2 zum Forschungsstand die Unterscheidung zwischen quantitativen und computergestützten Methoden in der Literaturwissenschaft aus historischer Sicht perspektiviert, um daran anschließend auf konkrete Referenzwerke wie Markus (2015), Brüning (2022), Van Hulle (2022) sowie Ketzan und Schöch (2021) einzugehen, die aus unterschiedlichen Blickwinkeln die Fragestellung und Herangehensweise der Arbeit unterstützen. Während sich Markus spezifisch auf die Bearbeitungstendenzen in der Textgenese von Ilse Aichingers Gedichten fokussiert und Brüning sowie Van Hulle auf der theoretischen als auch praktischen Ebene die Vorgehensweise zur Analyse von Textgenese beschreiben, entwickeln Ketzan und Schöch ihr eigenes Kollationswerkzeug, um voneinander abweichende Teststellen nach ihrer Besonderheit zu kategorisieren.

Im Anschluss daran wird in Kapitel 3 sowohl auf die Merkmale der Lyrik von Friederike Mayröcker eingegangen als auch auf ihre Schreib- und Arbeitsweise, die durch eigens verfasste Texte (*MAIL ART*, 1983), Interviews (Kastberger 2000b, Hell (1987)) und die textgenetische Analyse des Prosawerks *Reise durch die Nacht* von Kastberger (2000c) näher beleuchtet wird. Damit wird die Hypothesenbildung in Hinblick auf mögliche Bearbeitungstendenzen möglich.

Das vierte Kapitel befasst sich mit dem editionswissenschaftlichen Kontext, welcher zunächst mit einen kurzen historischen Abriss (Bosse und Fanta 2019) eingeleitet wird. Infolgedessen werden zwei Textkonzeptionen diskutiert und für diese Arbeit produktiv eingesetzt. Denn während Martens (1971; 2008) den Fokus auf die Textdynamik und das ‚Gemachtsein' der Texte bei der textgenetischen Analyse legt, bietet Sahle (2013) die Möglichkeit, sechs Grundkonzeptionen von Text nach unterschiedlichen Forschungsperspektiven und -interessen in seinem Textrad zu vereinen. Die Definition einschlägiger editionswissenschaftlicher Begriffe wie ‚Fassung' und ‚Textstufe' bilden den Abschluss dieses Kapitels.

Im anschließenden Kapitel 5 zum Textkorpus wird die Begründung für die Auswahl und die fünf Gedichte selbst beschrieben. Insgesamt 29 Typoskripte stehen von den Gedichten *Alp-*

Traum (12.01.1981), *Newtonringe* (03.03.1981), *Rosenfragment* (02./03.10.1981), *Wassersachen, nach Mimmo Paladino* (11.10.1981) und *vielleicht weil das Auge zuerst bricht* (23./25.10.1981) zur Verfügung.

Welchen Ablauf der Aufbereitung diese Textzeugen erfahren, wird detailliert im Kapitel 6 zur Vorgehensweise und den Methoden erläutert. Auf die Transkription der Texte folgt die Tokenisierung und das POS-Tagging durch die Python Library *SpaCy*. Die Modellierung in XML nach TEI-Standards bedient sich einem ausgewählten Set an Tags, Attributen und Attributwerten, um die maschinellen und insbesondere handschriftlichen Eingriffe in die Typoskripte angemessen auszuzeichnen. Um die Auswirkungen dieser Revisionen im Kontext des Texten zu analysieren und mit anderen Typoskripten vergleichen zu können, werden einzelne Textstufen pro verwendeter Stiftfarbe mithilfe von *XPath* extrahiert. Sowohl die Transkription als auch die Modellierung erfordern einheitliche Richtlinien, die in enger Anlehnung an das vorliegende Material beschrieben und transparent gemacht werden.

Die Analyse fokussiert sich auf die Genese der einzelnen Gedichte auf primär deskriptive Weise. Das Vorkommen der Tags, die die Eingriffe beschreiben, die Häufigkeitsverteilungen der Parts of Speech über alle Typoskripte und Textstufen hinweg sowie häufige Sequenzen werden ausgewertet und mit einem qualitativen Vergleich der Textzeugen kontextualisiert. Die Zusammenfassung und Diskussion erfolgen im achten Kapitel, in welchem auch erstmals in einem Vergleich der Ergebnisse die Gemeinsamkeiten und Unterschiede der Gedichte ermittelt werden. Darüber hinaus werden rückblickend die Vorgehensweise und die Methoden besprochen, mögliche Hindernisse und etwaige Optimierungen diskutiert. Ebenso wird ein Ausblick auf weitere methodologischen Möglichkeiten und die Ausweitung der Analysekriterien für die textgenetische Analyse geboten.

2 Forschungsstand

2.1 Quantitative Methoden in der Literaturwissenschaft

Diese Arbeit ist aufgrund der Materialgrundlage und der methodologischen Ausrichtung sowohl in der Editionswissenschaft als auch in den Bereichen der quantitativen und digitalen Literaturwissenschaft anzusiedeln. Letztere werden seit dem Aufstieg der Digital Humanities unter eben diesem Feld vermehrt subsumiert (vgl. Bernhart 2018, S. 216). Jedoch ist zunächst historisch gesehen zu unterscheiden zwischen dem Aspekt des Quantifizierens von Texten und der „Literaturwissenschaft im Kontext der Digitalisierung und Vernetzung“ (Jannidis 2022b, S. 1), wie Jannidis das Merkmal der Digitalität präzisiert. Denn Ansätze für eine quantitative Literaturwissenschaft, wenngleich bis dato nicht eingehend aufgearbeitet, lassen sich bis auf vereinzelte, mögliche Vorläufer mit Sicherheit bis ins frühe 19. Jahrhundert zurückverfolgen, so Bernhart (vgl. 2018, S. 210). Generell sind darunter nach Bernhart all jene „zählende[n], messende[n], mathematische[n], statistische[n], geometrische[n], empirische[n], computergestützte[n] und informatische[n] Verfahren“ zu rechnen, „sofern sie für die Analyse und Interpretation von Literatur Verwendung finden“ (Bernhart 2018, S. 208).

Seit den Anfängen seien drei weitere Zeiträume festzustellen, in denen es zu intensiverer Auseinandersetzung mit quantitativen Verfahren in Kombination mit literarischen Texten kommt. Aufgrund geringer Rezeption kann sich jedoch lange keine literaturwissenschaftliche Tradition etablieren. Um 1900 beschäftigen sich mehrheitlich Personen ohne philologischen Hintergrund damit, Literatur als auch Sprache mit quantitativen Mitteln zu untersuchen, ohne aber literaturwissenschaftliche Fragen zu beantworten. Ein zweiter Aufschwung verortet Bernhart ab der zweiten Hälfte des 20. Jahrhunderts bis 1980, da in dieser Zeit literaturwissenschaftliche als auch interdisziplinäre Fragestellungen diskutiert werden. Dies geschieht mitunter in Bezug auf literaturtheoretische Überlegungen und computergestützte, mathematische Verarbeitungen, welche sowohl zur wissenschaftlichen Auswertung als auch für die Fertigung von Computerkunst herangezogen werden. (Vgl. Bernhart 2018, S. 212–215)

Der Wechsel dieser produktiven Phase hin zur „Tabuisierung quantitativer Verfahren in der Literaturwissenschaft“ (Bernhart 2008, S. 57) begründet Bernhart unter anderem mit der Entfremdung der Literaturwissenschaft und der Linguistik. Letztere strebe seitdem nach quantitativ-empirischer Untersuchung von ‚Sprache‘, während in der Literaturwissenschaft „qualitativ-hermeneutische […]“ (ebd.) Ansätze dominieren.

Die dritte Zeitspanne, in der literaturwissenschaftliche „Quantifizierung als eingeführte und anerkannte Methode gilt, gewinnbringend angewendet und eng mit qualitativen und hermeneutischen Verfahren verknüpft wird“ (Bernhart 2018, S. 216), ist ab den 2000er Jahren durch den Einfluss der Digital Humanities zu beobachten.

In dieser letzten Phase taucht neuerdings der Begriff ‚Distant Reading‘ auf, welcher mit Moretti (*Distant Reading*, 2013) in Verbindung gebracht wird und oftmals fälschlicherweise als insgesamt neue Methode gefasst wird. Moretti stellt aus einer primär literaturgeschichtlichen Perspektive Untersuchungen von Textkorpora an, beispielsweise in Hinblick auf die Entwicklung und Etablierung der Charakteristika von Genres, wie der Einsatz von Hinweisen in Kriminalromanen (vgl. Moretti 2000, 212ff). Underwood (2017) geht dem Begriff als auch der implizierten Methode nach und kommt zu dem Schluss, dass ‚Distant Reading‘ insgeheim kein neuer Zugang zu literarischen Texten ist:

> I have characterized distant reading as a tradition continuous with earlier forms of macroscopic literary history, distinguished only by an increasingly experimental method, organized by samples and hypotheses that get defined before conclusions are drawn. The interdisciplinary connections that mattered most for this tradition were, until recently, located in the social rather than computational sciences. (Underwood 2017, Absatz 29)

Dabei verortet Underwood erste Auseinandersetzungen dieser Art Ende des 19. Jahrhunderts (ähnlich wie Bernhart in Bezug zur quantitativen Literaturwissenschaft) und sieht zunächst eine enge Verbindung zu soziologischen Fragestellungen. Erst mit dem Aufkommen der Digital Humanities werde mit der Formulierung des Begriffs ‚Distant Reading‘ der Konnex zu digitalen Mitteln und der Generierung von Daten geschlossen. (Vgl. Underwood 2017, Absatz 1-5)

Insofern betont Underwood wie Jannidis die Unterscheidung zwischen Herangehensweisen an Texte und den digital-computergestützten Möglichkeiten dieser. ‚Distant Reading‘ nach Underwood ist im Vergleich zur quantitativen Literaturwissenschaft nach Bernhart vager und offener formuliert und könnte womöglich als ein Teilaspekt der quantitativen Literaturwissenschaft eingeordnet werden.

In Hinblick auf die Entwicklung einer digitalen Literaturwissenschaft sind nach Jannidis‘ Ausführungen erste Schritte ab den 1960er Jahren zu verorten. Es sind hierfür vier Aspekte in Betracht zu ziehen, welche maßgebliche Veränderungen aufgrund der Digitalisierung durchliefen. Zunächst betrifft dies das Forschungsgebiet selbst aufgrund der vielfältigen Neuerungen in Bezug auf die Produktion und Verbreitung von literarischen Inhalten. Als

zweiter Punkt nennt Jannidis die „Edition und Annotation“, da hier ab 1990 Varianten computergestützt kollationiert, Stemmata erstellt, Editionen angefertigt und mithilfe des XML/TEI-Formats vermehrt digital veröffentlicht werden. Als dritten Teilaspekt wird die quantitative Analyse „mit Verfahren der Statistik und des maschinellen Lernens“ (Jannidis 2022b, S. 6) aufgeführt, welche Jannidis im Vergleich zu Bernhart erst ab 1960 etabliert sieht. Der Fokus liegt damals auf Autorschaftsattribution und Analysen stilistischer Merkmale kleiner Textkorpora aufgrund der geringen Rechenleistungen. Mit der Demokratisierung des Computers und digital verfügbaren Korpora entsteht im Kontext quantitativer Literaturanalyse der bereits diskutierte Begriff ‚Distant Reading‘, während mittlerweile das Fachgebiet ‚Computational Literary Studies‘ alle Forschungsfragen aus dieser Richtung umfasse. Mit dem vierten Punkt spricht Jannidis die Schnittstellen zwischen den Literaturwissenschaften und anderen Institutionen wie Bibliotheken und Archive an, welche „eine wesentliche Rolle bei der Herstellung, Pflege, Verwaltung und Erschließung von digitalen Informationen“ (Jannidis 2022b, S. 9) einnehmen. (Vgl. Jannidis 2022b, S. 1–7)

Die Etablierung quantitativer und digitaler Ansätze innerhalb der Literaturwissenschaft blicken auf eine wechselhafte Vergangenheit zurück, die sowohl mit der Entwicklung der Linguistik als auch der Editionswissenschaft und deren Errungenschaften verbunden werden kann. Im Zuge dieser Arbeit zeigen sich die Verknüpfungen dieser Aspekte, indem mithilfe linguistischer Ausrichtung (POS-Tagging) und editionswissenschaftlicher Aufbereitung (XML/TEI-Format) eine Annäherung an die quantitative Untersuchung von literarischer Textgenese versucht wird. Trotz der verstärkten Integration von quantitativen und computergestützten Methoden in der Literaturwissenschaft sieht sie sich gleichzeitig mit Kritik diesbezüglich konfrontiert. Jannidis führt beispielsweise Einwürfe auf, dass keine neuen Erkenntnisse generiert würden, es sich um überholte Fragestellungen und Themenbereiche handle und die quantitative Herangehensweise das Kernstück literarischer Texte nur verfehlen könne. Dementgegen zu setzen ist, dass es nicht verwerflich ist, bestehende Erkenntnisse aus anderen Blickwinkeln und moderneren Methoden erneut zu bestätigen. Zudem sind divergierende Forschungsinteressen zu begrüßen. Gleichzeitig ermöglicht ein wiederkehrender Fokus mit quantitativen Mitteln, gewisse Tendenzen aufzuzeichnen. Darüber hinaus sind die quantitativen Ansätze – entgegen einer evaluativen Literaturauffassung – bestrebt, einen deskriptiven Literaturbegriff zu unterstützen und sich damit einer objektiveren Auseinandersetzung zu widmen, so Jannidis. (Vgl. Jannidis 2022b, S. 8–12)

Die Sammelbände *Quantitative Ansätze in den Literatur- und Geisteswissenschaften* (Bernhart et al. 2018) und *Digitale Literaturwissenschaft* (Jannidis 2022a) bieten einen Einblick in sowohl theoretische Überlegungen als auch praktische Auseinandersetzungen, welche die Vielfalt der gegenwärtigen Anwendungsbereiche als auch Desiderate aufzeigen. In letztgenannter Publikation sind zwei für diese Arbeit wesentliche Aufsätze zu finden, welche unter anderem zur Formulierung der Fragestellung und dem Aufbau der Vorgehensweise bzw. des Methodenteils beigetragen haben.

2.2 Referenzwerke

Brüning (2022), Van Hulle (2022) und Markus (2015) sind essenzielle Referenzwerke, welche die quantitative Analyse von Textgenese mit analogen als auch digitalen Verfahren theoretisch und methodologisch konkretisieren. Der Aufsatz von Ketzan und Schöch (2021) hat einen starken Fokus auf die Kollationierung von varianten Fassungen.

Brüning (2022) beschreibt anhand von Lessings Beobachtungen über die verschiedenen Varianten des Textes *Messias* von Friedrich Gottlieb Klopstock eine mögliche Herangehensweise, Lessings Annahmen für eine quantitative Analyse der Textgenese zu operationalisieren. Lessings Beobachtungen beziehen sich dabei auf „metrische Änderungen, Änderungen der Wortfügung […], Änderungen des Ausdrucks, Erweiterungen, Hinzufügungen und Streichungen“ (Brüning 2022, S. 316–317). Brüning betont die Möglichkeit, anhand von digitalen Editionen eine „computergestützte Analyse der Text- und Appartdaten zu ermöglichen“ (Brüning 2022, S. 319) und kommt zu dem Schluss, es sei „möglich, textgeschichtliche Hypothesen wie diejenigen Lessings zumindest teilweise so zu operationalisieren, dass sie mit digitalen Analyseverfahren überprüft werden können“ (Brüning 2022, S. 326). Für dieses Beispiel bleibt Brüning jedoch nur bei den theoretischen Überlegungen, wie welche Analysen maschinell unterstützt bewerkstelligt werden könnten. Eine spätere Ergänzung zum Aufsatz weiht über die praktische Anwendung dieser Überlegungen ein, als die Fassung des von Franz Kafka verfassten *Process*-Texts mit der von Max Brods herausgegebenen Fassung verglichen wird. Auf die beschriebene Vorgehensweise wird im Kapitel 6 eingehender referiert.

Van Hulle (2022) schlägt einen „teleologischen Zugang“ zu textgenetischen Editionen vor, welche nach der Logik der Textentstehung aufgebaut sind und „digitale Werkzeuge“ wie „a

collation engine; […] a system to trace cuts and notes in the writing process; and […] a set of statistics applied to information in the XML encoding of the transcriptions“ (Van Hulle 2022, S. 342). In seinem Aufsatz stellt er dar, welche Bedeutung diese Werkzeuge für weitere Fragestellungen in Hinblick auf das literarische Werk haben können. Dabei geht er vom *Samuel Beckett Digital Manuscript Project*[1] aus, bei welchem über die Implementation von CollateX Sätze kollationiert mit ihren abweichenden Varianten dargestellt und damit vergleichbar werden (Vgl. Van Hulle 2022, S. 341). Zudem wird eine Art des „indirect or genetic reading“ (Van Hulle 2022, S. 344) erreicht, indem die XML annotierten Transkriptionen statistisch ausgewertet werden nach jenen Tags, welche Modifikationen (wie `<add>` und `<del>`) im Text auszeichnen. So wird ersichtlich, wie viel Prozent der Wörter über Versionen hinweg gleichbleiben, hinzugefügt oder gestrichen werden. Van Hulle ist es so möglich, erste Schlüsse in Hinblick auf „the author’s poetics“ zu ziehen, da Beckett in seinen Überarbeitungen beispielsweise mehr Wörter gestrichen als eingefügt habe. (Vgl. Van Hulle 2022, S. 344)

Die Dissertation von Markus widmet sich der Lyrik von Ilse Aichinger. Im Zuge einer Analyse der „Charakteristika und Entwicklungslinien sowie [der] Anlage und Struktur der Gedichte“, werden auch Aichingers „poetische Prinzipien“ (Markus 2015, S. 1) erfasst. Für einen maßgeblichen Teil der Dissertation liegt der Fokus auf den aufgefundenen Gedichten inklusive der Vorstufen aus dem Vorlass im Deutschen Literaturarchiv Marbach. Die Einbeziehung der Textgenese in die Analyse der Gedichte ermöglicht es, „Bearbeitungstendenzen mit ästhetischen Implikationen“ (Markus 2015, S. 24–25) herauszufiltern. Ihre umfangreiche Analyse wird im Sinne des ‚close readings‘ durchgeführt und umfasst 91 Kategorien, zu denen Klangfiguren, Reim, Rhythmus, Metrum, Struktur der Gedichte wie Syntax, rhetorische Figuren, Bildsprache und weitere augenfällige Merkmale gehören. Die Ergebnisse dazu werden in einer Excel-Tabelle festgehalten und ermöglichen über die Auswertung der Häufigkeiten und Verteilung der Kategorien eine Form des ‚Distant Readings‘. (Vgl. Markus 2015, S. 28–29, 44–45) Dies lässt schließlich Aussagen zu Aichingers Poetik in Hinblick auf ihre unterschiedlichen Schaffensphasen zu, dass bei jedem Wort „das Klangbild praktisch immer Anteil an der poetischen Aussage“ (Markus 2015, S. 309) habe, teils sei die „die Klanggestalt dem Sinngehalt sogar übergeordnet“ (Markus 2015, S. 174). Diese Erkenntnisse rühren daher, dass Aichinger einschlägige Veränderungen in der Textgenese aufgrund eines verstärkten Fokus auf den Klang

[1] Siehe https://www.beckettarchive.org, zuletzt geprüft am 15.05.2024.

der Wörter bewirkt und weniger auf die Bedeutung der Wörter achte, so Markus. Teils sehr komprimiert zeige sich dies durch gehäuftes Auftreten von Alliterationen, Assonanzen und Wortfiguren der Wiederholung, während die Semantik schwerer zu fassen sei.

Mit einem eigens erstellten Kollationsprogramm namens *Coleto* ist es Ketan und Schöch möglich, drei Fassungen des Romans *The Martian* von Andy Weir zu vergleichen (Websitefassung, Buchpublikation, Schulbuchfassung). Darüber hinaus liegt das Ziel bei *Coleto* daran, die divergierenden Stellen zu kategorisieren und zu kontextualisieren. Hierfür wird unterschieden zwischen Kürzungen und Erweiterungen („condensation", „expansion"), Einfügungen und Löschungen („insertion", „deletion"), Großschreibungen, Veränderungen bei Leerzeichen, Kursivsetzungen, Satzzeichen, Silbentrennung und Veränderungen bezüglich Zahlenangaben. In der Anwendung, welche auf Python basiert, werden zwei Texte miteinander verglichen. Ketzan und Schöch kommen in ihren Vergleichen zu dem Ergebnis, dass sowohl durch die Korrekturen für die Buchpublikation als auch in größerem Ausmaß für die Schulbuchfassung vulgäre Ausdrücke gestrichen oder ersetzt werden. (Vgl. Ketzan und Schöch 2021) Die Ausrichtung von *Coleto* für größere Textmengen ist besonders sinnvoll, wenn die Streichungen und Hinzufügungen nicht offensichtlich sind. Wenngleich im Zuge dieser Arbeit diese Art des Vergleichs über die Kollationierung nicht intensiver verfolgt wird, regt die Idee von Ketzan und Schöch an, über Adaptionen der Kategorisierung von Varianten mit linguistischer Ausrichtung wie POS-Tagging oder Lemmatisierung nachzudenken.

3 Friederike Mayröcker: Lyrik, Schreib- & Arbeitsweise

Friederike Mayröcker (1924-2021) gilt als „virtuose Sprachartistin“ (Kastberger 2000a, S. 158) und „bedeutende Verfasserin experimenteller Gedichte und Prosatexte“ (Combrink 2016, S. 61). In Hinblick auf ihre Lyrik wird Mayröckers Schreiben zumeist als „assoziativ“ (Combrink 2016, S. 61) bezeichnet, wobei dies, so Combrink, weniger auf arbiträre Verbindungen der „(oft elliptischen) Sätze, Satzreste und Wörter“ zurückzuführen sei, als auf „akustische oder semantische Ähnlichkeiten“ (Combrink 2016, S. 62). Auch Kastberger spricht von einer erkennbar „poetologisch genau kalkulierte[n] Form“ (Kastberger 1998a, S. 22), wenn auch in Bezug auf ihre Prosa. Combrink fasst die Merkmale Mayröckers Lyrik, die er in nachvollziehbare Bearbeitungen von Erlebnissen und abstraktere Behandlungen eines Themas oder Geschehens einteilt, zusammen:

> häufig finden sich Passagen in ihren Texten, in denen sich syntaktisch vollständige Sätze durch die Verse ziehen. Teilweise sind Satz- und Verslänge auch identisch. Fast kontrapunktisch gegen diese Sätze gestellt sind Wortballungen – meistens handelt es sich um Anhäufungen von Substantiven oder Reihungen gleichgebauter Satzteile. Diese Reihungen sorgen für eine Verlangsamung des Lesetempos; hier wird eine erhebliche Menge semantischer Materie auf engstem Raum zusammengezogen. (Combrink 2016, S. 64)

Nicht selten kommen Neologismen sowie Komposita und die Montagetechnik zum Einsatz, bewusst gesetzte Satzzeichen ersetzen die Rolle von Reim und Metrum in Hinblick auf die Strukturierung der Gedichte (vgl. Combrink 2016, S. 64).

Über das Zustandekommen solcher Merkmale in Mayröckers Literatur und Lyrik im Speziellen werden im Folgenden die vorgefundenen Quellen zusammengetragen, um sich der Thematik anzunähern. Die Schreib- und Arbeitsweise wird von Mayröcker selbst auf unterschiedliche Art aufgegriffen, literarisch in Lyrik und Prosa aber auch in Interviews. In ihrem Text *MAIL ART*, welcher in der Publikation *Magische Blätter* (1983b) erschienen ist, äußert sich Mayröcker gegenüber Stimmen, die nach ihrem Schreiben und ihrem Arbeitsprozess fragen würden. Dabei merkt sie an, dass sie nicht leicht über ihre literarische Arbeit sprechen kann: „daß es mir schwerfällt, mich zur Idee meiner Poesie zu äußern“ (Mayröcker 1983b, S. 21). Ihren Prozess des Schreibens und Überarbeitens beschreibt sie insofern, als dass „man […] lange korrigieren, feilen, Korrekturen von Korrekturen machen“ (Mayröcker 1983b, S. 14) müsse. An späterer Stelle im Text holt Mayröcker weiter aus und versucht, konkreter über die Vorgänge während ihres Schreibprozesses zu berichten:

> Ich könnte Ihnen etwas von random-Elementen in meinen Texten erzählen, von ästhetischen Verdichtungs- und Verdünnungszonen, und daß ich mein Wortmaterial auflade, atomisiere, deformiere, daß ich, indem ich Collage auf Collage stülpe, solches Fremdmaterial meist nur als Vorlage für eigene Bildmotive herstelle, daß ich Verba substantiviere, Substantiva verbalisiere, daß ich eine Armee von Satzzeichen einsetze, um sie attackierend, schmetternd, beiseitesprechend, lockend, besänftigend, neutralisierend funktionieren zu lassen, daß ich Wiederholungen als Leitmotive verwende, daß eines meiner Hauptanliegen darin besteht, Disparatestes zu harmonisieren, gegensätzliche Verbalelemente zusammenzuknüpfen. (Mayröcker 1983b, S. 21–22)

Beschreibungen wie diese bleiben mehrheitlich ungreifbar, da uns bei Mayröcker, wie Kastberger ausführt, „die Rede über die Produktion als eine metaphorische entgegen [tritt]“ (2000, 16). Dennoch formiert sich eine Vorstellung dessen, was sich auf Mayröckers Schreibtisch vorgetragen haben mag.

Wenngleich Kastberger in mehreren Veröffentlichungen seinen Fokus auf die Textgenese Mayröckers Prosa legt, insbesondere auf das Werk *Reise durch die Nacht* (1984), geben seine Erkenntnisse dennoch Einblick in den Ablauf und der Art Mayröckers Schreibens. Inwiefern seine Schlüsse relevant sein können für die Hypothesenbildung zur textgenetischen Analyse Mayröckers Lyrik ist jedoch zu diskutieren.

Zu Beginn des Schreibprozesses stünden einströmende Sinneswahrnehmungen aus der unmittelbaren Umgebung. Diese „Rezeption von Wirklichkeit“ gehe über in ein „Notieren und Sammeln dieser Notizen […] als ein Vorgang der Versprachlichung“, welcher „den unabdingbaren Ausgangspunkt ihrer literarischen Werke“ (Kastberger 2000c, S. 28) darstelle. Dabei sei jedoch der Einbezug dieser Notizen in bestimmten Werkideen noch nicht zwingend festgelegt. (Vgl. Kastberger 2000c, S. 28)

In weiterer Folge würden die Notizen in ein Verhältnis zueinander gebracht und erhielten ihre erste gemeinsame Niederschrift mit der Schreibmaschine. Zudem werde vom vorhandenen Material aus weitergeschrieben, neue Ideen und Formulierungen werden eingefügt. Im Kontext von *Reise durch die Nacht* beobachtet Kastberger ein Anwachsen des Textmaterials sowie die Entstehung von ‚paradigmatischen Reihen‘, welche nicht vorläufig festgehalten werden, sondern bis zur Endversion bestehen bleiben. (Vgl. Kastberger 2000c, S. 66–72)

In einem von Kastberger geführten Interview mit Mayröcker erklärt die Autorin selbst, dass am Anfang „alles noch ganz üppig und struppig“ sei und „ganz bewußt in eine Form gebracht“ (Kastberger 2000b, S. 441) werden müsse. ‚Formlos‘ sei das Niedergeschriebene dabei jedoch nie gewesen: „Der Gedanke an das Formale ist von Beginn an vorhanden. Nur ist in den früheren Fassungen alles jederzeit verschiebbar und eliminierbar“ (ebd.), so Mayröcker.

Während die ersten Fassungen noch in Kleinschreibung entstehen und damit der Entwurfcharakter unterstrichen wird, kennzeichnet die bewusste Groß-/Kleinschreibung für Mayröcker die Konkretisierung des Textes (vgl. Kastberger 2000b, S. 441). Zwischen den Fassungen geschieht kein striktes Übernehmen des Verfassten, denn auf handschriftliche Revisionen folge „eine produktive Nachschrift" (Kastberger 2000b, S. 442).

In seiner Analyse der Genese von *Reise durch die Nacht* erkennt Kastberger Bearbeitungsabschnitte, die nach einem „Anwachsenlassen und […] Überborden" (Kastberger 1998a, S. 20) des Textmaterials einen „die vorhandene Textmenge komprimierende[n] und verdichtende[n] Vorgang" auslösen und den Fokus auf „verstärkt metrische, morphologische und phonetische Äquivalenzen" (Kastberger 1998b, S. 119–120) legen. Solche Äquivalenzen betonen wiederum das ‚Produziertsein' des Textes und damit dessen poetische Funktion[2] (vgl. Kastberger 2000c, S. 83–85). Insofern nähert Mayröcker sich der „Bewahrung der Künstlichkeit der Sprache" an und entfernt sich bewusst von einer „natürlichen Behandlung von Sprache" (Kastberger 2000b, S. 450).

Diese Erkenntnisse sind eng an den Schreibprozess von *Reise durch die Nacht* gebunden und können abstrakt für Mayröckers Prosaarbeiten vermutet werden. Es sind jedoch gattungsspezifische Unterschiede herauszufiltern. In Hinblick auf das Verfassen von Lyrik spricht Mayröcker von „einem emotionalen Anstoß" (Hell 1987, S. 181) bzw. von einem „kurzgeschlossene[n] Gedanke[n], ein[em] kurzgeschlosse[nem] Gefühl" (Mayröcker 1983a, S. 105), welche zum Schreiben anregen. Dabei sei das Schreiben von Gedichten weniger mühsam als jenes von Prosa, das ihr physisch und mental zusetzen könne, wie sie in einem Interview mit Hell preisgibt (vgl. Hell 1987, S. 181). In einem Gespräch mit Kastberger verweist sie dennoch darauf, dass auch bei Gedichten bzw. „kürzeren Sachen" „zuerst […] alles

[2] Die poetische Funktion geht auf Roman Jakobson zurück, welcher ein Kommunikationsmodell aus sechs Bestandteilen formuliert: Sender*in, Empfänger*in, Nachricht, Kontext, Kontaktmedium und Code. Demnach wird eine Nachricht mit Referenz auf einen Kontext in einem (sprachlichen) Code über ein Medium von einer sendenden zu einer empfangenden Person übermittelt. Über den Fokus auf jeweils einen Bestandteil aus diesem Modell wird je eine dominierende Funktion innerhalb einer Kommunikationssituation beschrieben. Ohne in Detail auf die anderen Funktionen eingehen zu können, fokussiert sich die poetische Funktion „auf die Nachricht, also auf das Zeichen selbst" (Sexl 2004, S. 166). Diese zeichnet sich dadurch aus, dass „das Prinzip der Äquivalenz von der Achse der Selektion auf die Achse der Kombination" (Jakobson 2007, S. 170) übernommen wird. Beispielsweise wird bei der Wortwahl den semantischen oder phonetischen Ähnlichkeiten größeren Wert beigemessen als im alltäglichen Gebrauch. Dadurch können rhetorische Figuren wie Alliterationen oder Parallelismen entstehen. (Vgl. Sexl 2004, S. 166–167) Die Betonung der poetischen Funktion geschieht in Mayröckers Lyrik durch die von Kastberger erwähnten Äquivalenzen, die Combrink auch als Reihungen erwähnt.

ganz turbulent und wild“ (Kastberger 2000b, S. 441) sei und von ihr noch weiter bearbeitet werden müsse. Insofern verlieren die Aussagen zu Mayröckers Bearbeitungstendenzen nicht zwingend ihre Aussagekraft im Kontext der Analyse ihrer Lyrik für diese Arbeit.

Neben Kastberger (2000c) gibt es keine textgenetischen Analysen von Mayröckers Prosatexten. Einzeltextanalysen sind hingegen mit spezifischen Perspektiven sowohl auf ihre Prosa als auch nunmehr vermehrt auf ihre Lyrik vorhanden.[3] Zur Genese von Mayröckers Gedichten gibt es kaum weitere Ausführungen. Das Kapitel „Die Entstehung eines Textes. »Küste bei Marathon«“ aus dem Sammelband von Schmidt (vgl. 1984, S. 311–320) enthält den Abdruck aller Textzeugen, ohne diese zu kontextualisieren oder zu kommentieren. In dieser Hinsicht nimmt diese Arbeit einen ersten Schritt in die Richtung einer textgenetischen Analyse von ausgewählten Gedichten vor. Da Mayröcker ihre Gedichte mehrmals bearbeitet hat (vgl. Combrink 2016, S. 62), liegen davon zumeist zahlreiche Fassungen bzw. Textstufen vor, welche im Rahmen dieser Arbeit als Fallbeispiele fungieren und im Kapitel 5 zum Textkorpus näher beschrieben werden.

In Hinblick auf die Fragestellungen dieser Arbeit sind aus den zusammengetragenen Zitaten und Beschreibungen zu Mayröckers Gedichten und Schreiben im Allgemeinen Hypothesen abzuleiten, die Anhaltspunkte und eine Ausrichtung innerhalb der Analyse der POS-Tagging- und Modellierungs-Ergebnisse bieten. Angesichts des Textumfangs ist anzunehmen, dass sich die einzelnen Fassungen teilweise stark unterscheiden können, wenn es nach großzügigem, assoziativem Ausformulieren zu Kürzungen und Komprimierungen kommt. Insofern ist ein Auge auf die Gesamtanzahl der Tokens pro Fassung zu legen sowie auf den Einfluss von handschriftlichen Eingriffen (wie Einfügungen und Streichungen) auf die Veränderungen der Token-Anzahl. Im Kontext der Wortartenfrequenzen ist mit gewissen Auffälligkeiten zu rechnen. So ist nach der Etablierung der angesprochenen Reihungen bzw. Äquivalenzen Ausschau zu halten, bei denen dieselbe Wortart hintereinander oder dieselbe Folge von Wortarten vermehrt im Gedicht auftauchen können. Wie eingangs erwähnt, verweist Combrink beispielsweise auf gehäuftes Auftreten von Substantiven. Der Einsatz von elliptischen Sätzen lässt ebenso auf eine schrittweise erfolgende Auswahl und/oder Auslassung von bestimmten Wortarten schließen, die ermittelt werden könnte. Möglicherweise könnte dies ein geringes Vorkommen von Verben zur Folge haben. Da Mayröcker erwähnt, dass sie beispielsweise

[3] Vergleiche hierzu beispielsweise die umfangreichen Sammelbände von Kastberger und Schmidt-Dengler (1996), Melzer und Schwar (1999), Kühn (2002), Strohmaier (2009), Lughofer (2017) sowie Arteel und De Felip (2020).

Nomen zu Verben umwandelt, wäre es möglich, dass sich bestimmte Wortartenhäufigkeiten in einem bestimmten Verhältnis zueinander verändern. Zuletzt ist die Zeichensetzung nicht zu ignorieren, die Mayröcker als eines der Werkzeuge für ihre Überarbeitungen bezeichnet und in Bezug auf den Rhythmus der Gedichte einen wesentlichen Einfluss hat. Diese Hypothesen werden mehrheitlich auf Basis von Ergebnissen der textgenetischen Analyse eines Prosawerks und anhand allgemeineren Äußerungen von Mayröcker in Gesprächen formuliert. Inwiefern diese Annahmen hilfreich und zu verifizieren sind, wird sich im Laufe der Analyse und der Diskussion zeigen.

4 Editionswissenschaftlicher Kontext

Indem der Fokus auf den Möglichkeiten liegt, Textgenese quantitativ nachzuvollziehen, besteht aus theoretischer Sicht ein Zusammenhang zur Philologie bzw. Editionswissenschaft im Speziellen. Letztere durchlief historisch gesehen eine Reihe an unterschiedlichen Konzeptionen von Text. Insbesondere mithilfe von Martens (1971) und Sahle (2013) können relevante Textbegriffe im Kontext dieser Arbeit diskutiert werden, um schließlich editorische Begriffe zu definieren.

Im 19. Jahrhundert wird insbesondere der Name Karl Lachmann mit jener textkritischen Praxis verbunden, in der aus den überlieferten Texten ein ‚Archetypus' durch logisches Ausschlussverfahren identifiziert wird, der dem Original am nächsten liegt. Auf Grundlage dessen wird anhand von Eingriffen im Sinne des Autors, kurz Emendation, der ‚originale' ‚Urtext' rekonstruiert. (Vgl. Wirth und Bremer 2010, S. 18–20)

Bosse und Fanta (vgl. 2019, S. VII) umreißen in Anschluss an diese Textkonzeption vier weitere Wenden. Zunächst vollziehe sich eine Fokussierung auf den Entstehungsprozess der Texte, woraufhin eine Perspektivierung auf die Materialität des vorliegenden Bestandes folge. Dann führe die neue Darstellungsform der Texte als Edition hin zu neuen Möglichkeiten der (digitalen) Repräsentation, Rezeption und Visualisierung. Über die computergestützte Verarbeitung und Generierung zusätzlicher Daten beziehe sich die vierte Wende auf „die maschinengesteuerte Erforschung von literarischen Schaffensprozessen" (Bosse und Fanta 2019, S. VII). Insbesondere im letzten Schritt lässt sich diese Arbeit verorten.

Nicht nur historisch betrachtet gibt es verschiedene Zugangsweisen zu Text, sondern generell je nach Disziplin bestehe die Frage „worauf *zielen* wir?", so Sahle (2013, S. 1). In seinem Versuch, verschiedenste Ansätze von Text zu benennen, geht er von der Pluralität des Textbegriffs aus und hält die sechs wesentlichsten Konzeptionen anhand eines Textrads fest. Dabei lassen sich teils auch die genannten Textverständnisse von Bosse und Fanta wiedererkennen.

Die fundamentale Idee von Text sei „(1.) Gedanken werden (2.) sprachlich formuliert und (3.) medial ausgeformt" (Sahle 2013, S. 7). Darauf baut Sahle die drei Hauptbegriffe des Textrads auf. Text_S steht dabei für „Text als sprachliches Gebilde, als sprachliche Äußerung", Text_I bezeichnet den „Text als Inhalt, Idee, Intention", während Text_D den Text „als Dokument, als Materialisierung in einem Medium" (Sahle 2013, S. 8) begreift. Mit Text_S wird von einem fixierten schriftsprachlichen Text ausgegangen, welcher jedoch nicht streng auf seine materielle

Form zurückgeführt wird. Diese idealistische Auffassung des Textbegriffs entspricht damit auch dem früheren Bemühen der Editionswissenschaft, den ‚richtigen' bzw. ‚originalen' Text zu rekonstruieren. (Vgl. Sahle 2013, S. 9–12) Dagegen ist Text$_I$ frei in Hinblick auf sprachliche oder materielle Fixierung, da es um den inhaltlichen Gedanken geht. Für die gegenwärtige Editionswissenschaft ist dieser Textbegriff in seiner Abstraktheit weniger relevant. (Vgl. Sahle 2013, S. 37–41) Text$_D$ lenkt den Fokus auf die Materialität eines Textes. Erst auf Basis des grundlegenden Mediums wird der Text als Dokument mit weiteren linguistischen und

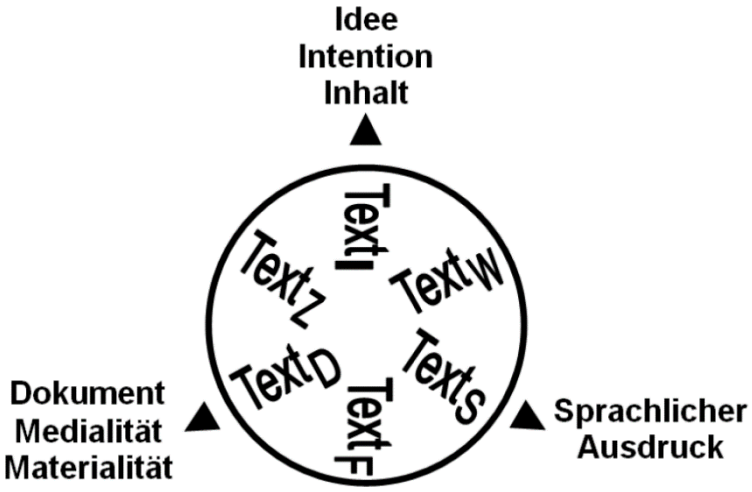

Abbildung 1: Textrad mit den sechs grundlegenden Textbegriffen nach Sahle: Text$_I$ als Idee bzw. Gedanke, Text$_S$ im Sinne einer sprachlichen Äußerung und Text$_D$ als Dokument sowie Text$_Z$ mit Fokus auf das Zeichen, Text$_W$ als Gesamtwerk und Text$_F$ als Fassungen (vgl. 2013, S. 46).

graphischen Codes lesbar. (Vgl. Sahle 2013, S. 28–32) Aus editionswissenschaftlicher Sicht wird dieser Textbegriff angewandt, wenn es um „dokumentarische Editionen (Archivausgaben), diplomatische Abschriften oder Faksimiles" (Sahle 2013, S. 34) geht.

Drei weitere Textbegriffe werden von Sahle dazwischengeschoben. Text$_W$ bezeichnet einen „Text als Werk, als sprachlich variable, aber bereits strukturierte Ausdruckformen eines Gedankens", während Text$_F$ eine „einzelne textliche Fassung, […] medial fixierte Äußerung" darstellt und Text$_Z$ den Fokus auf den Text „als zeichenhafter Ausdruck, als nicht-nursprachliche Äußerung" (Sahle 2013, S. 8) legt. Letzteres Konzept unterstreicht die komplexe visuell-semiotische Form von Texten (vgl. Sahle 2013, S. 43). Text$_W$ und Text$_F$ ergänzen sich insofern, als der Werkbegriff mit Text$_W$ mithin die Entwicklung und Reifung eines Plans bzw. einer (literarischen) Idee zusammenfasst, die sich materialistisch in unterschiedlichen Fassungen und Entwürfen als Texte$_F$ niederschlägt. Sahle ergänzt zu Text$_F$ zwei weitere Spezifikationen, nämlich eine kanonisierte Form Text$_K$ und Text$_V$ als die

Gesamtheit aller divergierenden Varianten. Dabei sind alle Ausformulierungen eines Werkgedankens in ihrer Verschiedenartigkeit gleichgestellt und zeugen vom dynamischen Prozess der Werkgenese. Sahle nennt die critique génétique als eine literaturwissenschaftliche Ausrichtung, welche $Text_W$ als Ausganglage für ihre Forschung heranzieht. (Vgl. Sahle 2013, S. 14–25)

Ein Vorläufer und Pendant zum Textbegriff der critique génétique ist jener von Martens, welcher gemeinsam mit Zeller 1971 durch die Herausgabe des Sammelbandes *Texte und Varianten* einen „Standard editionswissenschaftlicher Theoriebildung" (Plachta 2006, S. 60) festlegte. In seinem Beitrag *Textdynamik und Edition. Überlegungen zur Bedeutung und Darstellung variierender Textstufen* plädiert Martens in Bezug auf die historisch-kritische Lyrik-Edition von Georg Heyms für die Konzeption eines Textbegriffes, welcher die „Bewegung" bzw. die „Dynamik" bei der Entstehung jedes Textes betont. Wie Grésillon, die bekannteste Vertreterin der critique génétique (vgl. Bremer und Wirth 2010, S. 287), klarstellt, entwickelte sich der dynamische Textbegriff zunächst in Deutschland und wurde in Frankreich ohne weiteres Interesse am historisch-kritischen Edieren für die Erstellung eines „avant-texte" zur Interpretation des Schaffensprozesses herangezogen (vgl. Grésillon 1998, S. 52–53).

Sahles $Text_W$ ist durchaus in Martens Textbegriff wiedererkennbar, wenn Letzterer die „immanente Bewegung" des Texts betont sowie „die in ihm enthaltene Spannung, die auf ein intentionales Gebilde gerichtet ist – eine Dynamik, die nicht auf einen bestimmten Wortlaut festgelegt ist, sondern durch die kontextliche Bestimmung und durch die Variationsbreite ihre Ausrichtung offenbart" (Martens 1971, S. 169). Da es um den gesamten (unabschließbaren) Entstehungsprozess eines Textes geht, wird durch jede Variation innerhalb einer Fassung die Textdynamik sichtbar (vgl. Martens 1971, S. 170). Die Fassungen sind ähnlich zu Sahles $Texte_F$ bzw. $Text_V$ zu verstehen und ebenso als gleichwertig zu betrachten. Wird aber der Fokus nur auf eine spezifische Niederschrift gelegt, fordert Martens eine Begründung: „die Bevorzugung einer bestimmen Gestalt innerhalb der Entwicklung, bedeutet stets eine interpretierende Fixierung, die einer dezidierten Rechtfertigung bedarf" (Martens 1971, S. 171).

Martens warnt damit vor der Idee, es könne durch das Konsultieren früherer Fassungen der publizierte Text besser verstanden bzw. interpretiert werden. Dabei bleibe gerade über die Veränderungen und Eingriffe der Sinn nicht fixiert, sondern würde seine wandelnde Vieldeutigkeit aufzeigen. (Vgl. Martens 2008, S. 140–142)

Davon abgesehen befürwortet Martens die Analyse von Textgenese, welche „die Aufmerksamkeit auf das Produkt selbst, auf seine Textur, sein Gemachtsein [lenkt]; die in seiner Entstehung wirksamen Gesetzmäßigkeiten treten zum Vorschein“ (Martens 2008, S. 152). Insofern könne nachvollzogen werden, welchen Aspekten in der Überarbeitung eine gewisse Bedeutung zugemessen wurde, um „zumindest ansatzweise [zu] verstehen, was hinter den Veränderungsbewegungen steht“ (Martens 2008, S. 158). Im Zuge dessen kommt Martens auf die poetische Funktion nach Jakobson[4] zu sprechen, welche durch den Fokus auf das „Gemachtsein“ des Textes in Erscheinung trete (2008, S. 147).

In Analogie zu Martens Ausführungen werden die ausgewählten Gedichte von Mayröcker unter dem Aspekt der Wortarten quantitativ analysiert, um ebenso zu ermitteln, ob sich in dieser Hinsicht gewisse Veränderungsbewegungen herauskristallisieren. Insgesamt lassen sich in Hinblick auf das vorliegende Textkorpus die vorgestellten Begrifflichkeiten nun anwenden. Denn Sahles Textrad verweist über die divergierenden Textbegriffe auf Möglichkeiten der klaren Benennung von „unterschiedliche[n] editorische[n] Aufgaben“ und deren jeweiligen Fokus auf den Text (Vogeler 2019, S. 135). Für diese Arbeit sind insbesondere $Text_W$ und $Text_F$ sinnvoll zu übernehmen. $Text_W$ entspricht der dynamischen Entstehung einer literarischen Idee ($Text_I$) für ein Gedicht, die in verschiedenen Fassungen als $Texte_F$ formuliert werden bzw. in $Text_V$ gesammelt sind. Wie im Methodenkapitel 6 näher beschrieben wird, müssen die Fassungen (als $Texte_F$ bzw. $Text_V$) in gewisser Weise von ihrer materiellen Fixierung $Text_D$ gelöst und in Richtung eines $Textes_S$ normalisiert werden. Nur so ist es möglich, den Text maschinell lesbar zu machen und schließlich durch das Wortarten-Tagging die morphosyntaktische Dimension aufzudecken. Die an $Text_W$ anschließbare dynamische Textkonzeption von Martens ist darüber hinaus eine sinnvolle Herangehensweise, auf die literarische bzw. literaturwissenschaftliche Perspektive eines entstehenden Textes aufmerksam zu machen.

In Hinblick auf die editionswissenschaftliche Beschreibung des Textkorpus wird eine einfache Unterscheidung verfolgt, die an dieser Stelle kurz erläutert wird. Die Textzeugen liegen als verschiedene $Texte_D$ im Sinne Sahles (eingescannt) vor. Auf einem oder mehreren Textzeugen ist eine Fassung festgehalten, welche als „Ausführung eines (Kunst-)Werks (vollendet oder nicht vollendet), die von einer anderen Ausführung abweicht“ (Plachta 2006, S. 137),

[4] Vgl. dazu Fußnote 2.

verstanden wird. Innerhalb einer Fassung lassen sich insofern verschiedene Textstufen ausmachen, als sie „anhand graphischer Kriterien abgrenzbar“ (Zwerschina 1998, S. 185) sind und einer „Arbeitseinheit“ (Martens 1971, S. 180) entsprechen.

Führt man sich das Korpus vor Augen, so ist beispielsweise die Fassung eines Gedichts auf einem Textzeugen zunächst maschinenschriftlich festgehalten. Als weitere Textstufen sind die handschriftlichen Eingriffe miteinzubeziehen. Diese werden zudem in einzelne Stiftfarben zusammengefasst, da davon ausgegangen wird, dass ein Revisionsvorgang mit einem einzelnen Stift vorgenommen wurde. Auch Markus beschreibt in ihrer Vorgehensweise, dass „Änderungen in verschiedenfarbigen Stiften […] separaten Korrekturdurchgängen zuzuordnen sind“ (Markus 2015, S. 170). Da die Folge von Eingriffen nicht immer chronologisch bestimmt werden kann, werden die Stiftfarben bei der späteren Auswertung nicht miteinander kombiniert, sondern separat und gleichwertig betrachtet. Dass dies nicht dem komplexen Revisionsvorgang Mayröckers entsprechen kann, ist offensichtlich. Mutmaßungen und Interpretationen werden so jedoch vermieden.

5 Textkorpus

Die Materialgrundlage für die Arbeit bildet ein Korpus an Gedichten von Friederike Mayröcker, die im Literaturarchiv der Österreichischen Nationalbibliothek, Wien (LIT), in den Archivboxen W127 und W129 aufbewahrt sind. Dabei werden nur jene Gedichte in die Auswahl einbezogen, welche in Form von (teils mit handschriftlichen Eingriffen versehenen) Typoskripten vorliegen und dabei mindestens vier Zeugen haben. Rein handschriftliche Fassungen beschränken sich meist auf Notizen und Entwürfe, die nicht inkludiert werden. Denn diese Manuskripte verteilen sich unsortiert über teils mehr als eine Box und können aufgrund des Materialumfangs des Nachlasses schwer vollständig aufgesucht werden. Generell können die Inhalte seit Eingang des Nachlasses diesen Ausmaßes erst grob sortiert und schrittweise erfasst werden (vgl. Inguglia-Höfle und Rettenwander 2022). Damit kann kein Anspruch auf Vollständigkeit in Hinblick auf die Einbeziehung aller Textzeugen bestehen. Dennoch erscheint die Menge genügend, um die Vorgehensweise und Methodik zu veranschaulichen und zumindest einzelne Veränderungen zwischen den einzelnen Fassungen festzustellen.

Um die Schreibprozesse innerhalb einer Schaffensperiode einzufangen, bieten sich fünf Gedichte mit umfangreicherem Zeugenbestand aus dem Entstehungszeitraum 1981 an. In chronologischer Reihenfolge besteht der Textkorpus aus den Gedichten *Alp-Traum* (12.01.1981; W129/480), *Newtonringe* (03.03.1981; W129/480), *Rosenfragment* (02./03.10.1981; W127/471), *Wassersachen, nach Mimmo Paladino* (11.10.1981; W127/471 und W129/479) und *vielleicht weil das Auge zuerst bricht* (23./25.10.1981; W127/471).[5] Erstmals veröffentlicht wurden diese entweder im Gedichtband *Gute Nacht, guten Morgen. Gedichte 1978-1981* (1982)[6] oder im darauffolgenden Band *Winterglück. Gedichte 1981-1985* (1986)[7]. Als Quelle für die Transkriptionen der publizierten Fassungen dient der Sammelband *Friederike Mayröcker. Gesammelte Gedichte 1939 – 2003*, herausgegeben von Beyer (2004)[8].

[5] Fortan werden für die Gedichte die folgenden Kurztitel verwendet: *Alp-Traum* (*APTRM*), *Newtonringe* (*NWTRG*), *Rosenfragment* (*RSFGMT*), *Wassersachen, nach Mimmo Paladino* (*WSNMP*) und *vielleicht weil das Auge zuerst bricht* (*VWDAZB*).

[6] *NWTRG* auf S. 45 und *APTRM* S. 105.

[7] *WSNMP* auf S. 118, *VWDAZB*auf S. 127 und *RSFGMT* auf S. 136.

[8] *APTRM* auf S. 353f, *NWTRG* S. 362f, *RSFGMT* S. 412f, *WSNMP* auf S. 413f und *VWDAZB* S. 414.

Gedicht	Datum	Zeugen	Länge / Tokens
Alp-Traum	12.01.1981	1	71
		2	78
		3.1 & 3.2	84
		4	79
Newtonringe	03.03.1981	1	99
		2	105
		3	76
		4	77
		5	75
		6	83
		7.1 & 7.2	89
		8	86
Rosenfragment	02./03.10.1981	1	114
		2	95
		3	98
		4	76
		5	94
Wassersachen, nach Mimmo Paladino	11.10.1981	1	180
		2	151
		3	170
		4.1 & 4.2	183
		5.1 & 5.2	182
vielleicht weil das Auge zuerst bricht	23./25.10.1981	1	159
		2	135
		3	141
		4	181
		5	146
		6	147
		7	157

Tabelle 1: Auflistung der fünf Gedichte samt Entstehungszeitraum, Zeugenanzahl und Anzahl der Tokens (inkl. Satzzeichen).

Die Auswahl der einzelnen Typoskripte pro Gedicht hängt davon ab, was in den Archivboxen vorgefunden wird. Durchschläge von Typoskripten, dessen Originaltyposkript vorliegt und die selbst keine Eingriffe aufweisen, werden nicht in der Auswahl inkludiert. Über eingehendes

Studieren der ausgehobenen Typoskripte, konnten die Originale von den Durchschlägen unterschieden und deren Reihenfolge identifiziert werden.

Hilfreich ist dabei zu wissen, dass Mayröcker ihre ersten Entwürfe in Kleinschreibung formulierte (vgl. Kastberger 1998b, S. 120). Zudem ändern sich bei manchen Gedichten die Titel oder werden erst im Laufe der ersten paar Textzeugen festgehalten. Da Mayröcker zumindest stellenweise konsequent ihre handschriftlichen Änderungen in das nächste Typoskript übernimmt, erweisen sich diese Eingriffe als essenziell für die Identifikation der chronologischen Anordnung der Typoskripte. Mit einer professionellen Kollation hingegen ist es nicht zu vergleichen; dies würde jedoch den Umfang und die Ausrichtung der Arbeit übersteigen. Nichtdestotrotz stellt die Kollation einen wesentlichen Schritt dar, der beispielsweise von Van Hulle (vgl. 2022, S. 342) unternommen wird.

Schlussendlich besteht das Textkorpus aus 29 Typoskripten. Dabei sind vier Typoskripte dem Gedicht *APTRM* zuzuordnen. Acht Typoskripte repräsentieren den Schreibprozess des Gedichts *NWTRG*, das in den Typoskripten auch den Titel „Universum“ oder „Gabelsberger Universum“ trägt. Das Gedicht *RSFGMT* kann fünf Typoskripte vorweisen, ebenso das Gedicht *WSNMP*. Ein Sonderfall ist das Gedicht *VWDAZB* mit sieben Typoskripten, welche vermutlich von Mayröcker selbst in ihrer Reihenfolge nummeriert wurden und eine der umfangreichsten handschriftlichen Eingriffe aufweisen.

Hinzu kommt jeweils die veröffentlichte Fassung der Gedichte, welche im Rahmen dieser Arbeit als Endfassung zählt. In diese Richtung wird die Bewegung des Schreibprozesses interpretiert, wenn auch Martens in Hinblick darauf festhält, dass der Prozess eines Textes niemals als abgeschlossen gesehen werden kann, da jede Adaption oder divergierende Reproduktion durch den/die* Verfasser*in oder Dritte eine Weiterentwicklung des Textes bedeutet (vgl. Martens 1971, S. 170). Dennoch ist es für diese Fragestellung zumindest als vorläufige Endfassung aufzunehmen.

Mit Erlaubnis der Rechteinhaberin Edith Schreiber und in entgegenkommender Zusammenarbeit mit dem Personal des Literaturarchivs, wurden die ausgewählten Typoskripte als Scans für die Weiterverarbeitung bereitgestellt.[9] Welche weiteren Schritte daraufhin unternommen werden, ist im nächsten Kapitel zu lesen.

[9] Diese sind eingeschränkt in einer Collection des internen Repositoriums PHAIDRA der Universität Wien unter dem Link https://phaidra-local.univie.ac.at/o:45688 (vgl. dazu Daten & Code) zugänglich.

6 Vorgehensweise & Methoden

Die Vorgehensweise dieser Arbeit unterteilt sich in Transkription, POS-Tagging, Modellierung, Extrahierung der Textstufen und anschließende Analyse, Vergleich und Diskussion der Ergebnisse. Die Entscheidungen, die während der Vorgehensweise in Hinblick auf jeden einzelnen Schritt getroffen werden müssen, werden so detailliert wie möglich beschrieben und legitimiert. Insofern soll dem Spannungsfeld zwischen „Befund und Deutung" (Zeller 1971, S. 45) bzw. „Beobachtungsebene und […] Beurteilungsebene" (Wirth und Bremer 2010, S. 12) entgegen gewirkt werden. Dennoch kann die Interpretation gewisser textlicher Phänomene, insbesondere in Bezug auf handschriftliche Eingriffe, nicht vollkommen objektiv geschehen. Dies mag zum Teil auch an der Orientierung an dem ‚Prinzip Edition' nach Brüning liegen, welches im Unterkapitel 6.3 zur Modellierung näher vorgestellt wird. Es sieht von einer diplomatischen Abschrift der Textzeugen ab und fordert damit konkrete Deutungen der handschriftlichen Eingriffe der Autorin, welche nicht in allen Fällen mit absoluter Sicherheit gewährleistet werden können. Hiermit ist jedoch der Versuch eines Leitfadens entstanden, der auf den spezifischen Fall der ausgewählten Gedichte angepasst ist, und versucht, alle besonderen Vorkommnisse und den Umgang damit verständlich und transparent abzudecken.
Für alle weiterverarbeitenden Schritte werden eigene Jupyter Notebooks, welche im GitHub-Repository[10] der Verfasserin abgerufen werden können. Ziel dieser Python-Skripts ist es, für jeweils einen Schritt alle notwendigen Aktionen zu sammeln. Alle weiteren Materialien sind gemeinsam mit den Scans der Textzeugen in Collections des internen Repositoriums PHAIDRA der Universität Wien gesichert.[11]

6.1 Transkription

Brüning hält fest: „Um einer computergestützten Analyse zugänglich zu sein, müssen editorische Informationen zuallererst digital vorliegen" (2022: S. 327). Dieser Schritt geschieht durch die Transkription der eingescannten Textzeugen von Hand. Dabei ist in manchen Fällen keine einfache Trennung vom Arbeitsschritt der Modellierung möglich. Denn gewisse Aspekte wie Streichungen und Einfügungen müssen bereits hier mitbedacht werden, um beim POS-

[10] Zugang unter: https://github.com/johannaunterholzner/DH-Masterarbeit (vgl. dazu Daten & Code)
[11] Zugang zu den Collections ist eingeschränkt möglich, vgl. dazu Daten & Code.

Tagging erfasst zu werden. Folglich wurde die Transkription so vorgenommen, dass auch alle handschriftlichen Revisionen übernommen wurden. Dies hat zur Folge, dass die Transkription weder vollständig dem Original noch dem späteren modellierten Text entspricht. Es handelt sich um eine Zwischenstufe, in der alle Tokens vorkommen, jedoch manche Phrasen doppelt geschrieben werden oder Löschungen neben Einfügungen stehen, noch ohne als solche markiert zu sein. Für die spätere Auszeichnung der Revisionen sind damit schon alle Tokens als ganze Einheiten mit POS-Tags ausgezeichnet und benötigen nur die entsprechenden Tags der Eingriffe, um als solche erkennbar zu werden (vgl. dazu im Detail das Unterkapitel 6.3 zur Modellierung).

Gewisse Aspekte müssen bei der Transkription der Textzeugen vereinheitlicht werden, um die maschinelle Lesbarkeit zu gewährleisten. Anhand des Gedichts *VWDAZB* werden die folgende Angleichungen veranschaulicht:

- Wie Kastberger bemerkt und Mayröcker selbst erwähnt, schreibt sie zu Beginn ohne Rücksicht auf Groß- und Kleinschreibung (vgl. Kastberger 1998b, S. 120). Diese wird erst in einem Stadium beachtet, wenn der Text aus Mayröckers Sicht seine endgültigen Form annimmt. Zu beobachten ist beispielsweise, dass das erste Typoskript von *VWDAZB* beinahe vollständig in Kleinschreibung getippt wurde und bei allen späteren Textzeugen bereits die Groß- und Kleinschreibung angewendet wurde. Um die maschinelle Erkennung der Wortarten zu gewährleisten und die Fassungen vergleichbar zu machen, wird der Text auch in dieser Hinsicht der Rechtschreibung manuell angeglichen.
- Ein Merkmal Mayröckers Schreibens ist die „Eszett"-Schreibweise („sz") vom scharfen S („ß"). Alle Worte mit dieser Schreibweise werden entsprechend der Rechtschreibung

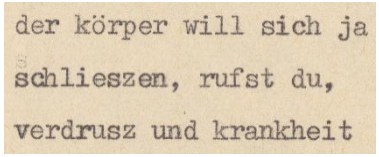
der körper will sich ja
schlieszen, rufst du,
verdrusz und krankheit

Abbildung 2: Ausschnitt aus Textzeuge 1 von *VWDAZB*,
Kleinschreibung und Verwendung von „Eszett".

angeglichen und damit wird „sz" je nach Schreibweise mit „ss" bzw. „ß" ersetzt. Im Gedicht *VWDAZB* (Textzeuge 1) wird so beispielsweise aus „schlieszen" „schließen" und aus „verdrusz" „Verdruss".

- Tipp- und Schreibfehler mit der Schreibmaschine werden angeglichen. So wird beispielsweise aus „Graumen“ „Gaumen“ (Textzeuge 3) und „Pompee“ bzw. „Pompjj“ wird als „Pompeji“ transkribiert (Textzeuge 2 in *VWDAZB*). Zudem gibt es Worte, die mit der Schreibmaschine geschrieben wurden und einen unerklärlichen Abstand zwischen den Buchstaben haben, welcher auf ein mögliches Vertippen zurückgeführt werden könnte. Hier wird der Abstand ignoriert. Zum Beispiel gibt es beim Textzeugen 1 von *VWDAZB* „übe rschwemmungen“, welche als „Überschwemmungen“ transkribiert werden. Einzelne Worte, die sowohl mit der Schreibmaschine geschrieben als auch wieder gestrichen werden, werden als sofortige Eingriffe angesehen und nicht einbezogen.
- Durch einen Bindestrich und über einen Zeilenumbruch getrennte Wörter werden ohne Bindestrich zusammengeschrieben. Weitere Normalisierungen löschen die Zeilenumbrüche schließlich. Ein Beispiel für solche Angleichungen aus dem Gedicht *VWDAZB* sind „zigeuner- spielkarten“ oder „Oktober- keule“, Textzeuge 1. Da es sich

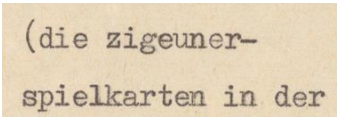

Abbildung 3: Ausschnitt aus Textzeuge 1 von *VWDAZB*, getrennte Substantive werden in der Transkription zusammengeschrieben.

eindeutig um zusammengefügte Nomen handelt, welche auch teils Neologismen sind und allein aufgrund der Entscheidungen über Worttrennungen am Zeilenende getrennt wurden, werden verbunden. Zudem scheint dies ein Aspekt zu sein, den Mayröcker über die verschiedenen Fassungen hinweg verändert. Sie trennt an manchen Stellen bestimmte Worte, um sie in einer weiteren Fassung zusammenzuschreiben. Da es in dieser Arbeit vielmehr um das computergestützte Erkennen von Wortarten geht, welches nur mit klar erkennbaren Nomen funktionieren kann, kann dieses „Erproben“ von Schreib- bzw. Trennweisen von Wörtern nicht mit einbezogen werden. Durch diese Normalisierung lassen sich die Worte passend POS-taggen und vergleichen.
- Handschriftliche Ergänzungen bzw. Eingriffe, die womöglich aufgrund von eiligem Schreiben nicht bis zum Ende ausgeschrieben werden oder abgekürzt werden, werden dem Kasus folgend „richtig“ ausgeschrieben und eingefügt. In Textzeuge 3 von *VWDAZB* wird „u.“ als „und“ transkribiert. Da es sich oftmals um Wörter oder Phrasen handelt, die bereits im Typoskript stehen, aber an eine andere Stelle verrückt werden,

ist damit bei den handschriftlichen Revisionen auch anzunehmen, dass es sich um denselben Wortlaut handelt. Zum Beispiel steht in Textzeuge 1 „dein B.“, welches in diesem Kontext als „deiner Beine“ transkribiert wird. Im Gegensatz dazu wird beim Textzeugen 1 die Einfügung: „verdruß + Krankheit“ zu „Verdruss + Krankheit“, da „+“ beim POS-Tagging als Konjunktor erkannt wird.

- Handschriftliche Eingriffe, die in Stenographie erfolgten, werden decodiert und in Langschrift wiedergeben, sofern über Recherchen im *Stenografie Wörterbuch* (vgl. Wawrczeck (2024)) und Vergleiche mit nachfolgenden Textzeugen eindeutige

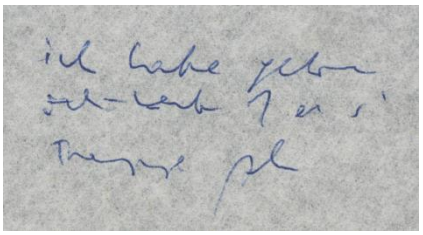

Abbildung 4: Ausschnitt aus Textzeuge 6 von *NWTRG*, Stenographie und Langschrift werden kombiniert in der Passage „ich habe geträumt ich habe mit dir auf der Treppe geschlafen“.

stenographische Zeichen und die entsprechenden Wörter in Langschrift bestimmt werden können. Ansonsten werden sie erst in der Modellierungsphase mit dem Tag <unclear> ausgezeichnet.

- Der Fokus liegt auf dem Gedicht, welches mit dem Titel beginnt und mit dem letzten Wort der Verse endet. Zusätzlich Niedergeschriebenes wie das Datum, der Name der Autorin (beispielsweise bei Textzeuge 5 von *RSFGMT*) oder Widmungen werden nicht mit einbezogen.

Das Transkribieren geschieht umgehend in einem XML-Dokument, um damit weiter arbeiten zu können. Die grundsätzliche Struktur in diesem XML-Dokument versucht minimalistisch aber präzise im <text>-Tag zwischen den einzelnen Textzeugen eines Gedichts zu unterscheiden. Die Hierarchie sieht es vor, zuerst ein Containertag mit den Attributen @type=“poem“ sowie einer ID @xml:id mit dem Gedichttitel als Wert zu erstellen (zum Beispiel: . Darauf folgt ein weiterer Containertag , welches über das Attribut @type und den Werten „witness“ oder „published“ bestimmt, ob es sich um einen Textzeugen handelt oder die publizierte Version des Gedichts. Die zusätzliche Vergabe von IDs @xml:id=“nr.1“ ermöglicht es, die einzelnen Zeugen individuell zu beschreiben, beispielsweise zu nummerieren und für später ansprechbar zu machen. In Hinblick auf

strukturelle Tags folgt innerhalb einer Linegroup <lg> eine Line <l>, in welcher zunächst der gesamte transkribierte Text eingefügt wird. Diese Hierarchie lehnt sich an die TEI-Guidelines für Verse an (vgl. TEI Consortium 2023, S. 231–246).
Nach dem Transkribieren aller Textzeugen wird zuletzt das publizierte Gedicht am Ende des XML-Dokuments eingefügt. Hierfür wird jene Version herangezogen, die im Band *Friederike Mayröcker. Gesammelte Gedichte 1939 – 2003* von Beyer (2004) herausgegeben wurde.

6.2 POS-Tagging

Das Tagging von Parts of Speech (POS, d.h. morphosyntaktische Wortklasse) erfolgt in Python mithilfe der Library *SpaCy* erfolgen. Dafür wird anhand der Python-Library *lxml* mit XPath auf das grundlegend codierte XML-Dokument und dessen Text in den Lines zugegriffen. Der Text wird tokenisiert und im nächsten Schritt mit Parts of Speech-Tags versehen, um schließlich in einem neuen XML-Dokument mit der ursprünglichen Struktur gespeichert zu werden. Zusätzlich wird ein CSV-Dokument mit denselben Informationen in tabellarischer Form gespeichert, um die Qualitätsprüfung durchführen zu können.
Schöch hält in Hinblick auf die Modellierungsmöglichkeiten linguistischer Aspekte von Texten fest, dass diese „zwar grundsätzlich möglich sind, aber insbesondere in Kombination mit textkritischen Auszeichnungen zu einer gewissen Unübersichtlichkeit führen und den Prozessierungsaufwand erhöhen können.“ (Schöch 2016, S. 339). Die TEI-Guidelines verfügen über ein eigenes Kapitel (17.4 Linguistic Annotation), in dem beispielsweise das Wort-Tag <w> mit dem Attribut @pos und den Wortarten als Attributwerte als Auszeichnungsmöglichkeit festgehalten werden (vgl. TEI Consortium 2023, S. 622-628). Ist jedoch für jedes Token ein Tag inklusive Attribut und Attributwert notwendig, wird es in der Tat schwer, weitere komplexe Eingriffe zu annotieren, ohne dabei den Überblick zu verlieren.
Aus diesem Grund werden für diese Arbeit die Tags nicht als spitze Klammern mit Attributen und Attributwerten geführt. Stattdessen werden die Tags in Form eines Suffixes an jedes Token angehängt, wie es einer ‚einfacheren‘ Art der Corpus-Annotation entspricht (vgl. Leech 2004). Dabei wird zwischen Token und Tag die visuelle Trennung bzw. Verbindung „_/“ eingefügt: f“{token.text}_/{token.pos_}“ bzw. konkret beispielsweise „Mond_/NOUN“. Über diese Zeichenverbindung, die in dieser Form nicht in einem schriftsprachlichen Text existiert, ist es möglich, die Tokens zu einem späteren Zeitpunkt wieder von den Tags zu trennen (vgl. Leech

2004). Beispielsweise können so die Tokens und Tags nach der Extrahierung der Textstufen in separaten Spalten einer Tabelle gespeichert werden. Die Tags selbst entsprechen der *SpaCy*-Library folgend den Universal POS-Tags (vgl. Universal POS tags).

Anschließend werden die Ergebnisse des POS-Taggings anhand einer Stichprobe überprüft, denn „linguistic annotation cannot be done accurately and automatically" (Leech 2004). Mit einer Evaluierung kann ermessen werden, in welchem Maße das Tagging erfolgreich funktioniert und in diesem Kontext insbesondere für lyrische Texte anwendbar und sinnvoll ist. Die Stichprobe ergibt sich aus 15 Wörtern pro Textzeuge, wodurch die Stichprobenanzahl je nach Anzahl der Textzeugen etwas schwanken kann. Das eigenständige POS-Tagging geschieht in einer Excel-Tabelle. Es werden zwei unterschiedliche Startpunkte in der ersten Gedichtversion gewählt, die dem Beginn des Gedichts und etwa der Mitte des Gedichts entsprechen. Von diesen Startpunkten aus werden jeweils sechs, sieben, acht oder neun Worte[12] hintereinander mit POS-Tags versehen, damit schließlich immer 15 Worte (6+9 oder 7+8) pro Version ausgezeichnet sind. Die Ergebnisse werden als CSV-Dokumente gespeichert und anhand eines Python-Scripts in eine Confusion-Matrix eingelesen, um mit der Library Scikit-Learn Accuracy, Precision und Recall zu berechnen.

Dabei gibt die Accuracy in Prozent an, wie vielen Tokens der richtigen Wortart zugeordnet werden (vgl. Leech 2004). Precision beachtet pro Wortart all jene Tokens, die im Kontext aller Tokens mit diesem Wortartentag getaggt wurden. Hierbei werden auch die fälschlicherweise mit diesem Tag versehenen Tokens in Betracht gezogen, während die unerkannten Tokens dieser Wortart nicht berücksichtig werden. Recall hingegen bezieht im Zusammenhang aller Tokens neben den richtig ausgezeichneten Tokens auch jene hinzu, welche fälschlicherweise einer anderen Wortart zugewiesen wurden. (Vgl. Klinke 2017, 269–270)

Für das erste aufbereitete Gedicht *VWDAZB* wurde ein Vergleich der Performance drei verschiedener Sprachmodelle vollzogen, welche alle in der *SpaCy*-Library abrufbar sind: „de_core_news_lg" und „de_dep_news_trf" von *SpaCy* selbst und „de-hdt" von *UDPipe*. In einem Vergleich der Sprachmodelle von *SpaCy* und *UDPipe* auf Basis der 120 Worte zeigt sich, dass das Modell „de_core_new_trf" von *SpaCy* am besten funktioniert mit 93,4% Accuracy, dicht gefolgt vom Modell „decore_news_lg" mit 92,5% Accurarcy. Das Modell von

[12] Es wurde hier bewusst die Entscheidung getroffen, nur Wörter in die Qualitätsüberprüfung einzubeziehen und die Satzzeichen in diesem Schritt auszulassen, da deren Erkennung als solche gewährleistet ist.

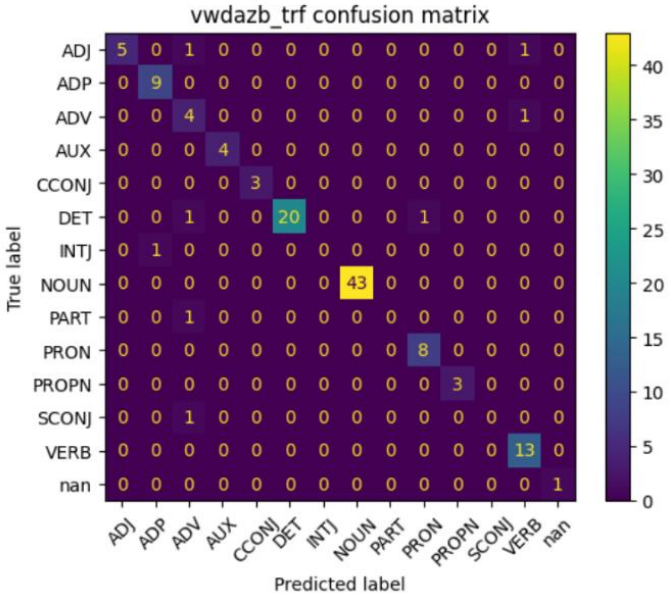

Abbildung 5: Confusion Matrix des Transformer-Modells von *SpaCy* zu den identifizierten Tokens aus dem Gedicht *VWDAZB*. Zugriff unter vwdazb/2_Confusion matrix.ipynb im GitHub-Repository, siehe Daten & Code.

UDPipe schnitt mit 83,33% im Vergleich am schlechtesten ab. Demzufolge werden alle weiteren POS-Taggings mit dem Transformermodell von *SpaCy* vollzogen.

Die Qualitätsprüfungen der weiteren vier Gedichte weisen jeweils eine Accuracy von über 91% auf. Die Gedichte *APTRM* und *WSNMP* erreichen 92% bzw. 92,2%. *NWTRG* befindet sich mit 91,85% am unteren Ende, während *RSFGMT* mit 94,44% die höchste Accuracy zeigt. Werden die Stichproben aller Gedichte kombiniert, liegt die Accuracy bei 92,76%. Da durch die zufällige Auswahl der Tokens keine gleich großen Wortartenklassen entstehen, können Precision und Recall lediglich in Hinblick auf die einzelne Wortart zufriedenstellende Ergebnisse liefern. Insgesamt zeichnet sich hier der Trend ab, dass Verben, Artikel, Präpositionen, Junktoren und Substantive am besten identifiziert werden können (Precision und Recall mit je über 95%). Bei Adverbien, Adjektiven und Eigennamen entstehen teils falsche Zuordnungen (Precision und Recall teils nur etwas über 50%). Letztlich scheinen die hohen Werte jedoch insgesamt dafür zu stehen, auch Lyrik mit dem Transformermodell von *SpaCy* einem POS-Tagging unterziehen zu können.

6.3 Modellierung

Die Modellierung der Textdaten hat im Sinne der Forschungsfrage in XML nach den TEI-Guidelines zu erfolgen. Schöch bezeichnet dies als die „am Besten geeignete Repräsentationsform“, sowohl „für das Gebiet der digitalen Textedition, mithin für den Bereich der Textkonstitution selbst“ als auch „für das Gebiet der Textanalyse und Interpretation“ (2016, S. 327).

In Hinblick auf alle textinternen Varianten betont Brüning (2022) die Relevanz, diese maschinell lesbar auszuzeichnen (vgl. Brüning 2022, S. 325). Hierfür unterscheidet er zwischen zwei verschiedenen Zugängen, dem ‚Prinzip Edition‘ und dem ‚Prinzip Transkription‘. Im Grunde plädiert er für das ‚Prinzip Edition‘, welches im Gegensatz zum ‚Prinzip Transkription‘ nicht an einer diplomatischen Transkription festhält. Denn dieses wendet Tags auch „unterhalb der Token-Ebene“, sprich auf Buchstaben-Ebene an, wodurch „automatisierbare Prozesse der Analyse (angefangen mit der Tokenisierung) und der davon abhängigen Anreicherung (z. B. Lemmatisierung, POS Tagging)“ (Brüning 2022, S. 331) nur eingeschränkt möglich sind. Zudem lassen sich damit die bereits erwähnten, textinternen Varianten nicht sinnvoll aus dem Text extrahieren.

Als Beispiel eignet sich ein minimaler Eingriff von Mayröcker im Textzeuge 2 von *VWDAZB*, in dem „deiner Beine“ zu „seiner Beine“ wird, indem über dem „d“ ein kleines „s“ steht. Würde nach dem ‚Prinzip Transkription‘ modelliert werden, bleibt bei der Modellierung der Fokus auf

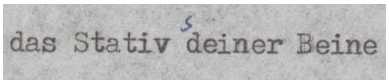

Abbildung 6: Ausschnitt aus Textzeuge 2 von *VWDAZB*, handschriftliche Revision der Passage „das Stativ deiner Beine“.

dem Buchstaben, in diesem Fall wird lediglich das „s“ mit dem „d“ ersetzt und die Worteinheit an sich wird aufgelöst: `<add>s</add><del>d</del>einer Beine`. Nach dem ‚Prinzip Edition‘ sollte die Einheit beibehalten werden, wodurch derselbe Modellierungsvorgang sowohl von Menschen als auch später maschinell besser gelesen werden kann.

Insofern wird sich die Modellierung in dieser Arbeit an dem ‚Prinzip Edition‘ anlehnen und dabei „die genaue Wiedergabe des Befunds *und* die Eignung der Daten für computergestützte Analysen“ (Brüning 2022, S. 329–331) gewährleisten. Aus diesem Grund wurde im Unterkapitel 6.1 zur Transkription festgehalten, dass die Tokens von Streichungen und Einfügungen als vollständige Einheiten bereits in die Transkription übernommen werden.

Die verwendeten Tags fokussieren sich auf eine minimale Auswahl, die sich in den Kapiteln „3.5 Simple Editorial Changes“ (vgl. TEI Consortium 2023, S. 95-103) und „11.7 Identifying Changes and Revisions“ (vgl. TEI Consortium 2023, S. 433-437) finden lassen. Dazu gehören die Tags <add> für Einfügungen, <del> für Streichungen und <choice> für komplexere Eingriffe in dem Text. Innerhalb des Tags <choice> ist es möglich, zwei verschiedene Tagpaare zu verwenden. Zum einen <sic> für eine falsche Textversion und <corr> für den korrigierten Text. Zum anderen steht <orig> für die „originale“ Fassung eines Texts und <reg> für einen Text, „which has been regularized or normalized in some sense“ (TEI Consortium 2023, S. 98). Da es sich bei Mayröckers Eingriffen womöglich eher nicht um Korrektur einer *fehlerhaften* Passage handelt, sondern um eine wertfreie Revision, wird das letztere Tagpaar aus <orig> und <reg> für all jene Phänomene herangezogen, in denen beispielsweise eine Änderung der Reihenfolge teils inklusive Streichungen oder Einfügungen erfolgt.

Um auf die unterschiedlichen farblichen Eingriffe in einem Text einzugehen, wird das Attribut @change herangezogen, welches bei jedem „Eingriffs“-Tag zur Anwendung kommt. Als Attributwerte dienen die augenscheinlichen Farben des verwendeten Stifts, beispielsweise @change=“#red_pen“. Im teiHeader wird innerhalb der <profileDesc> eine eigene <listChange> angelegt, in der die einzelnen Farben als <change resp=“#red_pen“> inklusive weiterer Spezifikationen oder Erläuterungen festgehalten werden. Bei der Extrahierung der Textstufen wird speziell auf das @change Attribut geachtet, um bestimmte Tags anzusprechen oder gerade nicht.

Im Zuge der Modellierung müssen Entscheidungen über gewisse Phänomene getroffen werden, um diese einheitlich behandeln zu können. Diese werden an dieser Stelle transparent gemacht und betreffen die folgenden Punkte:

- Eingriffe, die die Reihenfolge der Worte nicht verändern, werden nicht in Betracht gezogen. Denn es gibt Fälle, in denen die Autorin ein Wort streicht, nur um es an das Ende der vorherigen Zeile einzufügen. Es wird damit also lediglich der Zeilenumbruch verändert, aber nicht die Wortfolge. Dies gilt auch für Fälle, in denen ein Wort eingekreist und über einen Pfeil an das vorherige Zeilenende gesetzt wird, ohne gestrichen zu werden. Handelt es sich nicht um eine exakte Übernahme bei der Einfügung des gestrichenen Worts, wird es verzeichnet. Beispielsweise beim Textzeugen 1 von *VWDAZB*, als „deiner Beine“ anstelle von „(deiner Beine)“ ans vorige Zeilenende geschoben wird.

- Streichungen von mehreren Worten, wenn auch einzeln durch Striche ausgeführt, werden mit einem einzelnen <del> ausgezeichnet, wenn alle Wörter davon sichtlich betroffen sind. Im Textzeuge 3 von *VWDAZB* werden die Worte „abgehoben“ und

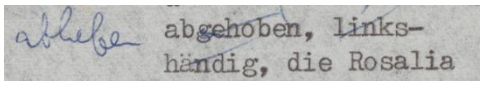

Abbildung 7: Ausschnitt aus Textzeuge 3 von *VWDAZB*, zwei Streichungen nebeneinander werden als eine verzeichnet.

„links-händig“, welche nebeneinanderstehen, einzeln gestrichen, aber mit einem <del> ausgezeichnet, da es mit derselben Stiftfarbe geschieht.

- Es gibt Fälle von Änderungen, die sowohl mithilfe der Tags <add> und <del> ausgezeichnet werden könnten als auch durch die Tags <orig> und <reg> innerhalb eines <choice>-Tags. Hier gilt es eindeutig zu unterscheiden, ob es sich um einen klaren Akt der vollständigen Streichung bzw. Einfügung handelt oder ob ein komplexerer Revisionsvorgang abzulesen ist. Auf <add> oder <del> wird zurückgegriffen, wenn ganze Wörter bzw. Wortgruppen gestrichen und eingefügt werden oder beides an einer Stelle geschieht. In allen anderen Fällen, in denen nicht vollständige Löschungen bzw. Hinzufügungen passieren, sondern nur Teile von Wörtern oder Wortgruppen verändert (und beispielsweise mit weiteren Einfügungen kombiniert) werden, werden die Tags <orig> und <reg> eingesetzt. Wird zum Beispiel ein Wort in seiner Schreibweise adaptiert, in dem das Affix gestrichen und verändert wird, handelt es sich um einen Fall, der mithilfe der Tags <orig> und <reg> innerhalb eines <choice>-Tags beschrieben wird. Um an das Beispiel im Kontext mit dem Prinzip Edition zu Beginn des Unterkapitels anzuschließen (siehe Abbildung 6), würde die Annotation wie folgt aussehen: <choice><reg>seiner</reg><orig>deiner</orig></choice> Beine.
- Eingriffe in die Schreibweise der Worte werden nicht in die Modellierung einbezogen, wenn das Wort lediglich über einen Zeilenumbruch getrennt geschrieben werden soll. Dies gilt aus denselben Gründen, die im Unterkapitel 6.1 zur Transkription bereits dazu genannt wurden. Im Textzeuge 3 von *VWDAZB* wird in das Wort „Kälberleib“ mit einem Trennungsstrich eingegriffen.
- Eingriffe in den Text, die durch Symbole oder Ziffern Einfügungen signalisieren, jedoch ohne ersichtliche Referenz zu einem konkreten Wort oder einer Phrase außerhalb des getippten Fließtexts stehen, werden nicht in die Modellierung einbezogen. Umgekehrt

werden auch Worte und Phrasen am Seitenrand nicht in Betracht gezogen, wenn ihre Position im Text nicht eindeutig markiert ist oder die Passagen selbst nicht zu entziffern sind (allerdings werden sie in diesem Fall mit <unclear> markiert). In Textzeuge 3 von *VWDAZB* gibt es sowohl handschriftlich eingefügte Ziffern („2“) im Text, welche aber bei keinen handschriftlich verfassten Passagen am Seitenrand aufscheinen, als auch die handschriftliche Einfügung „3 zu allererst“ am Seitenende, ohne eine Referenz dazu im Text zu haben.

- Eingriffe, welche beispielsweise an unterschiedlichen Stellen dieselbe Einfügung bewirken würden, jedoch revidiert werden, können nicht in ihrer Vielzahl abgebildet werden, da angenommen werden darf, dass eine solche Einfügung nur einmal erfolgen sollte, jedoch an der Platzierung herumgespielt wurde. Ähnliches geschieht im Textzeuge 5 von *VWDAZB*, wenn die Phrase „mein Gaumen krümmt sich“ über vier Pfeile an unterschiedliche Positionen im Gedicht eingefügt wird. Drei dieser Pfeile werden jedoch im Nachhinein durch dieselbe und eine andere Stiftfarbe gestrichen. Damit wird nur die letzte Entscheidung in Betracht gezogen.
- Der Umgang mit runden Klammern erscheint nicht einheitlich zu sein. Im Vergleich der Textzeugen zeigt sich, dass runde Klammern sowohl als Kennzeichnung für Löschungen genutzt werden als auch für Einfügungen dieser Satzzeichen selbst. Im Textzeuge 2 von *VWDAZB* wird beispielsweise „ist verbrannt“ in Klammern gesetzt und taucht im Textzeuge 3 nicht mehr auf. Ebenso wird „der Dürer“ in Klammern

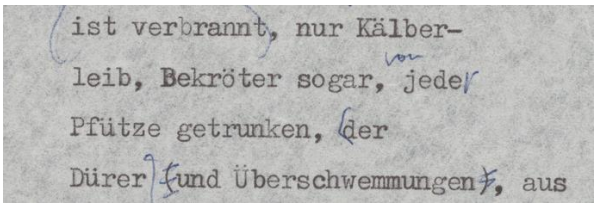

Abbildung 8: Ausschnitt aus Textzeuge 2 von *VWDAZB*, unterschiedlicher Einsatz von runden Klammern.

gesetzt und mit den Klammern in Textzeuge 3 auch fortgeführt. In diesem Fall basiert die Interpretation der Funktion dieser Eingriffe auf den Vergleich zwischen den Textzeugen.

- Eingriffe mit derselben Stiftfarbe (und damit demselben Attributwert von @change) können sich insofern aufheben, als eine Einfügung beispielsweise wieder gestrichen wird. Da die Extrahierung der Korrekturschichten nicht jede Variante innerhalb eines

Revisionsvorgangs mit einer Stiftfarbe berücksichtigen kann und damit in Grenzen gehalten werden muss, kann nur das Endergebnis eines gesamten Revisionsvorgangs mit einer Stiftfarbe einbezogen werden. Ein Beispiel hierfür wäre im Textzeuge 1 von *VWDAZB*, als bei der Einfügung „Brust, auf die Finger mit der linken Hand 3 mal geteilt“ die Wortfolge „auf die Finger“ mit demselben Stift gelöscht wird. Korrekt modelliert würde auch die Streichung verzeichnet werden, jedoch würde dies bei der Extrahierung weit komplexere Logiken erfordern, deren Umsetzbarkeit nicht innerhalb dieser Arbeit geprüft werden können. Dasselbe gilt für Eingriffe, die aus einer Kombination von mehreren Stiftfarben bestehen. Mehrere Beispiele dazu findet man im Textzeuge 3 des Gedichts. Wenn einzelne verschiedene Versionen ausgemacht werden können, werden sie als solche verzeichnet. Ansonsten kann nur die Kombination mit dem Attributwert für die dominantere Stiftfarbe übernommen werden.

- In Hinblick auf Eingriffe mit derselben Stiftfarbe tritt zudem ein zweites Phänomen, beispielsweise bei den Textzeugen zum Gedicht *NWTRG*, auf, bei dem sich die Eingriffe weder aufheben noch eine dominantere Schreibweise ausgemacht werden kann. Dabei stehen zwei unterschiedliche Revisionen zur selben Textstelle in derselben Stiftfarbe nebeneinander. In Textzeuge 1 wird die Phrase „gabelsberger schrift?“ sowohl innerhalb der Zeile mit „gabelsberger handschrift?“ als auch daneben mit

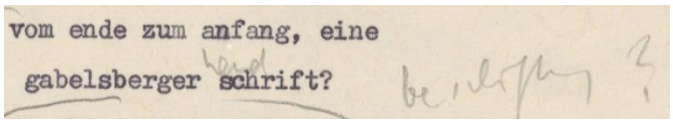

Abbildung 9: Ausschnitt aus Textzeuge 1 von *NWTRG*, zwei verschiedene Revisionen mit demselben Stift erfordern eine Entscheidung, um Redundanzen zu verhindern.

„gabelsberger beschriftung?“ ausgebessert. Die Unterscheidung in zwei verschiedene Textstufen wäre logisch, jedoch müsste hierzu entschieden werden, welche Eingriffe derselben Farbe zu einer Textstufe gehören und welche nicht. Umgekehrt würde der Einbezug beider Varianten innerhalb einer Textstufe die Anzahl der Tokens als auch der POS-Verteilungen verfälschen. Im Fall von *NWTRG* ist in Textzeuge 2 „gabelsberger beschriftgung?“ zu lesen. Um im Sinne der Textdynamik die Vielfalt der Revisionen wiederzugeben, wird jener Eingriff verzeichnet, der (noch) nicht der darauffolgenden Textstufe entspricht. Somit wird „gabelsberger handschrift?“ als Revision verzeichnet. Ähnliche Entscheidungen werden getroffen, wenn mit der

Schreibmaschine mehrere Versionen eines Absatzes hintereinander verfasst werden, wie es bei Textzeuge 3 von *APTRM* der Fall ist.

- Entgegen der typischen diplomatischen Transkription und Modellierung, werden visuelle Hervorhebungen wie Kursivsetzungen, reine Großschreibung von Worten oder Unterstreichungen nicht in Betracht gezogen, da es nicht dem Prinzip Edition entspricht und keinen Mehrwert bei der Fragestellung dieser Arbeit hat.
- Die Satzzeichen und Eingriffe in ihre Platzierungen werden mit derselben Genauigkeit übernommen wie alle anderen Tokens. Denn Mayröcker setzt selbst, wie bereits zitiert, während ihrer Überarbeitungsvorgänge stark die Zeichensetzung ein, um ein Gedicht rhythmisch zu formen.

6.4 Extrahierung der Textstufen

Da sich wie bereits erwähnt in den vorgefundenen Typoskripten unterschiedliche handschriftliche Eingriffe der Autorin finden lassen, gilt es nach Brüning zu unterscheiden zwischen „Textfassungen ohne Varianten“ und „Transkriptionen einzelner Zeugen mit Binnenvarianten infolge handschriftlicher Änderungen“ (Brüning 2022, S. 327). Wie Van Hulle in Hinblick auf Kollationsprogramme festhält, ist es noch nicht möglich, sogenannte „in-text variation[s]“ (Van Hulle 2022, S. 342) gesondert zu behandeln, ohne sich eigens darum zu kümmern. Auch im Fall dieser Arbeit erscheint es sinnvoll, alle möglichen Fassungen innerhalb eines Textzeuges gesondert in Hinblick auf die POS-Verteilung zu betrachten. Insofern muss jede Fassung jedoch zunächst extrahiert werden. Dabei wird festgelegt, dass ein Revisionsvorgang aus allen handschriftlichen Eingriffen in den Text mit derselben Stiftfarbe besteht und damit eine Fassung ausmacht. Zudem setzt sich der ‚originale‘ Text aus dem reinen Typoskript ohne handschriftliche Eingriffe zusammen.

Zunächst wird auf das Extrahieren von Fassungen durch einen Revisionsvorgang eingegangen. Dazu wird das XML-Dokument mit den modellierten Textzeugen in einem Python-Skript mithilfe der Library *lxml* und XPath-Abfragen nach Tags mit einem `@change`-Attribut abgesucht. Für jeden einzigartigen Attributwert (sprich Stiftfarbe) in einem Textzeuge wird der Text mit den entsprechenden Änderungen extrahiert und als separate Fassung im XML-Dokument gespeichert.

In Hinblick auf die Revisionstags gilt es spezifische Regeln zu beachten, die auch in XPath formuliert werden müssen. Die Fassung eines Revisionsvorgangs entspricht zunächst dem zugrundliegenden Typoskript, also dem nicht weiter ausgezeichneten Text des modellierten

```
text = tree.xpath(f"""//t:div[@type='{divtype}' and @xml:id='{divid}']//t:l/text() |
//t:div[@type='{divtype}' and @xml:id='{divid}']//t:l//t:add[@change='{attributevalue}']/text() |
//t:div[@type='{divtype}' and @xml:id='{divid}']//t:l//t:del[not(@change='{attributevalue}')]/text() |
//t:div[@type='{divtype}' and @xml:id='{divid}']//t:l//t:choice/t:reg[@change='{attributevalue}']/text() |
//t:div[@type='{divtype}' and @xml:id='{divid}']//t:choice[(t:reg[not(@change='{attributevalue}')])]/t:orig/text()""",
                  namespaces=ns)
```

Abbildung 10: Screenshot aus dem Python-Skript zur Extrahierung der Revisionsfassungen, hier spezifische XPath-Abfragen für die Fassung durch eine Stiftfarbe bzw. einen Attributwert von @change. Zugriff unter im GitHub-Repository, 3_Extrahieren der Textstufen.ipynb, siehe Daten & Code.

Textzeuges im Line-Tag <l> (siehe 1. Zeile in der Abbildung). Zudem werden die Tags <add> und <reg> mit dem Text, den sie einschließen, einbezogen (siehe Zeile 2 und Zeile 4). Denn jede Einfügung oder Veränderung einer Reihenfolge mit dieser Stiftfarbe gilt als neues, hinzugefügtes Element im ‚originalen' Text. Umgekehrt verhält es sich mit den Texten innerhalb von <del> Tags (Zeile 3), denn diese enthalten Wörter, die im ‚originalen' Text zwar enthalten waren, aber handschriftlich gestrichen wurden. Damit sind sie nicht mehr Teil der Fassung, es sei denn, es sind Streichungen durch andere Stiftfarben. Innerhalb des <choice> Tags muss immer eines der beiden Tags <orig> oder <reg> einbezogen werden. Wie schon erwähnt, wird die regulierte Version eines Textabschnitts dann einbezogen, wenn sie den spezifischen gemeinsamen Attributwert (selbe Stiftfarbe) hat. Ist dies jedoch nicht der Fall und hat das <reg> Tag einen anderen Attributwert als jener, nach dem gerade gesucht wird, muss der Text aus dem <orig> Tag einbezogen werden. Dies wird in der letzten Zeile formuliert als „wenn <reg> nicht den entsprechenden @change-Wert hat, wird der ‚originale' Text einbezogen". Vereinfacht ausgedrückt, werden regulierte Versionen einer anderen Stiftfarbe nicht einbezogen, aber ‚originale' Textstellen innerhalb von Tags schon.

Geht es andererseits darum, die reine Typoskript-Fassung aus dem modellierten Text zu extrahieren, müssen die zuvor ausgeführten Regeln ins Gegenteil umformuliert werden. Das

```
text = tree.xpath(f"""//t:div[@type='{divtype}' and @xml:id='{divid}']//t:l/text() |
//t:div[@type='{divtype}' and @xml:id='{divid}']//t:del/text() |
//t:div[@type='{divtype}' and @xml:id='{divid}']//t:orig/text()""", namespaces=ns)
```

Abbildung 11: Screenshot aus dem Python-Skript zur Extrahierung der Revisionsfassungen, hier spezifische XPath-Abfragen für die Extrahierung der ursprünglichen Typoskript-Fassung.

bedeutet, dass sowohl Einfügungen (<add>) als auch andere komplexere Eingriffe (<reg>) nicht

in Betracht gezogen werden. Stattdessen werden alle ursprünglichen Passagen aus den <orig> Tags einbezogen (letzte Zeile in der Abbildung) sowie alle Streichungen (<del>, siehe 2. Zeile). Basis bildet ebenfalls der nicht weiter annotierte Text (1. Zeile).
Um die extrahierten Fassungen schließlich für die Analyse auswerten zu können, werden die Tokens von ihren POS-Tag-Suffixes getrennt und in ein CSV-Dokument überführt mit zwei separaten Spalten, je eine für die Tokens und die POS-Tags.

7 Analyse

Für die Analyse werden die modellierten und extrahierten Texte in vier Schritten ausgewertet, auf die im Folgenden Bezug genommen wird. Hierzu dient ein Jupyter Notebook, in welchem die Verarbeitung und Visualisierung der Daten mit Python in der nachfolgenden Reihenfolge geschieht. In allen Fällen werden dazu die Ergebnisse der Häufigkeitsverteilungen in absoluten und relativen Zahlen als Diagramme visualisiert.[13] Anhand dessen folgt die deskriptive Analyse der chronologischen Verfassung der Gedichte, wie sie im Kapitel 5 zum Textkorpus bereits erwähnt wurde.

Im ersten Schritt gilt es, den Blick auf die modellierten, handschriftlichen Eingriffe in die Typoskripte zu richten. Über XPath-Abfragen wird auf die einzelnen Tags und deren Häufigkeit in den einzelnen Textzeugen zugegriffen, insbesondere mit dem Fokus auf die verschiedenen Arten von Eingriffen bzw. Bearbeitungen. Damit wird in etwa an das Vorgehen angeknüpft, welches Van Hulle (2022) beschreibt und im Kapitel 2.2 zu den Referenzwerken ausführlicher dargelegt wurde. Dabei ergibt sich ein Einblick in all jene Eingriffe, die Auswirkungen auf die Textstruktur haben und als solche ausgezeichnet werden. Differenziert wird bei den Häufigkeiten zwischen den verwendeten Tags (<add>, <del>, <choice>, <orig>, <reg>), den Attributen der Tags (in diesem Fall @change), den Attributwerten (den Stiftfarben wie #black_pen) und den verschiedenen Kombinationen dieser drei Aspekte (bspw. <add change="#blue_pen">). Dies zeigt auf, in welchen Ausformungen sich die Revisionen in den Typoskripten finden lassen und wie intensiv bestimmte Textzeugen bearbeitet sind. Die Tags <choice>, <orig> und <reg> tauchen immer in Kombination miteinander auf und gelten insgesamt für einen Eingriff. Der Einfachheit halber wird in den Ausführungen der Fokus nur auf die Anzahl und Verteilung der <reg>-Tags gelegt, da diese allein über das Vorkommen komplexerer Eingriffe genügend informieren.

Im Zuge des zweiten Schritts liegt der Blick auf die Wortartenverteilung innerhalb der Textstufen. Wie am Ende des Methodenkapitels 6 erwähnt, werden die Tokens der extrahierten Textzeugen in einer Tabelle gespeichert, um so die Häufigkeiten der Parts of Speech pro Textzeuge hochzurechnen. Dabei sei daran erinnert, dass die Zuordnungen nicht vollständig korrekt sind und besonders für Adverbien, Adjektive und Eigennamen teils Falschzuweisungen

[13] Diese können im jeweiligen Gedichtordner auf dem GitHub-Repository im Jupyter Notebook „5_Auswertung“ eingesehen werden, siehe Daten & Code.

existieren, wie die Qualitätsüberprüfung im Unterkapitel 6.2 zum POS-Tagging ergeben hat. Die Visualisierung stellt alle Ergebnisse nebeneinander, um so einen Vergleich innerhalb der Fassungen mit allen Textstufen als auch über alle Fassungen mit der publizierten Fassung anstellen zu können. Die Revisionsstufen der jeweiligen Textzeugen sind in den Visualisierungen zufällig angeordnet und entsprechen keinem klaren Verlauf, sondern sind als gleichwertig nebeneinander zu betrachten. Die Farbkodierung der POS-Tags ist in allen Visualisierungen zu Vergleichszwecken dieselbe (siehe Abbildung 13). Neben dieser rein deskriptiven Vorgehensweise wurde über den Chi-Quadrat-Test und Cramér's V aus der Inferenzstatistik versucht zu ermitteln, ob es signifikante Unterschiede zwischen den Häufigkeitsverteilungen gibt und wie stark diese Unterschiede einzuordnen sind (vgl. Meindl 2011, 162ff, 234). Wie sich bei allen POS-Verteilungen der Gedichten jedoch herausstellt, ist kein p-Wert des Chi-Quadrat-Tests unter 0,05, um für einen entscheidenden Unterschied zu sprechen. Ebenso verhält es sich mit den Resultaten von Cramér's V, welche alle auf eine mittlere bis geringe Differenz der POS-Häufigkeiten hinweisen. Dies legt nahe, dass die Revisionen mehrheitlich minimale Auswirkungen auf die Gesamtstruktur des jeweiligen Textes haben. Dies bestärkt die deskriptive Analyse, um die Feinheiten der Genese im Kontext des vorliegenden Korpus zu erschließen. Zudem gebührt aus einer literarischen und literaturwissenschaftlichen Perspektive den Minimalverschiebungen wesentliche Aufmerksamkeit. Für alle weiteren Datensätze zu den Eingriffen, auch in Kombination mit den eingeschlossenen POS, wurde ebenfalls der Chi-Quadrat-Test angewandt. In all jenen Fällen, in denen signifikante Unterschiede ausgemacht werden können, wird dem weiter nachgegangen, die Größe der Unterschiede mittels Cramer's V berechnet und die Ergebnisse in der Analyse besprochen.

Im dritten Schritt werden die beiden vorherigen Perspektiven (Revisionen und POS-Verteilung) kombiniert. Hierfür werden nicht nur die Eingriffs-Tags aus dem XML-File mit den modellierten Textzeugen abgerufen, sondern auch der eingeschlossene Text. Pro Tag und Textzeuge werden so die Wortartenverteilungen aufgeschlüsselt, die bei den Revisionen involviert sind.

Zu guter Letzt werden im Skript die Tabellen aus Tokens und Parts of Speech zum jeweiligen Gedicht in Hinblick auf sich wiederholende Sequenzen abgesucht. Dabei wird nach den dominantesten Sequenzen Ausschau gehalten, welche sowohl aus mindestens drei Tokens bestehen als auch mindestens dreimal innerhalb einer Textstufe auftauchen. Gibt es zwei

Sequenzen unterschiedlicher Länge, die exakt dieselbe Häufigkeit aufweisen sowie dieselben Anwendungsfälle beschreiben, wird nur die umfangreichere Sequenz in Betracht gezogen.
Insgesamt könnten diese statistischen Analyseergebnisse bereits erste Rückschlüsse auf Bearbeitungsmuster innerhalb der Gedichtbeispiele zulassen sowie bei auffallenden Gemeinsamkeiten auf allgemeinere Tendenzen hinweisen. Bei der Auswertung werden die formulierten Hypothesen aus dem Kapitel 3 zu Mayröckers Arbeits- und Schreibweise berücksichtigt. Die beispielhaften Zitate richten sich nach der normalisierten Transkription und nicht nach dem (teils in Kleinschreibung verfassten) Original. Kursivsetzungen von Worten werden in den Zitaten vorgenommen, um auf geänderte bzw. neue Aspekte hinzuweisen. Während in den folgenden Unterkapiteln die Ergebnisse zum jeweiligen Gedicht im Zentrum stehen, geschieht eine Zusammenfassung, Diskussion sowie ein Vergleich der Ergebnisse im Kontext aller Gedichte erst im darauffolgenden Kapitel 8.

7.1 *Alp-Traum*

Das Gedicht *Alp-Traum* (12.1.1981; W129/480) weist in allen vier Typoskripten

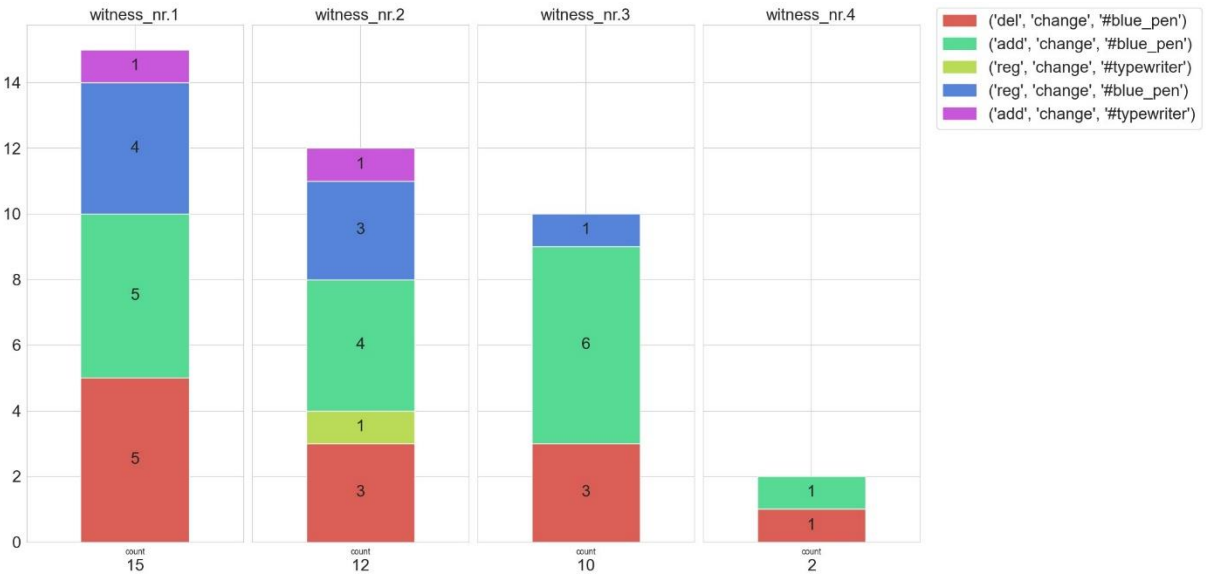

Abbildung 12: Absolute Häufigkeitsverteilungen der Kombinationen aus Eingriffsart und Stiftfarbe für alle vier Textzeugen des Gedichts *APTRM*.

handschriftliche Eingriffe auf. Dabei zeigt sich schrittweise eine Verringerung in der Anzahl der Revisionen pro Textzeuge. Nur wenige Eingriffe werden mit der Schreibmaschine vorgenommen, alle anderen @change-Attributwerte verweisen auf Revisionen mit blauem Stift.

Insofern gibt es wenige Kombinationen von Tag, Attribut und Attributwert. Zu beobachten ist, dass es eine absteigende Zahl von <del>-Tags sowie <reg>-Tags gibt. Dies bedeutet, dass es zu weniger Löschungen und komplexeren Bearbeitungen kommt, je fortgeschrittener der Gedichttext ist. Die Anzahl der <add>-Tags ist zwischendurch bei Textzeuge 2 etwas geringer und hat damit einen Fall mehr als die <reg>-Tags. Bei Textzeuge 3 sind doppelt so viele Fälle von Einfügungen verzeichnet als von Löschungen. In Textzeuge 4 gibt es nur mehr zwei verzeichnete Eingriffe.

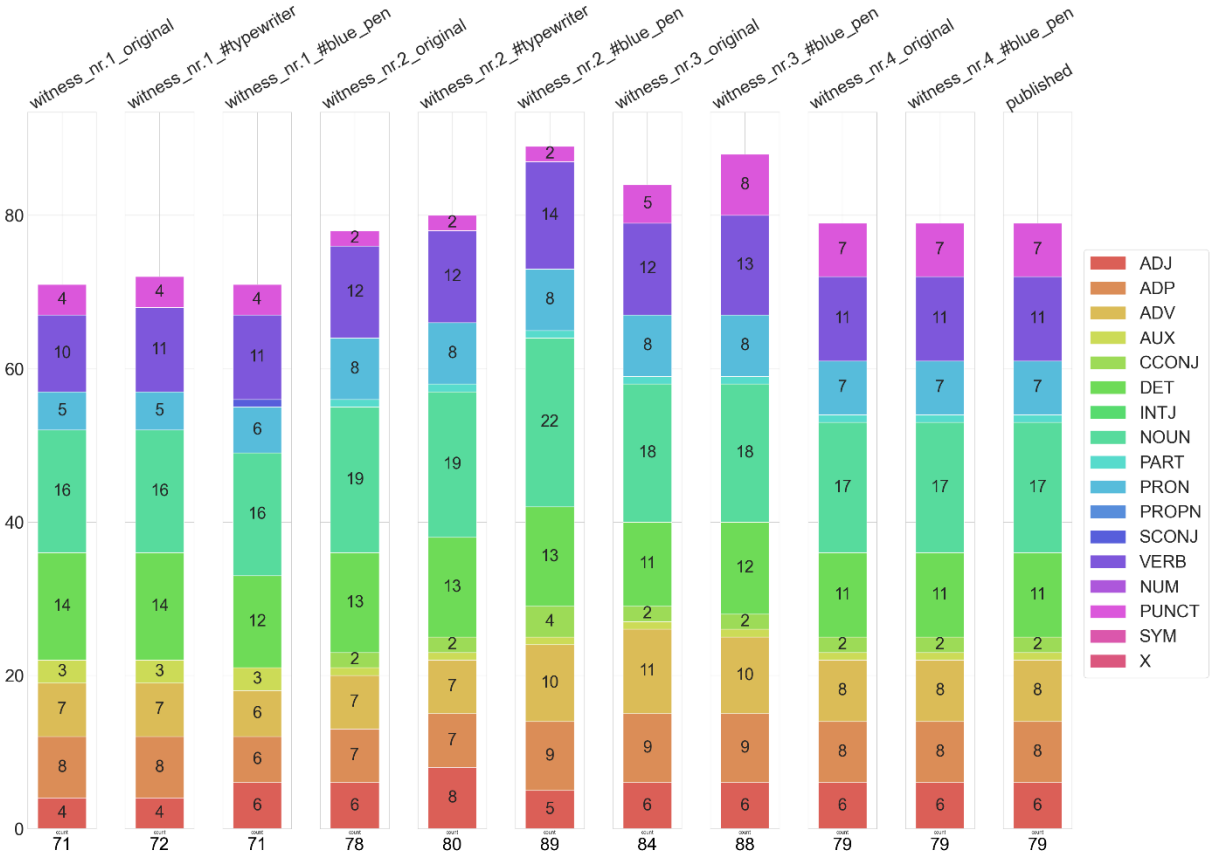

Abbildung 13: Absolute Häufigkeitsverteilung der Parts of Speech im Gedicht *APTRM* über alle Textzeugen und deren Revisionsstufen hinweg bis zur publizierten Fassung.

Jeder Textzeuge hat neben dem Typoskript ein bis zwei weitere Textstufen, die sich aus den maschinengeschriebenen Eingriffen und handschriftlichen Veränderungen eines Stifts ergeben. Interessant ist hier, dass sich die Textstufen innerhalb eines Textzeugen teils nicht markant in der Tokenanzahl unterscheiden. Vergleicht man die Tokenanzahl der Typoskripte miteinander, so zeigt sich, dass beginnend bei 71 Tokens bei Textzeuge 1 ein Anstieg zu 78 Tokens bei Textzeuge 2 und schließlich 84 bei Textzeuge 3 zu erkennen ist, um dann auf 79 ab Textzeuge 4 zu sinken. Durch dieses Ansteigen und Absinken der Tokenanzahl lässt sich zur Hypothese

zurückkehren, dass es zu einem Ausformulieren und Ausprobieren kommt, um schließlich durch eine produktive Abschrift, wie Mayröcker im Kapitel 3 zu ihrer Schreibweise zitiert wird, eine Komprimierung vorzunehmen.

Bei Textzeuge 1 kommt es durch die Revisionen zu insgesamt einem Token mehr. Während durch die Schreibmaschine lediglich ein Verb hinzukommt, zeigen sich in der Verteilung der Wortarten durch die handschriftlichen Eingriffe mehrere Veränderungen. Die Anzahl an Artikeln sinkt, wie auch jene der Präpositionen und Adverbien. Die Zahl der Adjektive, Pronomen und Verben steigt leicht. Die Menge an Nomen und Satzzeichen bleibt gleich. Mit Blick auf Textzeuge 1 wird deutlich, dass insbesondere die zweite Hälfte des Gedichts bearbeitet wird. Zwei umfangreichere Einfügungen als auch Streichungen bewirken die kleinen Verschiebungen. Es ist zu beobachten, dass einzelne Wörter in Hinblick auf ihre Semantik geändert werden, wie „die jungen Fremden“ anstelle der „lauten Fremden“ oder „Dachtraufen“ statt „Dächern“. Solche Revisionen haben keinen Einfluss auf die Wortartenverteilung. Umgekehrt bewirken Eingriffe wie „knirschen“ zu „knirschend“, „Schnee“ zu „verschneit“, „dolchen“ zu „Dolche“ eine Veränderung der Wortart. Mit den Eingriffen wird aber beispielsweise eine Vorstufe der Alliteration in den letzten Zeilen des Gedichts etabliert, auf die an späterer Stelle erneut verwiesen wird.

Von der Revisionsstufe von Textzeuge 1 zum Typoskript von Textzeuge 2 nimmt die Summe an Präpositionen, Adverbien und Artikeln um je eins zu, während zwei Pronomen und drei weitere Nomen verzeichnet sind. Allein zwei Satzzeichen werden in der Fassung von Textzeuge 2 eingesetzt. Es sind semantische Veränderungen in sonst gleichbleibenden Wortgruppen zu beobachten, so wird „das gelbe Schilf nickte am Ufer“ zu „das vergilbte Schilf schwankte am Ufer“. Die alliterative Formulierung „lauthals und lässig“ taucht hier das erste Mal auf. Während der Beginn des Gedichts bereits stark der Struktur der Endfassung gleicht und eine erste Komprimierung erfährt, wird in der zweiten Hälfte an bestimmten Formulierungen intensiv gefeilt. Die Tokenanzahl steigt durch beide Revisionen. Mit der Schreibmaschine werden die Adjektive „kreuzweise“ sowie „weiße“ eingefügt und damit wird nur die Anzahl der Adjektive erhöht. Mit einem blauen Stift fügt Mayröcker mehr hinzu als sie streicht. Im Vergleich zur Gesamtzahl der Tokens in den <orig>-Tags sind den <reg>-Tags doppelt so viele Tokens enthalten. Das bedeutet, dass in den komplexeren Bearbeitungen, unter anderem die Änderung der Reihenfolge von vier Zeilen, auch neue Verben, Nomen, Präpositionen und

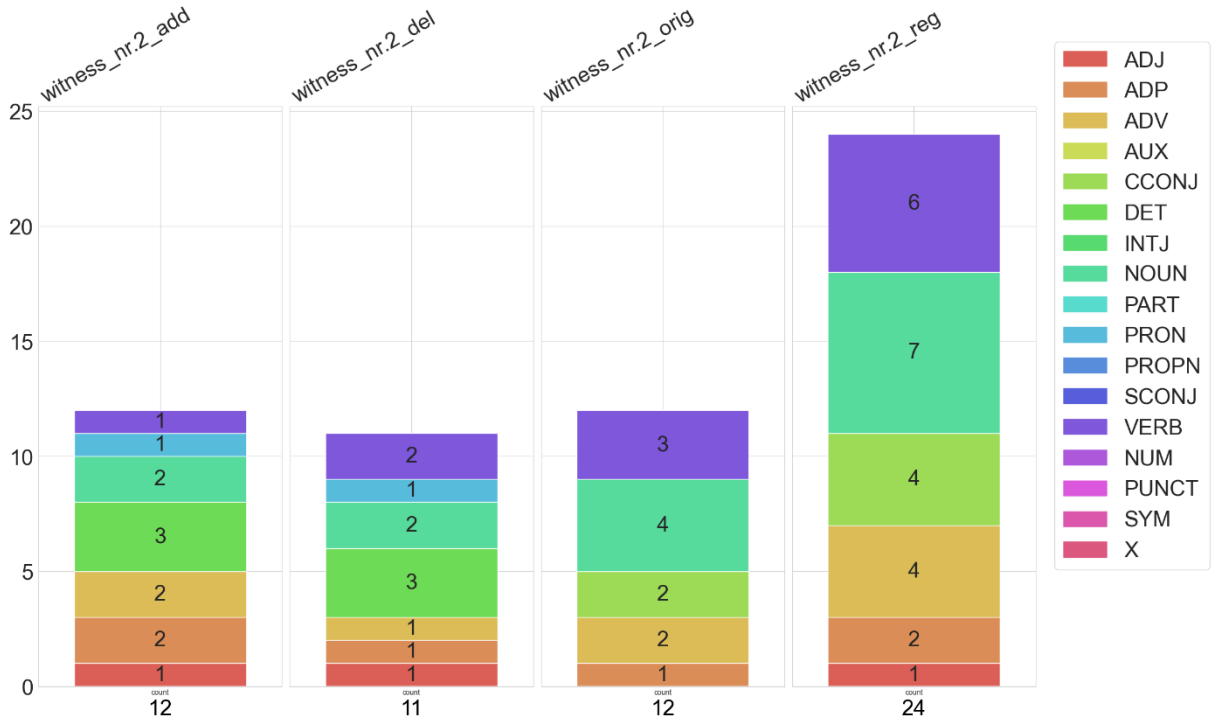

Abbildung 14: Absolute Häufigkeitsverteilung der Parts of Speech innerhalb der Eingriffs-Tags für Textzeuge 2 des Gedichts *APTRM*.

Adverbien hinzukommen. Dagegen sinkt allein die Zahl der Adjektive. Insgesamt enthält diese Textstufe die meisten Tokens aller Textzeugen für dieses Gedicht.

Am Seitenende von Textzeuge 2 stehen verschiedene Versionen zu einem Satz, welcher in abgeänderter Form in der nächsten Fassungen auftaucht. Aufgrund des starken Entwurfcharakters konnte diese Stelle nicht zufriedenstellend annotiert werden. Zu sehen ist hier jedoch, wie Mayröcker verschiedene Verben gegenüberstellt und, wie es scheint, sich zunächst nicht für eines entscheiden kann. Neben dem Satz „ich hielt jetzt auf die Mitte der Straße zu“, sind verstreut auch teils durchgestrichen die Wörter „benutzte“, „wanderte“, „steuerte“, „wechselte“ und „Straßenmitte“ zu finden.

Im Typoskript von Textzeuge 3 werden einige Streichungen durch den blauen Revisionsvorgang von Textzeuge 2 übernommen. Streckenweise sind kaum Veränderungen bzw. neue Formulierungen zu finden. Allein das Verb im ersten Satz wird von „rügte“ zu „melde“, in der Mitte findet sich „Schilfrohr“ anstelle von „Schilf“, welches zudem als „gebüschelt schwankend“ charakterisiert wird. Die größte Neuerung ist der Satz „ich spähe angstvoll nach oben / benutze die Fahrbahnmitte“. Mit dieser Abschrift geschieht der Wechsel vom Präteritum zum Präsens. Insgesamt ist die Tokenanzahl mit 84 niedriger als beim blauen Revisionsvorgang des letzten Textzeugen. Durch die Eingriffe mit dem blauen Stift verändert

sich die Gesamtzahl um vier weitere Tokens. Dabei nimmt die Menge an Artikeln und Verben aufgrund der Ersetzung von „durchs Zimmertelefon“ zu „über das Zimmertelefon“ sowie der Änderung von „schwankend“ zurück zu „schwankt“ um je einen Fall zu. Der Wechsel von „riesige Dolche“ zu „schreckliche Dolche“ sowie die kleine Streichung in der erst etablierten Alliteration „knirschender Kirchenvorplatz“ zu „knirschender Kirchenplatz“ geschieht ohne eine Veränderung bei den Häufigkeiten der Wortarten. Besonders die Zunahme der Satzzeichen ist zu erkennen.

Der Rückgang der Gesamtzahl der Tokens zwischen den Textstufen von Textzeuge 3 zum Typoskript von Textzeuge 4 ist auf das Fehlen jenes Satzes zurückzuführen, welcher im Typoskript von Textzeuge 3 neu aufscheint. Insofern ist weniger von einer Komprimierung als vom Prüfen dieses einen Satzes zu sprechen. Neben geringen Veränderungen in der Wahl der Zeichensetzung wird ohne Veränderung der Wortartenverteilung lediglich der Ausdruck „wie Mondfahrer ausgestattet“ zu „wie Mondfahrer ausgerüstet“ umgeschrieben. Die Revisionen mit dem blauen Stift bei Textzeuge 4 haben keinerlei Einfluss auf die Gesamtzahl der Tokens sowie auf die Wortartenverteilung. Sowohl das Typoskript als auch die bearbeitete Textstufe unterscheiden sich in dieser Hinsicht auch nicht von der publizierten Fassung des Gedichts. Jedoch wird in Textzeuge 4 das Wort „bedrückt“ mit „verdrossen“ ersetzt, was letztlich für die veröffentlichte Gedichtfassung wieder revidiert wird.

Mit Blick auf die Revisions-Tags und den darin eingeschlossenen Tokens über alle Textzeugen hinweg ist auszumachen, dass mehr Tokens durch Mayröckers Eingriffe hinzugefügt werden als gelöscht. Während die <add>- und <del>-Tags in etwa dieselben Wortarten betreffen, ist bemerkenswert, dass dennoch mehr Artikel, Adverbien und Präpositionen gelöscht als eingefügt werden. Umgekehrt ist die Zahl der ergänzten Verben doppelt zu hoch wie jene der gestrichenen. In Hinblick auf die Zeichensetzung sind lediglich Zusätze vermerkt. Die Wortartenverteilungen zu den <orig>- und <reg>-Tags enthalten keine Artikel, dafür Konjunktoren. Komplexere Verarbeitungen scheinen in diesem Fall insgesamt zu mehr Tokens zu führen.

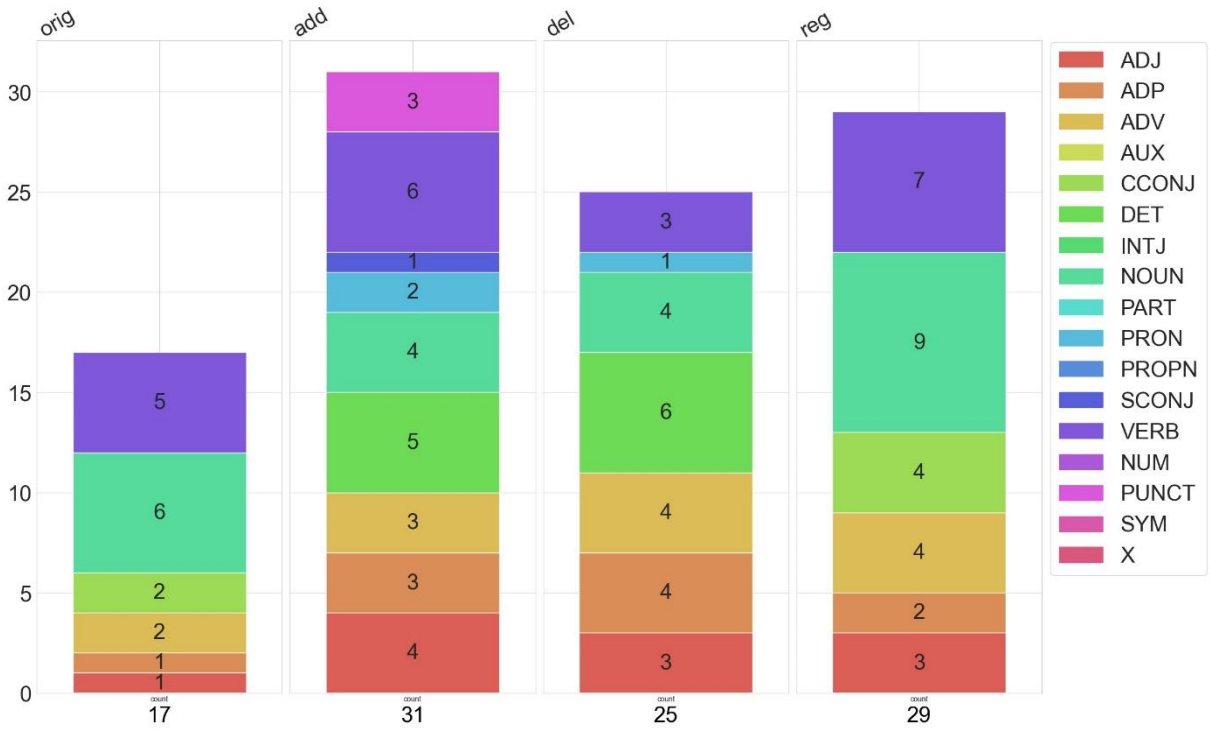

Abbildung 15: Absolute Häufigkeitsverteilung der Parts of Speech in allen Eingriffs-Tags des Gedichts *APTRM*.

Auf besonders signifikante Unterschiede weisen die p-Werte des Chi-Quadrat-Tests beim Vergleich der POS-Verteilungen in den Tags <add> und <reg> (0,0423) sowie <del> und <reg> (0,036). Während sich die POS-Verteilung in den <add>-Tags insbesondere durch die

	del	reg	add	orig
del	0.000000	**0.036313**	0.728317	0.091455
reg	**0.036313**	0.000000	**0.042277**	0.991981
add	0.728317	**0.042277**	0.000000	0.138812
orig	0.091455	0.991981	0.138812	0.000000

Abbildung 16: Matrix für die paarweise ausgeführten Chi-Quadrat-Tests zur Häufigkeitsverteilung der Parts of Speech in allen Eingriffs-Tags des Gedichts *APTRM*.

Satzzeichen und den Subjunktor absetzt, ist beim <reg>-Tag die hohe Zahl an Substantiven auffällig. Die <del>-Tags zeichnen sich stattdessen durch wenige Verben aber ein Pronomen aus. Dabei stehen die Werte von Cramer's V mit etwas über 0,3 für die größere Differenz zwischen den Verteilungen der genannten Tags.

In *Alptraum* sind wenig elliptische Stellen zu finden, da es sich um beinahe vollständige syntaktische Sätze handelt, bei denen nur selten das Subjekt fehlt. Die häufigste Sequenz mit

vier Vorkommen besteht aus „Artikel – Adjektiv – Substantiv – Präposition“. Hierzu gehört beispielsweise „den zugefrorenen See entlang“, da „entlang“ hier in Verbindung mit dem Substantiv im Akkusativ präpositional eingesetzt ist. Diese Formulierung ist bereits in Textzeuge 1 vorhanden. Mit je drei Fällen tauchen die Sequenzen „Artikel – Substantiv – Verb“ und „Artikel – Substantiv – Artikel – Substantiv“ auf. Zu Ersterem gehört beispielsweise „die Rezeption bedauert“ ab Textzeuge 2, zu Letzterem Genitivkonstruktionen wie „das Zimmer des Berghotels“ oder „das Fehlen eines Flaschenöffners“ seit Textzeuge 1. Nicht wenige Sequenzen sind bereits in den ersten Textzeugen angelegt, jedoch kommen die meisten erst mit Textzeuge 2 und 3 vollständig auf.

7.2 *Newtonringe*

Einen umfangreichen und durchwachsenen Bearbeitungsprozess ist beim Gedicht *Newtonringe* (3.3.1981; W129/480) mit acht Typoskripten nachzuverfolgen. Werden die Verteilungen der Revisionen am Beispiel der Vorkommen des @change-Attributs vor Augen gehalten, fallen die teils sehr divergierenden Zahlen auf. Werden in Textzeuge 1 und Textzeuge 2 zehn Eingriffe

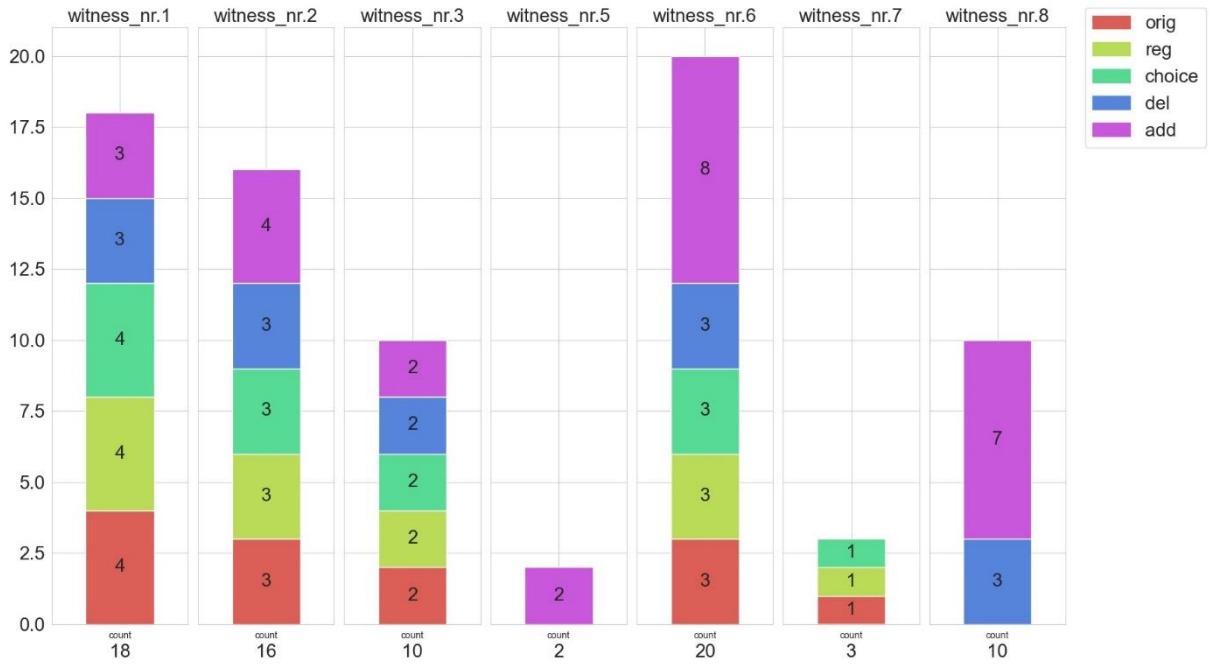

Abbildung 17: Absolute Häufigkeitsverteilungen aller Tags für alle Textzeugen des Gedichts *NWTRG*. Textzeuge 4 weist keine Vorkommen auf.

vorgenommen, sinkt die Anzahl von Revisionen bei Textzeuge 3 auf sechs, um bei Textzeuge 4 keine relevanten aufzuweisen, da es sich um die Änderungen der Zeilenumbrüche handelt.

Textzeuge 5 hat zwei Revisionen, während in Textzeuge 6 ganze vierzehn aufzuzeichnen sind. Textzeuge 7 zählt einen einzelnen handschriftlichen Eingriff, während Textzeuge 8 zehn aufweist.

Im Gegensatz zu den Ergebnissen des vorherigen Gedichts, zeigt sich hier zwar die dominante Stiftfarbe grau, jedoch wurde auch mit einem blauen Stift an den letzten drei Textzeugen gearbeitet, insbesondere jedoch bei Textzeuge 6. Kommen bei Textzeuge 1 je drei Streichungen und Einfügungen vor, gibt es vier komplexere Revisionen. Zu gleich vielen Streichungen kommt es bei Textzeuge 2, wobei es dagegen zu mehr (vier) Einfügungen kommt, während drei Fälle mit <reg> ausgezeichnet sind. Je zwei Fälle pro Tag weist Textzeuge 3 auf. Gibt es bei Textzeuge 4 keine Eingriffe, sind bei Textzeuge 5 nur zwei Einfügungen zu finden. Im Gegensatz dazu ist bei Textzeuge 6 die Anzahl von Einfügungen am höchsten mit acht Fällen, gefolgt von drei <reg>-Tags und drei <del>-Tags. Während Textzeuge 7 lediglich eine komplexere Revision aufweist, zeigen sich bei Textzeuge 8 sieben Einfügungen und drei Löschungen.

Zu ersichtlichen Abweichungen führen die Verteilungen der Eingriffs-Tags von Textzeuge 8 im Vergleich zu Textzeuge 1 und 7 (p-Wert 0.015 und 0,011), womöglich aufgrund der vielen

	witness_nr.1	witness_nr.2	witness_nr.3	witness_nr.5	witness_nr.6	witness_nr.7	witness_nr.8
witness_nr.1	0.000000	0.866878	0.965069	0.186374	**0.000217**	**0.011726**	0.125689
witness_nr.2	0.866878	0.000000	0.965069	0.301194	**0.000217**	**0.011726**	0.231078
witness_nr.3	0.965069	0.965069	0.000000	0.263597	**0.001250**	0.071898	0.254773
witness_nr.5	0.186374	0.301194	0.263597	0.000000	**0.001134**	0.665006	0.753004
witness_nr.6	**0.000217**	**0.000217**	**0.001250**	**0.001134**	0.000000	0.229188	**0.001723**
witness_nr.7	**0.011726**	**0.011726**	0.071898	0.665006	0.229188	0.000000	**0.026564**
witness_nr.8	0.125689	0.231078	0.254773	0.753004	**0.001723**	**0.026564**	0.000000

Abbildung 18: Matrix für die paarweise ausgeführten Chi-Quadrat-Tests zu den Häufigkeitsverteilungen der Kombinationen aus Eingriffsart und Stiftfarbe für alle Textzeugen des Gedichts *NWTRG*.

Einfügungen und dem Fehlen von komplexeren Bearbeitungen bei Textzeuge 8. In der Betrachtung aller Kombinationen von Eingriffstags und Stiftfarben zeigen sich signifikante Unterschiede durch die Chi-Quadrat-Tests fast aller Textzeugen (Ausnahme ist Textzeuge 7) mit Textzeuge 6. Auch hier erweisen sich Textzeuge 7 und Textzeuge 8 im Test als signifikant verschieden. Die Werte des Cramer's V mit zumeist 0,447 weisen zudem auf große Unterschiede.

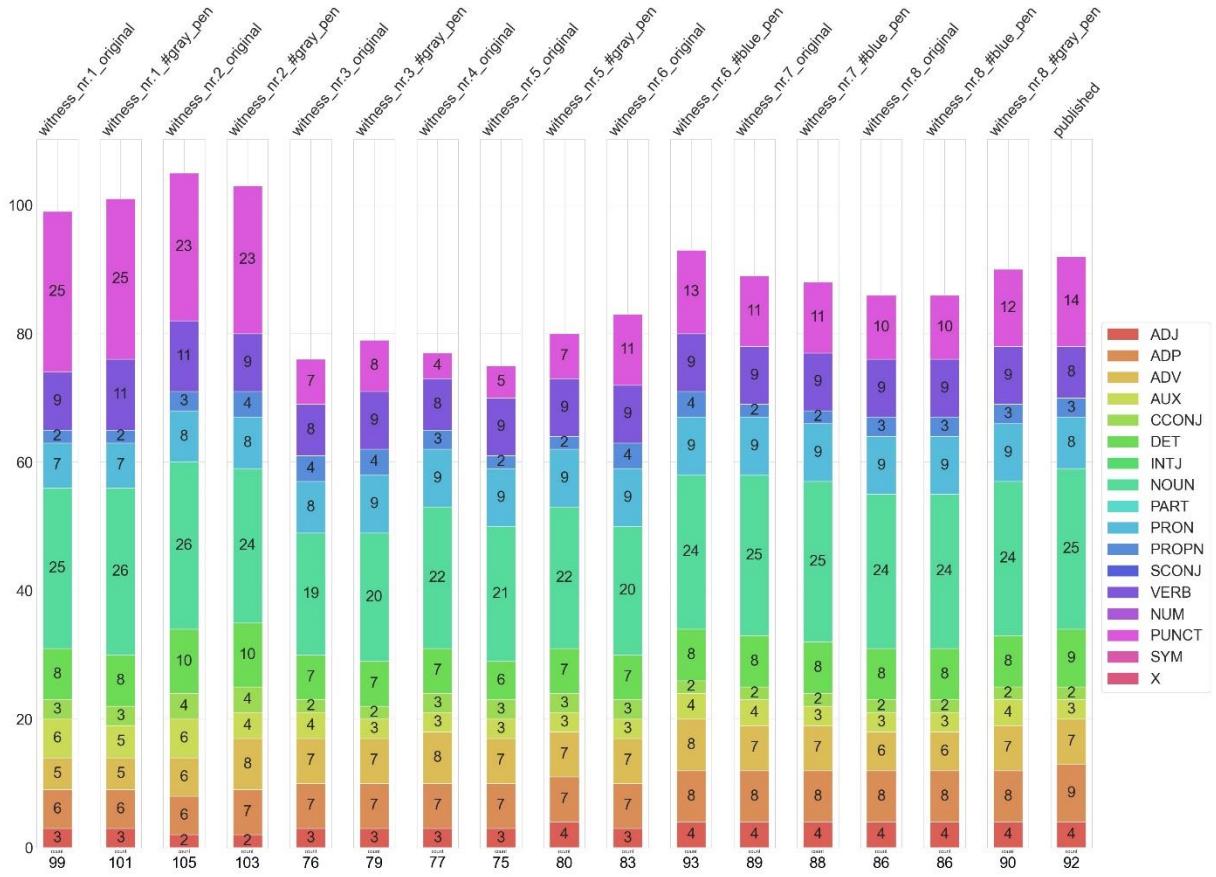

Abbildung 19: Absolute Häufigkeitsverteilung der Parts of Speech im Gedicht *NWTRG* über alle Textzeugen und deren Revisionsstufen hinweg bis zur publizierten Fassung.

Der Vergleich der Gesamtanzahl der Tokens bei den Typoskriptfassungen der Textzeugen zeigt eine wechselnde Bewegung mit teils größeren Sprüngen. Umfasst Textzeuge 1 zunächst 99 Tokens, steigt sie auf 105 bei Textzeuge 2, um schließlich auf 76 bei Textzeuge 3 zu sinken. Textzeuge 4 verzeichnet lediglich ein zusätzliches Token, während bei Textzeuge 5 das Minimum von 75 Tokens erreicht wird. Daraufhin steigt die Anzahl auf 83 Tokens bei Textzeuge 6 und bei Textzeuge 7 auf 89, um bei Textzeuge 8 auf 86 zu fallen. Die publizierte Gedichtversion ist schließlich bei 92 Tokens.

Zwischen dem Typoskript von Textzeuge 1 und der Textstufe durch die handschriftlichen Eingriffe ergibt sich eine Steigerung um 2 Tokens. Insgesamt verschieben sich dabei die POS-Verteilungen am markantesten bei den Verben um zwei zusätzliche. Zudem gibt es ein Substantiv mehr aber ein Hilfsverb weniger. Mit Blick auf den eingescannten Textzeugen ist die umfangreichste Einfügung „Steine gesammelt + weggeworfen“ dominant. Bei insgesamt fünf Fällen ist zwar keine Veränderung der Wortart, aber inhaltlicher Natur zu beobachten. So wird „Lamelle ist abgegangen“ zu „Lamellenheft abgesprungen“, „Gabelsberger Schrift?“ zu

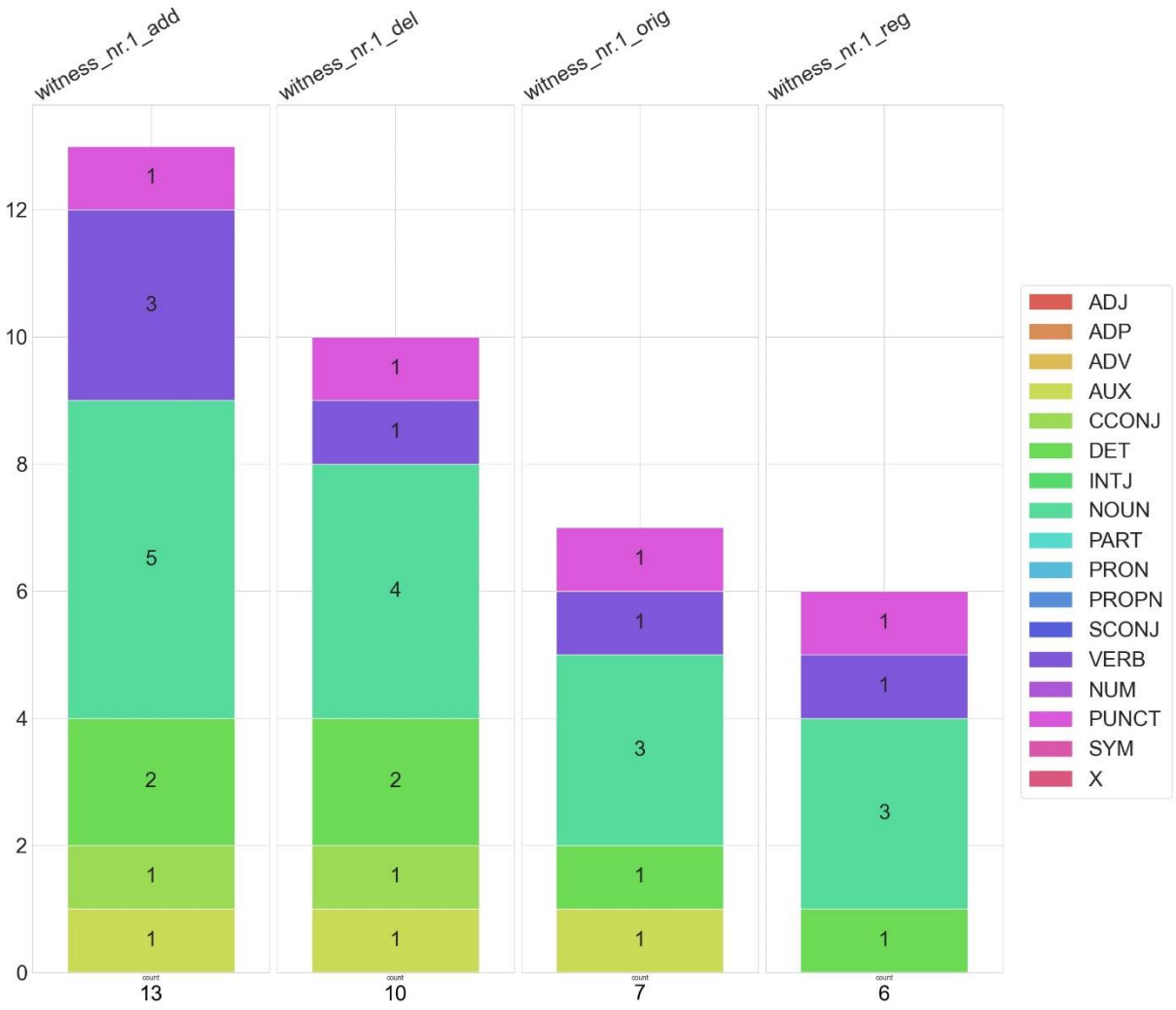

Abbildung 20: Absolute Häufigkeitsverteilung der Parts of Speech innerhalb der Eingriffs-Tags für Textzeuge 1 des Gedichts *NWTRG*.

„Gabelsberger Handschrift?“ und „den kahlen Bäumen“ zu „der kahlen Decke“. Insofern ist es wenig verwunderlich, dass sich bei der Verteilung der eingefügten, gelöschten und veränderten Wortarten in etwa dieselben Parts of Speech finden lassen.

Im Typoskript von Textzeuge 2 bleibt die Anzahl von Präpositionen, Substantiven und Verben gleich. Hingegen vermindert sich die Anzahl von Satzzeichen und Adjektiven. Eine Steigerung ist bei Adverbien, Hilfsverben, Konjunktoren, Artikeln, Pronomen und Eigennamen zu verzeichnen. Abgesehen von stärkeren Veränderungen in Hinblick auf die Reihenfolge mancher Passagen hat sich im Wortlaut kaum etwas verändert. Dies zeigt sich auch beim Vergleich der Wortartenverteilung für die <orig>-Tags und <reg>-Tags, die sich sehr ähneln. Kleinere Einfügungen sind erkennbar. Spezifiziert wird insbesondere der letzte Satz in Textzeuge 2 im Vergleich zur grauen Revisionsstufe von Textzeuge 1: „es ballerte *dann* von

der Decke *des Zimmers*“. Durch die Eingriffe in Textzeuge 2 wird es noch zusätzlich erweitert: „es ballerte dann von der Decke des Zimmers *mächtig auf uns herab*“. Insgesamt führen die Revisionen des grauen Stiftes zu kleineren Verschiebungen. Zunahmen zeigen sich bei Präpositionen, Adverbien und Eigennamen. Dagegen sinken die Zahlen bei Verben, Substantiven und Hilfsverben. Das ist das Resultat der Streichung der Passage „die Steine gesammelt und weggeworfen“ und der Einfügung von „die kahlen Wipfel“. Wieder gibt es Stellen, bei denen die Eingriffe keinerlei Auswirkungen auf die Wortartenverteilung haben, sondern die Inhalte verschieben: „abgesprungen“ zu „zersprungen“ oder „Wipfel“ zu „Wimpel“. Insbesondere bei „Baums“ – „Baumes“ wird auf den Klang geachtet.

Der Sturz der Tokenanzahl bei Textzeuge 3 lässt sich auf das Fehlen umfangreicherer Passagen zurückführen (wie „die Gesetze des Gebens und Nehmens sind aufgehoben“, und „ist es eine Gabelsberger Beschriftung?“) als auch auf eine Komprimierung („Decke des Zimmers“ zu „Zimmerdecke“) und eine radikale Reduktion der Zeichensetzung um sechzehn Fälle. Bei den Substantiven ist eine größere Verringerung um fünf Token zu beobachten. Verben, Konjunktoren, Adverbien und Artikel werden um ein bis maximal drei Tokens weniger, während ein zusätzliches Adjektiv vorkommt. Gleich bleibt die Zahl der Präpositionen, Hilfsverben, Pronomen und Eigennamen. Mit der Übernahme der Revision von Textzeuge 2 wird in Textzeuge 3 der Titel des Gedichts zu „Gabelsberger Universum“ gewechselt. Durch die Revisionen im Typoskript von Textzeuge 3 gibt es nur kleine Verschiebungen. Innerhalb der Tags für komplexere Eingriffe bleiben die Verteilungen sogar gleich. Die Zahl der `<add>`-Tags ist jedoch größer als jene der `<del>`-Tags, je ein Token mehr bei den Satzzeichen, den Verben, den Pronomen und den Nomen, eines weniger bei den Hilfsverben. Neben der Einfügung der Passage „Steine gesammelt + weggeworfen“ hat der Wechsel von „getrennt sein“ zu „sich trennen“ Einfluss auf die Wortartenverteilung und ermöglicht so auch die Etablierung der Anapher „sich umarmen sich trennen“.

Wenngleich in Textzeuge 4 Eingriffe auszumachen sind, so betreffen sie allein die Zeilenumbrüche, nicht aber die Wortfolge. Die Satzzeichen werden weiter zurückgenommen, es gibt minimale Veränderungen im Vergleich zur vorigen Textstufe. Bis zu Textzeuge 6 zeigen sich kleine Verschiebungen in der Verteilung. Neben der Zeichensetzung wird insbesondere im Mittelteil des Gedichts gefeilt, so wird „die Ofenlamellen die kahlen Wipfel“ zu „die zersprungene Ofenlamelle“ von Textzeuge 4 zu 5. In Textzeuge 5 ist insbesondere die Einfügung „eine Gabelsberger Beschriftung“ ausschlaggebend. Zwischen Textzeuge 5 und 6

erhält das Gedicht den Titel vom Schreibbeginn und der Verlust der Wortfolge „die kahlen Wipfel“ ist zu verzeichnen.

Spätestens ab Textzeuge 4 sind gewisse Trends zu beobachten, dass sich die Anzahl gewisser Wortarten nur minimal oder gar nicht über die Fassungen hinweg verändert. Dies betrifft beispielsweise die Menge der Pronomen ab der Revisionsstufe von Textzeuge 3 und jene der Verben ab Textzeuge 5 bis zur Revisionsstufe von Textzeuge 8, um schließlich bei der publizierten Version auf acht zu sinken. Die Revisionsstufe von Textzeuge 6 scheint zudem ein Wendepunkt zu sein, ab welchem sich gewisse Mengen wie jene der Adjektive (vier), Präpositionen (acht), Konjunktoren (zwei) und Artikel (acht) nicht mehr ändern, ungeachtet der Wortartenverteilung des publizierten Gedichts. Diese Eingriffe mit dem blauen Stift in Textzeuge 6 bleiben stellenweise prägend für weitere Textzeugen. Dabei handelt es sich um wesentliche Einfügungen wie „die Pollution“ zu Beginn des Gedichts und „die perlmuttfarbene Frühe im Fenster“ am Ende, wodurch gleich zwei Alliterationen eingeführt werden. Am bestehenden Wortschatz wird zudem gefeilt, ohne sich auf die Wortartenverteilung auszuwirken. So wird der Titel von „Universum“ zu „Newtonringe“ geändert, „eine Gabelsberger Beschriftung“ zu „Gabelsberger Notiz?“, „Wimpel des Baumes“ zu „Leporello des Baumes“ und „Brauenbogen“ zu „Brauenvogel“. Mit 93 Tokens ist es zwischenzeitlich die umfangreichste Revisionsstufe.

Mit dem Typoskript von Textzeuge 7 werden alle Eingriffe übernommen, gleichzeitig wird stärker an der Zeichensetzung gearbeitet. Insbesondere auf das Ende der ersten Strophe fokussiert sich Mayröcker von Textzeuge 6 bis Textzeuge 8. Bei Textzeuge 7 vollzieht sich durch die Revisionen der Wechsel von Perfekt („ich habe geträumt“) zu Präteritum („ich träumte“). Damit wird in der Wortartenverteilung lediglich die Zahl der Hilfsverben kleiner. In Textzeuge 8 wird mit blauem Stift allein das „habe“ zu „hätte“, wodurch sich bei dieser Revisionsstufe vom Typoskript in Hinblick auf die POS-Häufigkeiten nichts tut. Spannend ist der Wechsel von Subjekt und Objekt („ich habe mit dir“ zu „du hast mit mir“) durch den Eingriff mit dem grauen Stift, was jedoch ebenfalls keine weitere Auswirkungen hat. Jedoch bewirkt es die Alliteration „mit mir“.

In der veröffentlichten Gedichtfassung ist schließlich zu erkennen, dass bei dieser Passage noch eine weitere Veränderung stattgefunden haben muss, nämlich steht anstelle von „ich träumte“ nun „von deinem Traum“. Damit ist der Anstieg der Wortartenverteilung in Hinblick auf die Präpositionen und Substantive sowie die Verringerung der Verben zu erklären. Mayröckers

Aussage, sie substantiviere Verben, scheint hier zutreffend. Die Satzzeichen erfahren einen kleineren Ausbau.

Zusammengefasst ist in der Graphik für die Wortartenverteilungen innerhalb der Eingriffs-Tags abzulesen, dass mehr Tokens hinzugefügt werden als gelöscht. Jedoch ist die prozentuelle Aufteilung teilweise sehr ähnlich, beispielsweise bei den Artikeln 9,1% bzw. 9.5%, oder den Substantiven 21,2% und 21,4%. Die absoluten Häufigkeiten in den <orig>-Tags und <reg>-Tags sind teilweise ebenso gleich, so sind je vier Adverbien und sieben Hilfsverben auf beiden Seiten. In manchen Fällen ist der Unterschied minimal, wie bei den Adjektiven, den Konjunktoren, den Verben und Satzzeichen. Um zwei Tokens mehr haben die Artikel und Nomen, starke Unterschiede sind bei den Präpositionen und Pronomen zu erkennen. Insgesamt sind bei komplexeren Änderungen mehr Tokens involviert als bei Streichungen oder

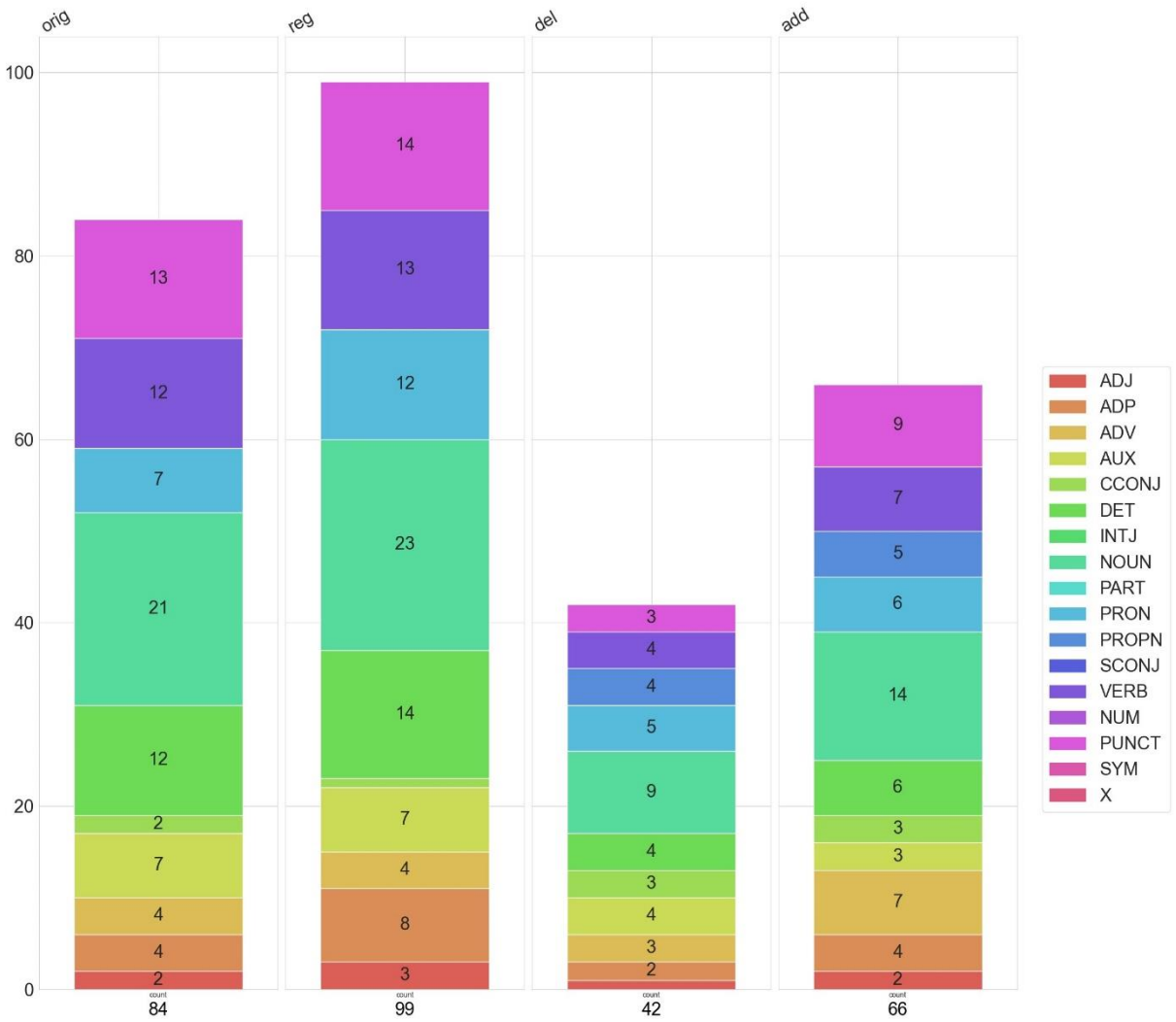

Abbildung 21: Absolute Häufigkeitsverteilung der Parts of Speech in allen Eingriffs-Tags des Gedichts *NWTRG*.

Löschungen. Gleichzeitig geschehen einige Streichungen auch zwischen den Typoskripten und sind damit nicht in den hier visualisierten Daten offensichtlich.

Wenngleich das größtenteils elliptische Gedicht über einige Reihungen („Blendung Verblendung“, „sich umarmen sich trennen“) verfügt und beispielsweise mit wenigen Verben auskommt, gibt es ein geringes Vorkommen an wiederkehrenden Sequenzen. Viele davon etablieren sich erst im Laufe der Revisionen, wenn nicht erst mit der publizierten Fassung. Die drei Sequenzen mit je drei Vorkommen sind „Substantiv – Artikel – Substantiv“ wie „Leporello des Baums“ auf Textzeuge 6 (zuvor schon als „Wipfel“ und „Wimpel des Baums“ präsent), „Präposition – Artikel – Substantiv“ wie „auf der Treppe“ seit Textzeuge 1 oder „von der Zimmerdecke“ ab Textzeuge 3 und zuletzt „Substantiv – Präposition – Substantiv“ wie im Fall von „Frühe im Fenster“ in Textzeuge 6.

7.3 *Rosenfragment*

Im Vergleich zum vorherigen Gedicht gibt es zu *Rosenfragment* (2./3.10.1981; W127/471) fünf Textzeugen, von denen wiederum nur drei von handschriftlichen Revisionen betroffen sind. Dabei sind diese Eingriffe bei Textzeuge 1 und 3 mit dreizehn und sechzehn verzeichneten @change-Attributen intensiver als bei Textzeuge 2 mit lediglich 7 Bearbeitungen. Bei allen bearbeiteten Typoskripten wurde mit mehr als einem Stift vorgegangen, wobei sich für Textzeuge 1 und 2 der blaue Stift als der dominantere erweist, während in Textzeuge 3 primär

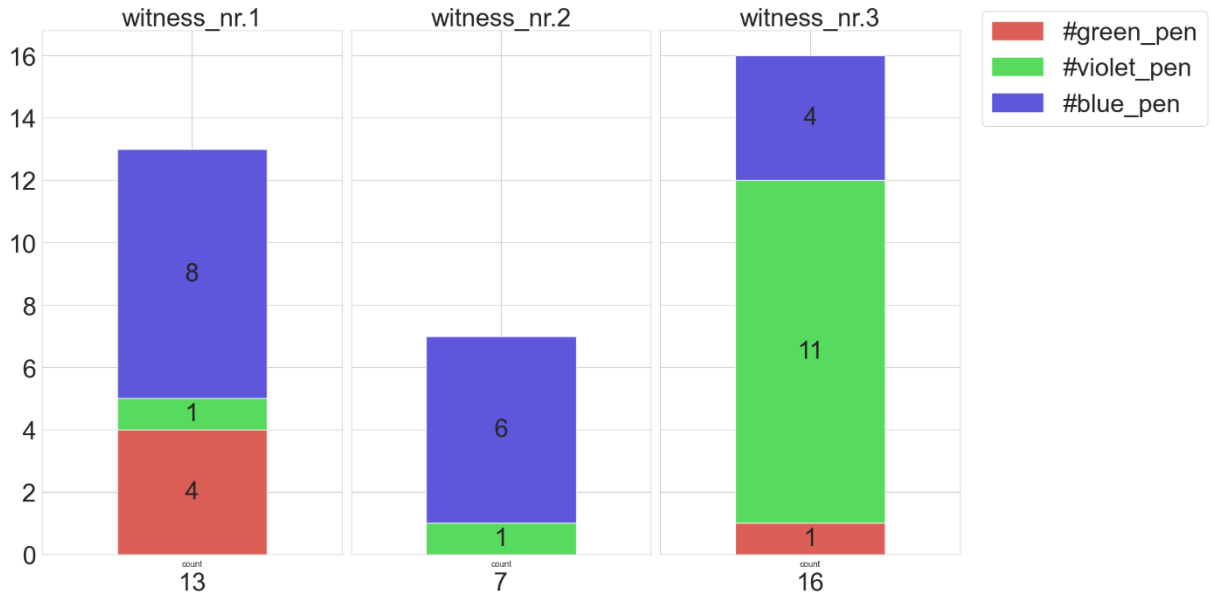

Abbildung 22: Absolute Häufigkeitsverteilung aller Attributwerte (sprich verwendete Stiftfarben) für drei Textzeugen des Gedichts *RSFGMT*.

mit einem violetten Stift gearbeitet wird. Zudem kommt ein grüner Stift in Textzeuge 1 und 3 zum Einsatz. Aufgrund der anderen Nutzung der Stiftfarben bei Textzeuge 3 zeigt sich jeweils ein markanter Unterschied bei den p-Werten der Chi-Quadrat-Tests im Vergleich zu den

	witness_nr.1	witness_nr.2	witness_nr.3
witness_nr.1	0.000000	0.255182	**0.003558**
witness_nr.2	0.255182	0.000000	**0.025544**
witness_nr.3	**0.003558**	**0.025544**	0.000000

Abbildung 23: Matrix für die paarweise ausgeführten Chi-Quadrat-Tests zur Häufigkeitsverteilung aller Attributwerte für die Textzeugen des Gedichts *RSFGMT*.

Verteilungen der Textzeugen 1 und 2 (0,003 und 0,025). Werden alle möglichen Kombinationen von Eingriffs-Tags und Attributwerten (Stiftfarben) einbezogen, zeigt sich neben einem signifikanten p-Wert zwischen Textzeuge 1 und Textzeuge 3 auch eine deutliche Diskrepanz bei einem Cramer's V-Wert von 0,522. In allen drei bearbeiteten Typoskripten gibt es einige komplexere Fälle, in denen <reg> eingesetzt wird.

Zu beobachten ist, dass bei allen Typoskripten mehr Streichungen vollzogen werden als Einfügungen, teils mit beachtlichem Unterschied. Acht <del>-Tags und drei <add>-Tags sind es in Textzeuge 1, während im Vergleich dazu in Textzeuge 2 nur in zwei Fällen etwas gestrichen

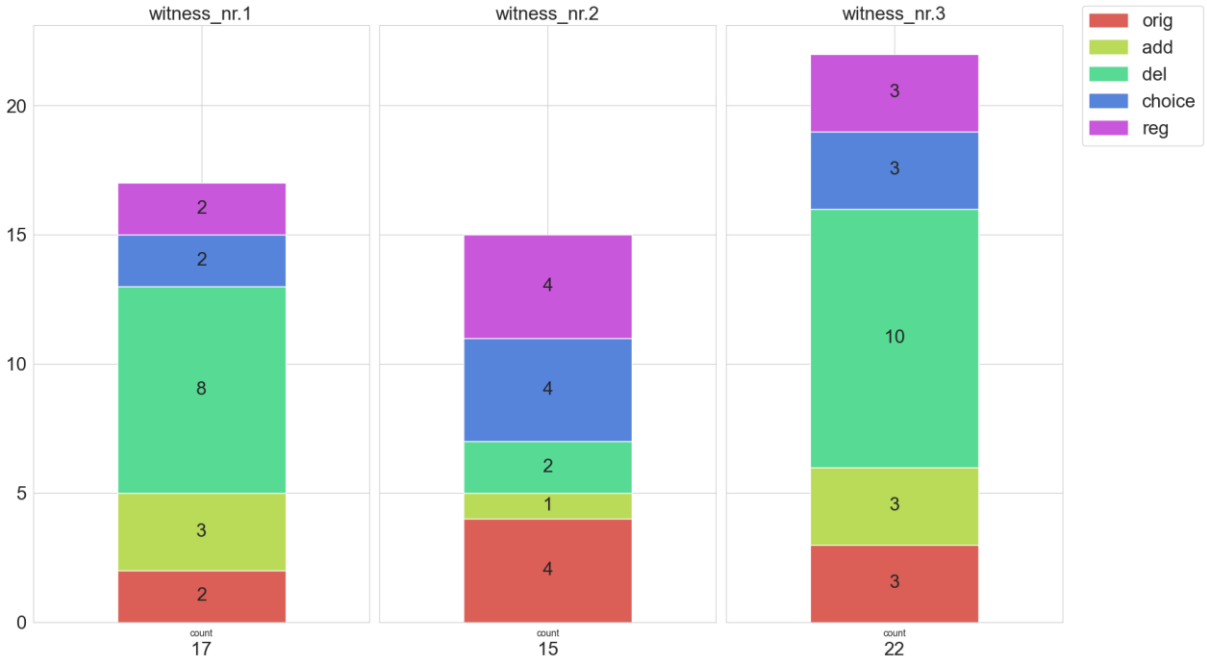

Abbildung 24: Absolute Häufigkeitsverteilungen aller Tags für drei Textzeugen des Gedichts *RSFGMT*. Textzeuge 4 und Textzeuge 5 weisen keine Eingriffe auf.

wird und in einem Fall eine Einfügung verzeichnet ist. In Textzeuge 4 gibt es drei Erweiterungen und zehn Tilgungen.

Da es in Textzeuge 1 Revisionen in drei verschiedenen Farben und in Textzeuge 3 Bearbeitungen in vier verschiedenen Farben gibt, wurden dementsprechend viele Textstufen extrahiert. Insofern finden sich trotz geringerer Anzahl an handschriftlich bearbeiteten Textzeugen dennoch einige Textstufen zum Vergleich der POS-Verteilung. Das Typoskript von Textzeuge 1 unterscheidet sich in der Tokenanzahl und der Häufigkeiten der POS wesentlich von den Textstufen durch die Revisionen mit grünem und violettem Stift. Diese beiden sind sich jedoch in einigen Aspekten sehr ähnlich. Das Typoskript wird von 114 Tokens auf 112 Tokens durch Eingriffe mit einem grünen Stift reduziert, zählt die Revision mit einem violetten Stift zehn zusätzliche Tokens. Während die Anzahl beinah aller Tokens zwischen Typoskript und grüner Revisionsstufe gleichbleibt, ist je ein Token bei den Konjunktoren und den Partikeln zu erkennen. Tatsächlich geschieht die folgenlose Verschiebung der Passage „und funkelnd mit gepolsteter Hand“ sowie die Löschung von „und“ und „zu“. Bei der violetten Revisionsstufe

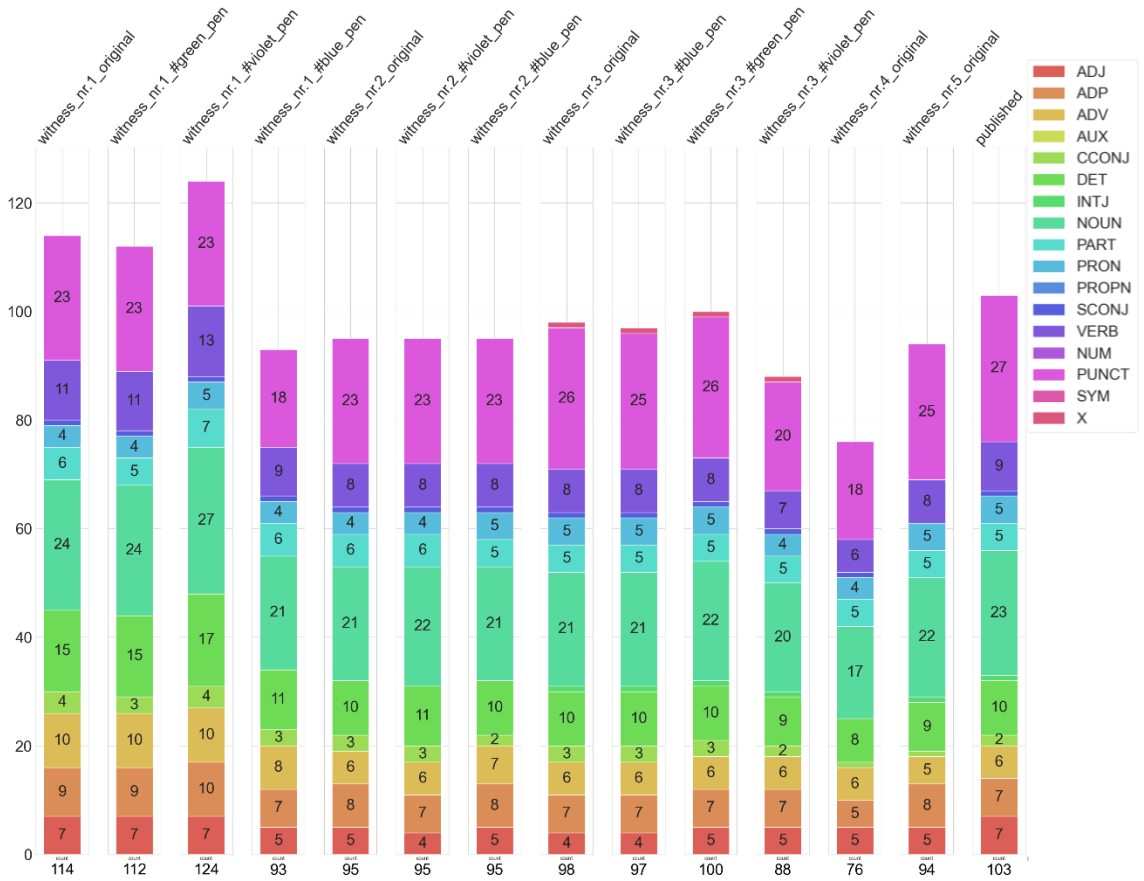

Abbildung 25: Absolute Häufigkeitsverteilung der Parts of Speech im Gedicht *RSFGMT* über alle Textzeugen und deren Revisionsstufen hinweg bis zur publizierten Fassung.

handelt es sich um die umfangreichere Einfügung „Schüttelhaltung wir essen den Mond unter der Schläfe zu liegen“, wodurch sich im Vergleich zum Typoskript die Zahl der Präpositionen, Partikel und Pronomen um je ein Token, der Artikel und Verben um je zwei Token und jene der Substantive um drei Token erhöht. Durch die Revisionen mit dem blauen Stift geschieht eine substanzielle Kürzung auf insgesamt 93 Tokens. Zwischen der Typoskriptfassung und dieser Textstufe sind dadurch alle POS-Kategorien in verringerter Menge präsent, abgesehen von der Menge der Subjunktoren, Partikel und Pronomen.

Nur zwei Tokens mehr verzeichnen alle Textstufen des Textzeugen 2 im Vergleich zur blauen Revisionsstufe von Textzeuge 1. Insofern kann angenommen werden, dass insbesondere die Eingriffe für die neue Abschrift relevant erscheinen. In der Verteilung bleiben dabei beim Typoskript die Adjektive, Konjunktoren, Substantive, Partikel, Pronomen und Subjunktoren gleich. Neu hinzu kommt beim Typoskript des Textzeugen 2 das Substantiv „Fremdflocke“, welches in der grünen Revisionsstufe des vorherigen Textzeugen lediglich am Rand ohne Positionierung zu finden ist und damit nicht mit Gewissheit ausgezeichnet werden konnte. Hiermit entsteht auch die Alliteration „Farbstich, Fremdflocke“.

Die Eingriffe mit dem violetten Stift in das Typoskript fokussieren sich auf eine Stelle im Text, bei welcher aus „mit gepolsteter Hand“ „Innenpolster der Hand“ wird. Folglich steigt die Zahl der Nomen sowie Artikel und sinkt jene der Präpositionen und Adjektive um je 1 Token. Relevant für die minimalen Verschiebungen in der Textstufe durch den blauen Stift ist die Einfügung „rufe ich“ und die Löschungen eines Konjunktors und Partikels. Während zwei Eingriffe („entschleierte“ zu „enthüllte“ sowie „Wolkenschleiern“ zu „Wolkenblättern“) an der Wortartenverteilung nicht rütteln, hat die Umformung des Suffixes von „erkennen“ zu „erkennbar“ durchaus einen Einfluss.

Mit Blick auf die gesamte Tokenanzahl steigt diese von Textzeuge 2 auf Textzeuge 3 nur um drei Tokens, neben leichten Veränderungen nimmt vor allem die Zeichensetzung zu. Über alle drei Textstufen bleibt die Anzahl der Präpositionen, Adverbien, Interjektion, Partikel sowie des Subjunktors gleich. Über die handschriftlichen Eingriffe steigt diese Anzahl auf 100 durch einen grünen Stift, auf 97 Tokens sinkt sie durch einen blauen Stift sowie auf 88 durch einen violetten Stift. Die POS-Verteilung beim blauen Revisionsvorgang ist dieselbe wie beim Typoskript, allein ein Beistrich wird gestrichen. Die weiteren Tags mit dem Attributwert #blue_pen sind für die Änderung des Numerus zuständig, von „die Vliese“ zu „das Vlies“, schlagen sich aber nicht in der Häufigkeitsverteilung nieder. Durch die Eingriffe mit dem

grünen Stift im Vergleich zum Typoskript zeigt sich ein Wachstum der Zahlen von den Adjektiven, Substantiven und Satzzeichen. Dies kann neben dem einflusslosen, aber starken Umbau in der Reihenfolge zu Beginn des Gedichts auf die Einfügung „anarcher Kunstverstand“ zurückgeführt werden. In der Bearbeitung durch den violetten Stift gibt es einen Zuwachs um ein bis zwei Tokens bei den Adjektiven, ansonsten fallen die Tokens der Konjunktoren, Artikel, Substantive, Pronomen, Verben sowie der Satzzeichen. Während der Austausch des Worts „Traumgesicht“ mit „Schattenlocke“ keinen Einfluss hat, kommt es zu Streichungen einzelner Wortfolgen und vielen Beistrichen sowie einer weiteren Wandlung des Ausdrucks „Innenpolster der Hand“ hin zu „gepolsteter Innenhand“. Die Zahl der gestrichenen Satzzeichen ist damit am höchsten, während sich beispielsweise die POS-Verteilung zwischen originaler und regulierter Schreibweise nur minimal in der Zahl der Artikel und Adjektive verändert, da sich hier das letzte Beispiel niederschlägt. Bei allen Textstufen konnte das Teilwort „Lebens-“

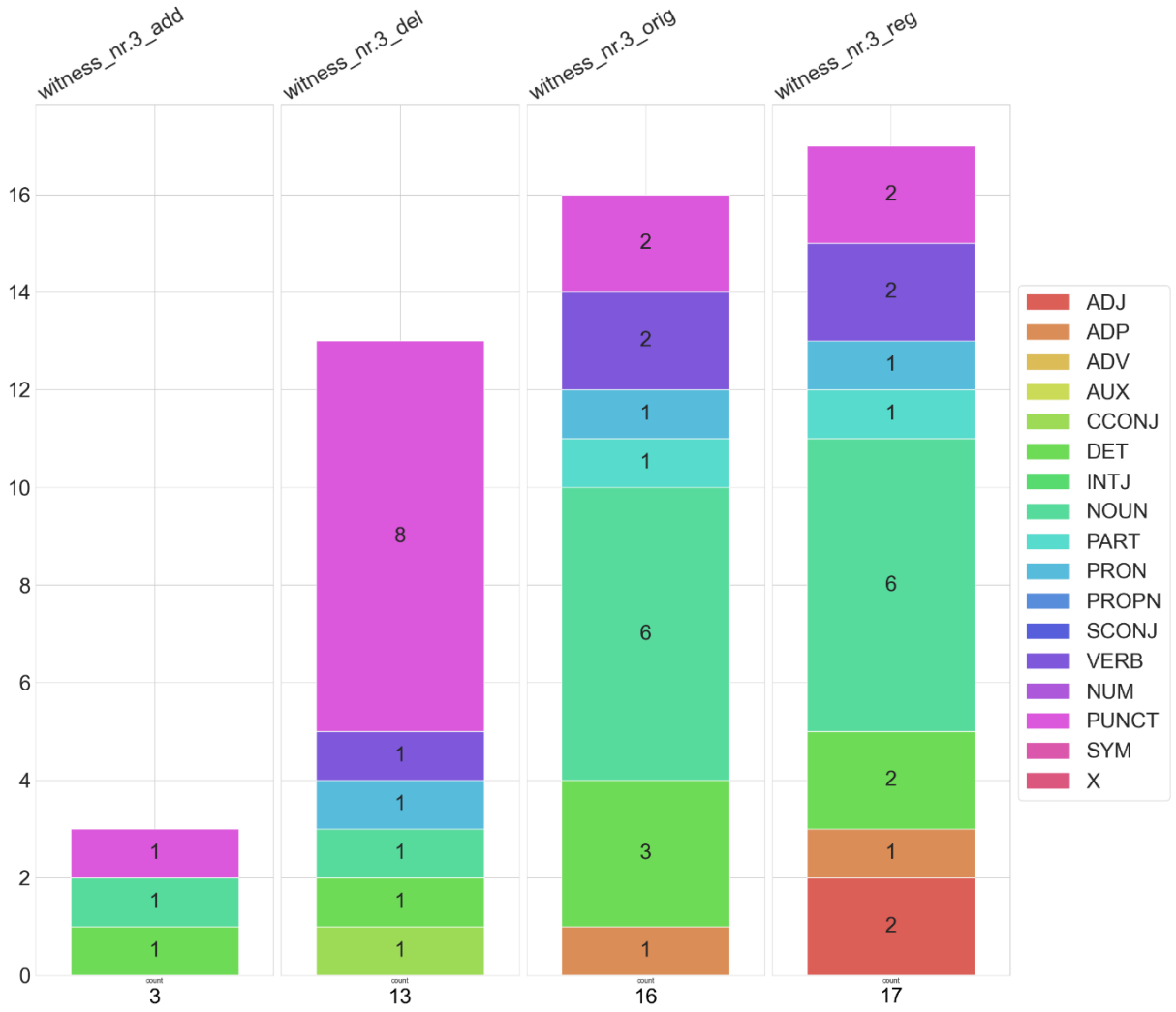

Abbildung 26: Absolute Häufigkeitsverteilung der Parts of Speech innerhalb der Eingriffs-Tags für Textzeuge 3 des Gedichts *RSFGMT*.

nicht von vom POS-Tagger erkannt werden und wird deswegen mit „X“ ausgezeichnet, in allen anderen Fällen hat es das Problem nicht gegeben.

Beim Textzeugen 4 ist eine starke Reduktion zu beobachten, die jedoch daher führt, dass es sich mit großer Gewissheit nicht um ein vollständiges Typoskript handelt. Im Gegensatz zum Textzeugen 5 fehlen in etwa zwei Verse. Dennoch wird sichtbar, dass mit der Abschrift neben den Revisionen von Textzeuge 3 auch weitere Aspekte leicht verändert sind. So heißt es anstelle von „magisches Werk“ in Textzeuge 4 „magischer Schattenlippe“ und ein Artikel sowie ein Konjunktor werden ausgelassen. Ohne handschriftliche Revisionen gibt es in Hinblick auf den vorhandenen Großteil kaum Änderungen, die weiter übernommen werden. Hingegen ähnelt die Fassung von Textzeuge 5 mehr jener von Textzeuge 3. Eine Reduktion im Vergleich zu den meisten Textstufen von Textzeuge 3 ist bei Textzeuge 5 in verminderter Form zu beobachten. Neben dem geringeren Vorkommen an Satzzeichen ist eine Verschiebung der Verteilung auf den Wechsel der Wortart von „funkelnd“ hin zu „funkeln“ zurückzuführen.

Die publizierte Fassung besteht schließlich aus 103 Tokens. Damit ist vergleichsweise ein Anstieg um je ein Token bei den Verben, Subjunktoren, Substantiven, Artikeln, Konjunktoren und Adverbien zu erkennen. Je zwei Tokens mehr verzeichnen die Satzzeichen und Adjektive. Eine Präposition weniger ist verzeichnet. Gründe für diese Veränderungen sind zum einen die umfangreiche Einfügung „zwischen kleinem und großem Liebfrauentag“ gegen Ende des Gedichts, wo zu anderen „die zur Hälfte enthüllte Schattenlippe“ zu „das halb erloschene Traumgesicht“ umgeschrieben ist und damit sowohl inhaltliche als auch syntaktische Änderungen bewirkt. Mit genannter Einfügung wird die Alliteration erweitert zu „Liebfrauentag, ein Lebens- und Liebesfreund“.

Bezeichnend ist insgesamt, dass Substantive am meisten von den Eingriffen betroffen sind und bei den komplexeren Überarbeitungen eine höheres Vorkommen aufweisen. Es gibt in etwa doppelt so viele Streichungen als Einfügungen. Damit werden in zwanzig <del>-Tags über fünfzig Tokens gestrichen. Während genauso viele Pronomen gelöscht wie hinzugefügt werden, sind bei allen anderen POS größere Unterschiede zu beobachten. Auffallend ist, welche dominante Stellung die Satzzeichen innerhalb der <del>-Tags einnehmen, während in allen anderen Eingriffs-Tags diese gering ausfallen.

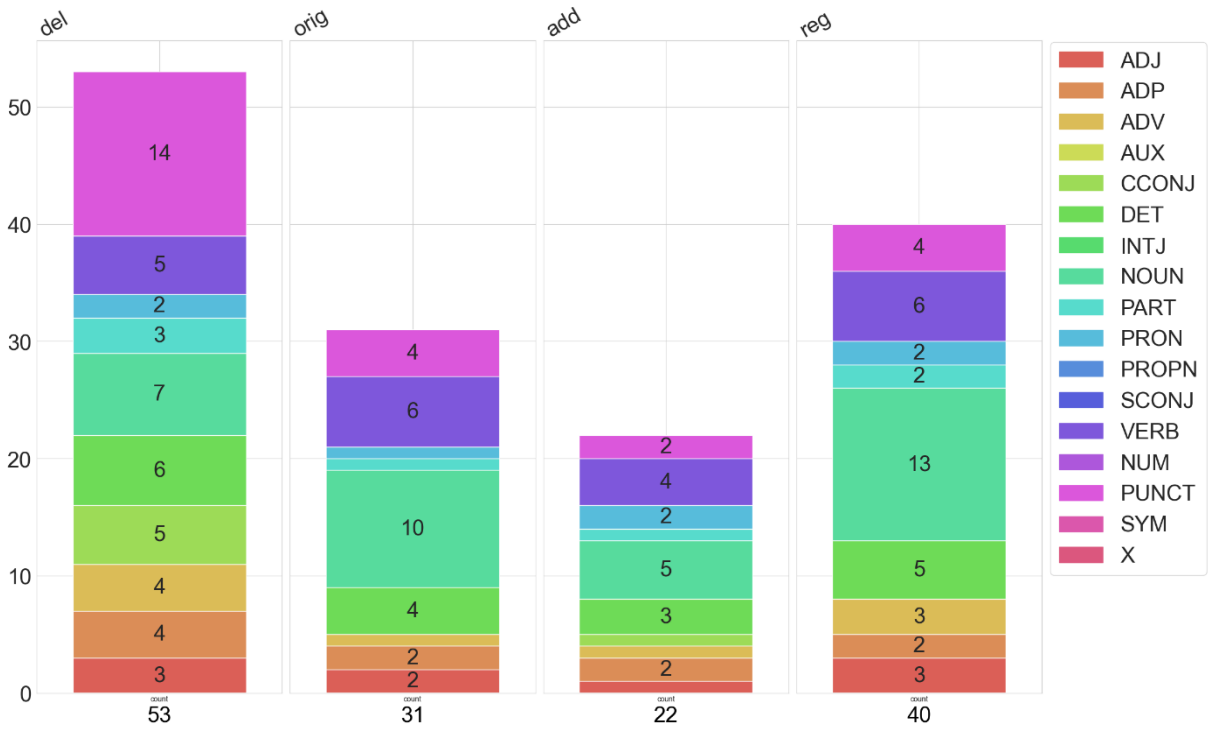

Abbildung 27: Absolute Häufigkeitsverteilung der Parts of Speech in allen Eingriffs-Tags des Gedichts *RSFGMT*.

Im Gedicht fallen Reihungen mit vielen Substantiven ins Auge, allein der Anfang des Gedichts ist hierfür bezeichnend („mit Farbstich, Fremdflocke, magisches Werk, Schüttelhaltung, anarcher Kunstverstand“). Wie im vorigen Gedicht ist mit fünf Vorkommen die Sequenz aus „Substantiv – Artikel – Substantiv“ im Zuge von Genitivkonstruktionen dominant, wie bei „Innenpolster der Hand“ oder „Sattel des Bergs“. Die häufige Sequenz „Substantiv – Adjektiv – Substantiv“ funktioniert aufgrund der Aufzählungen, wie bei „Vlies, leuchtendes Lächeln“ ab Textzeuge 4 und zweimal zu Beginn des Textes (siehe oben). Dies zeigt eine Tendenz auf, wie Mayröcker diese Reihungen in diesem Gedicht gestaltet. Jedoch etabliert sich diese Sequenz erst im Laufe der Revisionen von Textzeuge 3 in dieser Dominanz, zuvor ist die Sequenz „Artikel – Adjektiv – Substantiv“ („*das* leuchtende Lächeln“) häufiger vertreten. Hiermit zeigt sich zumindest in diesen Fällen ein Rückgang der Artikel.

7.4 *Wassersachen, nach Mimmo Paladino*

Von den fünf Textzeugen des am 11.10.1981 entstandenen Gedichts *Wassersachen, nach Mimmo Paladino* (W127/471 und W129/479) sind vier Typoskripte mit handschriftlichen Revisionen versehen. Während in Textzeuge 1 und 3 relativ viele Eingriffe zu finden sind mit zwanzig und neunzehn @change-Attributen, scheinen bei Textzeuge 2 lediglich dreizehn und bei Textzeuge 5 fünf solcher auf. Alle Bearbeitungen finden mit einem blauen Stift statt. Im

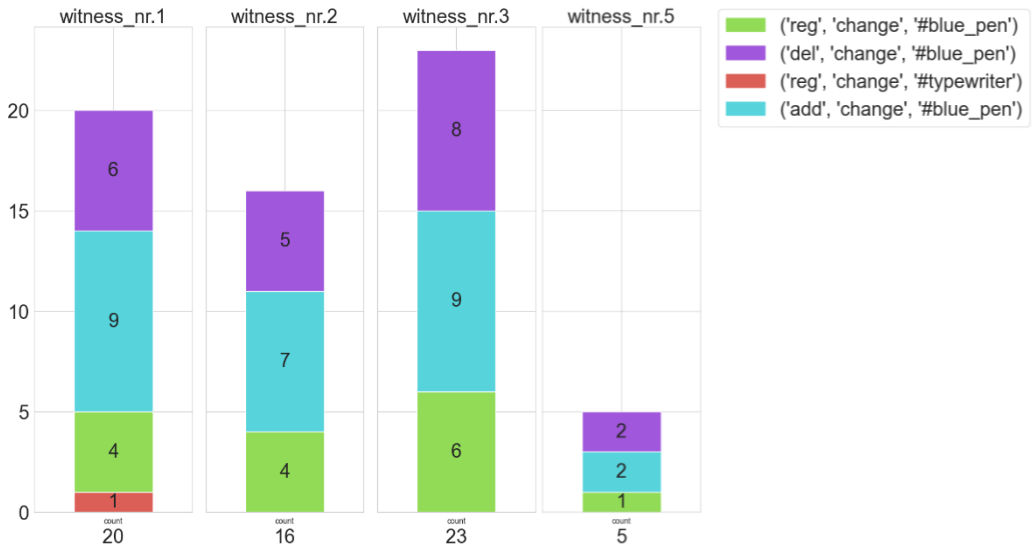

Abbildung 28: Absolute Häufigkeitsverteilungen der Kombinationen aus Eingriffsart und Stiftfarbe für alle vier Textzeugen des Gedichts *WSNMP*.

ersten Textzeugen gibt es mehr Hinzufügungen (9) als Streichungen (6), am niedrigsten sind jene Fälle, für die ein <reg>-Tag (5) verwendet wird. Dagegen gibt es lediglich drei Löschungen im zweiten Textzeugen und gleichviele Einschübe wie komplexere Revisionen (5). Auch bei Textzeuge 3 ist die Zahl der Kürzungen (5) geringer als jene der Ergänzungen (8). Mit sechs Vorkommen eingehender Eingriffe ist es die höchste Anzahl im Vergleich aller Textzeugen. Gleichviele Streichungen wie Einfügungen fallen in Textzeuge 5 auf. Hingegen scheint nur eine Verwendung des <reg>-Tags auf.

Im Vergleich der Tokenanzahl über alle Textstufen hinweg ist zu beobachten, dass die Revisionsstufen der Textzeugen jeweils etwas höher sind als die der Typoskripte. Dies mag auch mit dem Ergebnis korrelieren, dass die Menge an Streichungen in drei von vier bearbeiteten Textzeugen niedriger ist als jene der Einfügungen. Folglich hat das Typoskript von Textzeuge 1 180 Tokens, durch die Korrektur mit der Schreibmaschine 181 Tokens und durch die handschriftlich revidierte Textstufe 193 Tokens. Bei Textzeuge 2 ist ein Wechsel von 151

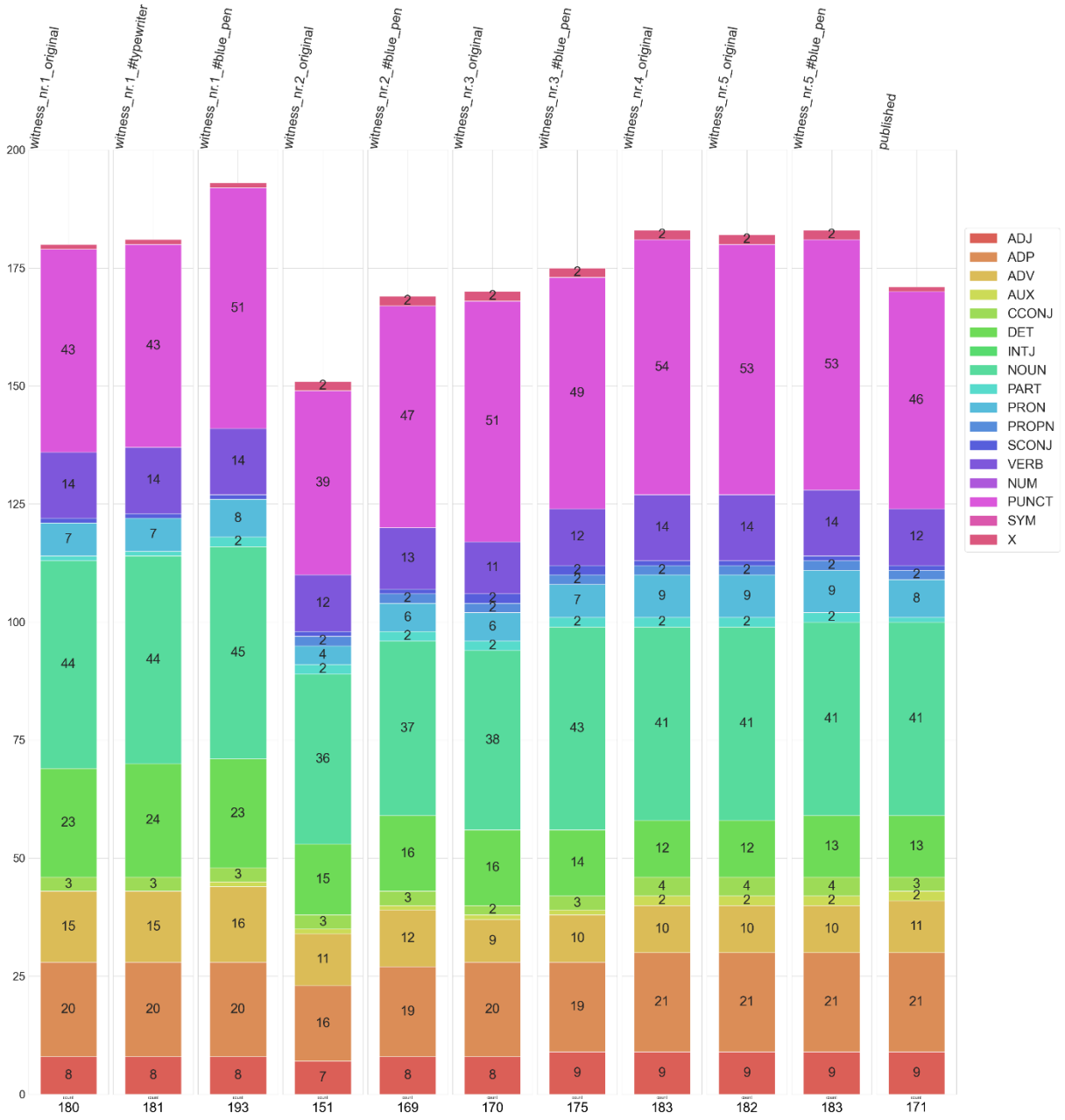

Abbildung 29: Absolute Häufigkeitsverteilung der Parts of Speech im Gedicht *WSNMP* über alle Textzeugen und deren Revisionsstufen hinweg bis zur publizierten Fassung.

Tokens auf 169 Tokens abzulesen und bei Textzeuge 3 erhöht sich die Tokenanzahl von 170 um weitere 5 Tokens. Darauf folgt Textzeuge 4 mit 183 Tokens. Textzeuge 5 zählt nur 1 Token weniger, bei der Revisionsstufe wieder gleichviel. Die veröffentlichte Gedichtfassung zählt letztendlich 171 Tokens.

Mit Blick auf die Verteilung der Parts of Speech verändert sich bei Textzeuge 1 durch die Korrektur mit der Schreibmaschine die Passage „das Echo des Hundegebells im Tal“ zu „aus dem Tal“, wodurch allein ein Artikel mehr zu verzeichnen ist. Im Vergleich mit dem Typoskript

ist bei der Revisionsstufe von Textzeuge 1 ein Anstieg von je einem Token bei den Adverbien, Hilfsverben, Artikeln, Substantiven, Partikeln und Pronomen festzustellen. Insbesondere die Anzahl der Satzzeichen nimmt um acht weitere Fälle zu. Unverändert dagegen bleibt die Summe von Adjektiven, Präpositionen, Konjunktoren, Subjunktor und Verben. Konkret sind zwei Einfügungen mehrerer Worte („was abfällt vom Tisch“ und „kann nicht mehr auffliegen“) zu beobachten sowie kleinere Streichungen. Gemeinsam mit ausschlaggebenden Ersetzungen wie „voll Grießbrei“ anstelle von „riecht nach Grießbrei“ oder „im Gegenlicht“ statt „von

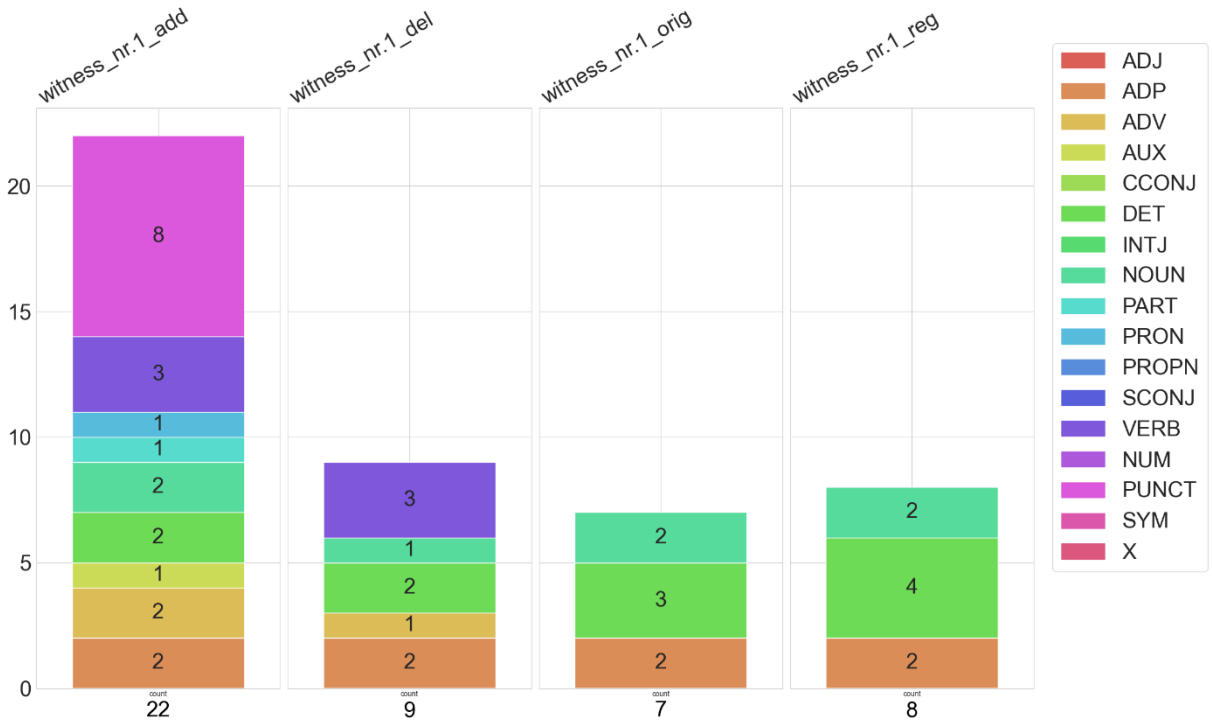

Abbildung 30: Absolute Häufigkeitsverteilung der Parts of Speech innerhalb der Eingriffs-Tags für Textzeuge 1 des Gedichts *WSNMP*.

hinten“ kann sich aber beispielsweise die Anzahl der Verben vollständig ausgleichen. Zudem erweisen sich einige Ersetzungen ohne Einfluss auf die POS-Verteilung, wie „Ziegen im Tal“ zu „Ziege im Fels“ als auch der Wechsel von der 1. Person Singular hin zur 3. Person Singular oder zur 1. Person Plural bei diversen Pronomen („dein Blick“ zu „sein Blick“).

Über die starke Reduktion der Tokens mit der Verfassung des zweiten Typoskripts bleibt die Anzahl der Hilfsverben, Konjunktoren, Partikel und Subjunktoren bestehen und hinzukommen zwei Eigennamen. Letzteres hängt mit dem erstmaligen Anführen eines Teils des schlussendlichen Titels („nach Mimmo Paladino“) zusammen. Die massivsten Verminderungen sind bei den Satzzeichen mit 12 weniger Vorkommen aber auch bei den

Substantiven und Artikeln mit je 9 weniger Tokens. Insbesondere das Fehlen einiger Verse aus dem Mittelteil der Fassung von Textzeuge 1 ist hierfür als Grund heranzuziehen (bspw. „eigentlich interessieren mich die Bildunterschriften immer mehr als die Bilder“). Zu weiten Teilen bleibt ansonsten die Wortwahl dieselbe, allein „Grießbrei“ wird zu „Hirsebrei“ und „Feststück“ wird zu „Felsstück“. Teilweise wiederholen sich die Eingriffe, als wären die handschriftlichen Änderungen in Textzeuge 1 bei der maschinellen Abschrift nicht intensiv beachtet worden.

Durch den Revisionsvorgang an Textzeuge 2 wiederum geschieht ein Zuwachs der Satzzeichen um acht Tokens, während die Verben, Nomen, Artikel, Adverbien und Adjektive um je ein Token wachsen. Ein Anstieg der Satzzeichen geschieht aufgrund der in Klammer gesetzten direkten Rede, die vervollständigt und um eine Frage erweitert wird („welches Feld ..?“). Die Präpositionen nehmen um drei Tokens zu, während die Pronomen zwei Tokens mehr verzeichnen. Zurückzuführen ist dies durch die Einfügungen „vor mir, von hinten“ sowie „dein Blick weicht mir aus“. Die Präpositionen sind mit fünf Fällen die am meisten eingefügte Wortart in allen `<add>`-Tags. Unverändert bleibt die Anzahl des Hilfsverbs, der Konjunktoren, Partikel, Eigennamen, des Subjunktors sowie der beiden unerkannten französischen bzw. französisch inspirierten Begriffe („allez!“ und „tamponte“). Insgesamt gibt es mehr als dreimal weniger gestrichene als eingefügte Tokens. Ersetzungen wie „Tag“ anstelle von „Morgen“ oder „mild“ statt „lässig“ sowie leichte Veränderungen in Tempus, Numerus oder Genus vereinzelter Wörter wirken sich nicht weiter auf die Häufigkeiten aus.

Um nur ein Token verändert sich die Anzahl zwischen der Revisionsstufe des Textzeugen 2 und des Typoskripts des Textzeugen 3. Dabei wandelt sich in Hinblick auf die Wortartenverteilung einiges. So steigt die Menge an Präpositionen, Nomen und Subjunktoren um je ein Token. Die Summe der Satzzeichen nimmt um 4 Tokens zu. Verringert werden die Vorkommen der Adverbien, Konjunktoren und Verben. Insbesondere im Mittelteil wird über die maschinelle Abschrift teils sowohl gekürzt als auch erweitert. So heißt es beispielsweise anstelle von „Klebestreifen über Wasserfarbe“ nun „Klebestreifen über Wassersachen wie Mühlen, Fischereien, etcetera“.

Mit der Revision des Typoskripts von Textzeuge 3 wird nicht nur diese Substantiv-Reihung mit zwei Alliterationen („wie *Träne und Meer, auch* Mühlen, *Fluss* Fischereien“) weiter ausgebaut, der Titel erhält sein Anfangswort und der Beginn des Gedichts wird mit Einfügungen und Angleichungen insbesondere inhaltlich spezifiziert („mit weißen Schwingen in den violetten

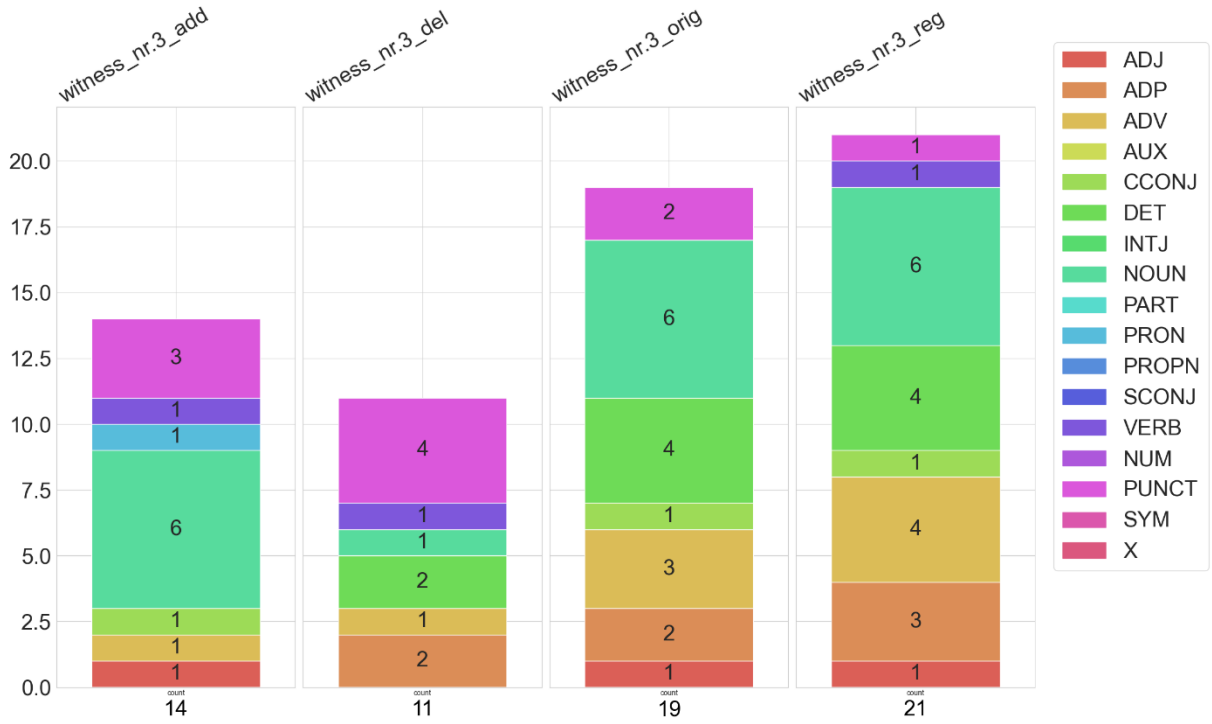

Abbildung 31: Absolute Häufigkeitsverteilung der Parts of Speech innerhalb der Eingriffs-Tags für Textzeuge 3 des Gedichts *WSNMP*.

Tag, die Luft voll Hirsebrei“ zu „mit blendenden weißen Schwingen ins violette Morgenlicht des Tags, die Lüfte riechen nach Hirsebrei“). Zudem sind weitere kleinere Ergänzungen, Änderungen des Numerus und Wortwechsel auf paradigmatischer Ebene (bspw. „wenn ich mit einer Frau spreche“ zu „wenn ich mit Frauen rede“) über den Rest den Gedichts verteilt. Durch diese Eingriffe bleibt die Anzahl der Adjektive bis zur publizierten Fassung bestehen, jene der Adverbien für vier Textstufen. Zunahmen um je ein Token ist bei den Adjektiven, Adverbien, Konjunktoren, Partikeln und Verben zu beobachten. Die Substantive verzeichnen fünf Token mehr. Um je zwei Tokens nehmen die Satzzeichen und die Artikel ab, die Präpositionen um eines.

Durch den Sprung von der Revisionsstufe des Textzeugen 3 zum unbearbeiteten Textzeugen 4 sind mehrheitlich Zunahmen augenfällig. Um je ein Token steigt die Anzahl der Hilfsverben und Konjunktoren, um je zwei Token jene der Präpositionen, Pronomen und Verben. Mit 5 weiteren Fällen steigt die Zahl der Satzzeichen. Reduktionen gibt es um zwei Token bei den Substantiven und Artikeln, um eines bei den Subjunktoren. Die komplexeren Veränderungen, wie zu Beginn in Textzeuge 3, werden nicht in ihrem vollen Ausmaß übernommen. Einzelne Detaillierungen werden vorgenommen „vor mir *dein wächsernes* Ohrenpaar“. In der zweiten

Hälfte des Gedichts passieren mit dem Verfassen des Typoskripts Ersetzungen, so wie anstelle von „die grünen Halme am Schuh, Reste von klumpigem Kot, von nahe, der Klötze Form“ nun „die klumpigen Halme noch am Schuh vom fliehenden Lauf durchs gemähte Feld, von nahe“ zu lesen ist. Gegen Ende wird die direkte Rede um eine Antwort auf die zuletzt eingefügte Frage erweitert („Feststück, das Schwarze hat sie dir abgeschaut..“).

In Hinblick auf Textzeuge 4 ist festzustellen, dass sich bestimmte Wortarten entweder bis zur Revisionsstufe von Textzeuge 5 oder gar bis zur veröffentlichten Gedichtversion bezüglich ihrer Anzahl nicht verändern. Über drei Textstufen bleibt die Menge an Verben, Pronomen, Partikel und Konjunktoren unverändert. Bis zur Endfassung erhalten bleibt die Summe an Präpositionen, Hilfsverben, Substantiven, Eigennamen und des Subjunktors.

Die Verminderung der Gesamtzahl der Token um eines bei Textzeuge 5 schlägt sich allein in der Verteilung bei der Anzahl der Satzzeichen nieder. Ansonsten bleiben die Häufigkeiten dieselben, da sich mit dem Verfassen dieses Typoskripts abgesehen von der Strukturierung des Layouts keine Neuerungen ergeben. Die Revision bei Textzeuge 5 äußert sich durch nur einen weiteren Artikel, da mit der folgenden Ersetzung die Anpassung des Genus einen Artikel fordert: „durchs gemähte Feld“ wird zu „durch den gemähten Acker“ korrigiert.

Wird dazu die Verteilung der Endfassung des Gedichts verglichen, zeigt sich die Reduktion der Tokens insbesondere durch weniger Vorkommen von Satzzeichen um sieben Fälle. Je ein Token weniger verzeichnen die Pronomen, Partikel und Konjunktoren. Die Verbenanzahl vermindert sich um zwei Tokens. Allein die Zahl der Adverbien nimmt um ein Token zu. Zwischen Textzeuge 5 und der Endfassung ist also vor allem ein Komprimierungsvorgang anzunehmen. So werden einzelne Worte und auch eine Infinitivgruppe („mich zu nehmen“) ausgelassen, aus „den gemähten Acker“ wird „die Stoppelfelder“ und „Waschzimmer (Zettel)“ wird zu „Waschzimmerzettel“. Eine markante Erweiterung ist gegen Ende des Gedichts, als die Antwort aus direkter Rede in zwei Sequenzen unterteilt wird („Felsstück und Fest..“) – („..das Schwarze hat sie dir abgeschaut..“). Damit beginnen die alliterativen Formulierungen in diesem Teil des Gedichts in enger Folge von fünf Substantiven mit dem Anfangsbuchstaben F. Viele Alliterationen bestehen jedoch bereits mit dem ersten Textzeugen wie „Kuss Klebestreif“, „Federchen“ und „französisch“ oder die Assonanz „kehren / queren“.

Um vier Tokens unterscheiden sich die Gesamtzahlen der Tags <orig> und <reg>. Dabei handelt es sich um dieselben Wortarten, von denen etwas mehr Adjektive, Präpositionen, Adverbien, Artikel und Verben eingefügt werden. Der Vergleich der Tags <del> und <add> zeigt ebenso,

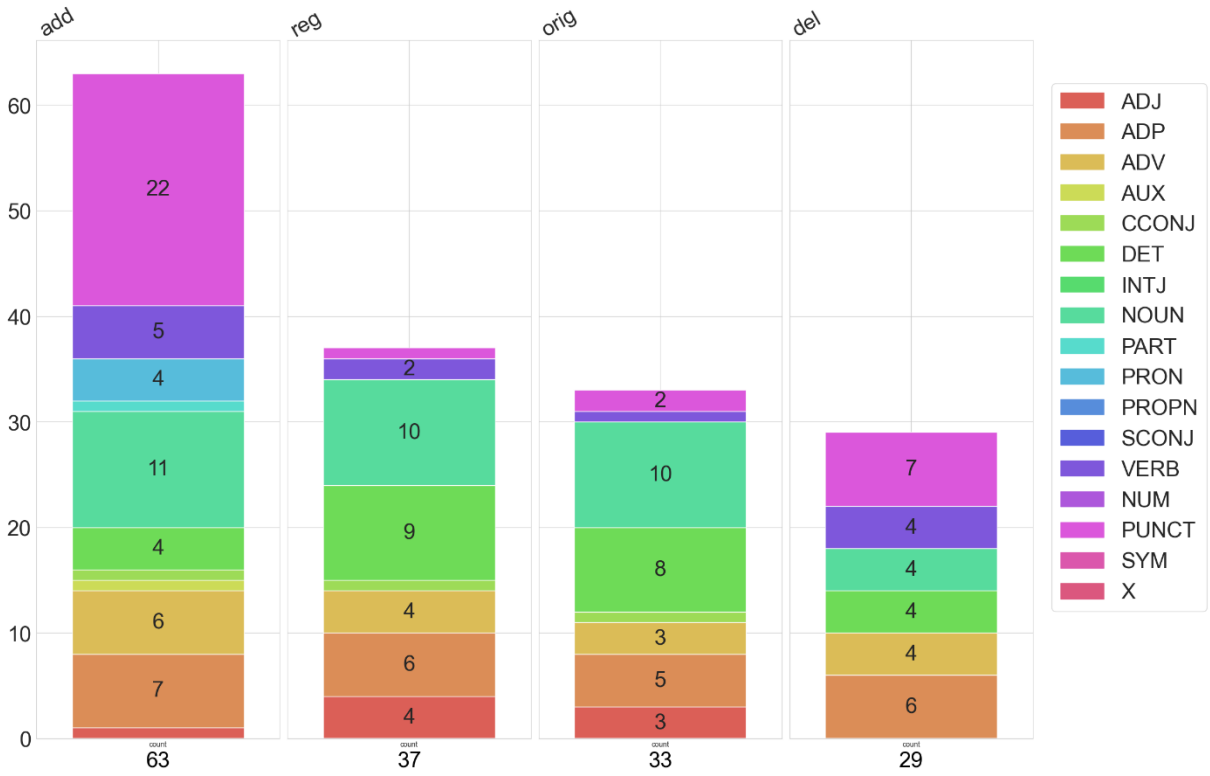

Abbildung 32: Absolute Häufigkeitsverteilung der Parts of Speech in allen Eingriffs-Tags des Gedichts *WSNMP*.

dass doppelt so viele Tokens hinzugefügt als gelöscht werden, insbesondere Substantive, Pronomen und eine beachtlich Zahl an Satzzeichen. Während keine Adjektive gelöscht werden, werden Hilfsverben, Konjunktoren und Partikel nur hinzugefügt. Die Zahl der Artikel bleibt gleich. Der p-Werte des Chi-Quadrat-Tests zu den Eingriffsarten im Vergleich verweist auf

	add	orig	reg	del
add	0.000000	**0.016147**	**0.004190**	0.612920
orig	**0.016147**	0.000000	0.996584	0.084977
reg	**0.004190**	0.996584	0.000000	0.053069
del	0.612920	0.084977	0.053069	0.000000

Abbildung 33: Matrix für die paarweise ausgeführten Chi-Quadrat-Tests zur Häufigkeitsverteilung der Parts of Speech in allen Eingriffs-Tags des Gedichts *WSNMP*.

signifikante Unterschiede zwischen den Tags <add> und <reg> (0.00429) sowie <add> und <orig> (~0.01615). Die ungleichen Häufigkeiten der Satzzeichen, Pronomen und Artikel in den genannten Eingriffs-Tags könnten darauf Einfluss haben.

In der publizierten Gedichtfassung werden zwölf Sequenzen gefunden. Teilweise schließen gewisse Sequenzen in einzelnen Sätzen aneinander an oder überlappen sich. Beispielsweise sind die Sequenzen „Artikel – Substantiv – Verb“ und „Verb – Präposition – Substantiv“ in „die Luft riecht nach Hirsebrei“ vorhanden. Ebenso verhält es sich bei den Sequenzen „Präposition – Adjektiv – Substantiv“, „Adjektiv – Substantiv – Präposition“ und „Substantiv – Präposition – Adjektiv – Substantiv“, die in „mit weißen Schwingen ins violette Licht“ seit Textzeuge 1 (wenn auch nicht mit durchgehend derselben Wortwahl) auftauchen. Gleichfalls augenscheinlich sind die Sequenzen „Artikel – Adjektiv – Substantiv“ (bspw.: „dein wächsernes Ohrenpaar“, „der erkaltete Garten“), „Präposition – Artikel – Substantiv“ (wie „durch die Stoppelfelder“, „aus dem Tal“) und „Substantiv – Präposition – Substantiv“ (in „Ziege im Tal“, „Halme am Schuh“). Ein weiterer Teil der Sequenzen besteht aus Wortfolgen, die aus dem Kontext gerissen inhaltlich unvollständig sind, teils aufgrund von Aufzählungen. Hierzugehören „Präposition – Substantiv – Präposition“ („als Vogel im“), „Artikel – Substantiv – Präposition“ („das Echo von“) und wider Erwarten keine Genitivkonstruktionen bei der Sequenz „Substantiv – Artikel – Substantiv“ (wie bei „Licht, die Luft“). Während beinahe die Hälfte dieser Sequenzen bereits in Textzeuge 1 aufscheinen, wenn auch teils in anderen Wortfolgen, kommt es zu einer schrittweisen Verdichtung mit wesentlichen Veränderungen in Textzeuge 3 und 4 hin bis zur publizierten Fassung. Im Vergleich zu den anderen Gedichten ist es der umfangreichste Text in Hinblick auf die Tokens und die Sequenzen, dicht gefolgt vom letzten Gedicht dieser Analyse.

7.5 *vielleicht weil das Auge zuerst bricht*

Wie bereits im Kapitel 5 zum Textkorpus erwähnt, erweist sich die Textgenese des Gedichts *vielleicht weil das Auge zuerst bricht* (23.-25.10.1981; W127/471) als sehr umfangreich und vielfältig. Auf sieben Typoskripten werden mit fünf verschiedenen Farben und der Schreibmaschine Bearbeitungen vorgenommen. Nur zwei Typoskripte werden mit einer Stiftfarbe verändert, nämlich Textzeuge 1 und 6 mit rein schwarzen Eingriffen. In Textzeuge 2, 3 und 7 greift Mayröcker vermehrt mit Blaustift ein. Eine Revision ist in Textzeuge 2 einem violetten Stift zuzuordnen. Die zweitmeisten Bearbeitungen sind in Textzeuge 3 festzustellen mit fünf Fällen in schwarzer Farbe, zusätzlich wurde einmal in grün an dem Text gefeilt.

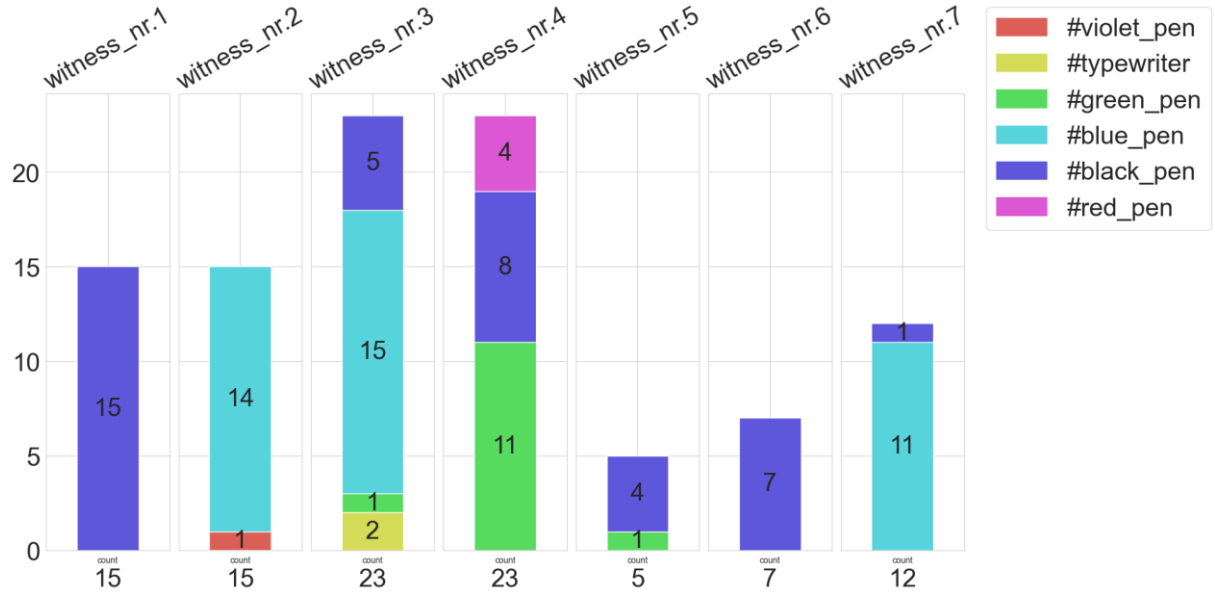

Abbildung 34: Absolute Häufigkeitsverteilung aller Attributwerte (Stiftfarben) für die Textzeugen des Gedichts *VWDAZB*.

Darüber hinaus gibt es zwei Stellen, an denen eine schreibmaschinelle Revision verzeichnet ist. Für Textzeuge 4 ist der Grünstift die dominante Farbe der Eingriffe, aber auch mit rot und schwarz wurden Veränderungen vorgenommen. In Textzeuge 5 ist neben einer Veränderung in grünem Stift mehrheitlich mit einem schwarzen Stift gearbeitet worden. Umgekehrt wird der schwarze Stift in Textzeuge 7 nur in einem Fall gezückt.

Bei fünf der sieben Textzeugen scheinen mehr Fälle von Einfügungen auf als Streichungen (Textzeugen 1, 3, 5, 6, 7). Zumeist weniger Vorkommen sind für komplexere Revisionsvorgänge zu verzeichnen. In Textzeuge 1 befindet sich die einzige Stelle, die nicht eindeutig zu entziffern ist und damit mit dem <unclear>-Tag ausgezeichnet wurde.

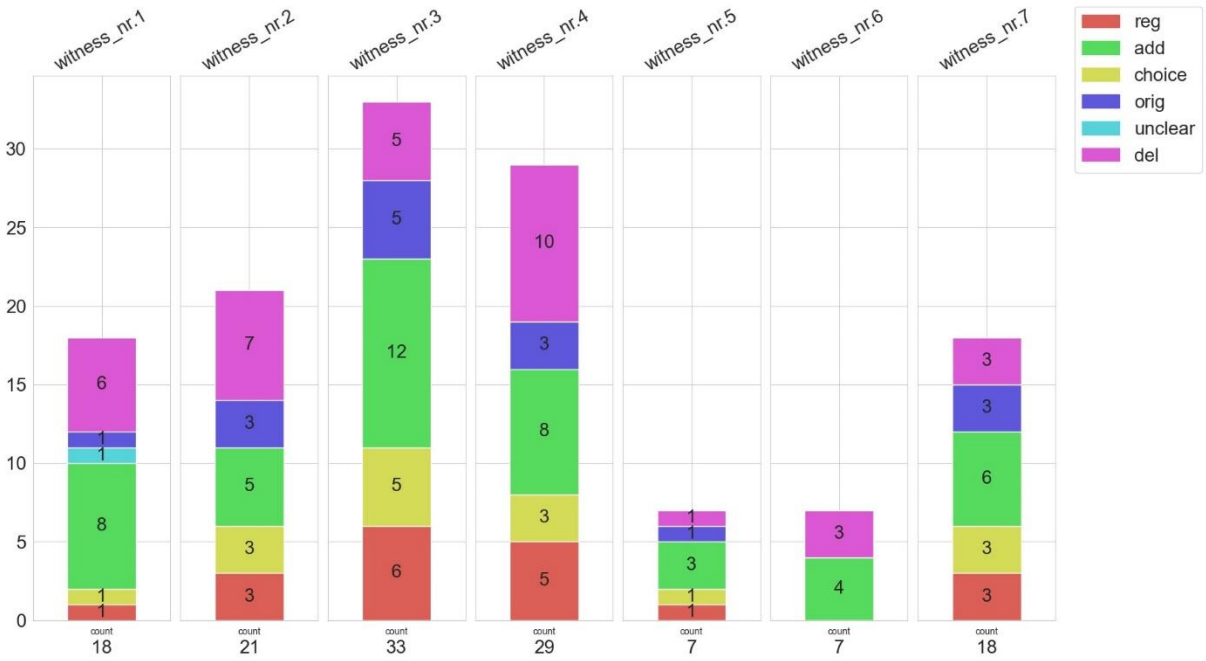

Abbildung 36: Absolute Häufigkeitsverteilungen aller Tags für die Textzeugen des Gedichts *VWDAZB*.

Wird ein Blick auf alle möglichen Kombinationen aus Eingriffsart und Stiftfarbe geworfen, weichen die Textzeugen teils markant voneinander ab. Somit gibt es einige als signifikant einzustufende Unterschiede zwischen den Textstufen in Hinblick auf die p-Werte der Chi-Quadrat-Tests. Direkt aufeinanderfolgende Textzeugen mit signifikantem sowie starkem

	witness_nr.1	witness_nr.2	witness_nr.3	witness_nr.4	witness_nr.5	witness_nr.6	witness_nr.7
witness_nr.1	0.000000	**0.000039**	**0.001716**	**0.017981**	0.284794	0.783139	**0.000283**
witness_nr.2	**0.000039**	0.000000	0.423861	**0.000154**	**0.002769**	**0.000524**	0.609215
witness_nr.3	**0.001716**	0.423861	0.000000	**0.000752**	0.115183	**0.015853**	0.832644
witness_nr.4	**0.017981**	**0.000154**	**0.000752**	0.000000	0.277867	0.113107	**0.000751**
witness_nr.5	0.284794	**0.002769**	0.115183	0.277867	0.000000	0.394312	**0.020006**
witness_nr.6	0.783139	**0.000524**	**0.015853**	0.113107	0.394312	0.000000	**0.003667**
witness_nr.7	**0.000283**	0.609215	0.832644	**0.000751**	**0.020006**	**0.003667**	0.000000

Abbildung 35: Matrix für die paarweise ausgeführten Chi-Quadrat-Tests zur Häufigkeitsverteilungen der Kombinationen aus Eingriffsart und Stiftfarbe für alle Textzeugen des Gedichts *VWDAZB*.

Unterschied sind Textzeuge 3 und 4 sowie Textzeuge 6 und 7 mit einem Cramer's V von je 0,36 bzw. 0,35.

Aufgrund der Verwendung dieser verschiedenfarbigen Schreibwerkzeugen sind teils bis zu vier Revisionsstufen zu extrahieren. Dementsprechend umfangreich stellt sich die Entwicklung der

Häufigkeitsverteilungen dar. Textzeuge 1 beginnt mit 159 Tokens, darauf folgen mit weniger Tokens Textzeuge 2 mit 135 und Textzeuge 3 mit 141. Textzeuge 4 birgt die insgesamt höchste Tokenanzahl mit 181, worauf Textzeuge 5 und 6 mit 146 bzw. 147 Tokens folgen. Im Vergleich zur publizierten Fassung mit 144 Tokens hat Textzeuge 7 noch 157 Tokens.

Die POS-Verteilung in Textzeuge 1 beginnt mit einer hohen Anzahl an Satzzeichen (36) als auch Substantiven (37), gefolgt von zwanzig Artikeln, vierzehn Verben, elf Pronomen, je acht Hilfsverben, Adverbien und Präpositionen, gefolgt von sechs Konjunktoren, vier Adjektiven und drei Eigennamen und zwei Partikeln und Subjunktoren. Durch die Eingriffe mit dem schwarzen Stift ist insgesamt eine Reduktion der Tokens zu beobachten, die sich auch bei den Satzzeichen, Partikeln, Substantiven, Artikeln und Adverbien um je ein Token niederschlägt. Die Pronomen weisen vier Token weniger auf, die Hilfsverben drei und die Konjunktoren zwei weniger. Gleichbleibende Zahlen sind bei den Verben, Eigennamen und Präpositionen zu

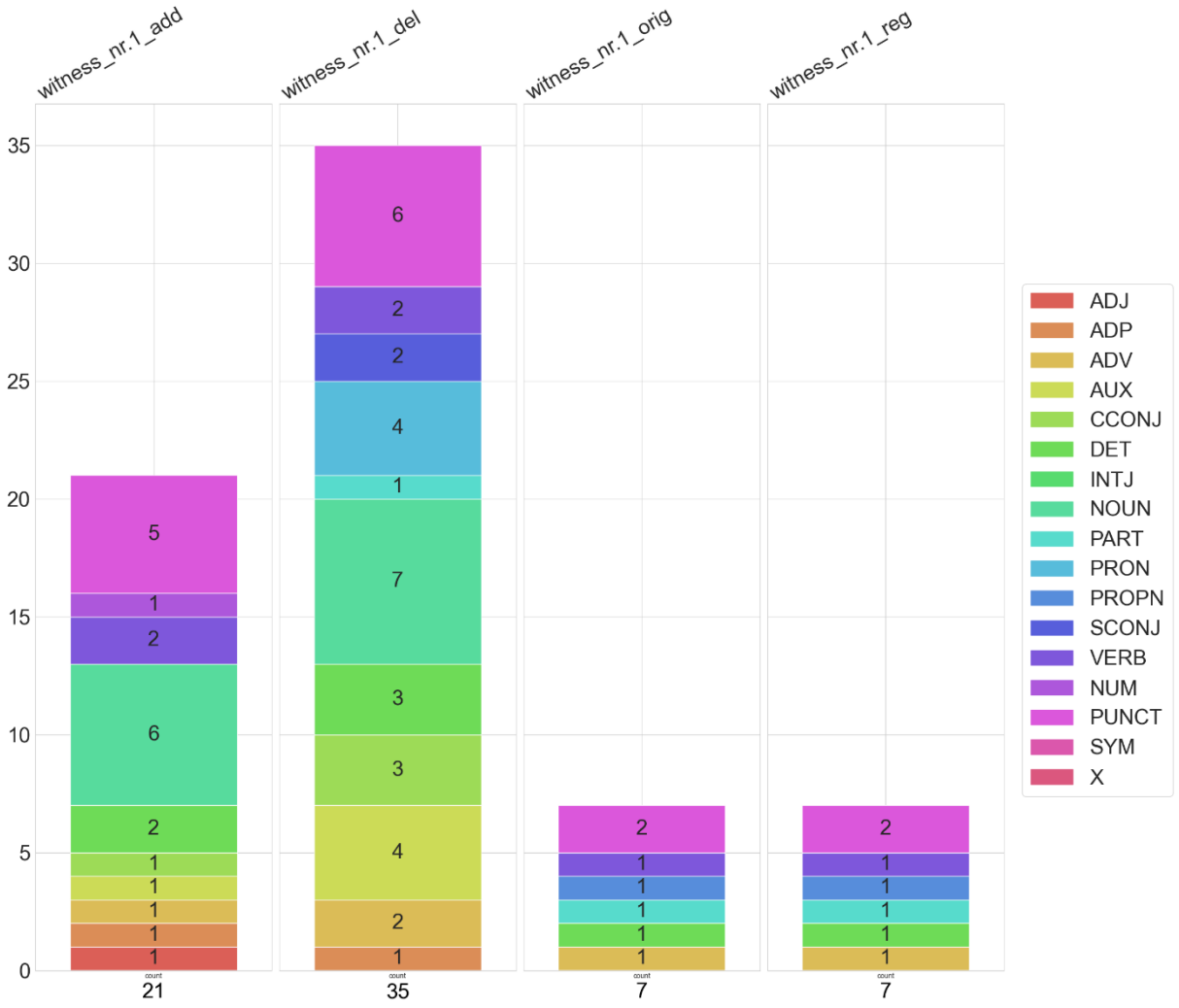

Abbildung 37: Absolute Häufigkeitsverteilung der Parts of Speech innerhalb der Eingriffs-Tags für Textzeuge 1 des Gedichts *VWDAZB*.

beobachten, während je ein Token mehr bei den Adjektiven und Ziffern zu verzeichnen ist. Insbesondere die Streichung von sechs Zeilen in der ersten Hälfte des Gedichts erscheint ausschlaggebend. Aufgrund dessen beträgt die Zahl der gelöschten Tokens 36, die POS-Verteilung ist dementsprechend vielfältig, wobei viele Pronomen, Substantive und Hilfsverben gestrichen werden. Darüber hinaus gibt es teils umfangreichere Einfügungen wie „mit der linken Hand 3 mal geteilt“ und „Verdruss + Krankheit“.

Auf einem niedrigeren Level der Anzahl von Tokens befinden sich die Textstufen von Textzeuge 2. Im Vergleich zur Revisionsstufe von Textzeuge 1 ist im Typoskript von Textzeuge 2 ein Verlust von je einem bis zwei Token(s) bei den Präpositionen, Adverbien, Hilfsverben, Konjunktoren, Artikeln und Verben zu erkennen. Drei Tokens weniger verzeichnen die Substantive. Während die Tokenanzahl der Adjektive, Partikeln, Pronomen und Eigennamen unverändert bleibt, nehmen die Satzzeichen um zwei weitere Fälle zu. An dieser Stelle trägt das Gedicht vorläufig den Titel „Werktreue!“. „*die Witwe*, als Göttin“ wird eingefügt. Über weite Strecken wird bei der Abschrift der Wortlaut von Textzeuge 1 beibehalten, vor allem gegen Ende des Gedichts werden neben Füllwörtern teils umfassendere Passagen ausgelassen wie „Säge des Laubs“ oder „unter dem Kinn gelegen, unter der Zunge“.

Durch die Streichung eines Beistrichs mit einem violetten Stift ist allein eines der Satzzeichen weniger verzeichnet. Durch die Eingriffe mit dem blauen Stift bleibt die Zahl der Satzzeichen stattdessen gleich, dagegen vermindert sich die Summe der Verben, Hilfsverben und Adjektive um je ein Token, jene der Artikel um zwei Token. Gleichzeitig steigt die Menge von Adverbien und Pronomen um ein Token. Mit Blick auf den Textzeugen ist die Löschung von „ist verbrannt“ augenscheinlich wie auch die Änderung „mit der linken Hand“ zu „mit der Linken“. Kleinere Ersetzungen („in der Brusttasche“ zu „aus der Brusttasche“, „deiner Beine“ zu „seiner Beine“) haben keinen Einfluss auf die Verteilung.

Abgesehen von den gleichbleibenden Zahlen für die Adjektive, Konjunktoren, Partikel, Pronomen und Eigennamen erfahren mit dem Typoskript von Textzeuge 3 alle anderen POS einen Zuwachs zwischen einem und maximal drei Tokens. Der Titel fixiert sich mit diesem Textzeugen vorerst auf „weil das Auge zuerst bricht“. Neben dem hinzugefügten Ende „wo die Schatten einander entgegen sind“, sind einzelne Tokens neu wie „Wahrträume“. An der Stelle „dreimal geteilt, mit der Linken“ geschieht eine Komprimierung zu „dreimal abgehoben, linkshändig“. Aufgrund zwei der maschinengeschriebenen Einfügungen („eingerastet *ist* das Stativ deiner Beine“, „Oktober*keule und* -schatten) erhöht sich die Anzahl von Hilfsverben,

Konjunktoren, Nomen und Satzzeichen um je ein Token. Ein Token weniger fasst die Verteilung der Revisionsstufe aufgrund des grünen Stifts, welcher sich lediglich auf die Reihenfolge einer Passage konzentriert. Hierbei reduziert sich allein die Anzahl der Adjektive, da eines („aus den *aufgeschlagenen* Karten“) mit der Neuanordnung gestrichen wird. Aufgrund der Revisionen mit einem blauen Stift nehmen die Satzzeichen um drei Fälle ab. Für die Adjektive bedeutet es einen Zuwachs von einem, für die Adverbien, Pronomen und Verben eine Zunahme um zwei und für die Artikel und Nomen eine Steigerung um drei Tokens. Wie in Textzeuge 2 wird wieder erweitert zu „mit der linken Hand“, umfangreiche Einfügungen wie „die Vögel picken zuerst die Augen aus“ sowie „*wie mit vielen Augen* von jeder Pfütze getrunken“ sind auffallend. Mit letzterem ist die Alliteration mit dem Begriff „Wahrträume“ erstmals möglich. Folgenlos bleiben die Kürzungen innerhalb der zwei Wörter „Zigeunerspielkarten“ zu „Zigeunerkarten“ und „Brusttasche“ zu „Tasche“. Die größte

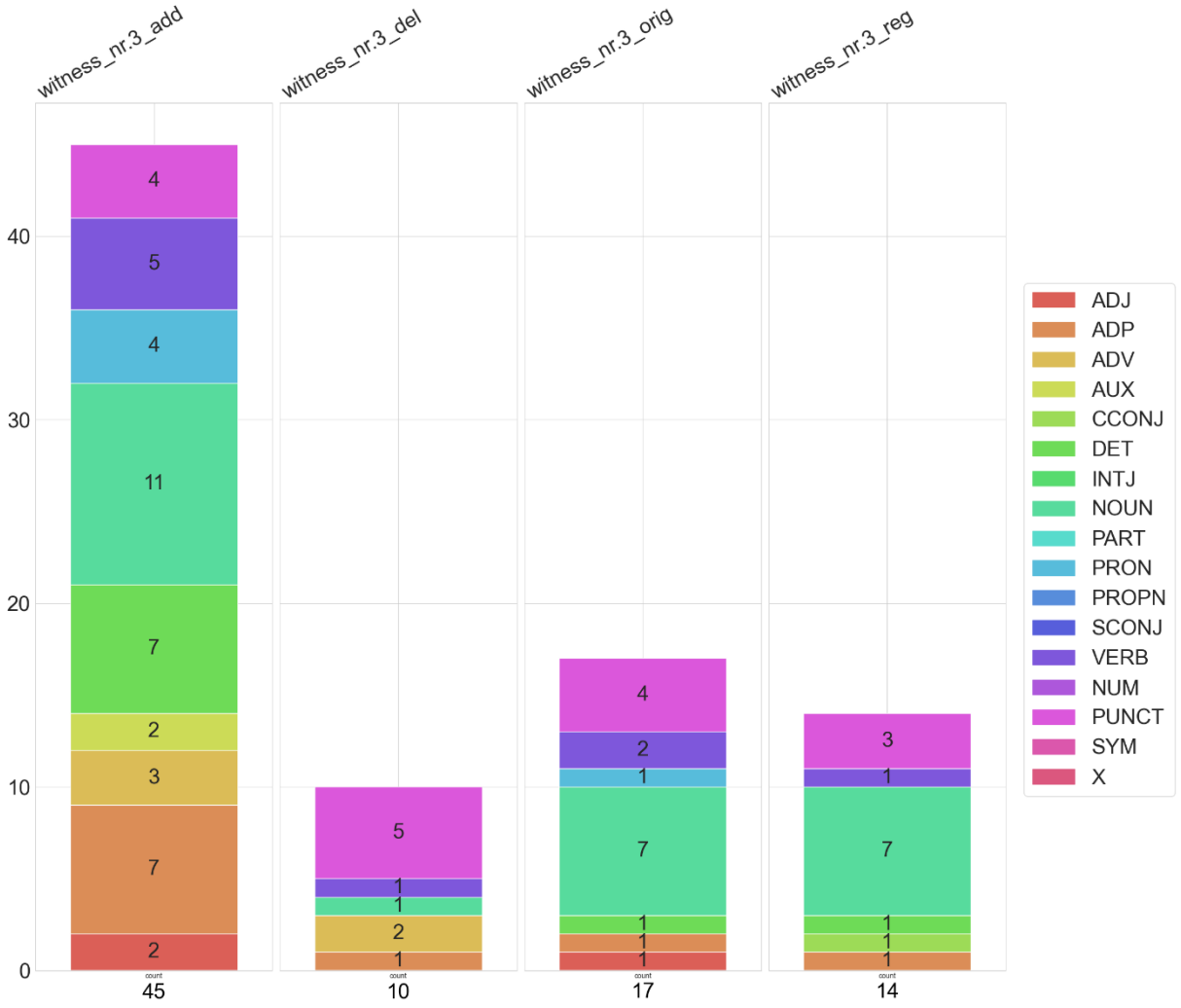

Abbildung 38: Absolute Häufigkeitsverteilung der Parts of Speech innerhalb der Eingriffs-Tags für Textzeuge 3 des Gedichts *VWDAZB*.

Textstufe formiert sich unter dem Einfluss eines schwarzen Stifts. Im Vergleich zum Typoskript verzeichnen Adjektive, Adverbien und Hilfsverben ein Token mehr. Um zwei Tokens steigt die Anzahl der Verben und Satzzeichen. Die Menge der Präpositionen und Artikel steigt um je vier weitere Fälle, während die Substantive ganze sieben weitere Tokens verzeichnen. Der schwarze Stift greift teils in jene oben angeführten Einfügungen mit dem blauen Stift ein und übernimmt sie in abgeänderter Form. Zudem wird die bereits bekannte Stelle aus Textzeuge 1 erneut eingeführt „Asche sein unter dem Kinn unter der Zunge".

Das Typoskript von Textzeuge 4 startet mit 181 Tokens. Damit hat es im Vergleich zum vorherigen Typoskript je ein Token mehr bei den Adverbien und Subjunktoren. Zwei Tokens mehr sind bei den Hilfsverben zu verzeichnen, drei mehr bei den Pronomen. Sowohl die Anzahl der Satzzeichen als auch jene der Verben steigt um je vier Tokens, während jene der Präpositionen und Artikel um sechs bzw. sieben zunimmt. Die größte Steigerung zeigt sich bei den Substantiven von 34 auf 47 Tokens. Insbesondere die Revisionen mit dem blauen und schwarzen Stift werden übernommen. Insgesamt ist die Tokenanzahl für alle Textstufen von Textzeuge 4 besonders hoch, da sich in der zweiten Hälfte des Typoskripts gewisse Phrasen wiederholen wie „aber es ist kein Schatz zugegen" oder „alles verflochten". Für die Modellierung ist an dieser Stelle jedoch nicht eindeutig auszumachen, ob und zu welchem Ausmaß hier eine Überarbeitung mit der Schreibmaschine stattgefunden hat. Um ein Token steigt die Gesamtzahl durch die Neuarrangierung der Wortfolgen „Stativ deiner Beine" und „schnarrendes Maulhorn" mit den roten Stift. Dabei zählen die Adjektive und Substantive ein Token mehr, während sich die Fälle der Zeichensetzung um eines vermindern. Über die Eingriffe mit einem schwarzen Stift erfährt das Typoskript eine Reduktion um neun Tokens. Während allein die Anzahl der Adjektive um ein Token steigt, sinkt jene der Adverbien, Artikel und Pronomen und je eines. Je zwei Tokens weniger verbuchen die Präpositionen und Nomen sowie drei der Satzzeichen. Gekürzt wird hierbei die Stelle „unter dem Kinn, unter der Zunge", während die letzte Zeile Aspekte früherer Zeilen zusammenführt zu „Oktoberkeule und schnarrendes Maulhorn, wo *unsere* Schatten einander entgegen sind". Die stärkste Verminderung im Vergleich zum Typoskript geschieht durch den grünen Stift, da in der zweiten Hälfte des Textes vier ganze Zeilen und drei teilweise gestrichen werden. Damit werden etwa auf jene Passagen verzichtet, die beim Typoskript mehrmals vorkommen. Am meisten trifft es hierbei die Zahl der Substantive und Satzzeichen mit neun bzw. sieben Tokens weniger. Um je drei Tokens reduziert sich die Anzahl der Pronomen und Artikel, um vier jene der Verben. Die

Präpositionen verlieren zwei Tokens, die Subjunktoren und Hilfsverben je eines. Allein die Menge an Adjektiven steigt um ein Vorkommen.

Das Typoskript von Textzeuge 5 zählt vergleichsweise wenig Tokens mit 146 und es kann eine gewisse Ähnlichkeit zur Revisionsstufe mit dem grünen Stift des vorigen Textzeugen ausgemacht werden. Denn die Anzahl der Satzzeichen, Verben, Subjunktoren, Eigennamen Pronomen und Adverbien ist dieselbe. In allen anderen POS-Verteilungen der Textstufen von Textzeuge 4 ist die Menge der Tokens jeweils geringer. Auch in Hinblick auf die Übernahme von Eingriffen scheint beispielsweise die Reihenfolge am meisten an den Änderungen mit dem Grünstift angelehnt. Keinen Einfluss auf die Tokenanzahl als auch auf die Häufigkeitsverteilung haben die Eingriffe vom grünen Stift, da hier lediglich die Anordnung

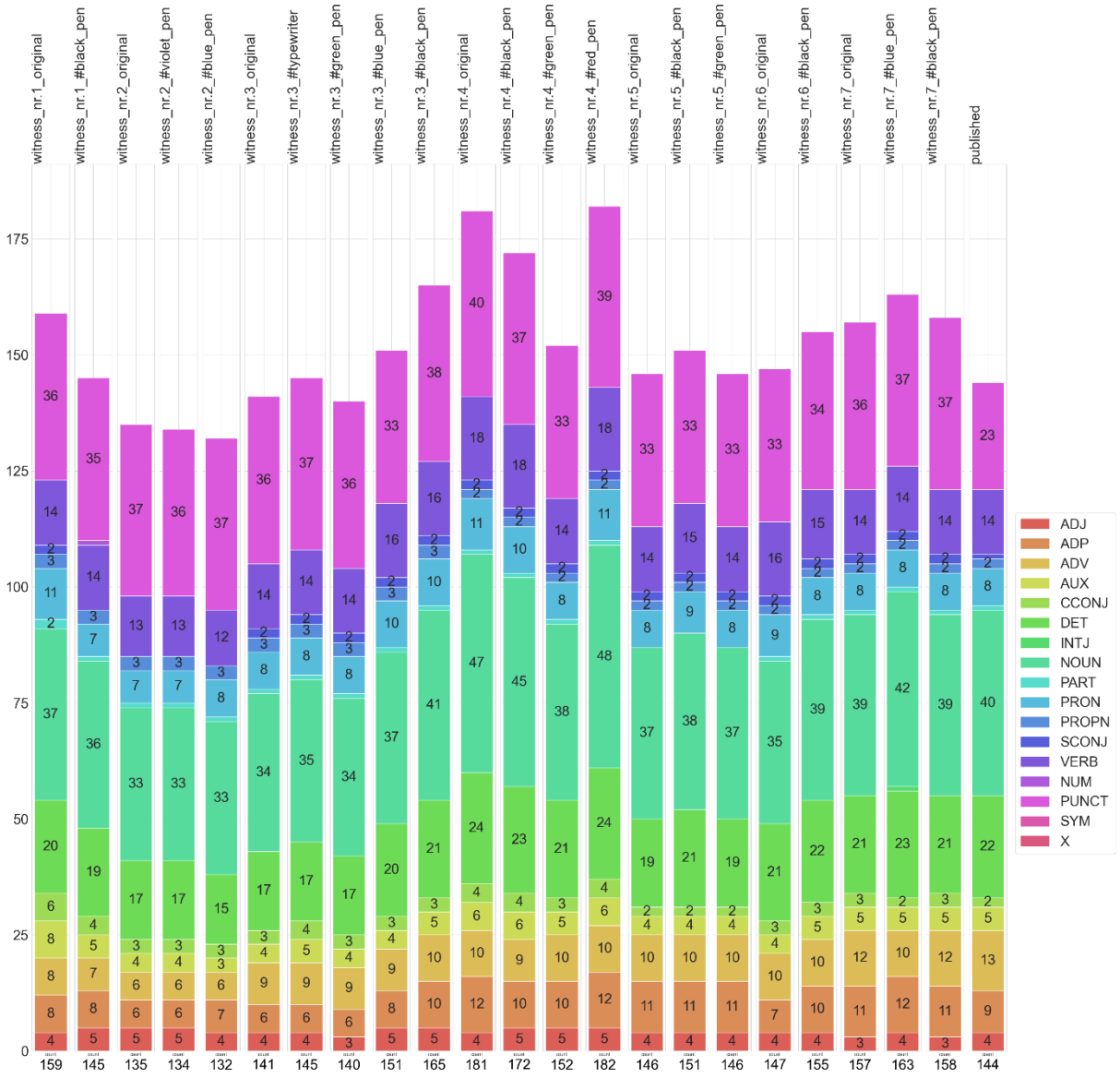

Abbildung 39: Absolute Häufigkeitsverteilung der Parts of Speech im Gedicht *VWDAZB* über alle Textzeugen und deren Revisionsstufen hinweg bis zur publizierten Fassung.

bestimmter Begriffe gegen Ende des Textes neu gewählt wird. Mit dem schwarzen Stift erbringen die Revisionen einen leichten Zuwachs der Tokenvorkommen. Um je ein Token erhöht sich die Summe der Substantive, Pronomen und Verben. Zwei Token mehr verzeichnen die Artikel. Mit Blick auf den Textzeugen zeigt sich, dass es sich um zwei Einfügungen handelt: „*ein* Kälberleib“ und „mein Gaumen krümmt sich“.

Im Vergleich zum Typoskript des Textzeugen 5 zählt jenes von Textzeuge 6 ein Token mehr. In Hinblick auf die Verteilung ist dabei aber zum einen eine Verminderung um vier Tokens bei den Präpositionen und um zwei Tokens bei den Substantiven zu erkennen. Zum anderen zählen die Artikel und Verben je zwei Tokens mehr, während die Konjunktoren, Partikel und Pronomen je ein Token mehr verbuchen. Grund hierfür ist das zweifache Vorkommen des Satzes „mein Gaumen krümmt sich“ als auch das Fehlen der Wortfolgen „Schläfe mit Kappe“, Asche sein“ und „unter der Zunge, unter dem Kinn“, welche jedoch allesamt durch den Eingriff des schwarzen Stifts wieder eingefügt bzw. auf ein Auftreten reduziert werden. Damit geschieht ein Zuwachs. Um je ein Token steigt die Anzahl der Satzzeichen, Artikel und Hilfsverben. Umgekehrt sinkt das Vorkommen der Verben und Pronomen um je ein Token. Größerer Anstieg erfahren die Präpositionen mit drei Tokens mehr sowie die Substantive mit vier zusätzlichen Tokens.

Während beim Typoskript von Textzeuge 7 im Vergleich zur vorigen Revisionsstufe nur zwei Tokens insgesamt hinzukommen, nehmen die Vorkommen von Adjektiven, Artikeln und Verben um je ein Token ab. Es gibt zwei Fälle weniger an Satzzeichen. Stattdessen nehmen die Präpositionen um ein, die Adverbien um zwei Token(s) zu. Es handelt sich bis auf eine Stelle Wort für Wort um denselben Text (ausgenommen der Zeichensetzung), dessen Tokens jedoch in vier Fällen beim automatischen POS-Tagging unterschiedlich kategorisiert wurden. Es handelt sich hierbei um die Begriffe „bewaldet“, „viel [Hitze]“, „schnarrend(es)“ und „entgegen“. „bewaldet“ wird in der Revisionsstufe von Textzeuge 6 als Verb gehandelt, während es in Textzeuge 7 als Adverb getaggt wird. Das Partizip ist im Gedicht durchwegs isoliert und wird nicht in einen direkten Kontext bzw. Satz eingebunden, weswegen eine klare Einordnung als Verb oder Adjektiv tatsächlich nicht möglich ist. In der Wortfolge „viel Hitze“ wird „viel“ als Artikel klassifiziert und in Textzeuge 7 als Adverb. Ob es sich um einen Fehler beim Abschreiben handelt, dass das Partizip bei „schnarrendes Maulhorn“ in Textzeuge 7 nicht flektiert wird und aufgrund dessen anstelle eines Adjektivs als Adverb eingestuft wurde, kann nicht ergründet werden. Im letzten Satz des Gedichts wird „entgegen“ erst als Adverb und

darauf als Präposition eingeordnet. In diesem Kontext scheint jedoch die adverbiale Verwendung vorzuliegen. Insgesamt zeigt dieses Beispiel auf, inwiefern es wichtig ist, die quantitative Analyse qualitativ zu begleiten und zu stützen. Bei der Arbeit mit Korpora kann das POS-Tagging nicht für jedes Token überprüft und korrigiert werden, zudem erweisen sich teils isolierte oder elliptische Passagen als Unsicherheitsmomente für die computergestützte Wortartenerkennung bei Gedichten.

Die Revision mit dem schwarzen Stift in Textzeuge 7 hat minimale Auswirkungen, da allein ein Beistrich eingefügt wird. Steigt die Menge an Satzzeichen auf gleiche Weise durch die Eingriffe mit dem blauen Stift, so dort auch nehmen die Adjektive, Präpositionen, Interjektionen und Satzzeichen um je ein Token zu. Die Zahl der Artikel steigt um zwei, jene der Substantive um drei Tokens. Im Gegenzug reduziert sich die Menge an Subjunktoren um einen, jene an Adverbien um zwei Fälle. Eingefügt wird der Begriff „Kunstwellen" in der Mitte des Texts, wodurch die Alliteration hin zur Assonanz „Kunstwellen Kälberleib Bekröter sogar" etabliert wird, und in der zweiten Hälfte die seit Textzeuge 1 nicht mehr vorkommende Genitivkonstruktion „Säge des Laubs" eingefügt wird. Neben kleineren Ergänzungen wird an anderen Ausdrücken weiter gefeilt wie „Spielkarten" zu „Zigeunerkarten", „gemischt" zu „vermischt", „und schnarrend Maulhorn" zu „ach dein schnarrendes Maulhorn".

Mit 144 Tokens ist die Gesamtzahl der publizierten Gedichtfassung vergleichsweise niedrig. Stark reduziert ist das Vorkommen der Zeichensetzung. Ansonsten gibt es kleinere Verschiebungen bei den Adjektiven, Präpositionen, Adverbien, Subjunktoren, Artikeln und Substantiven. Streckenweise sind die Änderungen der Revisionsstufe von Textzeuge 7 umgesetzt, gleichzeitig gibt es zwei kleinere Änderugnen in der Reihenfolge. Auffällig ist das Fehlen der Wortfolge „unter der Zunge, unter dem Kinn". An drei Stellen werden Worte ausgetauscht wie „aus jeder Pfütze" anstelle von „von jeder Pfütze" oder „Asche werden" statt „Asche sein". Der Titel ist schließlich „vielleicht weil das Auge zuerst bricht".

In Hinblick auf die POS-Verteilungen innerhalb der Tags für Eingriffe ist auffallend, dass diese teils relativ ähnliche Häufigkeiten enthalten. Die Gesamtzahlen der Tokens sind bei <del> und <add> allein um zwei verschieden. Obwohl es weniger Streichungen gibt, so sind beinahe die gleichen Wortarten involviert, wobei mehr Nomen, Artikel und Adverbien eingefügt, aber beispielsweise mehr Konjunktoren, Pronomen und Satzzeichen gelöscht werden. Die Tags <orig> und <reg> unterscheiden sich um neun Tokens, hingegen wird ersichtlich, dass im Fall

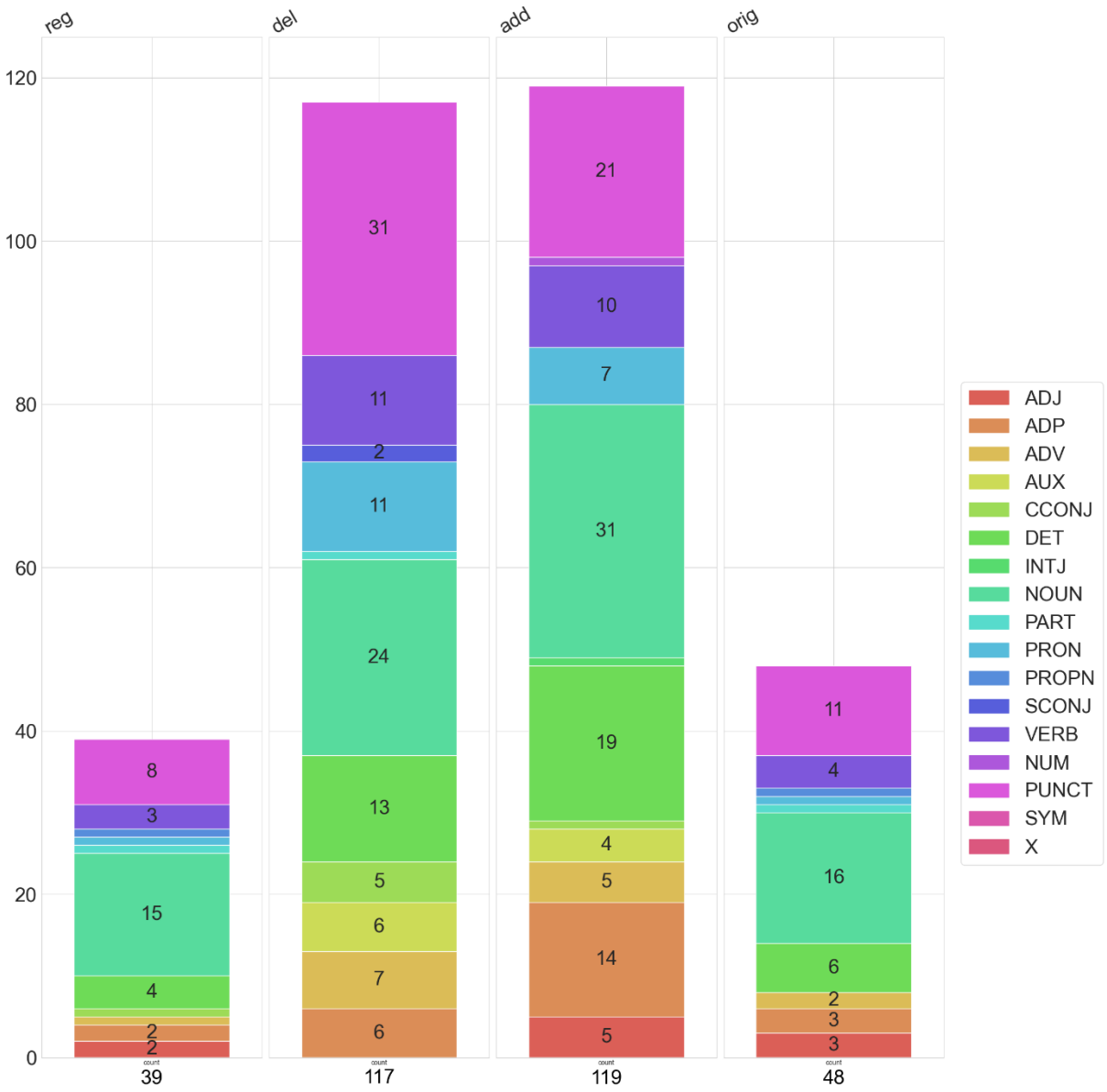

Abbildung 40: Absolute Häufigkeitsverteilung der Parts of Speech in allen Eingriffs-Tags des Gedichts *VWDAZB*.

dieses Gedichts die ursprünglichen Formulierungen umfangreicher gewesen sein müssen als die Überarbeitungen.

Von zehn dominanten Sequenzen überschneiden sich teilweise die Hälfte davon. Auffallend sind dabei insbesondere die Folgenden: „Präposition – Artikel – Substantiv“, „Präposition – Artikel – Substantiv – Verb“, „Artikel – Substantiv – Verb“, „Artikel – Substantiv – Verb – Artikel“ und „Substantiv – Präposition – Artikel – Substantiv“. Bezeichnend ist unter allen Beispielen die Stelle „Beständigkeit aus den Karten gelesen“. Von Relevanz erscheinen auch die Sequenzen „Substantiv – Artikel – Substantiv“ („Krankheit der Witwe“, „Säge des Laubs“) und „Artikel – Adjektiv – Substantiv“ („die grünen Zigeunerkarten“, „das menschliche Auge“).

Während manche Sequenzen mit Anbeginn in geringerer Zahl vorzufinden sind, sind bei einigen insbesondere die Veränderungen über Textzeuge 3 und Textzeuge 4 für die vollständige Etablierung wichtig. Beispielsweise ist die Sequenz „Präposition – Artikel – Substantiv“ von Anbeginn stark vertreten und auch jene mit dem zusätzlichen Verb am Ende taucht in zwei Fällen auf, jedoch werden erst mit den Revisionen des grünen Stifts in Textzeuge 3 drei Vorkommen daraus und werden weiterhin beibehalten.

8 Vergleich & Diskussion

Die in der Analyse erzielten Ergebnisse werden im Folgenden sowohl mit Fokus auf die Textgenese eines jeden Gedichts zusammengefasst als auch übergreifend für die erfassten Kategorien, wie Parts of Speech, die Art der Eingriffe, Gesamtzahl der Tokens und Sequenzen. Infolgedessen lässt sich abschätzen, ob trotz des geringen Beispielkorpus systematische Aussagen in Hinblick auf etwaige Bearbeitungstendenzen von Mayröcker getätigt werden können. Darüber hinaus wird auf Problemstellungen und mögliche Verbesserungen hingewiesen.

8.1 Ergebnisse

Dem Aufbau der Analyse folgend sind zunächst die handschriftlichen und maschinellen Eingriffe in die Typoskripte der Gedichte in den Blick zu nehmen. Obwohl beim Gedicht *APTRM* die Eingriffe von Textzeuge zu Textzeuge schrittweise weniger werden, stellt dies eine Ausnahme dar. So weisen die Bearbeitungsstufen der anderen vier Gedichte jeweils unterschiedlich viele Eingriffe auf. In drei der fünf Fälle (*NWTRG, RSFGMT, VWDAZB*) sind beim letztbearbeiteten Textzeugen mehr Eingriffe verzeichnet als im zweitletzten. Dies könnte daraufhin deuten, dass gegen Ende noch einmal intensivere Bearbeitungen vorgenommen werden. Die Häufigkeitsverteilung der Eingriffs-Tags gibt kaum weitere Hinweise auf spezifische Tendenzen, allein die Zahl an <reg>-Tags nimmt mehrheitlich im Laufe der Textzeugen ab (*APTRM*, *NWTRG*, *WSNMP*). Verallgemeinerungen sind insgesamt aber nicht möglich, es ist vielmehr naheliegend, dass jedes Gedicht in seiner Entstehung individuell intensive Bearbeitungsvorgänge beansprucht.

Mit Blick auf die involvierten Parts of Speech in den Eingriffs-Tags lassen sich ebenso zusammenfassende Aussagen formulieren. In vier von fünf Fällen sind mehr Tokens in den <add>-Tags als in den <del>-Tags. Große Ausnahme ist hierbei das Gedicht *RSFGMT*, bei welchem doppelt so viele Streichungen wie Einfügungen verzeichnet sind. Zudem sind in vier von fünf Fällen mehr Tokens in <reg>-Tags als in den <orig>-Tags. Ausnahme ist hier das Gedicht *VWDAZB*. Dies bedeutet insgesamt, dass durch die Eingriffe in den Textzeugen tendenziell mehr Tokens hinzugefügt als gelöscht werden.

Auch im Kontext der Gesamttokenzahl der Revisionsstufen im Vergleich zu jener ihrer Typoskripte zeigt sich mehrheitlich eine Tendenz zur Zunahme – mit Ausnahme des Gedichts *RSFGMT*. Zwischen den Typoskripten (ohne Einbezug der handschriftlichen Eingriffe) ist in drei der fünf Fälle häufiger eine Steigerung der Tokensumme zu entnehmen als eine Abnahme wie bei den Gedichten *RSFGMT* und *WSNMP*. Wird jedoch nur die Zahl der Tokens von Textzeuge 1 und der publizierten Gedichtfassung miteinander verglichen, ist mit Ausnahme von *APTRM* eine Verringerung der Tokens abzulesen. Dies ist mehrheitlich auf eine starke Reduktion (um etwa 30 Tokens oder mehr) zwischen zwei Typoskripten zurückzuführen. In diesem Ausmaß zumeist ist dies nur einmal in der Textgenese eines Gedichts auszumachen, jedoch hat es wesentlichen Einfluss auf die weitere Gesamtzahl der Tokens. Insofern ist die Beobachtung von Kastberger (1998a) in Bezug auf die Textgenese von *Reise durch die Nacht* hier zu bekräftigen, dass Mayröcker den Gedichttext sowohl ausformuliert und entfaltet als auch reduziert und verdichtet. Die Reihenfolge dieser Vorgänge ist bei den Gedichten jedoch nicht so einheitlich.

In der Kombination aus Gedichten (5) und der Art der Eingriffe (4) ergeben sich insgesamt 20 Fälle, aus denen die jeweiligen Wortartenverteilungen abgelesen werden können, die von den Revisionen betroffen sind. Am meisten involviert in die Eingriffe sind die Substantive. Diese machen generell bei allen Gedichttexten durchschnittlich etwa ein Viertel der Tokens aus. In 14 von 20 Fällen sind die Nomen die meistbetroffene Wortart der Revisionen. Dabei nehmen sie bis zu über 30% der Tokens pro Eingriffsart ein. Insbesondere in allen <reg>- und <orig>-Tags sind die Substantive an erster Stelle. Grund hierfür ist auch die editorische Wahl, komplexe Veränderungen an Wörtern mit dem <reg>-Tag auszuzeichnen. In allen anderen Fällen ist diese Wortart zumindest in den meistinkludierten Wortarten in Eingriffen vorhanden. Am wenigsten scheinen sie noch bei den <add>- und <del>-Tags des Gedichts *Alptraum* auf.

Sehr prominent bei den Revisionen sind die Satzzeichen. In 4 von 20 Fällen stellen die Satzzeichen die dominanteste Kategorie der POS von Revisionen dar. In drei Fällen tauchen sie in den <del>-Tags der Gedichte auf: *RSFGMT, WSNMP* und *VWDAZB*. Aber auch im <add>-Tag von *WSNMP* stehen sie an erster Stelle. Während die Satzzeichen prozentuell im Text sowie in den Eingriffstags bei dem Gedicht *APTRM* eine geringe Rolle spielen (8,9%), nehmen sie bei den Gedichten *NWTRG* und *VWDAZB* 15-16% ein. Bei *RSFGMT* und *WSNMP* nehmen etwa 26% des Textes die Satzzeichen ein. Ob dabei spezifische Muster dahinterstecken, wie das Vorkommen bestimmter Satzzeichen oder eine besondere Strukturierung, wäre noch weiter

zu ermitteln. Zudem wäre abzugleichen, inwiefern das Untergliederungssystem der Satzzeichen, welches Kastberger bei der Textanalyse von *Reise durch die Nacht* aufstellt, in ähnlicher Form bei Mayröckers Gedichten zu finden wäre (vgl. Kastberger 2000c, S. 83). Mit dem konkreten Fokus auf die POS konnte dieser Aspekt hingegen nicht weiter verfolgt werden. Die Artikel stehen nur in einem Fall, nämlich im <del>-Tag des Gedichts *APTRM*, im Vordergrund. Insgesamt ist diese Wortart zudem stark in den meistbeteiligten POS der Revisionen enthalten. Die Verben sind in 12 von 20 Fällen unter den POS mit den meisten Vorkommen in Eingriffen vorhanden. Aber wie bei den Artikeln sind die Verben nur in einem Fall die meistvertretene Wortart, nämlich im <add>-Tag des Gedichts *APTRM*. Hierzu ist zu erwähnen, dass die Verben etwa 14% des Gedichttexts von *APTRM* ausmachen, während bei allen anderen Gedichten nur 7% bis maximal 9,7% der Tokens Verben sind. Zudem ist *APTRM* dasjenige Gedicht, das über die wenigsten elliptischen Satzgefüge verfügt.

Öfters stärker ausgeprägt sind die Präpositionen. Während diese bei den <add>-Tags der Gedichte *RSFGMT* und *VWDAZB* jeweils an vierter Stelle stehen und beim <del>-Tag des Gedichts *APTRM* an zweithöchster, ist die Wortart bei allen Eingriffs-Tags des Gedichts *WSNMP* entweder am zweit- oder drittprominentesten und damit eine sehr präsente Wortart für die Bearbeitungen dieses Gedichts. Insgesamt machen bei der publizierten Fassung des Gedichts die Präpositionen 12,3% aus (im Vergleich haben die Gedichte *RSFGMT* und *VWDAZB* nur die Hälfte davon).

Sechsmal stärker präsent sind die Adverbien in den Revisionstags. In den <add>-Tags der Gedichte *NWTRG* und *WSNMP* kommen sie vor, in allen anderen Tags sind die Adverbien beim Gedicht *APTRM* sowie auch beim <reg>-Tag des Gedichts *WSNMP* vorhanden. Die Adjektive sind prinzipiell eher selten in Verwendung bei Mayröcker (zwischen vier bei *VWDAZB* und neun Fällen bei *WSNMP*). Zudem tauchen die Adjektive bei *WSNMP* an vierter Stelle in den <orig>- und <reg>-Tags auf. Des Weiteren sind sie an dritter Stelle bei den Einfügungen im Gedicht *Alptraum* vorzufinden.

Wenn Ersetzungen bzw. komplexere Überarbeitungen vonstattengehen, ist zu erkennen, dass sich diese auf die paradigmatische und nicht auf die syntagmatische Ebene einer Passage auswirken. Insofern haben solche Revisionen keinen Einfluss auf die Wortartenverteilung, da für die Position im Satz ein anderer oder leicht veränderter Ausdruck inhaltlicher Art gewählt wird. Als Beispiele sind hier Überarbeitungen zu nennen wie „[wie Mondfahrer] ausgerüstet“ (Textzeuge 4 von *APTRM*) anstelle von „ausgestattet“ (Textzeuge 3), die Entwicklung von

„Feld“ über „Acker“ hin zu „Stoppelfeld“ (Textzeuge 3, 4 und publizierte Fassung von *WSNMP*) oder die Entwicklung von Wechsel von „Westentasche“ zu „Brust“ hin zu „Brusttasche“ bis zu „Tasche“ (Textzeuge 1-4, *VWDAZB*). Inwiefern bei manchen dieser Ersetzungen sowohl auf die Spezifizierung des Sinngehalts als auch auf den Klang oder den Rhythmus geachtet wird, könnte genauer untersucht werden. Vereinzelt gibt es dennoch entscheidende Eingriffe, wie sie konkret in der Analyse beschrieben wurden. Hierbei lässt sich zusammenfassen, dass der Wechsel von Substantiven zu Adjektiven und umgekehrt auffallend erscheint (beispielsweise von „mit der linken Hand“ zu „mit der Linken“ in Textzeuge 2 von *VWDAZB*). Weiters zeigt sich in zwei Gedichten die Ersetzung von Verben durch adjektivisch oder adverbial verwendete Partizipien, so wird etwa in Textzeuge 1 von *APTRM* aus „knirschen“ „knirschend“ oder in Textzeuge 1 von *RSFGMT* „erkennen“ zu „erkennbar“. Nur einmal im Gedicht *NWTRG* tauchte ein Fall auf, der als Substantivierung eines Verbs bezeichnet werden kann, wie Mayröcker ihn in ihrem Text *MAIL ART* benannt hat. Insgesamt scheint die Beibehaltung des Wortstamms bei diesen Veränderungen teils kennzeichnend. Nicht wenige Fälle dieser Streichungen, Ersetzungen oder Veränderungen sind „Korrekturen von Korrekturen“ (Mayröcker 1983b, S. 14) geschuldet.

In Hinblick auf die Veränderungen der Wortartenverteilungen für jedes Gedicht zeigen sich oftmals geringe Unterschiede der Häufigkeiten. Dies wurde bereits zu Beginn des Analyseteils mithilfe der inferenzstatistischen Werte aufgezeigt. Im Gedicht *APTRM* bleiben viele Wortarten tendenziell gleich. Während die Zahl der Verben, Adjektive, Pronomen und Satzzeichen zunehmen, sinkt dagegen die Anzahl der Präpositionen und Adverbien oft. Beim Gedicht *NWTRG* sind es die Substantive und Satzzeichen, deren Summe steigt. In geringerem Ausmaß ist dies auch bei allen anderen POS der Fall, mit Ausnahme der Hilfsverben, welche abnehmen. Gleich verhält es sich für die Hilfsverben beim Gedicht *RSFGMT*, aber auch hier erhalten die Substantive Zuwachs. Während sich bei den restlichen Wortarten nur kleine Veränderungen absehen lassen, ist eine steigende Tendenz bei den Pronomen abzulesen, während sich die Zahl der Präpositionen, Adverbien, Artikel, Verben und Satzzeichen mehrheitlich vermindert. Umgekehrt zeichnen sich beim Gedicht *WSNMP* kaum Abnahmen der Wortarten ab. Während die Menge an Adjektiven, Präpositionen, Hilfsverben und Partikeln nur teilweise steigt, so nimmt sie größtenteils bei den Adverbien, Artikeln, Substantiven, Pronomen und Satzzeichen zu. Für das Gedicht *VWDAZB* zeigen sich für einige POS nur geringe Veränderungen. Darunter ist für die Zahl der Adverbien und Verben eine leicht steigende, für die Zahl der Konjunktoren

und Partikel eine sinkende Tendenz zu erkennen. Während die Summe an Adjektiven, Präpositionen, Artikel und Substantive zumeist steigt, nimmt jene der Satzzeichen vornehmlich ab.

Für die Wortarten im Einzelnen lassen sich anhand der Analyse von fünf Gedichten keine eindeutigen Muster ablesen, jedoch zeigen sich zumindest einzelne Gruppierungen mit Tendenzen. Die Zahl der Adjektive steigt im Laufe der Gedichtbearbeitungen in zwei Fällen merklich, in zwei Fällen leicht und in einem Fall ist sie gleichbleibend. Die Präpositionen werden in drei Gedichten mehr, umgekehrt werden sie in zwei Gedichten weniger. Bei den Adverbien ist die Tendenz dieselbe. Gleichbleibende Zahlen verzeichnen die Hilfsverben in drei Fällen, in den anderen zwei Fällen ist jedoch eine Abnahme zu beobachten. Teilweise ist dieser Abnahme auch der Wechsel von Perfekt zu Präteritum geschuldet. In den drei Gedichten, in denen Konjunktoren Anwendung finden, ist ebenfalls kaum eine Veränderung zu beobachten, wenn auch die Tendenz eher sinkend ist. Während in zwei Fällen die Artikel eine Zunahme erfahren, bleiben in den restlichen Fällen die Veränderungen gering. Eindeutig steigt die Anzahl der Substantive bei vier der fünf Gedichte, allein bei *APTRM* gibt es kaum Änderungen bzw. ist es vereinzelt ausgleichend. Die gering vorkommende Wortart der Partikel ist wenig von Eingriffen betroffen und wird dennoch in Einzelfällen vermehrt gestrichen als eingefügt. Für die Pronomen ist ein leichter Anstieg der Anzahl in vier Fällen zu beobachten, eine eindeutige Zunahme ist beim Gedicht *WSNMP* auszumachen. Subjunktoren sind kaum Ziel nennenswerter Revisionen. Dabei ist für die Gedichte *NWTRG* und *VWDAZB* eine leicht steigende, für das Gedicht *APTRM* eine eindeutig steigende Tendenz auszumachen, für die zwei anderen Gedichte das Gegenteil. Die Vorkommen der Verben bleiben mehrheitlich unverändert, nur teilweise ist eine Tendenz zur Zunahme abzulesen wie im Gedicht *APTRM*. Zuletzt ist für die Satzzeichen mit drei Fünftel eine Zunahme der Anzahl auszumachen, bei den Gedichten *VWDAZB* und *RSFGMT* verringern sich die Summen. Wie Mayröcker ihre „Armee von Satzzeichen“ (Mayröcker 1983b, S. 21–22) einsetzt, scheint dementsprechend vom individuellen Schreibprozess abhängig.

Die Anzahl der Gedichte ist wenig aussagekräftig und müsste erweitert werden, um überprüfen zu können, ob sich die Dreifünftelmehrheit vieler Tendenzen bestätigen könnte. Die gleichbleibenden Zahlen für unterschiedliche POS kann damit in Verbindung gebracht werden, dass nicht wenige Stellen der Gedichttexte teils von Anbeginn unberührt bleiben, während andere Passagen mehr Revisionen erfahren. Genauer zu überprüfen wäre dies in Anlehnung an

Van Hulle (Vgl. Van Hulle 2022, S. 344), welcher durch die Kollationierung der Fassungen ermittelt, wie viel Prozent des Textes der vorhergehenden Fassung unverändert bleibt.

In Hinblick auf die Chi-Quadrat-Tests, die für alle Datensätze (POS-Häufigkeitsverteilung, Eingriffs-Tags etc.) durchgeführt wurden, zeigten sich keine signifikanten Unterschiede zwischen den POS-Verteilungen der Textzeugen und Textstufen. In erster Linie ist dies damit vereinbar, dass die Häufigkeitsverteilungen der Fassungen sehr ähnlich sind und es sich zumeist nur um wenige Veränderungen handelt. Zudem könnte dieses Ergebnis darauf zurückgeführt werden, dass manche Kategorien teils einen geringen Umfang haben. Beispielsweise sind POS wie Konjunktoren, Subjunktoren, Pronomen und Partikel bei der Verteilungen spärlich vorhanden und lassen damit die Ergebnisse weniger aussagekräftig werden. Denn die Voraussetzung, dass pro Zelle eine Häufigkeit von mindestens 5 erreicht werden muss (vgl. Rasch et al. 2014, S. 129), ist bei vielen der vorhandenen Datensätze nicht gegeben. In welchem Maß dies einen Einfluss auf die Ergebnisse hat, könnte weiter ermittelt werden.

Die Chi-Quadrat-Tests bei Datensätzen zu den Revisionen weisen dagegen auf teils signifikante Unterschiede hin. So ist zu beobachten, dass bei den Gedichten *NWTRG*, *RSFGMT* und *VWDAZB* sowohl die Verteilungen der Attributwerte (Stiftfarbe) als auch die Kombination aus Tag (Eingriffsart) und Attributwert markante Differenzen bergen. In den Fällen, in denen inferenzstatistisch klar ersichtlich ist, dass diese Unterschiede nicht zufällig sein können, scheint es naheliegend, die Textzeugen als Interpretationshintergrund heranzuziehen. Es kann vermutet werden, dass der Wechsel des Arbeitswerkzeugs und der dominierenden Revisionsweise des Texts beispielsweise zeitliche sowie geografische Veränderungen des Arbeitsumfelds implizieren könnte. Textzeugen bzw. Textstufen, welche nacheinander keine signifikanten Unterschiede aufweisen, sind womöglich eher zeitlich und geografisch nahe verfasst worden als Textzeugen, die sich durch eine stark abweichende Verteilung von Stiftfarben auszeichnen. Insgesamt könnten in dieser Hinsicht weitere Auseinandersetzungen zu weiteren Schlussfolgerungen kommen, die im Rahmen dieser primär deskriptiven Analyse nicht weiterverfolgt werden konnten.

In Bezug auf die Alliterationen in den Gedichten ist zu schlussfolgern, dass diese zwar vereinzelt bereits in den ersten Textzeugen vorhanden sind, sich aber alle weiteren erst im Laufe des Schreibprozesses etablieren. Dabei sind es vorrangig Einfügungen, die die rhetorische Figur wirksam machen. Neue Abschriften oder Einfügungen in Kombination mit Streichungen sind seltener Grund für die Bildung von Alliterationen. Die weitaus präsenteste Wortart in den

Alliterationen ist jene der Substantive. Artikel, Verben, Adjektive, Adverbien und Präpositionen machen weniger als die Hälfte der mitwirkenden POS bei Alliterationen aus. Zumindest teilweise sind die Alliterationen (mit mehrheitlich Substantiven) auch in den Sequenzen wiederzufinden.

Insgesamt werden dreizehn Sequenzen identifiziert, die bei mehr als einem Gedicht in der publizierten Fassung vorkommen. Jedoch gibt es zumeist mindestens ein zusätzliches Gedicht, bei welchem diese Sequenz in den Typoskripten bzw. Revisionsstufen ebenfalls mehrmals präsent ist, jedoch für die veröffentlichte Fassung nicht weiter ausgebaut wird.

Die weniger stark vertretenen Sequenzen ergeben sich häufig innerhalb von Aufzählungen bzw. durch das Aufeinandertreffen von Ende und Beginn zwei unterschiedlicher, inhaltlicher Einheiten. Diese ausschnitthaften Stellen zeigen sich bei den Sequenzen „Adjektiv – Substantiv – Artikel“ (wie „violette Licht, die“ in *RSFGMT* und „schwankenden Sattel des“ in *WSNMP*), „Substantiv – Adjektiv – Substantiv“ (im Fall von „Fremdflocke, magisches Werk“ in *RSFGMT* und „Baums .. singender Brauenvogel“ in *NWTRG*) und „Substantiv – Artikel – Adjektiv – Substantiv“ (beispielsweise „Schamanenlied / die perlmuttfarbene Frühe“ in *NWTRG* und „Oktoberkeule dein schnarrendes Maulhorn“ in *VWDAZB*). Während die aufgezählten Sequenzen jeweils in den zwei aufgeführten Gedichten prominenter vertreten sind, tauchen die zwei erstgenannten Sequenzen auch in den Textzeugen 1, 2 und 4 des Gedichts *VWDAZB* und letztere Sequenz im Textzeugen 1 von *RSFGMT* mehr als einmal auf. Drei weitere Sequenzen sind jeweils in drei publizierten Gedichtfassungen vertreten, zudem in Arbeitsstufen weiterer Gedichte. Auffallend scheint die Präsenz der Präpositionen und damit teils inhaltlich vollständigeren Wortfolgen. Zu diesen Sequenzen gehören „Adjektiv – Substantiv – Präposition“ (zum Beispiel „zugefrorenen See entlang“ in *APTRM* aber auch „erloschene Traumgesicht, hinter“ in *RSFGMT*), „Substantiv – Präposition – Substantiv“ (wie „Frühe im Fenster“ in *NWTRG* sowie „Halme am Schuh“ in *WSNMP*) und mit dem teils erforderlichen Artikel in der vorigen Folge „Substantiv – Präposition – Artikel – Substantiv“ (etwa „Göttin aus der Maschine“ in *VWDAZB* und „Lauf durch die Stoppelfelder“ in *WSNMP*). Abgesehen vom Textzeugen 1 des *RSFGMT*s taucht die Sequenz „Artikel – Adjektiv – Substantiv“ in den anderen vier Gedichten mehrmals auf, so beispielsweise in Form von „die jungen Fremden“ in *APTRM* oder „eine zersprungene Ofenlamelle“ in *NWTRG*. Teilweise überschneidend sind die Sequenzen „Präposition – Artikel – Substantiv“ und „Artikel – Substantiv – Verb“, wobei erstere in allen fünf veröffentlichten Gedichtfassungen zu finden ist und zweitere nur in den

Fassungen von *NWTRG* kaum aufscheint. Die angesprochenen Überschneidungen zeigen sich beispielsweise in „von den Dachtraufen dräuen“ in *APTRM*, „aus jeder Pfütze getrunken“ in *VWDAZB* und „auf der Treppe geschlafen“ in *NWTRG*. Häufiger sind es jedoch einzelne Wortfolgen wie „die Zeit weiterging“ in *RSFGMT* oder „aus der Maschine“ in *VWDAZB*.

Vier Sequenzen bestehen aus unterschiedlichen Kombinationen bzw. Längen der Abfolge von Artikeln und Substantiven. Während in zwei Gedichten die Folge „Substantiv – Artikel – Substantiv – Artikel“ (beispielsweise „Liebfrauentag, ein Lebens- ein“ in *RSFGMT*) identifiziert wird, wird abgesehen von *WSNMP* in allen anderen Gedichten die Abfolge „Artikel – Substantiv – Artikel – Substantiv“ sowie dieselbe Sequenz exklusive des letzten Substantivs gefunden (etwa „des Laubs, mein [Gaumen]“ in *VWDAZB*). Allein die Form „Substantiv – Artikel – Substantiv“ ist in allen fünf Veröffentlichungen vermehrt zu finden. Zwar sind dies beispielsweise zusammenhängende Beschreibungen wie „Zimmertelefon des Berghotels“ in *APTRM*, „Stativ deiner Beine“ in *VWDAZB* oder „Sattel des Bergs“ in *RSFGMT*, aber insgesamt betrachtet sind nicht einmal die Hälfte davon Genitivkonstruktionen. Die Mehrheit der konkreten Wortfolgen dieser Artikel-Substantiv-Sequenzen entsteht aus den bereits erwähnten Aufzählungen bzw. aneinander anschließende Sinneinheiten, wie „den Mond, das Land“ in *VWDAZB* und „der Schellenbaum die Pollution“ in *NWTRG*.

Damit sind in den häufigsten Sequenzen aller Gedichte Substantive, Artikel, Adjektive, Präpositionen und Verben in eben dieser Reihenfolge vertreten. Mit diesen Wortarten werden essentielle Sequenzen und Inhalte formuliert. Bei Mayröckers Erwähnung von „ästhetischen Verdichtungs- und Verdünnungszonen“ (Mayröcker 1983b, S. 21–22) scheinen insbesondere Sequenzen mit Substantiven auch komprimierten Sinngehalt zu fassen. Darüber hinaus können Sequenzen auch als eine Form von „Wiederholungen als Leitmotive“ (Mayröcker 1983b, S. 21–22) gesehen werden, da die gemeinsamen Sequenzen meist vermehrt in jedem Gedicht zu finden sind.

Die Gedichte bewegen sich, trotz ihrer unterschiedlichen Länge, im selben Rahmen in Hinblick auf die Anzahl jener Sequenzen, die sie mit mindestens einem anderen Gedicht teilen. Die Gedichte *APTRM*, *NWTRG* und *WSNMP* teilen mit allen Gedichten je neun Sequenzen, *RSFGMT* zehn und *VWDAZB* elf. Insofern hat *APTRM* trotz der geringsten Tokenzahl (97) dennoch im Vergleich mit anderen Gedichten wie *WSNMP* mit 171 Tokens viele (gemeinsame) Sequenzen vorzuweisen. Die Anzahl der vermehrt vorkommenden Sequenzen wechselt bei jedem Gedicht und spezifischen Textzeugen teils stark. Nicht wenige Sequenzen werden im

Laufe der Überarbeitungen wieder zurückgebaut. Im Vergleich der letzten Textzeugen mit den publizierten Fassungen zeigt sich in vier von fünf Fällen jedoch ein Anstieg an identifizierbaren Sequenzen. Wie in der Analyse aufgezeigt, bestehen einige Sequenzen bereits seit Anbeginn, zudem werden weitere im Laufe der Typoskripte und deren Überarbeitungen weiterentwickelt und ausgebaut.

8.2 Vorgehensweise & Methoden

Die Auswahl des Textkorpus mit dessen Umfang hat sich in Hinblick für den Rahmen der Arbeit als passend und ausreichend erwiesen. Die Ergebnisse zeigen ein mehrheitlich gemischtes Bild in Bezug auf die Analysekategorien, wobei sich die Gedichte oftmals in einer Gegenüberstellung von 40 zu 60 Prozent befinden. Insofern wäre für konkretere Ergebnisse in Zusammenhang mit den Bearbeitungstendenzen eine Ausweitung des Korpus erforderlich. Um den Aufwand gering zu halten, erwies sich der Fokus auf wenige Archivboxen für die Auswahl der Gedichte und Textzeugen dennoch als erfolgreich und für das Ausmaß der Arbeit als notwendig.

Die Anzahl der Textzeugen pro Gedicht ist den Umständen geschuldet und wird bzw. kann nicht den tatsächlich entstandenen Typoskripten entsprechen. Es konnten die Entwicklungsstufen nachgezeichnet werden, die teils lückenlos erscheinen. Allein die Sprünge zwischen dem ermittelten ‚Endtyposkript' und der publizierten Gedichtfassung lassen vermuten, dass es vor der Veröffentlichung zu kleineren Bearbeitungen gekommen sein muss, dessen Textzeugen jedoch nicht gefunden wurden – und womöglich verloren gegangen sind. Da teilweise die Notizzettel mit den ersten niedergeschriebenen Einfällen im Konvolut erhalten blieben, könnte genauer untersucht werden, inwiefern diese Notizen im Text weiterbearbeitet wurden, oder ob sie jenen Teil der Typoskripte ausmachen, der unverändert bleibt. Eigenständig wurde die Reihenfolge der Textzeugen ermittelt, wobei es auch hier computergestützte Hilfsmittel wie Stemmaweb[14] gäbe, die nach der Transkription der Typoskripte angewandt werden könnten. Dabei ist offensichtlich, dass die Transkription vorher stattfinden muss. Abgesehen vom Stemma sind die sogenannten ‚Alignment Tables' oder andere Formen der

[14] Vgl. https://stemmaweb.net/, zuletzt geprüft am 20.09.2024.

Variantendarstellung von CollateX[15], PyCoviz[16], TRAViz[17] oder Coleto von Ketzan und Schöch (2021) hilfreich, um beispielsweise bei der qualitativen Analyse Abweichungen oder Veränderungen zwischen den Typoskripten zu erkennen. Aufgrund des geringen Korpus erschien es nicht zwingend notwendig, diese Darstellung einzubeziehen. Bei einer Ausweitung des Textumfangs scheint es jedoch ein unterstützendes Mittel für die Analyse. Inwiefern nicht auch Möglichkeiten zur erweiterten Darstellung der Textstufen mit Einbezug der Parts of Speech (beispielsweise farbkodiert wie in den Diagrammen) realisiert werden könnten, wäre eine weiterführende Aufgabe. Erst eine solche Variante der kombinierten Darstellung aus farbkodierten POS und variierender Textstufen zum Beispiel in Form eines Alignment Tables würde eine wesentliche Unterstützung für die Analyse bilden.

Die Transkription und alle damit einhergehenden editorischen Entscheidungen zur Vereinheitlichung erwiesen sich als passend und es gab keine Probleme bei der Verarbeitung der Tokens, die auf die Transkription zurückgeführt werden könnten. Teils sind die handschriftlichen Bearbeitungen in den Typoskripten stenographisch verfasst. Dabei konnten alle Stellen erkannt werden bis auf ein Fall in Textzeuge 1 von *VWDAZB*, der mit einem <unclear>-Tag ausgezeichnet ist. Der Abgleich mit den vorherigen und nachfolgenden Typoskripten und Revisionen erwies sich als essentiell, um zusammen mit dem *Stenografie Wörterbuch* von Wawrczeck (2024) die Wörter als Laiin zu ermitteln. Abgesehen davon könnte bei allgemeinen Schwierigkeiten bei der Handschriftenentzifferung auch die Erkennung mithilfe von Transkribus[18] unterstützt werden. Dabei ist der Export des erkannten Texts auch als XML-Dokument möglich und erleichtert damit die Weiterbearbeitung.

Das Tagging der Parts of Speech erfolgte in dieser Arbeit direkt auf die Transkription. Die Entscheidung, die POS-Tags als Suffixe an die Tokens anzuhängen, bewies sich als passende Lösung. Insgesamt ist der Schritt des POS-Taggings an dieser Stelle der Vorgehensweise einfach zu bewerkstelligen, da nicht eigenhändig auf all jene Textstellen eingegangen werden muss, die sich innerhalb von Revisions-Tags befinden. Andererseits erwies sich die XML-Modellierung teils schwieriger, da die POS-Suffixe das Lesen etwas beeinträchtigen und den Überblick eingrenzen. Wird in einem editionswissenschaftlichen Kontext mit einem bereits

[15] Vgl. https://collatex.net/, zuletzt geprüft am 20.09.2ß24.
[16] Vgl. https://github.com/enury/collation-viz, ebd.
[17] Vgl. http://www.traviz.vizcovery.org/index.html, ebd.
[18] Vgl. https://www.transkribus.org/, zuletzt geprüft am 15.09.2024.

modellierten Text gearbeitet, müsste für diesen Fall ein nachträgliches POS-Tagging eingeführt werden. Ein schwierigeres Unterfangen könnte dabei sein, auf alle Tokens in bereits existierenden Tags für das Tagging zuzugreifen, ohne die existente Struktur zu verändern.

Insgesamt zeigte die Stichproben-Evaluierung, dass auch das Tagging bei Gedichten mehrheitlich funktioniert. Auf Ausnahmen und Grenzfälle wie das isolierte Wort „bewaldet“ wurde bereits eingegangen, ebenso auf die Schwierigkeit, teils Adjektive und Adverbien zu unterscheiden, wie dies im Unterkapitel 6.2 zur Qualitätsprüfung des POS-Taggings ausgeführt ist. Beinahe ohne Probleme wurden Neologismen von Mayröcker zumeist als Substantive erkannt und nur selten etwa als Eigennamen kategorisiert. Neben der Erfassung der morphosyntaktischen Wortklassen (POS), könnte das Hinzuziehen weiterer morphologischer Merkmale der Tokens wie Lemma, Tempus, Numerus, Genus, Kasus und Person zu weiteren Erkenntnissen führen. Da mit der Python Library *SpaCy* diese Merkmale ebenso ermittelt werden können, wäre es keine große Aufgabe, das Python-Skript zum POS-Tagging sowie das resultierende CSV-Dokument um diese Aspekte zu erweitern. Beispielsweise ließe sich damit näher auf jene Ersetzungen bzw. Eingriffe eingehen, die keinen Einfluss auf die Wortartenverteilung haben und somit dennoch genauer beschrieben werden können. Die Annahme, dass das Lemma in manchen Ersetzungen eine Rolle spielt, könnte hiermit konkreter verfolgt werden.

Für die Modellierung sind die eigens formulierten Editionsrichtlinien mit Zuschnitt auf die vorliegenden Gedichte wichtig, um im Falle von Grenzfällen eine begründete Entscheidung zu treffen. Ebenso überzeugte das ‚Prinzip Edition‘ nach Brüning (2022). Insgesamt scheint die Auswahl der Tags zwar einfach, aber durchaus ausreichend. Aus diesem Grund erschien es auch nicht relevant und notwendig, im Rahmen dieser Vorgehensweise ein eigenes Schema für die Modellierung zu erstellen. Mit der klaren Befolgung der Einsatzregeln der TEI-Tags kann auch so sichergestellt werden, dass es sich um valide XML-Dokumente handelt. Bestimmte Aspekte könnten noch detaillierter modelliert werden, falls beispielsweise die Art bzw. die Position der Eingriffe genauer spezifiziert werden sollte. Dies würde allerdings den Fokus vom Text weg hin zu weiteren Besonderheiten des Layouts lenken, die im Fall dieser Arbeit nicht als relevant eingestuft wurden. Bei einer solchen Erweiterung würde sich mit einhergehender Ausweitung des Datenmaterials das Erstellen eines Schemas für die Modellierung durchaus anbieten. Daraus würden sich auch weitere Auswertungsmöglichkeiten ergeben, wie Zeilenanzahl und -position, stenographische Abfassung, Kapitalisierung und Unterstreichung

von Wörtern. Das Attribut @change ist wesentlich für die klare Zuweisung eines Stifts zur handschriftlichen Revision. Ebenso könnten aber @resp oder @hand hierfür verwendet werden. Es ist sinnvoll, die unterschiedlichen Stifte voneinander zu unterscheiden, zu überdenken sind hingegen die damit einhergehenden Richtlinien in Hinblick auf die Textstufen. Beispielsweise zeigte sich, dass Änderungen mit der Schreibmaschine zum einen gering ausfielen und zum anderen als Grundlage für weitere handschriftliche Revisionen dienen könnten. Eine eigene Textstufe allein für einen zusätzlichen Beistrich oder eine kleine Einfügung kann hinterfragt werden, wenn alle weiteren Eingriffe diese Änderungen nicht weiterbearbeiten, sondern darauf aufbauen zu scheinen. Insofern sind die Textstufen nicht unabhängig voneinander zu behandeln, sondern haben Verbindungen, die durch die Vorgehensweise dieser Arbeit vernachlässigt wurden. Gleichzeitig erschwert der Interpretationsspielraum eine klare Entscheidung zur Anwendung der verfassten Editionsrichtlinien.

Die Extrahierung der Textstufen mithilfe der XPath-Abfragen ist sinnvoll und an sich gut umsetzbar. Dennoch könnte stellenweise die Logik verfeinert werden und muss immer dem Verständnis des Textes, der Bearbeitungsstufen und den Editionsrichtlinien entsprechen. Insgesamt erscheint es eine Herangehensweise, die für alle nach einem klaren Schema modellierten Texte funktionieren könnte.

Wenngleich der Korpus zu klein ist, um gefestigte Aussagen zu Mayröckers Bearbeitungstendenzen zu tätigen, zeigt bereits der Ergebnisreichtum, wie viel auf der Basis von Parts-of-Speech-Tagging analysiert werden kann. Mit der Erweiterung der Analysekategorien durch die XML/TEI-Tags für die Revisionsarten und die Sequenzen werden sowohl quantitative als auch qualitative Herangehensweisen unterstützt, um die Textgenese unter diesen Aspekten nachzuvollziehen. Inwiefern diese Aspekte auf linguistischer Ebene weiter ausgebaut werden könnten, wurde bereits angedeutet. Wie literaturwissenschaftliche Aspekte in computergestützte Analysen eingebaut werden können, ist bei Anwendungsfällen wie Popescu et al. (2015) beispielsweise in Hinblick auf Assonanzen, Alliterationen und Reim zu sehen. Die Vorgehensweise und Methodik dieser Arbeit erwies sich in ihren einzelnen Punkten als gut durchsetzbar und für weitere Analysekategorien durchaus erweiterbar, um sich verschiedenen anderen Perspektiven der Schreib- und Arbeitsweise Mayröckers anzunähern.

9 Resümee

Diese Masterarbeit zeigt Möglichkeiten und Methoden auf, um die Textgenese von ausgewählten Gedichten Friederike Mayröckers anhand quantitativer Textanalyse, speziell mittels Parts-of-Speech-Taggings, auf diverse Arten nachzuvollziehen. Die Fragestellungen legten den Fokus bei der Analyse von Textgenese auf zwei unterschiedliche Aspekte. Zum einen lag die Gewichtung auf der Herangehensweise, um den Prozess rund um die Aufbereitung der Textzeugen, die Weiterverarbeitung der Textdaten bis hin zur Auswertung zu dokumentieren. Damit wurden mögliche Arbeitsschritte und Methoden erarbeitet, um die Textgenese von Gedichten Friederike Mayröckers auf quantitativer Ebene ergründen zu können. Zum anderen bestand das Vorhaben, die Analyseergebnisse mit den handschriftlichen Eingriffen in den Textzeugen zu kontextualisieren und mit verschiedenen Aussagen zu Mayröckers Lyrik, Arbeits- und Schreibweise abzugleichen, um mögliche Bearbeitungstendenzen durch den Vergleich der Resultate aller Gedichte herauszufiltern.

Die mehrheitlich theoretischen Kapitel zu Beginn der Arbeit legten dabei die Grundlage für das gesamte Herangehen. Die Diskussion des Forschungsstands verweist auf Referenzwerke wie Brüning (2022), Van Hulle (2022), Markus (2015) sowie Ketzan und Schöch (2021), welche einen wesentlichen Beitrag und Anhaltspunkte zur Etablierung eines Leitfadens für die Vorgehensweise und die Methoden lieferten. Dies wird durch editionswissenschaftliche Perspektiven von Sahle (2013) und Martens (1971; 2008) ergänzt, die Begrifflichkeiten zur Aufbereitung und Analyse der Genese literarischer Texte bieten. Zur Hypothesenbildung in Hinblick auf mögliche Bearbeitungstendenzen von Mayröcker waren Combrink (2016) mit Merkmalen ihrer Lyrik und Kastberger (2000c) zur Genese eines Prosawerks dienlich. Am Anschluss an die Vorstellung des Textkorpus aus fünf Gedichten aus dem Jahr 1981 mit insgesamt 29 Textzeugen wurden die einzelnen Schritte von Transkription, über POS-Tagging und Modellierung hin zur Extrahierung der Textstufen mit allen editorischen Entscheidungen beschrieben. Während bei der Transkription neben den notwendigen Vereinheitlichungen die stenographisch verfassten Passagen eine Herausforderung darstellten, zeigte sich das automatische POS-Tagging mithilfe des Transformermodells der Python-Library *SpaCy* mit hoher Accuracy auch als passend für die Anwendung auf lyrische Texte. In Kauf genommen werden mussten dabei geringere Erkennungswerte für Adverbien, Adjektive und Eigennamen. Die Modellierung geschah in Anlehnung an das ‚Prinzip Edition' nach Brüning (2022) mit fünf wesentlichen Tags für Eingriffe (<add>, <del>, <choice>, <orig>, <reg>) und dem Attribut @change

für die Spezifizierung der Stiftfarbe des jeweiligen Eingriffs. Pro Stiftfarbe wurde eine Textstufe deklariert und mittels der Python Library *lxml* und XPath-Abfragen extrahiert.
In der Analyse wurden die Veränderungen von Häufigkeitsverteilungen der Parts of Speech in allen Typoskripten und weiteren Revisionsstufen eines Gedichts im Vergleich mit den handschriftlichen Eingriffen auf den Textzeugen ermittelt und kontextualisiert. Darüber hinaus lag der Blick auf den Gesamtzahlen der Tokens der Textstufen, den Vorkommen der verschiedenen Eingriffs-Tags sowie den darin enthaltenen Tokens und POS-Verteilungen. Die Etablierung von Alliterationen und Sequenzen wurde ebenso beobachtet.
Im Kapitel 8 Vergleich und Diskussion wurden die Ergebnisse der genannten Analysekategorien schließlich im Kontext aller Gedichte als mögliche Bearbeitungstendenzen Mayröckers ausgelegt. Im Zuge dessen zeigte sich womöglich aufgrund des kleinen Beispielkorpus eine Verteilung von 40% zu 60% zu vielen Aussagen, weswegen Verallgemeinerungen schwierig sind. Die Häufigkeitsverteilungen allein zeigen nur nuancierte Tendenzen. Nichtsdestotrotz kann zum Beispiel in Hinblick auf die Eingriffe gesagt werden, dass die Gesamtzahl der Tokens im Laufe der Genese durch vereinzelte, aber umfangreiche Reduktionen abnimmt, wenngleich sich zwischen den Textstufen mehrheitlich eine Zunahme abzeichnet. Mayröcker wechselt somit zwischen Kürzungen und Erweiterungen, wobei die Abfolge dieser Vorgänge je nach Gedicht unterschiedlich verläuft. In vier der fünf Gedichte werden mehr Tokens eingefügt als gestrichen, wobei die Substantive die am meisten betroffene Wortart von handschriftlichen Eingriffen sind. Gleichzeitig machen sie zumeist etwa ein Viertel aller Wortarten eines Gedichttextes aus. Umgekehrt ist der Anteil an Verben im Text bis auf ein Gedicht unter 10% und spricht in Kombination mit der Dominanz der Substantive für das häufige Auftreten von Ellipsen. Diese sind teils auch den Sequenzen geschuldet, deren Hauptbestandteil die Substantive ausmachen, wie bei der Folge „Substantiv – Artikel – Substantiv“. Diese kommt in allen Gedichten vor und steht weniger für Genitivkonstruktionen als für Aufzählungen bzw. das Aufeinandertreffen von unterschiedlichen Inhalten. Die Sequenzen und Alliterationen als zwei Varianten von Reihungen bzw. Äquivalenzen entstehen mehrheitlich im Laufe der Genese und sind vor allem auf Einfügungen zurückzuführen. Nicht wenige Eingriffe ändern die inhaltliche Dimension, ohne damit einen Einfluss auf die Wortartenverteilung haben, da das Feilen am Ausdruck im Vordergrund zu stehen scheint. Während von Mayröcker in einer Textstelle zwar den Austausch von Verben und Substantiven benennt, zeigt sich bei diesem Korpus eher der Wechsel zwischen Adjektiven und Substantiven.

Die Zeichensetzung scheint wichtiger Bestandteil der Gedichte, der beispielsweise in vereinzelten Fällen stark von Streichungen betroffen ist. Insgesamt sind die handschriftlichen Eingriffe in fast allen Textzeugen vorhanden und treten teils bei fortgeschrittenen Typoskripten noch einmal häufiger auf. Zudem scheint die Bearbeitung bestimmter Stellen intensiver als andere, welche seit Anfang an unberührt bleiben. Wenngleich der Korpus pro Gedicht nicht den vollständigen Prozess abbilden kann, da immer die Möglichkeit fehlender Textzeugen besteht, erwiesen sich die meisten Veränderungen dennoch als schlüssig. Schlussendlich ließ sich die von Martens (1971) formulierte Dynamik bzw. Bewegung dieser Gedichttexte über alle Analysekategorien abbilden und nachvollziehen, wenn es sich dabei auch um feindosierte Veränderungen handelt.

Die Hypothesen zu den Bearbeitungstendenzen konnten in der ein oder anderen Art beantwortet werden. Wenngleich sich mehrheitlich in Anlehnung an die Merkmale Mayröckers Schreibens von Prosa formuliert wurden, sind sie für ihre Lyrik ebenso hilfreiche Anhaltspunkte. Inwiefern sich diese Ergebnisse bei einem größeren Korpus ebenso finden lassen würden und damit zu eindeutigeren Erkenntnissen Mayröckers Schreib- und Arbeitsweise führen könnten, wäre zu ermitteln. Die Adaption und Erweiterung der Vorgehensweise und Methoden sowie der Analysekategorien scheint niederschwellig und weist auf viele weitere potentielle Perspektiven, die für eine quantitative Analyse von Textgenese bei Gedichten wie jenen von Friederike Mayröcker inkludiert werden könnten.

Literaturverzeichnis

Arteel, Inge; De Felip, Eleonore (Hg.) (2020): Fragen zum Lyrischen in Friederike Mayröckers Poesie. Stuttgart: J.B. Metzler Verlag.

Bernhart, Toni (2008): Die Vermessung der Farben in der Sprache. Zur Berlin-Kay-Hypothese in der Literaturwissenschaft. In: *Zeitschrift für Literaturwissenschaft und Linguistik* Vol. 38, S. 56–78.

Bernhart, Toni (2018): Quantitative Literaturwissenschaft: Ein Fach mit langer Tradition? In: Toni Bernhart, Marcus Willand, Sandra Richter und Andrea Albrecht (Hg.): Quantitative Ansätze in den Literatur- und Geisteswissenschaften. Systematische und historische Perspektiven. Berlin/Boston: De Gruyter, S. 207–219.

Bernhart, Toni; Willand, Marcus; Richter, Sandra; Albrecht, Andrea (Hg.) (2018): Quantitative Ansätze in den Literatur- und Geisteswissenschaften. Systematische und historische Perspektiven. Berlin/Boston: De Gruyter.

Beyer, Marcel (Hg.) (2004): Friederike Mayröcker. Gesammelte Gedichte 1939 - 2003. Frankfurt am Main: Suhrkamp Verlag.

Bosse, Anke; Fanta, Walter (2019): Vorwort. Wozu Textgenese in der digitalen Edition? Fragestellungen und Lösungsmodelle. In: Anke Bosse und Walter Fanta (Hg.): Textgenese in der digitalen Edition. Berlin/Boston: De Gruyter, S. VII–X.

Bremer, Kai; Wirth, Uwe (Hg.) (2010): Texte zur modernen Philologie. Stuttgart: Reclam.

Brüning, Gerrit (2022): Modellierung von Textgeschichte. Bedingungen digitaler Analyse und Schlussfolgerungen für die Editorik. In: Fotis Jannidis (Hg.): Digitale Literaturwissenschaft. DFG-Symposion 2017. Stuttgart: J.B. Metzler Verlag (Germanistische Symposien), S. 307–338.

Combrink, Thomas (2016): Friederike Mayröcker. In: Hermann Korte (Hg.): Kindler Kompakt Österreichische Literatur der Gegenwart. Stuttgart: J.B. Metzler Verlag, S. 61–67.

Grésillon, Almuth (1998): Bemerkungen zur französischen "édition génétique". In: Gunter Martens und Hans Zeller (Hg.): Textgenetische Edition. Tübingen: Max Niemeyer Verlag (Beihefte zur Editio, 10), S. 52–64.

Hell, Bodo (1987): es ist so ein Feuerrad. Bodo Hell im Gespräch mit Friederike Mayröcker in deren Winer Arbeitszimmer - am 28. September 1985. In: Friederike Mayröcker (Hg.): Magische Blätter II. Frankfurt am Main: Suhrkamp Verlag, S. 177–198.

Inguglia-Höfle, Arnhilt; Rettenwander, Susanne (2022): Friederike Mayröcker: FÜR ERNST. onb.ac.at. Online verfügbar unter https://www.onb.ac.at/mehr/blogs/detail/friederike-mayroecker-fuer-ernst, zuletzt geprüft am 14.05.2024.

Jakobson, Roman (2007): Linguistik und Poetik. In: Roman Jakobson (Hg.): Poesie der Grammatik und Grammatik der Poesie. Sämltiche Gedichtanalysen. Unter Mitarbeit von Hendrik Birus und Sebastian Donat. Berlin/New York: De Gruyter (Poetologische Schriften und Analysen zur Lyrik vom Mittelalter bis zur Aufklärung, I), S. 155–216.

Jannidis, Fotis (Hg.) (2022a): Digitale Literaturwissenschaft. DFG-Symposion 2017. Stuttgart: J.B. Metzler Verlag (Germanistische Symposien).

Jannidis, Fotis (2022b): Digitale Literaturwissenschaft. Zur Einführung. In: Fotis Jannidis (Hg.): Digitale Literaturwissenschaft. DFG-Symposion 2017. Stuttgart: J.B. Metzler Verlag (Germanistische Symposien), S. 1–18.

Kastberger, Klaus (1998a): Friederike Mayröcker (geb. 1924): Reise durch die Nacht. In: Klaus Kastberger und Bernhard Fetz (Hg.): Der literarische Einfall. Über das Entstehen von Texten. Wien: Paul Zsolnay Verlag, S. 20–26.

Kastberger, Klaus (1998b): Zu einer Philologie des Schaffensaktes. Anhand Friederike Mayröckers »Reise durch die Nacht«. In: *Sichtungen online* (1). Online verfügbar unter http://purl.org/sichtungen/kastberger-k-2a.html, zuletzt geprüft am 19.02.2024.

Kastberger, Klaus (2000a): Friederike Mayröcker: Modell einer genetischen Analyse. In: Klaus Kastberger und Bernhard Fetz (Hg.): Von der ersten bis zur letzten Hand. Theorie und Praxis der literarischen Edition. Wien/Bozen: Folio Verlag, S. 158–163.

Kastberger, Klaus (2000b): »Jedenfalls geht es bei mir nicht ganz anständig zu.«. Zwei Gespräche mit Friederike Mayröcker zur Technik ihrer literarischen Produktion (März 1994, April 1995). In: Klaus Kastberger (Hg.): Reinschrift des Lebens. Friederike Mayröckers 'Reise durch die Nacht'. Edition und Analyse. Wien u.a.: Böhlau Verlag, S. 439–452.

Kastberger, Klaus (Hg.) (2000c): Reinschrift des Lebens. Friederike Mayröckers 'Reise durch die Nacht'. Edition und Analyse. Wien u.a.: Böhlau Verlag.

Kastberger, Klaus; Schmidt-Dengler, Wendelin (Hg.) (1996): In Böen wechsel mein Sinn. Zu Friederike Mayröckers Literatur. Wien: Sonderzahl.

Ketzan, Erik; Schöch, Christof (2021): Classifying and Contextualizing Edits in Variants with Coleto: Three Versions of Andy Weir's The Martian. In: *Digital Humanities Quarterly* 15 (4).

Klinke, Harald (2017): Information Retrieval. In: Fotis Jannidis, Hubertus Kohle und Malte Rehbein (Hg.): Digital Humanities. Eine Einführung. Stuttgart: J.B. Metzler Verlag, S. 268–278.

Kühn, Renate (Hg.) (2002): Friederike Mayröcker oder "das Innere des Sehens". Studien zu Lyrik, Hörspiel und Prosa. Bielefeld: Aisthesis Verlag.

Leech, Geoffrey (2004): Adding Linguistic Annotation. In: Martin Wynne (Hg.): Developing Linguistic Corpora: a Guide to Good Practice: adhs literature, languages and linguistics. Online verfügbar unter https://users.ox.ac.uk/~martinw/dlc/chapter2.htm, zuletzt geprüft am 29.04.2024.

Lughofer, Johann Georg (Hg.) (2017): Friederike Mayröcker. Interpretationen Kommentare Didaktisierungen. Wien: Praesens Verlag.

Markus, Hannah (2015): Ilse Aichingers Lyrik. Das gedruckte Werk und die Handschriften. Berlin/Boston: De Gruyter.

Martens, Gunter (1971): Textdynamik und Edition. Überlegungen zur Bedeutung und Darstellung variiernder Textstufen. In: Gunter Martens und Hans Zeller (Hg.): Texte und Varianten. Probleme ihrer Edition und Interpretation. München: C. H. Beck'sche Verlagsbuchhandlung, S. 165–201.

Martens, Gunter (2008): Textgenese als Hilfsmittel zur Erschließung poetischer Texte. In: Françoise Lartillot und Axel Gellhaus (Hg.): Dokument / Monument. Textvarianz in den verschiedenen Disziplinen der europäischen Germanistik. Akten des 38. Kongresses des französischen Hochschulgermanistikverbandes (A.G.E.S.). Bern u. a.: Peter Lang, S. 139–160.

Mayröcker, Friederike (1982): Gute Nacht, guten Morgen. Gedichte 1978-1981. Frankfurt am Main: Suhrkamp Verlag.

Mayröcker, Friederike (1983a): Augenfalle / trompe-l'oeil. In: Friederike Mayröcker (Hg.): Magische Blätter. Frankfurt am Main: Suhrkamp Verlag, S. 102–108.

Mayröcker, Friederike (1983b): MAIL ART. In: Friederike Mayröcker (Hg.): Magische Blätter. Frankfurt am Main: Suhrkamp Verlag, S. 9–35.

Mayröcker, Friederike (1986): Winterglück. Gedichte 1981-1985. Frankfurt am Main: Suhrkamp Verlag.

Meindl, Claudia (2011): Methodik für Linguisten. Eine Einführung in Statistik und Versuchsplanung. Tübingen: Narr Verlag.

Melzer, Gerhard; Schwar, Stefan (Hg.) (1999): Friederike Mayröcker. Graz/Wien: Literaturverlag Droschl (DOSSIER. Die Buchreihe über österreichische Autoren, 14).

Moretti, Franco (2000): The Slaughterhouse of Literature. In: *Modern Language Quarterly* 61 (1), S. 207–228.

Moretti, Franco (2013): Distant Reading. London/New York: Verso.

Plachta, Bodo (2006): Editionswissenschaft. Eine Einführung in Methode und Praxis der Edition neuerer Texte. 2., ergänzte und aktualisierte Auflage. Stuttgart: Reclam.

Popescu, Ioan-Iovitz; Lupea, Mihaiela; Tatar, Doina; Altmann, Gabriel (2015): Quantitative Analysis of Poetic Texts. Berlin/Boston: De Gruyter.

Rasch, Björn; Friese, Malte; Hofmann, Wilhelm; Naumann, Ewald (2014): Verfahren für Nominaldaten. In: Björn Rasch, Malte Friese, Ewald Naumann und Wilhelm Hofmann (Hg.): Quantitative Methoden 2. Berlin/Heidelberg: Springer, S. 111–134.

Sahle, Patrick (2013): Digitale Editionsformen. Zum Umgang mit der Überlieferung unter den Bedingungen des Medienwandels. Teil 3: Textbegriffe und Recodierung. Norderstedt: BoD (Schriften des Insituts für Dokumentologie und Editorik, 9). Online verfügbar unter http://kups.ub.uni-koeln.de/5013/, zuletzt geprüft am 30.0.4.2024.

Schmidt, Siegfried J. (Hg.) (1984): Friederike Mayröcker. Frankfurt am Main: Suhrkamp Verlag.

Schöch, Christof (2016): Ein digitales Textformat für die Literaturwissenschaften. Die Richtlinien der Text Encoding Initiative und ihr Nutzen für Textedition und Textanalyse. In: *Romanische Studien* 4, S. 325–364.

Sexl, Martin (2004): Formalistisch-strukturalistische Theorien. In: Martin Sexl (Hg.): Einführung in die Literaturtheorie. Wien: Facultas, S. 161–190.

Strohmaier, Alexandra (Hg.) (2009): Buchstabendelirien. Zur Literatur Friederike Mayröckers. Bielefeld: Aisthesis Verlag.

TEI Consortium (Hg.) (2023): Guidelines for Electronic Text Encoding and Interchange. Verison 4.6.0. Online verfügbar unter https://tei-c.org/release/doc/tei-p5-doc/en/html/index.html, zuletzt geprüft am 20.06.2024.

Underwood, Ted (2017): A Genealogy of Distant Reading. In: *Digital Humanities Quarterly* 11 (2).

Universal POS tags. universaldepndencies.org. Online verfügbar unter universaldependencies.org/u/pos/all.html, zuletzt geprüft am 10.05.2024.

van Hulle, Dirk (2022): Digital Genetic Editions. Towards Macroanalysis across Versions. In: Fotis Jannidis (Hg.): Digitale Literaturwissenschaft. DFG-Symposion 2017. Stuttgart: J.B. Metzler Verlag (Germanistische Symposien), S. 339–352.

Vogeler, Georg (2019): Digitale Editionspraxis. Vom pluralistischen Textbegriff zur pluralistischen Softwarelösung. In: Anke Bosse und Walter Fanta (Hg.): Textgenese in der digitalen Edition. Berlin/Boston: De Gruyter, S. 117–136.

Wawrczeck, Jens-Christian (2024): Stenografie Wörterbuch. Deutsche Einheitskurzschrift (DEK) A - Z. Online verfügbar unter https://www.jens-wawrczeck.de/stenogenerator/stenografie_woerterbuch.htm, zuletzt geprüft am 15.06.2024.

Wirth, Uwe; Bremer, Kai (2010): Die philologische Frage. Kulturwissenschaftliche Perspektiven auf die Theoriegeschichte der Philologie. In: Kai Bremer und Uwe Wirth (Hg.): Texte zur modernen Philologie. Stuttgart: Reclam, S. 7–48.

Zeller, Hans (1971): Befund und Deutung. Interpretation und Dokumentation als Ziel und Methode der Edition. In: Gunter Martens und Hans Zeller (Hg.): Texte und Varianten. Probleme ihrer Edition und Interpretation. München: C. H. Beck'sche Verlagsbuchhandlung, S. 45–89.

Zwerschina, Hermann (1998): Die editorische Einheit 'Textstufe'. In: Gunter Martens und Hans Zeller (Hg.): Textgenetische Edition. Tübingen: Max Niemeyer Verlag (Beihefte zur Editio, 10), S. 177–194.

Daten & Code

Textzeugen. Collection für alle Textzeugen der Gedichte von Friederike Mayröcker. Eingeschränkter Zugang unter https://phaidra-local.univie.ac.at/o:45688

XML-Dokumente. Collection für alle XML-Dokumente zu den Textzeugen der Gedichte von Friederike Mayröcker. Eingeschränkter Zugang unter https://phaidra-local.univie.ac.at/view/o:45808

Sequenzen. Collection für alle CSV-Dokumente zu den Sequenzen der Textzeugen der folgenden Gedichte von Friederike Mayröcker. Eingeschränkter Zugang unter https://phaidra-local.univie.ac.at/view/o:45816

Repository DH-Masterarbeit. GitHub Repository für alle Jupyter-Notebooks in Python. Öffentlicher Zugang unter https://github.com/johannaunterholzner/DH-Masterarbeit

Tabellenverzeichnis

Abbildungsverzeichnis