



# Dem Batty Weber säin Abreißkalender (1913-1940)

**Theemen a Musteren**

*Anne-Marie Millim*



# **Eine LLM Analyse von Stilmitteln in Batty Webers Abreißkalender**

Peter Gilles  
Alistair Plum

**Dem Batty Weber säin Abreißkalender (1913-1940):**

**Theemen a Musteren**

Herausgegeben von Anne-Marie Millim

Publiziert in Melusina Press (Universität Luxemburg), 2026  
11, Porte des Sciences  
L-4366 Esch-sur-Alzette  
<https://www.melusinapress.lu>

Verlagsleitung: Niels-Oliver Walkowski, Johannes Pause  
Lektorat: Carolyn Knaup, Niels-Oliver Walkowski  
Gestaltung: Valentin Henning, Erik Seitz  
Umschlagbild: Raymon Mehlen, ‚Batty Weber‘ (ca. 1935) © All Rechte virbehal

Die digitale Version dieser Publikation steht unter <https://www.melusinapress.lu> frei zur Verfügung.

Bibliografische Information der Nationalbibliothek Luxemburg: Die Nationalbibliothek Luxemburg verzeichnet diese Publikation in der Luxemburgischen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über [bnl.public.lu](http://bnl.public.lu) abrufbar.

Druck: Libri Plureos GmbH, Friedensallee 273, 22763 Hamburg.

ISSN (Web): 2716-7518  
DOI (Publication): 10.26298/1981-5616-elvs  
DOI (Container): 10.26298/1981-5616  
ISBN (Web): 978-2-919815-61-6  
ISBN (PDF): 978-2-919815-62-3  
ISBN (Epub): 978-2-919815-86-9  
ISBN (Print): 978-2-919815-63-0

Das vorliegende Werk steht unter einer CC BY-SA 4.0 Lizenz. Informationen zu dieser Lizenz finden Sie unter <https://creativecommons.org/licenses/by-sa/4.0/deed.de>. Die in diesem Werk enthaltenen Bilder und Ressourcen unterliegen der selben Lizenz, sofern sie keiner anderen Quelle entnommen wurden oder mit einer anderen Lizenz versehen sind.





2026

# Eine LLM-Analyse von Stilmitteln in Batty Webers *Abreißkalender*

Peter Gilles & Alistair Plum

## Einleitung

Der Einzug von großen Sprachmodellen (*large language models*, LLMs) wie die GPT-Serie von OpenAI hat eine neue Ära für die Computerlinguistik und die Digital Humanities (DH) eingeleitet. Diese Modelle, die mit sehr großen Textdatenkorpora trainiert wurden, verfügen über eine noch nie dagewesene Fähigkeit, menschliche Sprache zu verstehen und zu erzeugen, was neue Wege für die automatisierte Analyse literarischer Texte bereitet. Im Bereich der DH bietet dies die Möglichkeit, die traditionellen Praktiken des *close reading* mit computergestützter Hilfe zu verbessern, insbesondere bei umfangreichen Textkorpora, die aufgrund von Ressourcenbeschränkungen nur schwer manuell analysiert werden können.

Batty Weber, eine Schlüsselfigur der luxemburgischen Literatur des frühen 20. Jahrhunderts, war Autor des *Abreißkalenders*, einer Sammlung täglicher Kolumnen, die über mehrere Jahrzehnte hinweg veröffentlicht wurden. Diese in deutscher Sprache verfassten Texte bieten einen umfassenden Einblick in die soziokulturelle Landschaft Luxemburgs in einer Zeit, die von bedeutenden historischen Ereignissen geprägt war, darunter beide Weltkriege. Der schiere Umfang und die historische Tiefe des *Abreißkalenders* machen ihn zu einer unschätzbaren Quelle für Wissenschaftler, stellen aber auch praktische Herausforderungen für ein gründliches *close reading*. Bisherige computergestützte Ansätze in der Literaturanalyse stützen sich oft auf Schlüsselwörter, Themenmodellierung und Stimmungsanalyse, um

Erkenntnisse aus großen Textsammlungen zu gewinnen. Diese Methoden sind zwar nützlich, können aber bei der Erfassung nuancierter Themen, kontextbezogener Feinheiten und der Intention des Autors versagen. Dies sind allerdings Elemente, die für die geisteswissenschaftliche Forschung von zentraler Bedeutung sind. Darüber hinaus haben herkömmliche Verfahren zur Verarbeitung natürlicher Sprache (*natural language processing*, NLP) möglicherweise Schwierigkeiten mit idiomatischen Ausdrücken, historischem Sprachgebrauch und den Besonderheiten weniger erforschter Sprachen oder Dialekte.

Diese Studie zielt darauf ab, diese Lücke zu schließen, indem sie die Wirksamkeit eines aktuellen LLM bei der Durchführung von automatisiertem *close reading* von Batty Webers *Abreißkalender* testet. Konkret untersuchen wir, wie ein GPT-Modell nuancierte textanalytische Aufgaben erledigen kann, wie beispielsweise die Identifizierung und Interpretation von Darstellungen nationaler Identität, Krieg oder technischem Fortschritt. Die Fähigkeit des Modells, nicht nur relevante Passagen zu erkennen, sondern auch Erklärungen für seine Auswahl zu liefern, bietet ein neuartiges Werkzeug für die Textanalyse in den Geisteswissenschaften.

Indem wir den *Abreißkalender* in das GPT-Modell einspeisen, bewerten wir dessen Leistung bei der Replikation eines menschenähnlichen Verständnisses komplexer literarischer Themen und Strukturen. Dieser Ansatz birgt ein erhebliches Potenzial für kleinere Forschungsprojekte, denen es an finanziellen Mitteln und Arbeitskräften mangelt. Automatisiertes *close reading* mit Hilfe von KI kann es Wissenschaftlern ermöglichen, sich effizienter mit großen Textsammlungen auseinanderzusetzen und so den Zugang zu umfassender Literaturanalyse zu demokratisieren.

Die Verbindung von *cultural analytics* mit modernen LLMs eröffnet zusätzliche Perspektiven, indem sie quantifizierbare und interpretative Ansätze in der Analyse kultureller Texte miteinander verknüpft. Insbesondere ermöglichen LLMs, kulturelle und linguistische Muster in großen Textkorpora zu erkennen und zu deuten – ein Ansatz, der in der *cultural analytics* und den Digital/Computational Literary Studies bereits zentrale Beachtung findet (vgl. Manovich 2020; Hatzel et al. 2023;

Mélanie-Becquet et al. 2024). Darüber hinaus bestätigen aktuelle Untersuchungen, dass LLMs nicht nur zur automatisierten Klassifikation, sondern auch zur Förderung eines tieferen Verständnisses kultureller Diskurse beitragen können (vgl. Bamman et al. 2024). Diese interdisziplinäre Methodik erlaubt es, literarische Artefakte in einen breiten kulturellen Kontext zu stellen und so neue Forschungsfragen zu generieren, die über traditionelle *close-reading*-Ansätze hinausgehen. Zudem fördert die Kombination von *cultural analytics* und LLMs den Aufbau ressourceneffizienter Analysemethoden, die insbesondere für die Untersuchung weniger erforschter literarischer Strömungen von großer Bedeutung sind.

Unsere Ergebnisse zeigen, dass der Einsatz von LLMs zwar mit Kosten verbunden ist, sowohl an Rechenleistung als auch finanziell, die Vorteile in Bezug auf analytische Tiefe und Effizienz jedoch erheblich sind. Das hier verwendete GPT-Modell zeigte eine starke Fähigkeit zur Unterstützung des *close reading*, indem es detaillierte Antworten lieferte und einschlägige Textstellen identifizierte, die mit wissenschaftlichen Interpretationen übereinstimmen. In den folgenden Abschnitten erläutern wir unsere Methodik, präsentieren die Ergebnisse unserer Analyse und diskutieren die Implikationen für die zukünftige Forschung in den DH. Wir argumentieren, dass die Integration von LLMs in die geisteswissenschaftliche Forschung die Forschungsergebnisse verbessert und den Weg für eine integrativere und ressourceneffizientere akademische Forschung ebnet.

## Fragestellung und Methodik

Kann ein GPT-Modell die feingliedrige literaturwissenschaftliche Analyse unterstützen? Konkret soll gezeigt werden, wie automatisiertes *close reading* für die Vorarbeiten für die inhaltliche und stilistische Analyse des Batty-Weber-Korpus eingesetzt werden kann. Methodisch orientieren wir uns an der Untersuchung von Underwood (vgl. 2023), der in einem Romankorpus das Verhältnis von erzählter Zeit und Erzählzeit mit Hilfe eines LLMs untersucht hat. Für die vorliegende Fallstudie widmen wir uns der Kriegsthematik, die in Webers Texten häufig wiederkehrt und sich größtenteils auf den Ersten Weltkrieg

bezieht. Des Weiteren wird die Kriegsthematik hinsichtlich ihrer Bildlichkeit analysiert: Dazu werden aus dem Korpus alle (potenziellen) Belege extrahiert, die die Stilmittel der Metapher, der Metonymie oder des Vergleiches beinhalten.

Das Korpus für diese Studie besteht aus Batty Webers *Abreißkalender*, einer Sammlung von täglichen Kolumnen in deutscher Sprache, die über mehrere Jahrzehnte im frühen 20. Jahrhundert verfasst wurden (online publiziert unter: <https://battyweber.uni.lu>). Das Korpus umfasst 6.817 Einzeltexte aus den Jahren 1913 bis 1940, wobei durchschnittlich 250 Kalenderblätter pro Jahr enthalten sind. Die Texte weisen eine durchschnittliche Länge von 656 Wörtern auf, wobei einzelne Sätze im Mittel 21,7 Wörter umfassen. Der Gesamtumfang des Korpus beträgt ca. 3,8 Millionen Wortformen (Tokens ohne Satzzeichen).

## Auswahl und Konfiguration des Sprachmodells

Für diese Studie verwendeten wir das von OpenAI entwickelte GPT-4o-Modell (vgl. Brown et al. 2020). Der Zugriff auf das Modell erfolgte über die OpenAI-Anwendungsschnittstelle und ermöglichte die Integration in unsere Datenverarbeitungspipeline mittels eines Python-Skripts. Aufgrund der historischen und sprachlichen Besonderheiten des Korpus zogen wir in Erwägung, das Modell an einer Teilmenge der *Abreißkalender*-Texte zu verfeinern, um sein Verständnis von Sprache und Kontext zu verbessern (vgl. Howard und Ruder 2018). In Anbetracht der begrenzten Ressourcen entschieden wir uns jedoch für die Verwendung des vortrainierten Modells mit sorgfältig ausgearbeiteten Aufforderungen, um die gewünschten Antworten zu erhalten.

## Aufgabendesign und Prompt-Engineering

Das Hauptziel bestand darin, die Fähigkeit des Modells zu bewerten, ein automatisiertes *close reading* durchzuführen, indem mittels unterschiedlicher Prompts kriegsrelevante Metaphern, Metonymien und Vergleiche aus dem Korpus extrahiert werden. Das System soll dazu jeden Text satzweise „lesen“ und jeweils die Identifikation eines Wortes oder einer Phrase sowie die Klassifikation durchführen. Ergänzend dazu

sollte eine Erklärung generiert werden, warum es sich um das jeweilige Stilmittel handelt. Für diese explorativen Studien wurde mit vier verschiedenen Prompts gearbeitet, um die bestmögliche Formulierung für die vorliegende Aufgabe zu ermitteln. Die vier Prompts unterscheiden sich hinsichtlich ihrer Zugangsweise und ihrer Komplexität. So enthält der erste Prompt z.B. nur die einfache Anweisung, die drei Stilmittel zu extrahieren. Die Aufforderungen wurden möglichst klar und kontextbezogen konzipiert, damit das GPT-Modell genügend Informationen erhält, um genaue und relevante Antworten zu geben (vgl. Liu et al. 2021).

| Prompt 1 (= zero-shot Prompt)

| Read the text passage below after these instructions. Search for all sentences containing comparisons, metaphors or metonymies for descriptions of war in the text. Consider only each sentence individually, do not go beyond the boundary of the sentence. Note that the texts are referring to WW 1.

| Now, analyze the following passage:

| <Es folgt der Text eines Kalenderblattes.>

Mit dem Prompt wurde also auch der Text des jeweiligen Kalenderblattes für die Analyse übergeben. Die Kompetenz des LLMs wurde also nicht im Sinne eines Wissensspeichers angezapft, sondern vielmehr soll es bei der Identifikation und Extraktion von Themen und Stilmitteln assistieren.

Für die Ausgabe wurde eine neue Funktionalität von OpenAIs Programmierumgebung genutzt, indem nämlich die Ausgabe in der Struktur eines JSON-Objekts erfolgt, wodurch jede der mehr als 6.000 Abfragen die gleiche tabellarische Struktur erzeugt. Diese hat schematisch das folgende Format:

- Satznummer
- Text des Satzes

- Kategorie („metaphor“, „metonymy“, „comparison“)
- Erklärung des Stilmittels im vorliegenden Satz
- Konfidenz des Systems für die Kategorisierung („low“, „medium“, „high“)
- Sentiment des Stilmittels („negative“, „neutral“, „positive“)

Durch dieses Ausgabeformat ist sichergestellt, dass jede Abfrage die gleiche Struktur erhält, sodass schließlich alle der mehr als 6.000 Abfragen einheitlich in eine Gesamttabelle überführt werden können.

| Prompt 2 (= just definition)

| Read the text passage below after these instructions. Search for all sentences containing examples for the figures of speech „metaphors“, „comparisons“ or „metonymies“ related to the descriptions of war in the text below. Consider only each sentence individually, do not go beyond the boundary of the sentence. Note that the text is a feuilleton text from a newspaper from the beginning of the 20th century. References to war mean WW 1. To help you with this task, I am providing also definitions and examples for the figures of speech you are going to identify. Take also the following definitions into account!

| Definition of a „metaphor“:

| Eine Metapher ist eine Stilfigur, ein „sprachliches Bild“, das auf einer Ähnlichkeitsbeziehung zwischen zwei Gegenständen bzw. Begriffen beruht, d.h., aufgrund gleicher oder ähnlicher Bedeutungsmerkmale findet eine Bezeichnungsübertragung statt (z.B. „der Himmel weint“ für „es regnet“). Häufig wird eine Metapher auch als verkürzter Vergleich beschrieben, wobei der Vergleich als solcher jedoch nicht ausgedrückt wird. Metaphern können in substantivischer, adjektivischer und verbaler Form im Satzkontext auftreten (z.B. „Fuchsschwanz“ für „Handsäge“, „spitze Bemerkung“ für „verletzende Bemerkung“, „sich zügeln“ für „sich zurückhalten“).

Definition of a „metonymy“:

Eine Metonymie ist eine Stilfigur, bei der ein Wort oder ein Konzept durch eine verwandte Bezeichnung ersetzt wird, die mit dem Gemeinten – im Gegensatz zur Metapher – durch einen sachlichen (z.B. räumlichen, zeitlichen, kausalen, instrumentalen) Zusammenhang bzw. durch semantische Kontiguität verknüpft ist. Beispiele: „Seide tragen“ für „Kleidung aus Seide tragen“, „Goethe lesen“ für „ein Werk von Goethe lesen“, „das Weiße Haus schweigt“ für „die Regierung der USA schweigt“, „das Zepter niederlegen“ für „die Regierungsgewalt abgeben“. Eine Spezialform ist die „Synekdoche“.

Definition of a „comparison“:

Ein Vergleich ist ein Stilmittel, bei dem mindestens zwei unterschiedliche Dinge, Personen oder Situationen direkt gegenübergestellt werden, um eine Ähnlichkeit zu betonen. Oft wird die Vergleichspartikel „wie“ dafür verwendet. Der Vergleich wird also auch grammatisch direkt ausgedrückt.

Dieser Prompt ist nun beträchtlich länger und die Intention ist, dem System Definitionen für die drei Stilmittel zu übergeben, wodurch im Vergleich mit dem „Zero shot“-Prompt exaktere Resultate generiert werden sollten. Die Definitionen stammen aus dem linguistischen Lexikon von Bußmann (vgl. 2008) und beinhalten auch Spezialformen, wie z.B. die Synekdoche, eine Form der Metonymie, wobei einem Objekt ein menschlicher Aspekt zugeschrieben wird. Mittels dieser Definitionen sollte das LLM also genauer für die Aufgabe instruiert werden.

Im nächsten Prompt 3 wurde anstelle der Definitionen auf prototypische Beispiele für die drei Stilmittel fokussiert, während der Rest des Prompts identisch blieb. Für Metapher, Metonymie und Vergleich wurden jeweils drei Beispiele übergeben. Ähnlich wie beim vorigen Prompt sollte das LLM dadurch in die Lage versetzt werden, anhand der Beispiele seine Aufgabe zu „lernen“ und sich besser darauf einzustellen (sogenannter „Few shot“-Prompt). Im folgenden Ausschnitt ist lediglich ein Beispiel für eine Metapher wiedergegeben. Im Beispielsatz wurde das identifizierte Stilmittel durch \*[…]\* markiert, und in einer Analyse wurde erläutert,

warum es sich um das jeweilige Stilmittel handelt und wie es expliziert werden kann.

| Prompt 3 (= few-shot)

| [...]

| To identify a metaphor and how to analyse them, orient yourself to these examples.

| [...]

| Example 3

| „In einem der armen, zerschossenen Wälder, in denen die schönsten Eichen nur noch verstümmelte Stümpfe sind, die im Frühjahr zu grünen versuchen, \*wie ein todwundes Wild\* noch den Kopf hebt oder mit den Läufen zuckt.“

| Analysis

| The sentence compares the damaged trees in the war-torn forest to a mortally wounded animal, using the phrase „wie ein todwundes Wild“. This comparison highlights the devastation of war.

Prompt 4 schließlich ist eine Kombination aus den beiden vorherigen (= few-shot-definition), enthält also sowohl die Beispiele mit Erklärungen als auch die Definitionen für die Stilmittel.

Unter Anwendung der beschriebenen Methodik haben wir das Sprachmodell beauftragt, Passagen im *Abreißkalender* zu identifizieren und zu analysieren, die den Krieg und damit verbundene Themen beschreiben. Das Modell lieferte detaillierte Ergebnisse in tabellarischer Form, einschließlich Seitenzahlen, Auszügen relevanter Passagen und Erläuterungen zu deren Bedeutung. Alle vier Prompts wurden kombiniert mit allen mehr als 6.000 Texten zur Generierung an das GPT-4o-Modell geschickt, welches zum Zeitpunkt der Abfassung dieses Beitrags das leistungsfähigste Modell war. Dieser Prozess dauert mehrere Stunden und es fallen Kosten von ca. 100 US-Dollar an. Da wir

mit verschiedenen Prompt-Verfeinerungen experimentiert haben, können sich die Kosten schnell aufsummieren. Für die Sinnhaftigkeit und Machbarkeit solcher Projekte gilt es also, den Kostenaspekt abzuwägen. Die verwendeten Prompts, das Python-Skript sowie die vollständige Ergebnisliste mit allen Stilmitteln werden über das GitHub-Repositorium <https://github.com/plumaj/batty-prompt> zur Verfügung gestellt.

## Ergebnisse

Die vier Prompts generierten teilweise recht divergierende Resultate, wie die Anzahlen in der folgenden Tabelle zeigen, was als ein Hinweis darauf gedeutet werden kann, dass die Prompts jeweils recht unterschiedlich interpretiert wurden. Hier fällt besonders der Prompt „just definition“ auf, durch den allein 3.177 Metaphern identifiziert wurden. Eine cursorische Durchsicht dieser Resultate ergibt Folgendes: Tatsächlich wurden hier überwiegend Metaphern identifiziert, allerdings nicht zwangsläufig zur Kriegsthematik, d.h., dieser Aspekt des Prompts wurde vom LLM ignoriert oder falsch interpretiert; für die folgende Diskussion wird dieser Prompt daher ausgeschlossen.

	<i>comparison</i>	<i>metaphor</i>	<i>metonymy</i>	<i>paradox</i>	<i>simile</i>	<i>sum</i>
few-shot	366	418	191	0	0	975
few-shot definition	533	818	216	0	0	1.567
just definition	1.712	3.177	722	0	1	5.612
zero-shot	508	1.300	337	1	56	2.002

Tabelle 1: Übersicht der extrahierten Stilmittel pro Prompt.

Die Resultate der übrigen drei Prompts liegen dichter beieinander, wobei der „Few shot“-Prompt die wenigsten und der „Zero shot“-Prompt die

## Ergebnisse

date	sentence	sentence_text	category	explanation
1916-01-01	9	Es ist keine Zeit mehr, es ist nur noch Krieg.	metaphor	The sentence uses the metaphor 'es ist nur noch Krieg' to convey that war has consumed all aspects of time.
1916-01-01	10	Das ist ein großer Abgrund voll Blut und Rauch, voll Gekrach und Wehgeschrei.	metaphor	The phrase 'ein großer Abgrund voll Blut und Rauch, voll Gekrach und Wehgeschrei' metaphorically depicts the war as a hellish abyss.
1916-01-01	13	Was bedeutet das auf und ab eines gewöhnlichen Jahres, wenn der Krieg alles in den Boden stampft, alle Träume vernichtet.	metaphor	The metaphor 'der Krieg alles in den Boden stampft' describes war as a force that crushes everything.
1916-01-01	24	Es ist eine kleine Zeit, und sie wird immer kleiner werden, je mehr sich der Krieg auf sein Endziel zuspitzt.	metaphor	The sentence uses the metaphor 'eine kleine Zeit' to describe the era as shrinking toward its end.
1916-06-20	12	Wenn wieder Friede ist, werden die Verträge auch wieder ausschlagen, wie die zerschossenen Bäume im April.	comparison	The sentence compares the resurgence of treaties after the war to the regrowth of trees in spring.
1916-06-20	27	Wir sollen uns in den Gedanken hineinleben, daß es der Tribut ist, den auch wir dem Kriegsgott zollen.	metaphor	The phrase 'Tribut dem Kriegsgott zollen' metaphorically represents war as a deity demanding tribute.
1916-06-20	34	Wenn der Ozean von Tod und Blut und Schrecken, der im Westen und Osten wogt, einmal eine Welle an unser Ufer wirft.	metaphor	The 'Ozean von Tod und Blut und Schrecken' metaphorically describes the war as a violent ocean.

## ABBILDUNG 1: Auszug aus den Ergebnissen

meisten Resultate liefert. Für Letzteren ist noch erwähnenswert, dass das System auch 56 *similes* (als Untergruppe zu *comparison*, wobei der Fokus auf Wörtern wie „wie“ und „als“ liegt) und ein Paradox („Wo die, die einen Krieg heute gewinnen, ihn ja doch verlieren!“, wobei es sich um eine widersprüchliche Aussage handelt) identifiziert, obwohl nicht danach gefragt wurde. Dies ist ein weiterer Beleg dafür, dass bei GPT-Systemen immer mit gewissen Unzuverlässigkeiten zu rechnen ist, die im vorliegenden Fall jedoch durchaus konzeptuell passend sind.

Abbildung 1 soll einen Eindruck vom „Leseprozess“ des LLMs vermitteln: Geordnet nach Datum und Satznummer sind die identifizierten Sätze mit den zugeordneten Stilmitteln und der Erklärung zu sehen. Auf diese Art wird erkennbar, wie die Stilmittel im jeweiligen Text sukzessive aufeinander folgen.

Aufgrund der schier unermesslichen Menge der Stilmittel in Batty Webers Kalenderblättern ist es kaum möglich, eine umfangreiche Analyse oder auch eine Evaluation der Qualität der verschiedenen Prompts zu geben. Vielmehr muss sich hier auf die Präsentation einiger ausgewählter Beispiele beschränkt werden.

Das Stilmittel des Vergleichs lässt sich relativ einfach über das Vorhandensein der Vergleichspartikel „wie“ identifizieren und das LLM scheint auch diese Strategie zu verwenden. In der Erklärung werden die Vergleichsphrasen korrekt zitiert (z.B. „wie Kräuter im Maien“, „wie [...] Mordstahl“, „wie [...] Menschen“) und die Struktur des Vergleichs wird erläutert. Da die Prompts weitestgehend auf Englisch formuliert sind, erfolgt auch die Antwort in derselben Sprache.

1924-02-08	
Glücklich das Land, dessen Militärmusik nie in den Fall kommt, seinen Söhnen in der Schlacht zum Sturmangriff die nötige Courage anzublasen, fünf Minuten bevor der Tod sie fällen wird „wie Kräuter im Maien“.	The phrase „wie Kräuter im Maien“ compares the soldiers being struck down by death to herbs being cut down in May, highlighting the suddenness and inevitability of death in war.

<b>1916-06-23</b>	
<p>Ich dachte: Wie aus einem fauchenden, surrenden Stück Mordstahl dieses friedfertige Blumenblatt werden konnte, so wird auch einmal aus dem Krieg wieder Frieden, aus dem Schlachtfeld wieder der gründende Acker, aus dem Haß wieder Versöhnung.</p>	<p>The sentence uses a series of comparisons to illustrate transformation: from „Mordstahl“ to „friedfertige Blumenblatt“, from „Krieg“ to „Frieden“, from „Schlachtfeld“ to „gründende Acker“, and from „Haß“ to „Versöhnung“. These comparisons highlight the hope for peace and reconciliation after the war.</p>
<b>1915-07-01</b>	
<p>Wir sind wie Menschen, die im Wartezimmer eines Zahnarztes sitzen.</p>	<p>The phrase „wie Menschen, die im Wartezimmer eines Zahnarztes sitzen“ compares the experience of waiting for the war to end to sitting in a dentist’s waiting room, highlighting the anxiety and anticipation associated with war.</p>

Tabelle 2: Beispielergebnisse few-shot

Während sich die Erläuterungen der Kriegsthematik für die ersten beiden Beispiele auf den jeweiligen Satz beziehen („Tod sie fällen“ > „wie Kräuter im Maien“, „Mordstahl“ > „friedfertige[s] Blumenblatt“) und die Vergleiche in ihrer Gänze in diesem Satz realisiert werden, verhält es sich anders mit dem dritten Beispiel: Aus dem isolierten Satz kann die Interpretation zur „anxiety and anticipation associated with war“ nicht direkt herausgelesen werden. Hier hat das LLM den weiteren Kontext des Kalenderblattes miteinbezogen, in dem es tatsächlich um den Ersten Weltkrieg geht. Die Interpretation des Vergleichs im Kriegskontext ist demnach also prinzipiell korrekt. Auch hier hat das LLM also eine Vorgabe des Prompts, nämlich nur innerhalb der Grenzen eines Satzes zu bleiben, ignoriert.

Der Prompt „few-shot definition“ identifizierte im Ganzen 818 kriegsbezogene Metaphern, von denen einige charakteristische

Exemplare im Folgenden aufgeführt sind. Insgesamt erhält man den Eindruck, dass Metaphern an sich recht gut erkannt werden. Auch wenn die Erklärungen teilweise pauschalisierende und unspezifische Äußerungen enthalten (z.B. „devastation of war“), so ist die hermeneutische Leistung der Metaphernbeschreibung doch positiv hervorzuheben. Das LLM ist also durchaus in der Lage, die Verbindung des mechanischen, eventuell industriellen „Schwungrades“ mit den allgemeinen Auswirkungen des Krieges oder die Naturereignisse „Winter“ oder „Hagel“ in Verbindung mit „Tod“ zu setzen, wodurch die prinzipielle Leistung der Metapher korrekt erfasst wird. Auch im letzten Beispiel erscheint eine Naturmetapher mit der „Blüte der Mannheit“, die dann wiederum mit dem Tod im Krieg in Verbindung gebracht wird.

<b>1919-03-14</b>	
Das bluttriefende Schwungrad des Krieges schwingt in der Heimat entsetzlich aus.	The phrase „das bluttriefende Schwungrad des Krieges“ metaphorically describes the war as a blood-soaked flywheel, suggesting the ongoing and horrific impact of war on the homeland.
<b>1917-01-31</b>	
Winter ist Tod.	The sentence „Winter ist Tod“ uses a metaphor to equate winter with death, suggesting the harshness and lifelessness associated with winter, akin to the devastation of war.
<b>1914-10-30</b>	
Und da hinein heult, kracht, zischt, wuchtet es weiter von betäubenden Schlägen und stählernem Hagel und riesigen Stichflammen.	The phrase „stählernem Hagel“ (steel hail) metaphorically describes the artillery fire as hail, emphasizing the destructive and relentless nature of war.

<b>1931-11-24</b>	
Die Pflastersteine langweilen sich zu Tod.	The phrase „Die Pflastersteine langweilen sich zu Tod“ metaphorically attributes human boredom to the cobblestones, indicating a lifeless and dull environment, which can be related to the desolation of war.
<b>1916-08-03</b>	
Der Tod reißt nicht Tag für Tag in langen Reihen die Blüte unserer Mannheit aus dem Volk, durch unsere Straßen schleppen sich keine Krüppel, von denen das Vaterland das Opfer ihrer gesunden Glieder gefordert hat.	The phrase „Der Tod reißt ... die Blüte unserer Mannheit aus dem Volk“ metaphorically describes death as tearing the best of the nation’s men from the people, attributing a violent action to death.

Tabelle 3: Beispielergebnisse few-shot definition

Mit 216 Belegen repräsentieren die Metonymien die kleinste Gruppe der erkannten Stilmittel. Aus den folgenden Beispielen geht hervor, dass das LLM Metonymien auf der Einzelwortebene verortet: „Rot“ > „Blut“, „Eingeweide“ > „Schützengräben“, „Feldgrau“ > „Soldaten“. Auch hier kann generell konstatiert werden, dass die Zuordnungen und Erklärungen des LLMs größtenteils zutreffend sind.

<b>1920-03-26</b>	
Und das Rauten-Aß daneben grinst: Rot ist Trumpf! Blut ist Trumpf!	The phrase „Rot ist Trumpf! Blut ist Trumpf!“ uses metonymy to associate the color red with blood, emphasizing the violent and deadly nature of war.

<b>1917-03-11</b>	
Meine Eingeweide waren Schützengräben und mein ganzes Innere ein Trommelfeuer.	The sentence uses metonymies to describe the internal turmoil of the narrator by comparing „Eingeweide“ (entrails) to „Schützengräben“ (trenches) and „mein ganzes Innere“ („my entire inner self“) to „Trommelfeuer“ (drumfire), both of which are elements associated with war.
<b>1923-10-14</b>	
Und an der Eisenbahn hinauf und hinunter zogen Feldgraue die Wachen auf.	The term „Feldgraue“ is a metonymy for German soldiers, referring to them by the color of their uniforms, which is associated with the military presence during the war.

Tabelle 4: Beispielergebnisse zero-shot

Die Abgrenzung zu Metaphern ist mitunter schwierig und müsste in einer manuellen Feinanalyse diskutiert werden: Man vergleiche etwa die obige Metapher „Winter ist Tod“ mit der Metonymie „Blut ist Trumpf“. Solche subtilen, oft an den Kontext gebundenen Unterschiede machen deutlich, dass heutige LLMs noch nicht als ganz zuverlässige, autonom agierende Instrumente zur Textanalyse eingesetzt werden können.

Abschließend soll die zeitliche Verteilung der Stilmittel über die gesamte Periode analysiert werden. Die folgende Abbildung zeigt die Häufigkeit der einzelnen Stilmittel über die Jahre hinweg und es zeigt sich sehr deutlich, wie die kriegsbezogenen Belege in den Jahren 1914 bis 1920 gegenüber den übrigen Jahren dominieren. In den Jahren danach nimmt die Anzahl beträchtlich ab. Diese Darstellung kann darüber hinaus auch als eine indirekte Bestätigung der weitgehenden Korrektheit der Zuordnungen interpretiert werden.

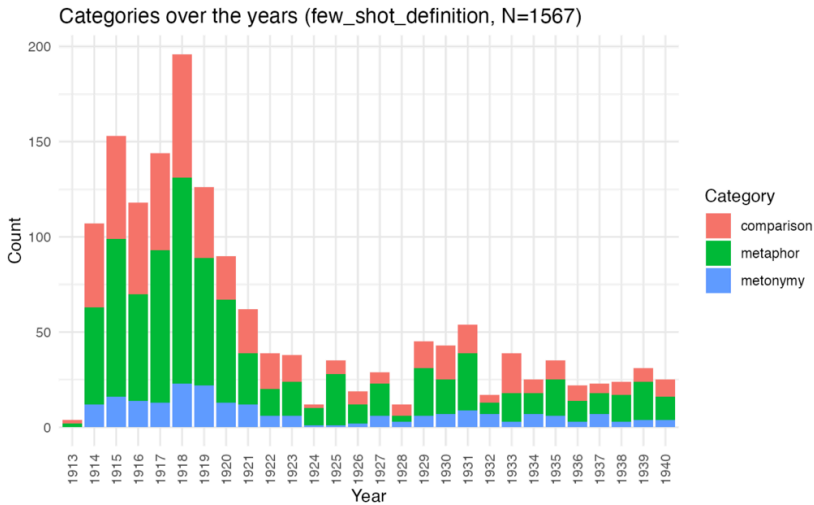


ABBILDUNG 2: Balkendiagramm zu den Ergebnissen

## Zusammenfassung der Ergebnisse

Insgesamt hat die Anwendung des LLM auf den *Abreißkalender* gezeigt, dass es ein erhebliches Potenzial für die Unterstützung bei Aufgaben des *close reading* hat. Das Modell war in der Lage, große Textmengen zu verarbeiten und aufschlussreiche Analysen zu den Stilmitteln zu liefern. Dieser Ansatz kann die Effizienz von Literaturstudien erheblich steigern, insbesondere bei Projekten mit geringen Ressourcen (vgl. Damai und Postema 2023). Das Modell identifizierte zahlreiche Belegstellen, an denen Batty Weber die Auswirkungen und Schrecken des Krieges auf die Gesellschaft, die Darstellung von Soldaten und den breiteren soziopolitischen Kontext der damaligen Zeit erörtert. Für unsere Fallstudie ließ sich das LLM überzeugend als hermeneutisches Hilfsmittel, als effizienter Leser und Exzerpierer einsetzen. Die Resultate stellen dann eine Vorstufe der eigentlichen literatur- oder kulturwissenschaftlichen Textanalyse dar.

## Diskussion

Die Anwendung eines aktuellen großen Sprachmodells zur Durchführung von automatisiertem *close reading* in Batty Webers *Abreißkalender* hat sowohl das Potenzial als auch die Herausforderungen der Integration fortschrittlicher NLP-Techniken in die Forschung im Bereich DH aufgezeigt. In dieser Studie haben wir festgestellt, dass das LLM große Textmengen effektiv verarbeiten und aufschlussreiche Analysen zu bestimmten Themen liefern kann. Bestimmte Einschränkungen und Überlegungen schreiben jedoch einen vorsichtigen Ansatz bei der Anwendung dieser Methoden vor.

Ob weitergehende hermeneutische Aufgaben an ein LLM übergeben werden können, müsste in weiteren Studien überprüft werden. Es erscheint uns allerdings offensichtlich, dass die LLM-gestützte Methode über eine korpusgestützte Methode hinausgeht bzw. sie sinnvoll ergänzen kann. Neben den angesprochenen Problemen, dass die Prompts hinsichtlich des Bezugs zur Kriegsthematik nicht immer einwandfrei funktionierten, liegt eines der noch zu lösenden Hauptprobleme in der Evaluation der Ergebnisse: Im vorliegenden Fall liegt kein manuell annotiertes Korpus für Vergleichszwecke vor. Es ist daher nicht möglich (oder nur durch hohen manuellen Aufwand) zu überprüfen, welche Belege falsch klassifiziert (*false positives*) und welche übersehen (*false negatives*) wurden.

Jüngste Fortschritte in der NLP, insbesondere die Entwicklung von transformerbasierten Modellen wie GPT-4 (vgl. Vaswani et al. 2018) haben die Möglichkeiten der automatischen Textanalyse erheblich verbessert. Diese Entwicklungen ermöglichen neue methodische Ansätze, die es erlauben, umfangreiche Korpora zu analysieren, die zuvor aufgrund von Ressourcenbeschränkungen kaum zugänglich waren (vgl. Terras 2022). Unsere Ergebnisse stimmen mit früheren Untersuchungen überein, die den Nutzen von LLMs bei der Aufdeckung thematischer Elemente in literarischen Texten belegen (vgl. Damai und Postema 2023; Underwood 2023). Gleichzeitig zeigen sich jedoch Limitationen, insbesondere in der genauen Interpretation nuancierter und historisch geprägter Sprache sowie in der Transparenz der Modellentscheidungen, da Verzerrungen und Befangenheit aus den

zugrundeliegenden, heterogenen Datensätzen in den Ergebnissen wiederauftauchen können (vgl. Rudin 2019; Shah et al. 2020).

Besonders hervorzuheben ist der kulturelle Kontext: Der *Abreißskalender* ist tief in der luxemburgischen Geschichte und Kultur verankert, was nicht nur neue Perspektiven für die Luxemburgistik eröffnet, sondern auch einen Beitrag zur Erhaltung der regionalen Literatur leisten kann (vgl. Krauwer 2003). Zudem ist das Verfahren hilfreich zur Exzerptesammlung und Belegstellenkompilation, was die Nachvollziehbarkeit der Ergebnisse unterstützt. Gleichzeitig übergeneralisieren LLMs oftmals, sodass – vermutlich – mehr Belegstellen identifiziert werden, als durch die ursprüngliche Konzeption des Prompts tatsächlich erwartet wird. Dies verdeutlicht, dass mehr Prompt-Engineering nötig ist, um sicherzustellen, dass alle Aspekte korrekt interpretiert werden, und um zu verhindern, dass bestimmte Aspekte, wie die Vorgabe, ein Stilmittel ausschließlich innerhalb des vorliegenden Satzes zu interpretieren und keinen darüberhinausgehenden Kontext zu berücksichtigen, ignoriert werden.

Insgesamt unterstreicht diese Studie die Notwendigkeit eines vorsichtigen, interdisziplinären Ansatzes, der menschliche Aufsicht, gezielte Modellanpassungen und eine kritische Reflexion der ethischen Implikationen vereint, um das Potenzial von LLMs in den DH optimal zu nutzen und ihre Schwächen zu adressieren.

## Konklusion

Zusammenfassend lässt sich sagen, dass die Integration großer Sprachmodelle in die DH trotz der verbleibenden Herausforderungen sehr vielversprechend scheint. Wenn man sich mit den Einschränkungen und ethischen Erwägungen auseinandersetzt, können diese Technologien die wissenschaftliche Arbeit erheblich bereichern und die detaillierte Analyse umfangreicher Textsammlungen leichter zugänglich und effizienter machen.

Trotzdem bedeutet dies nicht, dass manuelle Annotation „tot ist“ – also, dass man in wenigen Minuten einen Datensatz klassifizieren und offen

zugänglich machen kann. Zwar haben LLMs wie ChatGPT eine Debatte darüber ausgelöst, ob die traditionelle Datenbeschriftung obsolet ist, doch bleibt die menschliche Rückmeldung das Fundament zuverlässiger KI-Systeme. Reine synthetische Daten und KI-gestützte Beschriftungen weisen nachweisliche Mängel auf, etwa in Bezug auf Diversität, Genauigkeit und die Verstärkung von Vorurteilen. Anstatt also den Datenanmerkungsprozess vollständig durch KI zu ersetzen, könnte ein umgedrehter Ansatz, bei dem menschliche Experten die KI bei der Datenaufbereitung unterstützen, nicht nur die Qualität der Daten verbessern, sondern auch zu einer robusteren und faireren Modellentwicklung beitragen.

Diese Perspektive unterstreicht die Bedeutung der interdisziplinären Zusammenarbeit, in der Fachwissen aus den DH und den Computerwissenschaften zusammenfließt, um die Potenziale von LLMs optimal zu nutzen und gleichzeitig ihre Schwächen zu adressieren. Letztlich wird ein solcher Ansatz dazu beitragen, die Herausforderungen im Umgang mit automatisierten Textanalysen zu meistern und den Weg für eine ethisch verantwortungsvolle Anwendung von KI in der geisteswissenschaftlichen Forschung zu ebnen.

Abschließend zeigt das Fallbeispiel von Batty Webers *Abreißkalender*, wie der Einsatz von großen Sprachmodellen neue Wege in der Literatur- und Kulturforschung eröffnen kann. Batty Weber dient dabei als exemplarischer Beleg dafür, dass LLMs nicht nur die Effizienz bei der Analyse umfangreicher Texte steigern, sondern auch dazu beitragen können, kulturhistorische Fragestellungen differenzierter zu erfassen und zu interpretieren. Diese explorative Methode, die traditionelle *close-reading*-Ansätze mit innovativen, datenbasierten Analyseverfahren kombiniert, stellt einen vielversprechenden Ansatz dar, um die kulturelle Bedeutung und den historischen Kontext literarischer Werke neu zu beleuchten.

## Literaturverzeichnis

- Bamman, David, Kent K. Chang, Li Lucy und Naitian Zhou. „On Classification with Large Language Models in Cultural Analytics“. *arXiv:2410.12029*, 2024.
- Brown, Tom B., Benjamin Mann, Nick Ryder et al. „Language Models Are Few-Shot Learners“. *Advances in Neural Information Processing Systems* 33, 2020, S. 1877-1901.
- Bußmann, Hadumod (Hrsg.). *Lexikon der Sprachwissenschaft*. Kröner, 2008.
- Damai, Puspa, und Barbara Postema. „Humanities in the Time of ChatGPT and Other AI“. *Critical Humanities* 2 (1), 2023, S. i-vii.
- Hatzel, Hans Ole, Haimo Stiemer, Chris Biemann und Evelyn Gius. „Machine learning in computational literary studies“. *it – Information Technology* 65 (4-5), 2023, S. 200-217. doi: 10.1515/itit-2023-0041.
- Howard, Jeremy, und Sebastian Ruder. „Universal Language Model Fine-tuning for Text Classification“. *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, herausgegeben von Iryna Gurevych und Yusuke Miyao. Association for Computational Linguistics, 2018, S. 328-339.
- Krauwer, Steven. „The Basic Language Resource Kit (BLARK) as the First Milestone for the Language Resources Roadmap“. *Proceedings of SPECOM*, 2003, S. 8-15.
- Liu, Pengfei, Weizhe Yuan, Jinlan Fu, Zhengbao Jiang, Hiroaki Hayashi und Graham Neubig. „Pre-train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing“. *arXiv:2107.13586*, 2021.

- Mélanie-Becquet, Frédérique, Jean Barré, Olga Seminck et al. (2024): „BookNLP-fr, the French Versant of BookNLP. A Tailored Pipeline for 19th and 20th Century French Literature“. *Journal of Computational Literary Studies* 3 (1), S. 1-34. doi: 10.48694/JCLS.3924.
- Manovich, Lev. *Cultural Analytics*. MIT Press, 2020.
- Rudin, Cynthia. „Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead“. *Nature Machine Intelligence* 1 (5), 2019, S. 206-215.
- Shah, Deven Santosh, H. Andrew Schwartz und Dirk Hovy. „Predictive Biases in Natural Language Processing Models: A Conceptual Framework and Overview“. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, herausgegeben von Dan Jurafsky, Joyce Chai, Natalie Schluter und Joel Tetreault. Association for Computational Linguistics, 2020.
- Terras, Melissa. „Digital Humanities and Digitized Cultural Heritage“. *The Bloomsbury Handbook to the Digital Humanities*, herausgegeben von James O’Sullivan. Bloomsbury, 2022, S. 255-266.
- Underwood, Ted. „Using GPT-4 to measure the passage of time in fiction“. *The Stone and the Shell* [Blog], 19. März 2023. <https://tedunderwood.com/2023/03/19/using-gpt-4-to-measure-the-passage-of-time-in-fiction/>. Zugriff am 1. Mai 2025.
- Vaswani, Ashish, Noam Shazeer, Niki Parmar et al. „Attention Is All You Need“. *Advances in Neural Information Processing Systems 30. 31st Annual Conference on Neural Information Processing Systems (NIPS 2017)*, herausgegeben von Isabelle Guyon, Ulrike von Luxburg, Samy Bengio et al. NeurIPS, 2018, S. 5998-6008.