

Topic Modeling am Beispiel von Franz Kafka

Eine Arbeitsskizze

Laura Kraft (Literaturwissenschaft)

Franz Kafka ist einer der meistbeachteten Autoren in der literaturwissenschaftlichen Forschung. Es gibt unzählige Forschungsarbeiten zu seinem literarischen Werk, zu dessen Rezeption, seinem Leben, Leiden, Schreiben. Eine einfache Google-Anfrage zu *Kafka* ergibt 140 Millionen Treffer in 0,28 Sekunden, spezifiziert man das mit »*Franz Kafka*«, sind es immerhin noch 15,3 Millionen Ergebnisse. Doch noch immer gibt uns der Autor, der vor fast 100 Jahren starb, Rätsel wie kein anderer auf, sorgt er heute noch für Schlagzeilen und avancierte jüngst sogar zum Tik-Tok-Star. Dabei wollte er seinen schriftstellerischen Nachlass am liebsten verbrannt sehen und es ist allein dem Ungehorsam seines Freundes und Förderers Max Brodt zu verdanken, dass Kafka heute zum Kanon der Weltliteratur zählt und sein Name gar ein eigenes Adjektiv geprägt hat – *kafkaesk*.

Vermutlich ist es Kafka auf besondere und vielleicht ironische Weise angemessen, ihn mit quantitativen Analysemethoden zu untersuchen, hatte er doch neben seinem allnächtlichen Schriftstellerleben im Alltag als Angestellter und Jurist der Arbeiter-Unfall-Versicherungsanstalt für das Königreich Böhmen tagtäglich mit Statistiken und Zahlen zu tun. In seinem beruflichen Kontext als Jurist verfasste er im Büro etliche amtliche Schriften, in denen er mitunter durch die Verarbeitung jener Statistiken und Zahlen einer modernen Verwaltungsbehörde Ausdruck verlieh.¹ Mit seinem Brotberuf haderte er aber bekanntermaßen zeit seines Lebens »weil er meinem einzigen Verlangen und meinem einzigen Beruf das ist der Litteratur widerspricht«². Kafkas »schreckliches Doppelleben«³ zeigt sich aber nicht nur in der Diskrepanz zwischen der Amtstätigkeit im »Bureau«, wo er die für ihn ganz und gar sinnlose Amtstätigkeit verrichten musste, und dem

-
- 1 Zu Kafka und die Statistik empfiehlt sich auch: Wolf, Burkhardt (2006): Die Nacht des Bürokraten. Franz Kafkas statistische Schreibweise. In: Deutsche Vierteljahrsschrift für Literaturwissenschaft und Geistesgeschichte, 80.
 - 2 Kafka, F. (1990): Tagebucheintrag vom 21.8.1913. In: Ders.: Tagebücher. Schriften Tagebücher Briefe, Kritische Ausgabe, hg. v. H.-G. Koch, M. Müller & M. Pasley, Frankfurt a.M., S. 578f.
 - 3 Kafka, F. (1990): Tagebucheintrag vom 19.2.1911 In: Ders.: Tagebücher. Schriften Tagebücher Briefe, Kritische Ausgabe, hg. v. H.-G. Koch, M. Müller & M. Pasley, Frankfurt a.M., S. 29.

nächtlichen literarischen Schreiben. Der Erste Weltkrieg spielt mindestens eine genauso entscheidende Rolle für seinen Schaffensprozess wie auch die Identitätsproblematik im transkulturellen Raum als deutschsprachiger Jude in Prag.⁴ Kafka als Autor im Kontext des Ersten Weltkriegs zu betrachten, ist in der Forschung bisher weitestgehend vernachlässigt worden, herrschte dort bislang eine Lesart vor, die ihn mehr als den »ewigen Sohn« interpretierte, denn als einen Autor, der den Krieg diskursivierte. Dabei sind die meisten seiner literarischen Texte in eben jener Zeit entstanden.

Der Tagebucheintrag vom 2. August 1914 – »Deutschland hat Rußland den Krieg erklärt. – Nachmittag Schwimmschule«⁵ – wird gern als Beleg für Franz Kafkas vermeintliches Desinteresse am politischen Diskurs seiner Zeit und vor allem am Ersten Weltkrieg herangezogen. Erst in der neueren Forschung gibt es Ansätze, dieses vernachlässigte Themenfeld zu erschließen.⁶ Insbesondere der Begriff »Geopolitik«, der während des Ersten Weltkrieges Hochkonjunktur hatte, tauchte in der Forschung im Zusammenhang mit Kafka bisher nicht auf. Niels Werber begann mit einem Aufsatz 2017 erstmals dieses Thema zu erschließen und stellte dort exemplarisch Bezüge zwischen Kafkas literarischen Texten und Zeitungsartikeln aus dem Ersten Weltkrieg her unter besonderer Berücksichtigung des geopolitischen Diskurses.⁷

In der Kafka-Forschung fehlt nach wie vor eine umfassende Betrachtung Franz Kafkas als Zeitungsleser und dessen Einbettung und Operationalisierung des Zeitungsmediums in den Schreibprozess. Im Rahmen eines umfangreicheren Projektes sollen deshalb systematisch (geo)politische Semantiken in Kafkas Erzählungen aus dem Ersten Weltkrieg herausgearbeitet und untersucht werden, inwiefern der (geo)politische Diskurs und vor allem die Zeitungsberichterstattung für das Schreiben Kafkas von Bedeutung waren.⁸ Kafka als interessierter Zeitungsleser (und Abonnent des *Prager Tagblatts*)

-
- 4 Vgl. hierzu auch Deleuze, G./Guattari, F. (1976): *Kafka. Für eine kleine Literatur*. Frankfurt a.M.: Suhrkamp Verlag (8. Auflage, 2012) sowie Jeon, Yoo Jung (2017): *Zur Dynamik von Differenz und Minorität. Das Politische und Ethische der Minoritätsliteratur im besonderen Hinblick auf Franz Kafka und Emine Sevgi Özdamar*. Würzburg: Königshausen und Neumann.
 - 5 Kafka, F. (1990): *Tagebucheintrag vom 2.8.1914*. In: Ders.: *Tagebücher. Schriften*. Tagebücher, Kritische Ausgabe, hrg. v. Hans-Gerd Koch, Michael Müller & Malcolm Pasley, Frankfurt a.M., S. 543.
 - 6 Besonders zu erwähnen sind hier: – Engel, M./Robertson, R. (Hg.) (2012): *Kafka, Prag und der Erste Weltkrieg*, Würzburg: Königshausen und Neumann. – Anz, T. (1996): *Kafka, der Krieg und das größte Theater der Welt*. In: *Neue Rundschau*, Nr. 3, S. 131–143. – Wagner, B. (2009): *Kafkas Poetik des Unfalls*. In: *Kassung, C. (Hg.): Die Unordnung der Dinge: Eine Wissens- und Mediengeschichte des Unfalls*. Bielefeld: transcript, S. 421–254. – Werber, N. (2017): *Kafkas Geopolitik. 1917 – Schreiben am »Zeitungsrand der Weltgeschichte«*. *Neue Rundschau*, 128, S. 209–227 – Kittler, W. (1990): *Grabenkrieg – Nervenkrieg – Medienkrieg. Franz Kafka und der I. Weltkrieg*. In: *Hörisch, J./Wetzels, M. (Hg.): Armaturen der Sinne. Literarische und technische Medien 1870–1920*. München: Fink, S. 289–309. – Birgfeld, J. (2012): *Der Erste Weltkrieg im Prager Tagblatt: Zur Präsenz des Krieges als Kommunikationsereignis im Umfeld Kafkas*. In: Engel, M./Robertson, R. (Hg.): *Kafka, Prag und der Erste Weltkrieg*, Würzburg: Königshausen und Neumann, S. 19–35. – Neumann, B. (2014): *Franz Kafka und der Große Krieg: Eine kulturhistorische Chronik seines Schreibens*. Würzburg: Königshausen & Neumann.
 - 7 Vgl. Werber: *Kafkas Geopolitik*.
 - 8 Es handelt sich bei dem vorliegenden Beitrag um eine Auskopplung aus dem Dissertationsprojekt und dient vor allem der Reflexion methodologischer Herausforderungen.

konnte die Ereignisse um den Ersten Weltkrieg und den politischen Diskurs mitverfolgen und hat diese in seinen Tagebucheinträgen und Briefen kommentiert. Die Auswirkungen des Krieges auf Kafkas Leben und vor allem auf sein literarisches Schaffen lassen sich anhand vieler Briefe und Tagebucheinträge nachweisen. Kafka verhandelt den Krieg in seinen literarischen Texten allerdings nicht explizit, sondern indirekt durch Verfahrensweisen der Verschiebung, Verrätselung und Verdichtung.⁹ Die quantitative Analyse-methode »Topic Modeling« ist hierbei eine vielversprechende digitale Methode, um verborgene semantische Strukturen offenzulegen und eignet sich besonders für die Analyse literarischer Texte in einem mittelgroßen Korpus.¹⁰ Das Korpus für die Analyse besteht aus Kafkas Erzählungen aus der Zeit des Ersten Weltkriegs, seinen Tagebüchern und Briefen sowie Zeitungsausgaben vom *Prager Tagblatt* (1914–1918). Anhand dessen wird untersucht, ob es intertextuelle Bezüge zwischen den Zeitungen, die Kafka wohl gelesen hat, und seinen Erzählungen gibt und wie sich die literarischen Texte vor dem Hintergrund dieser organisieren. Darüber hinaus soll der Schreibprozess und Kafkas poetologische Verfahrensweisen in den Blick genommen werden.

So viel zum Themenfeld bzw. Desiderat in der Kafka-Forschung, das seitens der Verfasserin im Rahmen eines größeren Projektes bearbeitet wird. Der Fokus soll in dem vorliegenden Beitrag auf der methodologischen Erörterung der quantitativen Analyse-methode Topic Modeling liegen und die Chancen und Grenzen von Topic Modeling in Hinblick auf literaturwissenschaftliche Fragestellungen und dem hier beschriebenen Forschungsgegenstand darlegen. Dabei ist dieser Beitrag nicht als Sentenz einer abgeschlossenen Forschungsarbeit zu verstehen, sondern weist in diesem Stadium noch einen experimentellen Charakter auf. Vielmehr will dieser Beitrag aufzeigen, warum es für die Forschung interessant sein kann, wenn man mit Hilfe eines Algorithmus bei Franz Kafka zwischen den Zeilen lesen will.

Kafkas Schreibweise im Krieg

Kafka, der die Welt und das Leben schreibend verwand, »[d]a ich nichts anderes bin als Litteratur und nichts anderes sein kann und will«¹¹, hatte in den rund 28 Jahren literarischer Tätigkeit, von der rund 22 Jahre in Form von Texten überliefert sind, Phasen hoher literarischer Produktivität, die sich mit ruhigen Phasen und nahezu versiegender Tätigkeit abwechselten. Die Zeit während des Ersten Weltkriegs ist ausgerechnet die Phase seiner höchsten literarischen Produktivität. Zumindest entstehen dort die meisten Er-

9 Vgl. hierzu auch Kilcher, A. (2015): Politik und Parabolik. Kafkas Texturen des Krieges. In: Shahar, G. (Hg.): Texturen des Krieges. Körper, Schrift und der Erste Weltkrieg. Göttingen: Wallstein Verlag, S. 223–243.

10 Vgl. Weitlin, T. (2017): Scalable Reading. In: LiLi. Zeitschrift für Literaturwissenschaft und Linguistik. [Online] 47 (1), 1–6. Available from: <https://link.springer.com/article/10.1007/s41244-017-0048-4>

11 Kafka, Franz: Tagebucheintrag vom 21.8.1913 In: Ders.: Tagebücher. Schriften Tagebücher, Kritische Ausgabe, hrg. v. Hans-Gerd Koch, Michael Müller & Malcolm Pasley, Frankfurt a.M. 1990, S. 579.

zählungen.¹² Das lässt sich auch damit erklären, dass es gerade die widrigen Umstände zu sein scheinen, die Kafka erst das Schreiben ermöglichen, wobei es gleichermaßen eben jene sind, die ihn davon abzuhalten drohen.¹³ Die literarische Schaffenszeit während des Krieges lässt sich in drei Produktionsphasen einteilen, deren Beginn und Ende charakterisiert ist von Schreibenlässen, privaten Ereignissen und Veränderungen des Schreibortes oder -mediums.¹⁴ Zu erwähnen sind hier auch die Tagebücher und Briefe, die in untrennbarem Zusammenhang zum literarischen Schaffensprozess stehen. So weisen die Tagebücher mitunter einen Notiz- und Werkstattcharakter auf und bilden einen teils fließenden Übergang von biografischer Dokumentation zu fiktionalisierten Textentwürfen. Dies liegt nicht zuletzt in der (hybriden) Verwendung des Schreibmediums, welches in der Regel aus Heften im Quart- und Oktavformat bestand – die Begrenzung des Textumfangs durch das Heft tat sein Übriges dazu.

Die Differenzierung von Textsorten ist hier also kaum möglich. Dies wird umso deutlicher, stellt man konsequent den Schreibprozess ins Zentrum der Betrachtung: Kafkas Werk beschreibt ein »Spannungsfeld«¹⁵. Es ist »zu großen Teilen an der Schwelle zwischen dem Geschriebenen und dem Schreiben angesiedelt, es liegt zwischen Vollendung und Fertigstellung einerseits und Fragment und Verwerfung andererseits, zwischen Veröffentlichung und Schreibprozess.«¹⁶ Somit eröffnet sich eine komplexe Fragestellung nach Kafkas Schreibweise im Krieg. Die Ungewissheit darüber, ob »das im Text behandelte Motiv (Signifikant) auch dessen Thema (Signifikat)«¹⁷ sei, oder es sich um einen »relativ beliebige[n] Produktionsanlass, der im Schreibprozess metonymisch auf eine viel *allgemeinere* oder sogar metaphorisch auf eine *ganze andere* Bedeutungsebene überführt wird«¹⁸ handele, identifizieren Engel und Robertson als »Grundproblem der Kafka-Deutung«.¹⁹ Laut Kilcher ließe sich Kafkas methodische Verfahrensweise »zugleich als Arbeit der Transposition und der Kondensation verstehen, die die fokussierte Direktheit in eine allgemeine Indirektheit transferiert und dabei das Angesprochene verbirgt und verrät.«²⁰ In Kafkas literarischen Texten ist der Krieg kaum inhaltlich greifbar, bestenfalls metaphorisch vermittelt. Die beschriebene Deutungsproblematik könne man »[...] umgehen, wenn man [...] nur nach im weitesten Sinne kriegsbezogenen Diskurselementen sucht«,²¹ so schlussfolgern Engel und Robertson.

-
- 12 Zur Unterteilung von Kafkas Œuvre in Produktion- und Werkphasen vgl. Engel, Manfred (2010): Drei Werkphasen. In: Ders./Auerochs, B. (Hg.): Kafka. Handbuch. Leben – Werk – Wirkung. Stuttgart, Weimar: S. 88–90.
- 13 Vgl. dazu auch: Jahraus, O. (2006): Kafka. Leben, Schreiben, Machtapparate. Stuttgart: Reclam, S. 131ff.
- 14 Vgl. Engel: Drei Werkphasen.
- 15 Vgl. dazu auch: Kittler, W./Neumann, G. (1990): Kafkas »Drucke zu Lebzeiten«. Editorische Technik und hermeneutische Entscheidung. In: Kittler, W./Neumann, G. (Hg.): Franz Kafka. Schriftverkehr. Freiburg i.Br., S. 32.
- 16 Vgl. dazu auch: Jahraus: Kafka, S. 38.
- 17 Engel, M./Robertson, R. (2012): Vorwort/Preface. In: Engel, M./Robertson, R. (Hg.): Kafka, Prag und der Erste Weltkrieg, Würzburg: Königshausen und Neumann, S. 14.
- 18 Ebd.
- 19 Ebd.
- 20 Kilcher: Politik und Parabolik, S. 225.
- 21 Engel/Robertson: Vorwort/Preface, S. 14.

Kafka als Zeitungsleser

Der Presse kommt als bedeutsames Kommunikationsmedium während des Ersten Weltkrieges eine exponierte Rolle zu und die Entwicklung der Zeitung hin zum Medium der Massenkommunikation im 19. Jahrhundert ist in der Forschung gut untersucht.²² Wie intensiv Kafka während des Ersten Weltkrieges Zeitungen konsumiert oder welche Ausgaben er wann gelesen hat, ist nicht detailliert überliefert. Dass er aber ein regelmäßiger Zeitungskonsument und -abonnent war, lässt sich zuverlässig durch Selbstaussagen belegen. So erwähnt er u.a. in einem Brief an Felice Bauer 1912 seine Lektüregewohnheiten:

Übrigens lese ich selbst nur das Prager Tagblatt und dieses sehr flüchtig, und an Zeitschriften die Neue Rundschau und dann noch ›Palästina‹, das mir jedoch nicht mehr zugeschickt wird, trotzdem ich noch immer Abonnent bin.²³

Kafka hatte aber nicht nur Zeitungen und Zeitschriften abonniert – auch seine eigenen literarischen Texte wurden teilweise in eben jenen Zeitungen und Zeitschriften abgedruckt. So findet sich z.B. *Ein Traum* gleich auf der ersten Seite der Beilage zur Ausgabe des *Prager Tagblatts* vom 6. Januar 1917. Dass es einen nachvollziehbaren Bezug zwischen Kafkas literarischen Texten während des Krieges und dem geopolitischen Diskurs in der Tagespresse gibt, wurde in der Forschung wie erwähnt bereits exemplarisch dargelegt.²⁴ Welchen Einfluss die Presseberichterstattung u.a. auf die Genese von Kafkas *Ein Bericht für eine Akademie* hatte, lässt sich anhand der guten Quellenlage recht zuverlässig herleiten.²⁵ Hartmut Binder wies dies detailliert nach, indem er auf intertextuelle Bezüge zu u.a. E.T.A Hoffmann aufmerksam machte, insbesondere aber auch auf Diskurse in der deutschsprachigen Prager Presse, die die Zurschaustellung und Dressur von Menschenaffen thematisierten sowie vermehrt für Zirkus und Variété warben. Nicht nur anhand des Kontextes, sondern auch der Erzählstruktur zeichnet er minutiös nach, wer »Rotpeters Ahnen«²⁶ in Kafkas *Bericht für eine Akademie* waren.

Diese Diskurselemente helfen, um Schreibanlässe zu identifizieren und die Bedeutung der Briefe und Tagebücher für Kafkas Schaffensprozess unter besonderer Berücksichtigung der Referenzialität in den Blick zu nehmen. Die evidenten Zeitungsbezüge allein erklären jedoch noch nicht, wie Kafka eben diese produktiv macht und operationalisiert. Einen Hinweis auf eine Verfahrenstechnik gibt Franz Kafka wiederum selbst:

[...] Nun habe ich schon längst den Plan gehabt und nur aus Nachlässigkeit ihn immer wieder auszuführen unterlassen, verschiedene Zeitungsnachrichten, die mir aus

-
- 22 Zu erwähnen sei hier exemplarisch: Wilke, J. (2000): Grundzüge der Medien- und Kommunikationsgeschichte. Von den Anfängen bis ins 20. Jahrhundert, Köln/Weimar/Wien: Böhlau Verlag, 2. durchgesehene und ergänzte Auflage 2008, und auch Wilke, J. (2011): Von der frühen Zeitung zur Medialisierung. Gesammelte Studien II. Bremen: [edition lumière].
- 23 Kafka, F. (1999): Brief an Felice Bauer vom 24.11.1912 In: Ders.: Briefe 1900–1912. Schriften Tagebücher Briefe, Kritische Ausgabe, hg. v. H.-G. Koch, Frankfurt a.M., S. 261.
- 24 Vgl. Werber: Kafkas Geopolitik.
- 25 Vgl. Binder, H. (1983): Kafka. Der Schaffensprozess. Frankfurt a.M., S. 271–305.
- 26 Vgl. ebd.

irgendeinem Grunde überraschend waren, mir nahegingen und mir persönlich für nicht absehbare Zeit wichtig schienen [meistens waren es für den ersten Blick nur Kleinigkeiten, [...] auszuschneiden und zu sammeln. Fast jeden zweiten Tag finde ich in der Zeitung eine derartige förmlich für mich allein bestimmte Nachricht, aber ich habe nicht die Ausdauer eine solche Sammlung für mich anzufangen, wie erst für mich sie fortsetzen. Für Dich aber mache ich es mit Freuden, tu es doch, wenn es Dir gefällt, von Deiner Seite für mich. Solche Nachrichten, die nicht für alle Leser bestimmt sind, sondern nur auf bestimmte Leser hie und da zielen, ohne dass der unbeteiligte Beurteiler den Grund des besonderen Interesses herausfinden könnte, gibt es doch gewiß für jeden und solche kleine Nachrichten, die Dich besonders bekümmern, hätten für mich mehr Wert als meine eigene Sammlung, die ich Dir also ohne großes Bedauern schicken könnte. Verstehe mich recht, nur kleine Ausschnitte aus Tageszeitungen meine ich, meistens über wirkliche Ereignisse, Ausschnitte aus Zeitschriften wären nur seltene Ausnahmen, Du darfst nicht glauben, dass ich Deine schönen Hefte für mich zerreißen will.²⁷

Ein Beispiel lieferte Kafka gleich mit, indem er eine ausgeschnittene Notiz aus der Ausgabe des *Prager Tagblatts* vom 25. September 1912 über die *Seligspredung der Märtyrer von Uganda* beilegte.²⁸ Wenn Kafka einen nicht unwesentlichen Teil seines Briefes an Felice Bauer damit zubringt, die Praktik des Ausschneidens und Sammelns von Zeitungsnachrichten zu elaborieren, wird umso deutlicher, dass es sich hier nicht um eine Beliebigkeit, sondern um einen methodisch ausgerichteten Vorgang der Selektion handelt. Die Kriterien seiner Auswahl erwähnt er zugleich; es handelt sich um solche Notizen, die »überraschend waren«, ihm »nahegingen«, oder ihm »persönlich für nicht absehbare Zeit wichtig schienen«.²⁹ Dabei macht er deutlich, dass die Auswahl geeigneter Zeitungsnachrichten auf rein subjektiven Kriterien seinerseits beruht und er nimmt sogar an, dass Felice Bauer ihrerseits ebenso verfährt.

Indem Kafka Zeitungen nicht einfach nur liest, sondern selektiert, ausschneidet und archiviert, individualisiert er nicht nur das Massenmedium, sondern nutzt es auch ganz im Sinne seiner Zeit als »das Papierobjekt der Moderne«³⁰. Die Praktik des Zeitungsausschneidens ist auf mehreren Ebenen an die Schreibprozeduren Kafkas anschlussfähig und bildet ein, wenn nicht gar das konstituierende Element. Es zeigt sich, dass Kafkas Weltkriegs-Schreiben ein collagierendes und kein lineares Erzählen ist: das Verschieben hin zu Mehrdeutigkeit und die Verrätselung sind seinem Verfahren inhärent.³¹ Der Zeitungsausschnitt als »Chiffre des modernen Massenlebens«³² wurde nicht selten mit der »Idee der Verwertung«³³ aufbewahrt. Nicht zuletzt erfüllt er aber auch eine »cut and paste-Funktion«: »[...]In seinen variablen und mobilen Eigenschaften, im Trennen und Montieren, Loslösen und Verschieben, Neuordnen und Rekombinieren ist der Zeitungsaus-

27 Kafka, F. (1999): Brief an Felice Bauer vom 24.11.1912 In: Ders.: Briefe 1900–1912. Schriften Tagebücher Briefe, Kritische Ausgabe, hrg. v. Hans-Gerd Koch, Frankfurt a.M., S. 261.

28 Ebd.

29 Ebd.

30 te Heesen, A. (2006): Der Zeitungsausschnitt. Ein Papierobjekt der Moderne. Frankfurt a.M., S. 21.

31 Vgl. Kilcher: Politik und Parabolik, S. 225ff.

32 te Heesen: Der Zeitungsausschnitt, S. 13.

33 Ebd., S. 12.

schnitt ein für die Moderne konstitutives Objekt.«³⁴ Bei der Frage nach der Einbettung und Operationalisierung von Weltkriegsdiskursen der Prager Presse in Kafkas literarische Texte muss hier betont werden, was zunächst evident zu sein scheint, im Weiteren aber eine größere methodische Relevanz haben wird: Kafkas literarisches Verfahren ist keines im Sinne von »copy and paste«. Es weist aber Elemente von »cut and paste«³⁵ auf.

Topic Modeling als Methode

Auch wenn bereits exemplarische Beispiele für Bezüge zwischen Kafkas Zeitungskonsum und seinen literarischen Texten aufgezeigt wurden, steht eine systematische Untersuchung in der Forschung nach wie vor aus. Hierfür lohnt sich die Einbindung einer quantitativen Analyseverfahren. Nach wie vor ist in der Literaturwissenschaft eine gewisse Skepsis gegenüber quantitativen Analyseverfahren beobachtbar, da diese geeignet sein müssen, Antworten auf komplexe literaturwissenschaftliche Fragestellungen zu bieten – ihren Ursprung haben sie aber in anderen Fächern wie z.B. der Informatik, den Sozialwissenschaften oder auch der Computerlinguistik.³⁶

Das hier vorgestellte Kafka-Beispiel verdeutlicht, wo die Herausforderungen liegen: Es geht nicht um Worthäufigkeiten, sondern um Bedeutungsanalyse. Es gilt also, eine quantitative Methode auszumachen, mit der die klassische literaturwissenschaftliche Fragestellung im Rahmen einer hermeneutischen, intertextuellen und diskursanalytischen Betrachtung angemessen bearbeitet werden kann. In seiner Versuchsreihe zur digitalen Literaturgeschichte erörtert Thomas Weitin die Potenziale und Herausforderungen quantitativer Methoden in der Geisteswissenschaft und liefert einen wichtigen Beitrag zum aktuellen Methodendiskurs.³⁷ 2017 veröffentlichten Weitin und Herget einen Themenbeitrag in der Zeitschrift für Literaturwissenschaft und Linguistik, wo sie ihre Analyse des *Deutschen Novellenschatzes* von Paul Heyse anhand der Methode Topic Modeling vorstellten.³⁸ Diese Methode scheint geeignet zu sein, um ein mittelgroßes Korpus wie das hier zu betrachtende Kafka-Korpus nach semantischen Strukturen untersuchen zu können und das ist auch das Versprechen dieser Methode: Die Sichtbarmachung latenter semantischer Strukturen³⁹, oder auch »hidden structures«⁴⁰. Beim Topic Modeling handelt es sich um ein *distant reading*-Verfahren, das als Methode der quantitativen Bedeutungsanalyse geeignet ist, die Einzeltextanalyse mit der Analyse großer Textmengen zu kombinieren.⁴¹ Topic Modeling hat seine Basis nicht im Wörterzählen, sondern

34 Ebd., S. 21.

35 Vgl. ebd.

36 Vgl. Weitin, T. (2021): *Digitale Literaturgeschichte. Eine Versuchsreihe mit sieben Experimenten*. Berlin: J.B. Metzler/Springer, S. 3.

37 Vgl. ebd.

38 Weitin: Scalable Reading.

39 Weitin, T./Herget, K. (2017): Falkentopics. Über einige Probleme beim Topic Modeling literarischer Texte. In: Scalable Reading. LiLi. Zeitschrift für Literaturwissenschaft und Linguistik. [Online] 47 (1), S. 30.

40 Vgl. Blei, D. (2011): Probabilistic Topic Modeling. In: *Communications of the AMC* 55.4., S. 79.

41 Vgl. Weitin: *Digitale Literaturgeschichte*, S. 9.

beruht auf statistischen Modellen, die das gemeinsame Auftreten von Wörtern und deren unterschiedliche thematische Ausprägung in einem Text abbilden.⁴² Für das weitere Vorgehen ist es relevant, die Definition eines Topics näher zu betrachten:

Unter einem ›Topic‹ versteht man dabei eine Gruppe von Wörtern (wie zum Beispiel die Wörter »Theater«, »Schauspieler« und »Stück«), die in einem Text ungewöhnlich – d.h. statistisch auffällig – oft gemeinsam vorkommen. Ein ›Topic‹ ist also ein statistisches Phänomen und damit zwar eine Entsprechung, aber nicht exakt das Gleiche wie ein (inhaltlich definiertes) Thema.⁴³

Das Verfahren lohnt sich bei der Exploration größerer Textmengen, eine tiefere Kenntnis des Prozesses sowie des Korpus sind aber unumgänglich, um die Ergebnisse sach- und fachgerecht interpretieren zu können und Scheinevidenzen zu vermeiden. Zwar handelt es sich bei den Topics nicht um Themen im literarischen Sinne, sie sind aber als solche interpretierbar⁴⁴ und dienen somit als »Indikator für den jeweils verarbeiteten literarischen Stoff«. ⁴⁵ Topic Modeling weist demnach eher »den Charakter einer textanalytischen Heuristik«⁴⁶ auf. Die literaturwissenschaftliche Unterscheidung zwischen Thema, Stoff, Motiv und Topos zeigt hier ihre besondere Relevanz, wenn es um die Auslegung der Ergebnisse geht. Und es ist auch eben jene Differenzierung, die sich als für die Kafka-Deutung so herausfordernd erweist. Wo wir wieder bei Kafkas Weltkriegs-Schreiben wären und der Frage von Engel und Robertson, ob »das im Text behandelte Motiv (Signifikant) auch dessen Thema (Signifikat)«⁴⁷ sei.

Darüber hinaus handelt es sich beim Topic Modeling um ein probabilistisches, unüberwachtes Verfahren, das heißt, auch beim Setzen gleicher Parameter ist die Reproduzierbarkeit nicht exakt gegeben.⁴⁸ Das liegt auch darin begründet, dass man nur begrenzt Einblick in den Maschinenraum des Algorithmus und den Verfahrensprozess der Modellierung erhält.⁴⁹ Weitin schlussfolgert demzufolge:

Der Gedanke, dass bei der Analyse von Bedeutung die Textlektüre die größte Kontextsensitivität besitzt, wohingegen Topic Models abstrakte semantische Felder weit vom Text entfernt darstellen, ist sowohl semantisch als auch im technischen Sinne richtig, insofern es sich um ein *bag of words*-Verfahren handelt.⁵⁰

42 Vgl. ebd., S. 118ff. sowie J. Horstmann (2018): »Topic Modeling«. In: forTEXT. Literatur digital erforschen, § 2. URL: <https://fortext.net/routinen/methoden/topic-modeling> [Zugriff: 03. November 2023].

43 Ebd.

44 Vgl. ebd.

45 Ebd.

46 Ebd.

47 Engel, M./Robertson, R. (2012): Vorwort/Preface. In: Engel, M./Robertson, R. (Hg.): Kafka, Prag und der Erste Weltkrieg. Würzburg: Königshausen und Neumann, S. 14.

48 Vgl. Horstmann: Topic Modeling, § 13.

49 Vgl. ebd.

50 Weitin: Digitale Literaturgeschichte, S. 9.

Das bedeutet demzufolge, »die entsprechenden statistischen Ergebnisse lassen sich nicht direkt hermeneutisch rekontextualisieren, indem man einzelne, gut interpretierbare Wortfelder (Topics), die das Modell errechnet hat, im Text nachvollzieht.«⁵¹ Weitin ist es dennoch wichtig zu betonen, dass es sich bei Topic Models nicht um *black boxes* handelt, da zu jedem Topic-Wort eine Konkordanz⁵² ausgegeben werden kann.⁵³ Ein bestehendes Problem ist aber nach wie vor die Stabilität. Bei ihren Versuchen haben Weitin und sein Team Ernüchterndes zutage gefördert und als Lösungsansatz Kontrollmodelle einbezogen, die eine differenzierte Betrachtung der Ergebnisse ermöglichen.⁵⁴

Durch die vertiefte Recherche zum Topic Modeling und der aufschlussreichen Versuchsreihe der Darmstädter Literaturwissenschaftler:innen rund um Weitin sowie eigener Gehversuche mit der Methode, kristallisieren sich folgende wichtige Aspekte in Hinblick auf das Kafka-Korpus heraus: *Preprocessing*, Korpus-Kenntnis, Methoden-Kompetenz (Coding).

Preprocessing – Je sorgfältiger, desto besser

Das *Preprocessing* für einen Versuch mit Topic Modeling erweist sich als recht aufwendig, voraussetzungsreich und erfordert Sorgfalt. Es ist aber unumgänglich, wenn überhaupt verlässliche und interpretierbare Ergebnisse im Verfahren entstehen sollen. Hier liegen im Bearbeitungsprozess noch größere Hürden vor, die mit zusätzlichen Arbeitsschritten beseitigt werden sollen. Das ist auch der Grund, warum hier zu diesem Zeitpunkt noch keine bemerkenswerten Phänomene in den Topics oder aufwendige Visualisierungen abgebildet werden können und der experimentelle Charakter im Vordergrund steht.

Das Korpus setzt sich zusammen aus den Zeitungsausgaben des *Prager Tagblatts* von 1914–1918, Kafkas im Ersten Weltkrieg entstandene literarische Texte sowie seine Briefe und Tagebücher aus dieser Zeit. Die Zeitungsausgaben sind alle über das Projekt »ANNO« (AustriaN Newspaper Online) der Österreichischen Nationalbibliothek online abrufbar.⁵⁵ Es handelt sich dabei um Digitalisate historischer Zeitungen und Zeitschriften. Einzelne Ausgaben können in hoher Auflösung online betrachtet werden und stehen zudem als Download im PDF-Format und als TXT-Datei zur Verfügung. Eine Volltextsuche ist ebenso möglich. Für die Bereitstellung der Dateien arbeitet ANNO mit einem automatisierten Verfahren zur Texterkennung (OCR).⁵⁶ Eine optische Zeichenerkennung (OCR) ist hier auch notwendig, da die zu untersuchenden Ausgaben im Original in Fraktursatz gedruckt sind. Für die Zusammenstellung des Korpus können also die verfügbaren TXT-Dateien verwendet werden. Sämtliche Texte Kafkas stehen über das *TextGridRepository* digitalisiert als XML-Dateiformat zur Verfügung. Für die Tagebücher und Briefe

51 Ebd.

52 Anmerkung: Es handelt sich dabei um eine Belegliste (meist im Format *Key Word in Context*), die den unmittelbaren Kontext der untersuchten Wörter aufzeigt.

53 Vgl. ebd., S. 130f.

54 Vgl. ebd., S. 132f.

55 Abrufbar über: <https://anno.onb.ac.at>

56 Zur Digitalisierung bei ANNO: <https://anno.onb.ac.at/node/11>

kann man zudem (vorerst) auf weitere digitale Ressourcen wie Gutenberg.de zurückgreifen.

Bei dem vorliegenden Untersuchungsgegenstand handelt es sich also um ein mittelgroßes Korpus. Das ist von Vorteil, da eine Einzeltextanalyse nach wie vor möglich ist und das *too big to read*-Argument nur bedingt greift, das Korpus aber schon groß genug ist, um ausreichend Repräsentanz abbilden können. Topic Modeling liegt in der Regel ein Algorithmus mit dem Namen *Latent Dirichlet Allocation* (LDA) zugrunde, auf dem auch das Tool *MALLET* basiert.⁵⁷ *MALLET* kann zudem in der Programmierumgebung R bzw. R-Studio als Software genutzt werden. Weitin und sein Team arbeiten in ihren Versuchsreihen damit und an den Schritten und Parametern habe ich mich bei den ersten Versuchen orientiert.⁵⁸ Für Topic Modeling wird darüber hinaus eine Stoppwortliste benötigt, um Funktionswörter auszusortieren, die kaum einen semantischen Wert haben. Mit einer *Named Entity Recognition* (NER) können z.B. Personennamen, oder Namen von Orten markiert, extrahiert und gesondert gelistet werden.⁵⁹ Diese Listen können aber jederzeit modifiziert und auch händisch ergänzt werden. Zu Beginn habe ich mit den entsprechend aufbereiteten Datensätzen der Zeitungsausgaben gearbeitet, um mir zunächst einen Überblick über diese Topics zu verschaffen; das fusionierte Kafka-Korpus sollte dann in den nächsten Schritten durchlaufen werden. Als Zielgröße wurden 100 Topics á 10 Wörter festgelegt und wie vorgeschlagen mit 10.000 Iterationen gearbeitet.⁶⁰ Die Parameter Topicanzahl, Wörter pro Topic und Iterationen werden dann je im weiteren Durchlauf entsprechend angepasst, bis man stabile, reproduzierbare Ergebnisse hat. Hier traten bei meiner Durchführung erhebliche Schwierigkeiten auf, die sich auch mit veränderten Parametern nicht verbessern ließen.

Ein Blick in die Datensätze des Zeitungskorpus gibt hier Aufschluss: Das Korpus ist schlichtweg noch zu fehlerhaft, um damit eine quantitative Analyse durchführen zu können. Trotz OCR-Verfahren hat die Übertragung vom ursprünglichen Fraktursatz in eine maschinenlesbare Text-Datei nur mangelhaft funktioniert. Schaut man sich die Digitalisate bei ANNO in ihrer Bandbreite an, erkennt man, dass diese zwar eine hohe Qualität aufweisen, aber natürlich nur so gut wie die zu digitalisierenden Originale sein können. Und die sind sehr heterogen. Auch die recht zuverlässige Software *ABBYY FineReader* konnte hier zunächst keine wesentlichen Verbesserungen erzielen. Auch im Darmstädter LitLab arbeitet man mit OCR und einer händischen Nachkorrektur mit dem *ABBYY FineReader*.⁶¹ Um zu einem akzeptablen Volltextkorpus zu gelangen, führt an dem Schritt der händischen Korrektur wohl kaum ein Weg vorbei, was aber angesichts der Menge an Zeitungsausgaben für eine Einzelperson nur mit größtem Aufwand zu bewerkstelligen ist, denn es handelt sich um mehr als 1400 Ausgaben der Jahre 1914 bis 1918. Es stellt sich darüber hinaus auch die Frage, wie man bei der Aufbereitung des Korpus

57 Vgl. Blei, D., Ng, A., Jordan, M. (2003): »Latent Dirichlet Allocation«. In: *Journal of Machine Learning Research*. 3, 993–1022.

58 Vgl. Weitin: *Digitale Literaturgeschichte*, S. 120ff.

59 Vgl. Schumacher, M. (2018, § 1): »Named Entity Recognition (NER)«. In: *forTEXT. Literatur digital erforschen*. URL: <https://fortext.net/routinen/methoden/named-entity-recognition-ner> [Zugriff: 04. November 2023].

60 Weitin/Hergert: *Falkentopics*, S. 35.

61 Vgl. die Anmerkung bei Weitin: *Digitale Literaturgeschichte*, S. 38.

mit dem charakteristischen Aufbau einer Zeitung verfahren soll: Sollen beispielsweise Bildunterschriften mit aufgenommen werden? Oder wie am besten mit dem Anzeigenteil verfahren, der für die Untersuchung interessant sein, aber kaum sinnvoll in eine Textdatei transferiert werden kann?

Korpus-Kennntnis ist unverzichtbar

Gerade hier bestätigt sich, wie wichtig eine tiefe Kenntnis des Korpus bei quantitativen Analyseverfahren ist. Das *too big to read*-Argument trifft bei einem mittelgroßen Korpus nur bedingt zu und es erweist sich als äußerst vorteilhaft, dass Literaturwissenschaftler:innen ihre zu untersuchenden Korpora ohnehin gut kennen und darüber hinaus auch über vertiefte Kenntnisse des historischen Kontextes, der Intertextualität etc. verfügen. Ohnehin wird das Korpus vor dem Hintergrund einer bestimmten Forschungsfrage zusammengestellt und ein konkretes Erkenntnisinteresse geht der Entscheidung für den Einsatz einer quantitativen Methode voraus. Beim Topic Modeling setzt man natürlich auf Ergebnisse, die interpretierbar sind und die eigenen Erwartungen bestenfalls bestätigen. Die Topics sollen auch etwas über den Inhalt des Korpus aussagen und die verborgenen, thematischen Strukturen sichtbar machen, die man durch die Einzeltextlektüre bisher nicht erreichen konnte. Es ist grundsätzlich möglich, »ein *distant reading*-Ergebnis aus dem Topic Modeling durch gezieltes *close reading* zu evaluieren« und es ist auch legitim, »eine einzeltextbezogene These [...] mit quantitativer Evidenz zu erhärten«. ⁶² Dennoch gilt es etwaige Scheinevidenzen zu vermeiden. Auch wenn Abweichungen vom Makrotrend auf besonderes Interesse stoßen, warnt Weitin vor »hermeneutische[n] Schnellschüsse[n]« bei der Interpretation. ⁶³ So sollte bei etwaigen Abweichungen stets hinterfragt werden, »ob der hermeneutische Einzelbefund und das Ergebnis der statistischen Korpusanalyse überhaupt etwas miteinander zu tun haben.« ⁶⁴ Der *confirmation bias* erweist sich also als nicht zu unterschätzendes Problem.

Auch der Analyse des Kafka-Zeitungs-Korpus gehen ein paar Erwartungen voraus: Es handelt sich u.a. um eine erhoffte quantitative Evidenz der Thesen, die exemplarisch im *close reading*-Verfahren ausgemacht wurden. Weist das Modell also auch zuverlässig jene Themen in den Zeitungsausgaben in Bezug auf Kafkas literarische Texte als auffällig aus, die zuvor als solche identifiziert wurden? Wird beispielsweise der Diskurs rund um die Dressur von Menschenaffen, Zirkus und Varieté, die für Kafkas Erzählung vom Affen Rotpeter im *Bericht für eine Akademie* prägend waren, durch Topics aufgezeigt? Oder beruht diese Erwartung schon auf falschen Annahmen bezüglich der Methode? Und wie steht es eigentlich um die in den Zeitungen gedruckten Texte, die aus Kafkas Feder selbst stammen? Also jene literarischen Texte wie z.B. *Ein Traum*, der in der Ausgabe vom 6. Januar 1917 zu finden ist. Der Erwartung zu viel wäre es wahrscheinlich, würde man dann

62 Weitin/Herget: Falkentopics, S. 32.

63 Vgl. Weitin: Digitale Literaturgeschichte, S. 133 sowie S. 54. Vgl. auch Weitin/Herget: Falkentopics, S. 32.

64 Weitin/Herget: Falkentopics, S. 32.

noch darauf bauen, dass sogar die Beiträge im *Prager Tagblatt* als statistisch auffällig angezeigt werden, die zwar aus Kafkas Hand stammen, aber im Rahmen seiner Amtstätigkeit entstanden sind, wie der Aufruf »Helfet den Kriegsinvaliden!« in der Ausgabe vom 15. Dezember 1916.

Code-Kompetenz: Der Blick in den Maschinenraum

Neben einem sorgfältigen *Preprocessing* bei der Zusammenstellung des Korpus und einer profunden Kenntnis über das Korpus selbst, zeichnet sich außerdem ab, dass einschlägiges Wissen über die jeweilige digitale Methode sowie den ihr zugrunde liegenden Algorithmus mindestens hilfreich, aber auch nahezu unerlässlich ist. Man muss nicht zwangsläufig eine Programmiersprache beherrschen, um Topic Modeling durchführen zu können. Es werden auch vermehrt Analyse-Packages angeboten, die durch eine Benutzeroberfläche für Einsteiger:innen bedienbar sind. Wer aber Einblick in den »Maschinenraum« erhalten will, um den Prozess hinter den Ergebnissen zu verstehen, kommt um eine eigene Code-Kompetenz nicht herum. Wer quantitative Methoden in die eigene Forschung mit einbeziehen will, hat im Studium wahrscheinlich kaum entsprechendes Handwerkszeug gelernt, es sei denn, man ist durch eine Teildisziplin oder persönliches Erkenntnisinteresse bei entsprechenden Seminaren z. B. der Computerlinguistik oder Korpuslinguistik, Sozialwissenschaft etc. gelandet. Darüber hinaus muss man sich dann gezielt weiterbilden. Weitin konstatiert, dass – im Gegensatz zu den Digital Humanities – die »geisteswissenschaftlichen Fächer Buchkulturen geblieben«⁶⁵ seien. Und so käme in den klassischen Medien »Digitale Geisteswissenschaften vor allem als Debatte über Digitale Geisteswissenschaften vor«.⁶⁶ Zur Rolle der Literaturwissenschaft und ihrer Methoden im Zeitalter der Digitalisierung findet schon seit geraumer Zeit ein Fachdiskurs statt, der sich durch die Pandemie und den Einfluss von Künstlicher Intelligenz eine neue Dynamik bekommen hat. Interdisziplinäre Ansätze zeigen hier großes Potenzial und der vorliegende Band kann einen Beitrag zu diesem Fachdiskurs leisten.

Wer bei Kafka zwischen den Zeilen lesen will, kann dem mit traditionellen literaturwissenschaftlichen Methoden und einer sorgfältigen Einzeltextanalyse näherkommen. Eine quantitative Bedeutungsanalyse für eine größere Textmenge, die verborgene semantische Strukturen sichtbar machen kann, wie es die Methode Topic Modeling verspricht, kann einen wesentlichen Beitrag zur Forschung über Franz Kafka liefern. Dies gelingt aber nur, wenn man nicht mit der einen Methode die andere ersetzen will, sondern sie sinnvoll ergänzt. Franz Kafkas als Zeitungsleser während des Ersten Weltkriegs und dessen Einbettung und Operationalisierung des Zeitungsmediums in den Schreibprozess erweist sich als herausfordernde Fragestellung, die unter Berücksichtigung bestimmter Voraussetzungen vielversprechend mit Topic Modeling als quantitativer Methode analysiert werden kann.

65 Weitin: Digitale Literaturgeschichte, S. 3.

66 Ebd.