

Schulunterricht: Textanalyse mit CATMA unterrichten

Marie Flüh  ¹

1. Universität Hamburg

forTEXT

Thema:	Manuelle Annotation	DOI:	10.48694/fortext.3755
Jahrgang:	1	Ausgabe:	4
Erscheinungsdatum:	2024-08-07	Erstveröffentlichung:	2020-02-10 auf fortext.net
Lizenz:			open  access

Allgemeiner Hinweis: Rot dargestellte *Begriffe* werden im Glossar am Ende des Beitrags erläutert. Alle externen Links sind auch am Ende des Beitrags aufgeführt.

Eckdaten der Unterrichtseinheit

- Thema der Sitzung: Digitale Analyse der Schuldproblematik und ihrer Darstellung in Bernhard Schlinks Roman *Der Vorleser* (1995)
- Klassenstufe: Sekundarstufe II
- Phasen: Brainstorming zum Thema Schuld, Einführung in die Arbeit mit dem Textanalysetool CATMA, Wortsuche via *Analyze*-Modul, Visualisierung der Suchergebnisse, textnahe Analyse des semantischen Umfelds, Präsentation und Diskussion der Arbeitsergebnisse
- Sozialform(en): Lehrvortrag, Teamarbeit und Vortrag
- Medien/Materialien: *Der Vorleser* in digitaler Form, stabile Internetverbindung, die Schüler*innen arbeiten in Zweierteams an einem PC (2:1), für die Einstiegsphase benötigen Sie einen mit dem Beamer verbundenen Laptop
- Dauer der Unterrichtseinheit: 90 Minuten
- Schwierigkeitsgrad des Tools: einfach/mittel

Fachliche Lernziele:

- **Aufbau literarischer Kompetenz:** Die Schüler*innen lernen, in dem Beispieltexthaltene Dimensionen von Schuld zu verhandeln. Sie erlernen einen analytischen Umgang mit anspruchsvollen Texten und arbeiten an den Fähigkeiten, Texte selbstständig zu erschließen, eigenständig Interpretationshypothesen zu entwickeln und diese zu diskutieren.
- **Aufbau von Methodenkompetenz:** Die Schüler*innen lernen CATMA als Werkzeug zur Textanalyse und -erschließung kennen, das sich über den Literaturunterricht hinaus auf weitere Fachgebiete übertragen lässt.

KMK-Lernziele:

- Suchen, Verarbeiten, Aufbewahren: Arbeits- und Suchinteressen festlegen, Suchstrategien in CATMA nutzen, relevante Suchergebnisse identifizieren und zusammenführen, Textdaten analysieren, interpretieren und kritisch bewerten.
- Entwickeln und Produzieren: CATMA als technisches Bearbeitungswerkzeug kennenlernen und computergestützt Visualisierungen produzieren, die die Ergebnisse der Textanalyse beinhalten.
- Problemlösen und Handeln: Bedarfsgerechter Einsatz des Textanalysetools CATMA (*Analyze*-Modul als Möglichkeit der quantitativen Textanalyse).
- Analysieren und Reflektieren: CATMA als Werkzeug der quantitativen und qualitativen Textanalyse (vgl. *Distant Reading*) kennenlernen und sich kritisch mit den Grenzen von Ein-Wort-Analysen auseinandersetzen.

Bausteine

- Verlaufsraaster des Lehrmoduls: Aus welchen Phasen setzt sich die Unterrichtseinheit zusammen? Dem Verlaufsplan entnehmen Sie Inhalte und Schwerpunkte.
- Anwendungsbeispiel: Sie möchten mit Ihrer Lerngruppe die Schuldproblematik in Schlinks Roman untersuchen und textinterne Darstellungsweisen mit den individuellen Assoziationen der Lerngruppe zum Thema Schuld diskutieren und vergleichen? Leiten Sie Ihre Klasse dazu an, mit dem Textanalysetool CATMA bestimmte bedeutungstragende Begriffe, deren Häufigkeit und Semantik zu annotieren und genauer zu untersuchen.
- Verlauf der Unterrichtseinheit(en): Wie sieht die konkrete Ausgestaltung der Phasen aus und welche Arbeitsschritte werden vorgenommen? Erfahren Sie, wie die Unterrichtseinheit strukturiert ist und welche Beispielaufgaben Sie Ihren Schüler*innen stellen können
- Lösungen zu den Beispielaufgaben: Hat die Lerngruppe die Beispielaufgaben richtig gelöst? Hier finden Sie Antworten.

Verlaufsraaster des Lehrmoduls

Phase	Impulse des/der Lehrenden	Erwartete Aktivität der Lernenden	Sozialform	Medien / Materialien
Vorab und Einstieg (ca. 5 Min.)	Vorab: Einarbeitung in CATMA und Erstellen des Projekts; Einstieg: Schüler*innen in das gemeinsame Projekt einladen	CATMA-Tutorials durchlaufen, Projekt betreten	Einzelarbeit	Laptop, CATMA, Internetverbindung, digitale Version von <i>Der Vorleser</i>
Problematisierung (ca. 10 Min.)	Leiten des Brainstormings zum Thema Schuld	Sammeln von Assoziationen an der Tafel	Diskussion im Plenum	PCs (2:1), CATMA-Zugang, Tafel, Smartboard, Internetverbindung
Erarbeitung (ca. 30 Min.)	Aufgaben vorstellen, Fragen beantworten, ggf. durch die einzelnen Arbeitsschritte führen	Lösen der Aufgaben	Teamarbeit	PCs (2:1), CATMA-Zugang, Internetverbindung
Sicherung (ca. 25. Min.)	Durch die Fragen leiten	Vorstellung der Analyseergebnisse	Diskussion im Plenum	keine
Reflexion & Transfer (ca. 20 Min.)	Diskussionsleitung			
Vergleich von Mindmap und Analyseergebnissen	Kritische Reflexion der statistischen Auswertung literarischer Texte	Diskussion im Plenum	keine	

Das Verlaufsraaster finden Sie als PDF-Datei zum Download auf [Zenodo](#) (forTEXT 2020a).

1. Anwendungsbeispiel

Die hier skizzierte Unterrichtsstunde lässt sich an beliebiger Stelle einer Unterrichtseinheit eingliedern, in der Analyse und Interpretation von Bernhard Schlinks Roman *Der Vorleser* im Vordergrund stehen. Sie setzen das Textanalysetool CATMA (kurz für Computer Assisted Text Markup and Analysis, siehe Schumacher (2024)) ein, um digitale Verfahren der Textanalyse auszuprobieren und die inhaltliche Erschließung des Romans mit dem Aufbau digitaler Methodenkompetenz zu verbinden. *Der Vorleser* wird hier als exemplarisches Beispiel eines im schulischen Curriculum verankerten Werks der deutschsprachigen Gegenwartsliteratur herangezogen, die sich mit Mitteln und Folgen des Holocausts auseinandersetzt. Einer inhaltlichen Klassifikation folgend, wird der Roman auch als Werk der Nachkriegsliteratur bezeichnet (Feuchert und Hofmann 2017). I.d.R. fasst ein enger zeitlicher Rahmen unter dem Begriff der Nachkriegsliteratur all diejenigen Werke, die in den unmittelbaren Nachkriegsjahren bis zur Gründung der beiden deutschen Staaten entstand zuzüglich drei bis vier weiterer Jahre, die oft als „Ende der Nachkriegszeit“ oder „Literatur nach 1989“ benannt werden. Die Debatte darüber, welche Werke zu welchem Genre und welcher Epoche zählen, soll an dieser Stelle nicht aufgegriffen werden. Wichtig ist, sich für eine Klassifikation zu entscheiden und die Entscheidung auch für die Schüler*innen transparent zu machen, damit ein Epochen- und Überblickswissen ausgebaut werden kann. Für detaillierte Ausführungen zur Begriffsbestimmung (Gansel und Herrmann 2013). Dieser Text ist via [Internet Archive](#) uneingeschränkt verfügbar. Das Internet Archive ist eine US-amerikanische digitale Datenbank. Seit 2007 hat das auf die Langzeitarchivierung digitaler Daten abzielende Archiv den Status einer offiziellen Bibliothek inne. Im Textarchiv werden gemeinfrei gewordene Werke zum Herunterladen zur Verfügung gestellt. Das hier skizzierte Verfahren lässt sich darüber hinaus auch auf andere Werke übertragen.

2. Verlauf der Unterrichtseinheiten

2.1 Vorarbeiten und Einstieg

Die Unterrichtsstunde richtet sich an eine Lerngruppe mit grundlegenden Englischkenntnissen, da das Tool über ein englischsprachiges Interface verfügt. Die Stunde verfolgt – neben den oben genannten Lernzielen – drei zentrale Zielsetzungen.

(1) Auf einer übergeordneten Ebene und die gesamte Unterrichtseinheit betreffend sollen die Schüler*innen durch die Auseinandersetzung mit dem Roman für einen kritischen und reflektierten Umgang mit den Mitteln und Folgen des Holocausts sensibilisiert werden und Formen der literarischen Auseinandersetzung mit dem Nationalsozialismus kennenlernen.

(2) Zentrales Thema des Romans ist – neben der Inbezugsetzung von Erotik, dem Analphabetismus, der „Coming-of-Age“-Thematik, der Verarbeitung nationalsozialistischer Vergangenheit und dem Affekt der Scham (Hehl 2015, 471) – Schuld in unterschiedlichen Zusammenhängen. Die Auseinandersetzung mit Schuld umfasst mehrere Dimensionen: die historisch-moralische Schuld als eine von einer an die nächste Generation weitergegebene Schuld (Beziehung zwischen Kriegs- und Nachkriegsgeneration im Rahmen der 68er Bewegung, Beziehung zu Hanna, Beziehung zum Vater und generell zu Eltern, den „[...] Tätern, den Zu- und Wegsehern, Tolerierern und Akzeptierern“ (*Der Vorleser*, 126)) und deren Relativierung. Die Frage nach Täter*innen und Opfern und was sie dazu macht wird in den drei Etappen des Romans – vorrangig aber im zweiten und dritten Teil – verhandelt. Die Auseinandersetzung mit der juristisch-strafrechtlichen Schuld (und Hannas Übernahme der Hauptschuld) sowie die Reflexion über Gerechtigkeit dominieren v. a. den zweiten Teil des Romans, der die Wiederbegegnung beider Figuren im Zuge der Frankfurter Auschwitzprozesse, die Konfrontation mit Hannas Analphabetismus und Hannas Verurteilung zu 20 Jahren Haft thematisiert. Die Unterrichtsstunde zielt darauf ab, die unterschiedlichen Darstellungsweisen von Schuld im Roman zu erforschen und mit den Weltanschauungen und Vorstellungen der Schüler*innen zu vergleichen.

(3) Auf methodischer Ebene lernen die Schüler*innen eine Form der digitalen Textuntersuchung als systematische und übertragbare Form der Textanalyse kennen. Ein Mehrwert entsteht durch die digitale Textanalyse mit CATMA, indem der Text gezielt hinsichtlich des Stundenthemas ausgewertet wird (vgl. [Data Mining](#)) und Arbeitsergebnisse visualisiert und als Diskussionsgrundlage herangezogen werden. Auf Grundlage des Romantextes erstellen die Schüler*innen eigene Visualisierungen, die ihre Interpretationen widerspiegeln. Der Text wird zunächst aus der Ferne (um einen allgemeinen Überblick zu gewinnen) und dann genauer in seinen speziellen Beziehungen (vgl. [Scalable Reading](#)) betrachtet. Die Schüler*innen nehmen die Rolle von Forscher*innen ein, die zunächst gemeinsam Untersuchungskriterien festlegen (vgl. [Tagset](#)), den Untersuchungsgegenstand mithilfe des Tools in seine einzelnen Bestandteile zerlegen – in diesem Fall die einzelnen Wörter –, diese genau betrachten und daraus Analyseergebnisse und Interpretationshypothesen ableiten.

Zur eigenen methodischen Vorbereitung lesen Sie den Toolbeitrag zu CATMA (Schumacher 2024). Darüber hinaus sollten Sie die beiden Lerneinheiten „Manuelle Annotation mit CATMA“ (Horstmann 2024a) und „Analyse und Visualisierung mit CATMA“ (Horstmann 2024b) absolvieren. Um einen Überblick über den Aufbau und die Funktionen von CATMA zu gewinnen, schauen Sie das Tutorial-Videoreihe „Manuelle Annotation und Literaturanalyse“ an [forTEXT (2020b); forTEXTTutorialCATMAAnnotieren2020; forTEXT (2019c); forTEXT (2019b);

forTEXT (2019a)]. Die geführte Tour durch das Interface des Tools wird Ihnen dabei helfen, sich zurechtzufinden. Zwei weitere Tutorials zeigen Ihnen, wie Sie mit CATMA manuelle Annotationen anbringen können und wie Sie eigene Projekte anlegen können. Nachdem Sie beide Lerneinheiten durchgearbeitet und die Tutorial-Videos angesehen haben, verfügen Sie über einen eigenen CATMA-Account und kennen die grundlegenden Funktionen des Tools. Vorbereitend setzen Sie die folgenden drei Schritte um:

1. Laden Sie den Text *Der Vorleser* [hier](#) als PDF-Dokument herunter.
2. Loggen Sie sich bei CATMA ein und erstellen über den „CREATE NEW PROJECT“-Button das neue Projekt „Bernhard Schlink: Der Vorleser (1995)“. Nachdem Sie das Projekt per Doppelklick auf die graue Schaltfläche betreten haben, fügen Sie per „Add Document“ (zu finden hinter dem „+“-Symbol) die zuvor heruntergeladene Datei hinzu.
3. Nun erstellen Sie ein **Tagset** zum Thema Schuld, indem Sie in dem Tagset-Panel – erneut via Klick auf „+“ – auf „Add Tagset“ klicken. Nennen Sie das Tagset „Schuldproblematik“ und fügen anschließend mindestens die Tags „moralische Schuld“ und „juristische Schuld“ hinzu. Um Raum für weitere Formen von Schuld zu lassen, die sich nicht ad hoc einer der bereits erstellten Kategorien zuordnen lassen, legen Sie darüber hinaus den Tag „Ungewiss“ an.
Das für die Unterrichtsstunde angelegte Projekt ist nun fertig. Alle weiteren Arbeitsschritte werden im Rahmen des Unterrichts zusammen mit Ihren Schüler*innen durchgeführt. Zur inhaltlichen Vorbereitung sollten Sie natürlich in erster Linie den Roman gelesen haben. Textanalysen, Hintergrundinformationen und Interpretationsansätze finden Sie z. B. bei Corngold (1999), Gansel und Herrmann (2013), Hehl (2015), Knobloch (2001) oder Tommek (2015). Darüber hinaus sollten Sie für sich die Frage beantworten, welche Begriffe Sie mit „Schuld“ assoziieren. Die Schüler*innen bereiten sich videobasiert auf die Unterrichtsstunde vor, indem Sie einen kommentierten Rundgang durch CATMAs (forTEXT 2019a) einzelne Bestandteile und Funktionen anschauen und sich über einen Klick auf den „Sign In“-Button einen eigenen CATMA-Account anlegen.

2.2 Einstieg und Problematisierung

Um einen nahtlosen Übergang in die Textanalyse mit CATMA Vorschub zu leisten, bereiten Sie noch vor dem eigentlichen Stundeneinstieg – dem gemeinsamen Brainstorming – die Erarbeitungsphase mit CATMA vor. Auf diese Weise kann das System die Zeit während der Diskussionsphase nutzen, um die gesamte Klasse dem Projekt zuzuordnen. Hierfür leiten Sie Ihre Schüler*innen dazu an, sich mit ihrem vorab angelegten Account bei CATMA einzuloggen und auf den „JOIN PROJECT“-Button zu klicken. Zeitgleich öffnen Sie das vorab angelegte Bernhard Schlink-Projekt und klicken in der Team-Kachel auf das kleine Symbol mit den drei Punkten. Hier wählen Sie den „Invite someone to the Project“-Button, setzen ein Häkchen vor „Create one collection per Document and joined user“ sowie ein Häkchen vor „Name“, legen die Rolle „Student“ fest und bestätigen Ihre Auswahl per Mausklick auf „Invite“. Den nun generierten „invitation code“ geben Ihre Schüler*innen ein, um Zugriff auf das vorbereitete Projekt zu erhalten. Die technischen Vorbereitungen sind nun abgeschlossen. Der gemeinsame thematische Einstieg in die Unterrichtsstunde wird in Form eines Brainstormings zum Thema Schuld umgesetzt. Die Schüler*innen tragen ihre Ideen in dieser Phase der Stunde zunächst unkommentiert an der Tafel oder am Smartboard zusammen. Sichern Sie die Ergebnisse, indem Sie die Mindmap speichern oder abfotografieren.

Frage an das Plenum:

Was bedeutet Schuld für Sie und welche Assoziationen haben Sie zu diesem Thema?

Fordern Sie abschließend einige Stellungnahmen ein und fassen das Brainstorming kurz zusammen. Es sollte deutlich werden, dass der Begriff Schuld in unterschiedlichen Zusammenhängen verwendet wird und Ihre Schüler*innen, aller Wahrscheinlichkeit nach, ganz individuelle Erfahrungen mit Schuld gemacht haben. Mögliche Formen von Schuld sind z. B. Schuld im Zusammenhang mit Vorwurf und Anklage (v. a. im Zusammenhang der Debatte über gesellschaftliche und politische Missstände und deren Urheber*innen), Schuld als Tat und Verstoß gegen moralische oder gesetzlich geregelte Grundsätze, Folgen von Schuld z. B. in Form eines schlechten Gewissens und die mit Schuld in Verbindung stehenden negativen Gefühle sowie Schuld im finanziellen Sinn, als Form der abbezahlbaren Schuld, die eine konkrete Form der Wiedergutmachung beinhaltet. Leiten Sie nun in die nächste Phase des Unterrichts über: Ihre Schüler*innen sollen untersuchen, welche Bedeutung Schuld in *Der Vorleser* zukommt, ob ihre Konzepte und Ideen sich im Roman wiederfinden oder ob die Mindmap durch im Roman dargestellte Formen erweitert werden kann.

2.3 Erarbeitung

Nun beginnt die Erarbeitungsphase mit CATMA. Leiten Sie Ihre Schüler*innen dazu an, das Bernhard Schlink-Projekt zu betreten und per einfachem Mausklick im Resources-Panel auf „Bernhard Schlink – Der Vorleser“ den Text auszuwählen. Mit Klick auf die *Analyze*-Schaltfläche links gelangen sie in das entsprechende CATMA-Modul. Von dort aus startet die Textanalyse in Teamarbeit. Auf Grundlage der zu Hause erfolgten Vorbereitung im Umgang mit CATMA sollen die folgenden fünf Aufgaben gelöst werden:

Aufgabe 1: Wie häufig kommt der Begriff „Schuld“ im gesamten Roman vor? Nutzen Sie den Query-Builder, um den Text auszuwerten.

Aufgabe 2: Erstellen Sie einen Verteilungsgraphen, in dem die Verteilung des Begriffs „Schuld“ im gesamten Roman visualisiert wird. In welchem Teil des Romans kommt er besonders häufig vor?

Aufgabe 3: Zerlegen Sie den Text in seine Einzelteile, indem Sie eine Worthäufigkeitsliste generieren. Klicken Sie auf das „Wordcloud“-Symbol und erstellen eine Wortwolke, in der diejenigen Wörter enthalten sind, die Ihrer Meinung nach mit Schuld in Verbindung stehen (s. Mindmap). Exportieren Sie die Wordcloud.

Während Ihre Schüler*innen diese Aufgabe lösen, halten Sie sich im Hintergrund und stehen beratend zur Seite, sobald Fragen auftreten. Sollte es technische Probleme bei der Umsetzung der Aufgaben geben, können Sie einen exemplarischen Durchlauf – vom Betreten des *Analyze*-Moduls, von hier aus zur Wortliste und weiter bis zur **Wordcloud** – für die gesamte Lerngruppe vorstellen, indem Sie Ihren Laptop mit dem Beamer verbinden und die besagten Schritte exemplarisch durchgehen.

Aufgabe 4: Lassen Sie sich die gefundenen Wortvorkommnisse in der KWIC-Ansicht anzeigen. Um welche Formen von Schuld handelt es sich bei den gefundenen Suchergebnissen? Annotieren Sie die Wortvorkommnisse mithilfe des Tagsets „Schuldproblematik“ und ergänzen es – falls nötig – um weitere Tags.

An dieser Stelle kann es gut sein, dass Ihre Schüler*innen weitere Formen von Schuld im Text erkennen, die nicht in dem jetzigen Tagset enthalten sind. Führen Sie exemplarisch vor, wie sich während des Annotierens (vgl. **Annotation**) ein neuer Tag (vgl. **Tagset**) hinzufügen lässt (Klick auf den „+“-Button im rechts vom Text liegenden Tagsetbereich und Klick auf die „Add Tag“-Schaltfläche) und motivieren Ihre Schüler*innen dazu, eigene Tags zu entwerfen.

2.4 Sicherung

Die Sicherung besteht in diesem Fall in der mündlichen Diskussion der exportierten Arbeitsergebnisse und in der gemeinsamen Besprechung der Annotationen sowie neu erstellter Tags. Während die erste Aufgabe schnell beantwortet werden kann, bedarf die Diskussion der übrigen Aufgaben mehr Zeit. Nicht alle Teams werden in der Unterrichtsstunde ihre Arbeitsergebnisse vorstellen können. Drei bis vier Gruppen sollten aber mindestens präsentieren. Fordern Sie darüber hinaus auch eine methodische Reflexion ein und fragen nach Grenzen der digitalen Analysemethoden.

2.5 Transfer & Reflexion

Im letzten Teil der Stunde sollen die Schüler*innen abschließend diskutieren, ob ihre eigenen Assoziationen zum Thema der Stunde und die aus dem Text herausgearbeiteten Zusammenhänge übereinstimmen oder nicht. Übertragen Sie hierfür zunächst das von Ihnen angelegte CATMA-Projekt auf den Beamer und gehen Sie einige der Schuld-Annotationen (vgl. **Annotation**) durch. Fordern Sie Ihre Schüler*innen dazu auf, die zuvor gesammelten Gedanken und die in der Analyse herausgearbeiteten Aspekte miteinander in Beziehung zu setzen: Welche Annotationen stimmen überein und wo werden Differenzen sichtbar? Darüber hinaus können Mindmap und Wordclouds (vgl. **Wordcloud**) (s. Abb. 1) gegenübergestellt werden.



Wordcloud mit Schuldthematik: Der Vorleser

3. Lösungen zu den Beispielaufgaben

Die Lösungsvorschläge erhalten Sie auf [Zenodo](#) (forTEXT 2020a).

Externe und weiterführende Links

- CATMA: <https://web.archive.org/save/https://catma.de> (Letzter Zugriff: 03.07.2024)
- Download *Der Vorleser*: <https://web.archive.org/save/https://archive.org/details/DerVorleser> (Letzter Zugriff: 03.07.2024)
- Videos der Videoreihe „Manuelle Annotation und Literaturanalyse“ auf Zenodo (Letzter Zugriff: 03.07.2024):
 - CATMA 6 zur manuellen Annotation nutzen: <https://zenodo.org/records/10353556>
 - Projektmanagement in CATMA 6: <https://zenodo.org/records/10353713>
 - Tagsets in CATMA 6 anlegen: <https://zenodo.org/records/10377968>
 - In CATMA 6 annotieren: <https://zenodo.org/records/10353910>
 - Analysieren und visualisieren mit CATMA: <https://zenodo.org/records/10276637>
- Videoreihe „Manuelle Annotation und Literaturanalyse“ auf Youtube: <https://www.youtube.com/watch?v=d5yxEOQu2Dw&list=PLu-M0KuYw64q0Eb-7cZwdv485ywWG8ShP> (Letzter Zugriff: 03.07.2024)
- Verlaufsraster und Lösungsvorschläge: <https://web.archive.org/save/https://zenodo.org/records/12529635> (Letzter Zugriff: 03.07.2024)

Bibliographie

- Corngold, Stanley. 1999. 'Fürsorge beim Vorlesen': Bernhard Schlink's Novel 'Der Vorleser'. In: *Signatures der Gegenwartsliteratur. Festschrift für Walter Hinderer*, hg. von Dieter Borchmeyer, 247–257. Würzburg: Königshausen & Neumann.
- Feuchert, Sascha und Lars Hofmann. 2017. *Bernhard Schlink: Der Vorleser*. [Nachdruck] 2021. Reclam Lektüreschlüssel XL Nr. 15454. Ditzingen: Reclam.

- forTEXT. 2019a. Tutorial: CATMA 6 zur manuellen Annotation nutzen. Manuelle Annotation und Literaturanalyse. 28. Oktober. doi: 10.5281/zenodo.10353556, <https://zenodo.org/records/10353556>.
- . 2019b. Tutorial: Projektmanagement in CATMA 6. Manuelle Annotation und Literaturanalyse. 25. November. doi: 10.5281/zenodo.10353713, <https://zenodo.org/records/10353713>.
- . 2019c. Tutorial: Tagsets in CATMA 6 anlegen. Manuelle Annotation und Literaturanalyse. 23. Dezember. doi: 10.5281/zenodo.10377968, <https://zenodo.org/records/10377968>.
- . 2020a. Textanalyse mit CATMA unterrichten. 10. Februar. doi: 10.5281/zenodo.10519663, <https://zenodo.org/records/12529635>.
- . 2020b. Tutorial: Analysieren und visualisieren mit CATMA. Manuelle Annotation und Literaturanalyse. 17. Februar. doi: 10.5281/zenodo.10276637, <https://doi.org/10.5281/zenodo.10276637>.
- Gansel, Carsten und Elisabeth Herrmann. 2013. „Gegenwart‘ bedeutet die Zeitspanne einer Generation‘ – Anmerkungen zum Versuch, Gegenwartsliteratur zu bestimmen. In: *Entwicklungen in der deutschsprachigen Gegenwartsliteratur nach 1989*, hg. von Carsten Gansel und Herrmann Korte, 7–22. Göttingen: Vandenhoeck & Ruprecht. <https://doi.org/10.14220/9783847098881.7>.
- Hehl, Michael Peter. 2015. Schlink, Bernhard: Der Vorleser. Roman (Zürich: Diogenes). In: *Wendejahr 1995. Transformationen der deutschsprachigen Literatur*, hg. von Heribert Tommek, Matteo Galli, und Achim Geisenhanslüke, 470–475. Berlin, München, Boston: De Gruyter. <https://doi.org/10.1515/9783110419023-040> (zugegriffen: 7. November 2023).
- Horstmann, Jan. 2024b. Lerneinheit: Analyse und Visualisierung mit CATMA. Hg. von Evelyn Gius. *forTEXT* 1, Nr. 4. Manuelle Annotation (7. August). doi: 10.48694/fortext.3752, <https://fortext.net/routinen/lerneinheiten/analyse-und-visualisierung-mit-catma>.
- . 2024a. Lerneinheit: Manuelle Annotation mit CATMA. Hg. von Evelyn Gius. *forTEXT* 1, Nr. 4. Manuelle Annotation (7. August). doi: 10.48694/fortext.3750, <https://fortext.net/routinen/lerneinheiten/manuelle-annotation-mit-catma>.
- Knobloch, Hans-Jörg. 2001. Eine ungewöhnliche Variante in der Täter-Opfer-Literatur: Bernhard Schlinks Der Vorleser. In: *Schreiben nach der Wende. Ein Jahrzehnt deutscher Literatur 1989–1999*, hg. von Fischer Gerhard und David Roberts, 89–98. Tübingen: Stauffenburg-Verlag.
- Schumacher, Mareike. 2024. Toolbeitrag: CATMA. Hg. von Evelyn Gius. *forTEXT* 1, Nr. 4. Manuelle Annotation (7. August). doi: 10.48694/fortext.3761, <https://fortext.net/tools/tools/catma>.
- Tommek, Heribert. 2015. Der lange Weg in die Gegenwartsliteratur. Studien zur Geschichte des literarischen Feldes in Deutschland von 1960 bis 2000. In: *Studien und Texte zur Sozialgeschichte der Literatur*, hg. von Norbert Bachleitner, Christan Begemann, Walter Erhart, und Gangolf Hübinger, 140. Berlin, Boston: De Gruyter.

Glossar

- Annotation** Annotation beschreibt die manuelle oder automatische Hinzufügung von Zusatzinformationen zu einem Text. Die manuelle Annotation wird händisch durchgeführt, während die (teil-)automatisierte Annotation durch **Machine-Learning-Verfahren** durchgeführt wird. Ein klassisches Beispiel ist das automatisierte **PoS-Tagging** (Part-of-Speech-Tagging), welches oftmals als Grundlage (**Preprocessing**) für weitere Analysen wie Named Entity Recognition (NER) nötig ist. Annotationen können zudem deskriptiv oder analytisch sein.
- Browser** Mit Browser ist in der Regel ein Webbrowser gemeint, also ein Computerprogramm, mit dem das Anschauen, Navigieren auf, und Interagieren mit Webseiten möglich wird. Am häufigsten genutzt werden dafür Chrome, Firefox, Safari oder der Internet Explorer.
- Close Reading** Close Reading bezeichnet die sorgfältige Lektüre und Interpretation eines einzelnen oder weniger Texte. Close Reading ist in der digitalen Literaturwissenschaft außerdem mit der manuellen **Annotation** textueller Phänomene verbunden (vgl. auch **Distant Reading** als Gegenbegriff).
- Data Mining** Data Mining gehört zum Fachbereich **Information Retrieval** und bezieht sich auf die systematische Anwendung computergestützter Methoden, die darauf abzielt, in vorhandenen Datenbeständen Muster, Trends oder Zusammenhänge zu erkennen. Textbasierte Formen des Data Minings sind u. a. **Text Mining**, **Web Mining** und **Opinion Mining**.
- Distant Reading** Distant Reading ist ein Ansatz aus den digitalen Literaturwissenschaften, bei dem computergestützte Verfahren auf häufig große Mengen an Textdaten angewandt werden, ohne dass die Texte selber gelesen werden. Meist stehen hier quantitative Analysen im Vordergrund, es lassen sich jedoch auch qualitative **Metadaten** quantitativ vergleichen. Als Gegenbegriff zu **Close Reading** wurde der Begriff insbesondere von Franco Moretti (2000) geprägt.
- HTML** HTML steht für *Hypertext Markup Language* und ist eine textbasierte Auszeichnungssprache zur Strukturierung elektronischer Dokumente. HTML-Dokumente werden von **Webbrowsern** dargestellt und geben die Struktur und Online-Darstellung eines Textes vor. HTML-Dateien können außerdem zusätzliche **Metainformationen** enthalten, die auf einer Webseite selbst nicht ersichtlich sind.

- Information Retrieval** Die Teildisziplin der Informatik, das Information Retrieval, beschäftigt sich mit der computergestützten Suche und Erschließung komplexer Informationen in meist unstrukturierten Datensammlungen.
- Lemmatisieren** Die Lemmatisierung von Textdaten gehört zu den wichtigen **Preprocessing**-Schritten in der Textverarbeitung. Dabei werden alle Wörter (**Token**) eines Textes auf ihre Grundform zurückgeführt. So werden beispielsweise Flexionsformen wie „schneller“ und „schnelle“ dem Lemma „schnell“ zugeordnet.
- Machine Learning** Machine Learning, bzw. maschinelles Lernen im Deutschen, ist ein Teilbereich der künstlichen Intelligenz. Auf Grundlage möglichst vieler (Text-)Daten erkennt und erlernt ein Computer die häufig sehr komplexen Muster und Gesetzmäßigkeiten bestimmter Phänomene. Daraufhin können die aus den Daten gewonnen Erkenntnisse verallgemeinert werden und für neue Problemlösungen oder für die Analyse von bisher unbekannten Daten verwendet werden.
- Markup Language** Markup Language bezeichnet eine maschinenlesbare Auszeichnungssprache, wie z.B. **HTML**, zur Formatierung und Gliederung von Texten und anderen Daten. So werden beispielsweise auch **Annotationen** durch ihre Digitalisierung oder ihre digitale Erstellung zu Markup, indem sie den Inhalt eines Dokumentes strukturieren.
- Metadaten** Metadaten oder Metainformationen sind strukturierte Daten, die andere Daten beschreiben. Dabei kann zwischen administrativen (z. B. Zugriffsrechte, Lizenzierung), deskriptiven (z. B. Textsorte), strukturellen (z. B. Absätze oder Kapitel eines Textes) und technischen (z. B. digitale Auflösung, Material) Metadaten unterschieden werden. Auch **Annotationen** bzw. **Markup** sind Metadaten, da sie Daten/Informationen sind, die den eigentlichen Textdaten hinzugefügt werden und Informationen über die Merkmale der beschriebenen Daten liefern.
- Named Entities** Eine Named Entity (NE) ist eine Entität, oft ein Eigenname, die meist in Form einer Nominalphrase zu identifizieren ist. Named Entities können beispielsweise Personen wie „Nils Holgerson“, Organisationen wie „WHO“ oder Orte wie „New York“ sein. Named Entities können durch das Verfahren der Named Entity Recognition (NER) automatisiert ermittelt werden.
- Opinion Mining** Unter Opinion Mining, oder Sentiment Analysis, versteht man die Analyse von Stimmungen oder Haltungen gegenüber einem Thema, durch die Analyse natürlicher Sprache. Das Opinion Mining gehört zu den Verfahren des **Text Minings**.
- POS** PoS steht für *Part of Speech*, oder „Wortart“ auf Deutsch. Das PoS- **Tagging** beschreibt die (automatische) Erfassung und Kennzeichnung von Wortarten in einem Text und ist ein wichtiger **Preprocessing**-Schritt, beispielsweise für die Analyse von **Named Entities**.
- Preprocessing** Für viele digitale Methoden müssen die zu analysierenden Texte vorab „bereinigt“ oder „vorbereitet“ werden. Für statistische Zwecke werden Texte bspw. häufig in gleich große Segmente unterteilt (*chunking*), Großbuchstaben werden in Kleinbuchstaben verwandelt oder Wörter werden **lemmatisiert**.
- Scalable Reading** Die Kombination aus **Distant Reading**- und **Close Reading**-Methoden, angewandt auf einen Untersuchungsgegenstand, wird als Scalable Reading bezeichnet.
- Tagset** Ein Tagset definiert die Taxonomie, anhand derer **Annotationen** in einem Projekt erstellt werden. Ein Tagset beinhaltet immer mehrere Tags und ggf. auch Subtags. Ähnlich der **Type/Token**-Differenz in der Linguistik sind Tags deskriptive Kategorien, wohingegen Annotationen die einzelnen Vorkommnisse dieser Kategorien im Text sind.
- Text Mining** Das Text Mining ist eine textbasierte Form des **Data Minings**. Prozesse & Methoden, computergestützt und automatisch Informationen bzw. Wissen aus unstrukturierten Textdaten zu extrahieren, werden als Text Mining zusammengefasst.
- Type/Token** Das Begriffspaar „Type/Token“ wird grundsätzlich zur Unterscheidung von einzelnen Vorkommnissen (Token) und Typen (Types) von Wörtern oder Äußerungen in Texten genutzt. Ein Token ist also ein konkretes Exemplar eines bestimmten Typs, während ein Typ eine im Prinzip unbegrenzte Menge von Exemplaren (Token) umfasst.
Es gibt allerdings etwas divergierende Definitionen zur Type-Token-Unterscheidung. Eine präzise Definition ist daher immer erstrebenswert. Der Satz „Ein Bär ist ein Bär.“ beinhaltet beispielsweise fünf Worttoken („Ein“, „Bär“, „ist“, „ein“, „Bär“) und drei Types, nämlich: „ein“, „Bär“, „ist“. Allerdings könnten auch vier Types, „Ein“, „ein“, „Bär“ und „ist“, als solche identifiziert werden, wenn Großbuchstaben beachtet werden.
- Web Mining** Unter Web Mining versteht man die Anwendung von Techniken des **Data Mining** zur Extraktion von Informationen aus dem World Wide Web. Das Web Mining ist ein Teilbereich des Data Minings und zählt zu einem der wichtigsten Anwendungsgebiete für das **Text Mining**.
- Wordcloud** Eine *Wordcloud*, oder auch Schlagwortwolke, ist eine Form der Informationsvisualisierung, beispielsweise von Worthäufigkeiten in einem Text oder einer Textsammlung. Dabei werden unterschiedlich gewichtete Wörter, wie die häufigsten Wörter, i.d.R. größer oder auf andere Weise hervorgehoben dargestellt. Die horizontale/vertikale Ausrichtung und die Farbe der dargestellten Wörter hat meistens allerdings keinen semantischen Mehrwert.