




Lehrmodul: Digitale Annotation mit CATMA lehren

Mareike Schumacher ¹

1. Universität Regensburg

forTEXT

Thema:	Manuelle Annotation	DOI:	10.48694/fortext.3754
Jahrgang:	1	Ausgabe:	4
Erscheinungsdatum:	2024-08-07	Erstveröffentlichung:	2020-03-23 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 des Lehrmoduls

- Thema der Sitzung: Erzähltheoretische Analyse der Novelle *Der Striethast* von Emmy von Dincklage
- Lernziele: Kenntnis der Methode der manuellen, kollaborativen Annotation, sicherer Umgang mit CATMA, kritische Bewertung der Methode
- Phasen: Vorstellen des Tools und Einführung in dessen Funktionen, Gruppenarbeit, Gruppenpräsentation, Abschlussdiskussion
- Sozialform(en): Show-and-Tell der Toolfunktionen, Gruppenarbeit/Einzelarbeit in Gruppen, Vortrag, Diskussion
- Medien/Materialien: Alle Lernenden müssen einen Laptop mit stabiler Internetverbindung haben; Lehrende benötigen einen Laptop, einen Beamer und ggf. Adapter
- Dauer des Lehrmoduls: 2 x 90 Minuten
- Schwierigkeitsgrad des Tools: leicht

Bausteine

- Verlaufsraaster des Lehrmoduls: Aus welchen Phasen setzt sich das Lehrmodul zusammen? Dem Verlaufsplan entnehmen Sie Inhalte und Schwerpunkte.
- Anwendungsbeispiel: Anhand welcher Texte unterrichten Sie manuelle Annotation? Leiten Sie die Studierenden dazu an, narrative Muster in der Novelle *Der Striethast* von Emmy von Dincklage zu erkennen und zu analysieren.
- 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 Studierenden stellen können.
- Lösungen zu den Beispielaufgaben: Hat die Lerngruppe die Beispielaufgaben richtig gelöst? Hier finden Sie Antworten.

Verlafraster des Lehrmoduls

Phase	Impulse des/der Lehrenden	Erwartete Aktivität der Lernenden	Sozialform	Medien / Materialien
Vorab und Einstieg (ca. 20 Min.)	Was ist digitale manuelle Annotation und an welche Tradition knüpft diese Methode an?	Vorab: Methodeneinträge zur manuellen digitalen Annotation (Jacke 2024a) und zur kollaborativen Annotation (Jacke 2024b), Tooleintrag CATMA (Schumacher 2024a), Video-Tutorials der Reihe „Manuelle Annotation und Literaturanalyse“ (forTEXT 2019), Lerneinheit „Manuelle Annotation mit CATMA“ (Horstmann 2024a); Beteiligung an Diskussion; Einladung ins CATMA-Projekt	Diskussion im Plenum; gemeinsamer Einstieg ins CATMA-Projekt	Beamer, Laptop
Problemtisierung (ca. 15 Min.)	Tradition des Annotierens übertragen in die digitale Arbeitsumgebung; hermeneutisches Prinzip, Vergleiche zur linguistischen Annotation	Beteiligung an der Diskussion; Rückbezug auf Methodeneinträge	Diskussion im Plenum	Beamer, Laptop
Erarbeitung (ca. 100 Min.)	Vorstellung der Toolfunktionen; Betreuung der Kleingruppen	Hands-on digitale Annotation und Analyse in Kleingruppen; Vorbereitung einer Präsentation als Hausaufgabe	Lehrvortrag und Gruppenarbeit	Beamer, Laptop, CATMA
Sicherung (ca. 30 Min.)	Moderation der Gruppenpräsentationen	Präsentation der Gruppenarbeitsergebnisse	Präsentation	Beamer, Laptop, CATMA
Reflexion & Transfer (ca. 15 Min.)	Diskussion von Schwierigkeiten; Impulse für Transfer geben	Ergebnisse und Schwierigkeiten aus den Gruppenpräsentationen diskutieren	Diskussion im Plenum	Beamer, Laptops

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

1. Anwendungsbeispiel

In zwei aufeinanderfolgenden Sitzungen werden Sie die Methode der digitalen Annotation (Jacke 2024a) lehren. Die traditionell geisteswissenschaftliche Methode und Kulturtechnik der **Annotation** wird dabei auf eine digitale Forschungsumgebung übertragen. Die Studierenden werden die Methode anhand der Novelle *Der Striethast* von Emmy von Dincklage erproben und das Annotationstool CATMA (Schumacher 2024a) praktisch anwenden. Die Studierenden nutzen ein erzähltheoretisches **Tagset**, um Phänomene des *discours* (Jacke 2024c) zu analysieren. Die Annotationsdaten werden abschließend gemeinsam ausgewertet.

2. Verlauf der Unterrichtseinheiten

2.1 Vorarbeiten

Die Studierenden sollten vorab die Methodeneinträge „Manuelle Annotation“ (Jacke 2024a) und „Kollaborative Annotation“ (Jacke 2024b) sowie den Toolbeitrag zu CATMA (Schumacher 2024a) lesen. Sie sollten außerdem die Lerneinheit „Manuelle Annotation“ (Horstmann 2024a) mithilfe unserer Tutorial-Videos der Reihe „Manuelle Annotation und Literaturanalyse“ absolviert haben. Diese stehen sowohl auf [Youtube](#) (siehe auch [fortext.net](#)) als auch auf [Zenodo](#) für Sie zur Verfügung (forTEXT 2019). Bereiten Sie vorab ein CATMA-Projekt vor. Laden Sie die Text-Version der Novelle *Der Striethast* von Emmy von Dincklage vom deutschen Textarchiv (Horstmann und Kern) herunter und in Ihr CATMA-Projekt hoch. Laden Sie außerdem das narratologische Tagset (forTEXT 2020a; Jacke 2024c) zur *discours*-Ebene herunter und laden Sie es in den Tags-Bereich Ihres CATMA-Projektes hoch. Die hier ausgewählte Novelle ist eher kurz und weist dennoch interessante erzählerische Elemente auf. Wenn Sie trotzdem lieber einen anderen Text nutzen wollen oder mehrere Novellen von Ihren Studierenden annotieren lassen möchten, finden Sie im deutschen Textarchiv (Horstmann und Kern) zahlreiche Alternativen in hoher Qualität.

Für die hier beschriebenen Sitzungen benötigen alle Studierenden einen eigenen Laptop oder Computerplatz, wenn Sie im PC-Labor arbeiten. Bitte achten Sie darauf, dass auf keinen Fall mehr als zwei Studierende sich einen Rechner teilen müssen. CATMA ist weder für Smartphones noch für Tablets optimiert, kann aber mit einem Tablet, das mit einer Tastatur mit Touchpad verbunden ist, genutzt werden. Falls Sie es bevorzugen, für den Einstieg in Ihr Seminar eine Präsentation zu nutzen, haben wir ein paar Beispielfolien vorbereitet, die Sie auf [Zenodo](#) finden und an denen Sie sich orientieren können (forTEXT 2019).

2.2 Einstieg und Problematisierung

Zu Beginn berichten Sie kurz, worum es in den nun folgenden zwei Sitzungen gehen wird:

- In der ersten Sitzung vermitteln Sie die Grundlagen zur Methode der Annotation, betten sie in die geisteswissenschaftliche Tradition ein und beleuchten diese kritisch. Die Studierenden entwerfen Fallstudien zum Primärtext und beginnen, in Gruppen kollaborativ zu annotieren.
- In der zweiten Sitzung analysieren die Studierenden ihre Annotationsdaten. Im Plenum führen Sie die Interpretationen der Teilgruppen zusammen und gemeinsam mit Ihren Studierenden runden Sie die Fallstudie ab.

Bevor Sie nun kurz in die Methode einführen, empfehlen wir, dass Sie die Studierenden schon einmal in ihr vorbereitetes Projekt einladen. Gehen Sie dazu im Projekt-Modul auf der *Team*-Kachel ins Drei-Punkte-Menü und wählen Sie *Invite someone to the Project* (siehe Abb. 1).

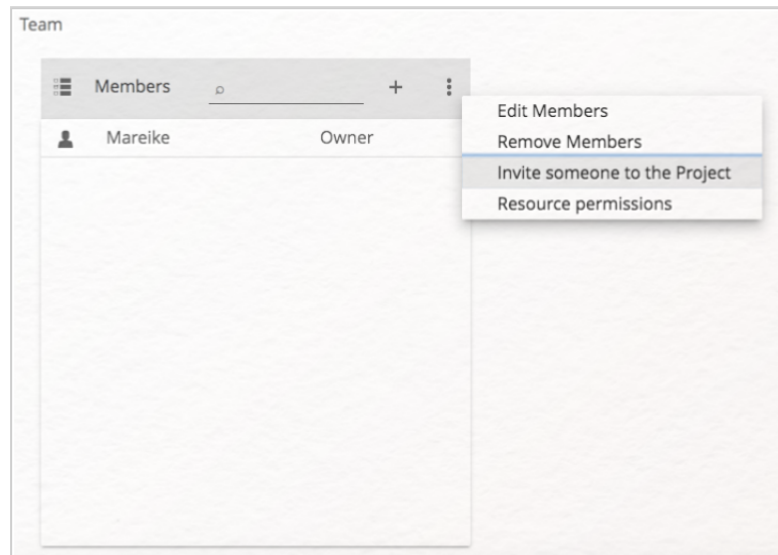


Abb. 1: Team-Kachel im CATMA Project-Modul

Es öffnet sich ein Fenster mit Spezifizierungs-Möglichkeiten für die Einladung. Sie können hier wählen, ob CATMA automatisch für jeden neuen am Projekt Teilnehmenden eine *Annotation Collection* anlegen soll. Da die *Annotation Collection* der Ort ist, an dem alle Annotationen eines Annotierenden gespeichert werden, empfehlen wir, hier einen Haken zu setzen. Sie können dann wählen, ob die Eingeladenen alle oder nur bestimmte Dokumente einsehen und annotieren können sollen. Über das Drop-Down-Rollen-Menü können Sie wählen, welche Rechte die Eingeladenen haben sollen. Eine Übersicht über die Rechte der einzelnen Rollen finden Sie [hier](#). Wir empfehlen für Seminare die Rolle „Assistant“, mit welcher Ihre Studierenden zwar Annotationen anlegen und löschen sowie Tagsets anlegen, aber keine Tags und Tagsets löschen können. Ihre Spezifikationen könnten also wie in Abb. 2 dargestellt aussehen.

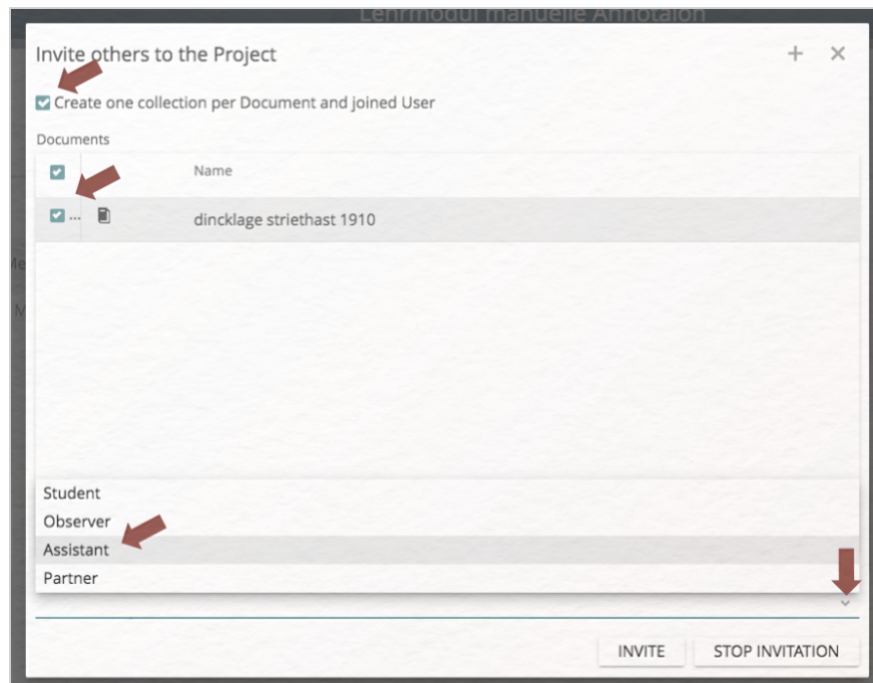


Abb. 2: Invite-Fenster

Wenn Sie nun auf *Invite* klicken, erhalten Sie einen Zugangscodes, den Sie über den Beamer an die Wand projizieren können. Schließen Sie dieses Fenster (siehe Abb. 3) erst, wenn sich alle Studierenden angemeldet haben.

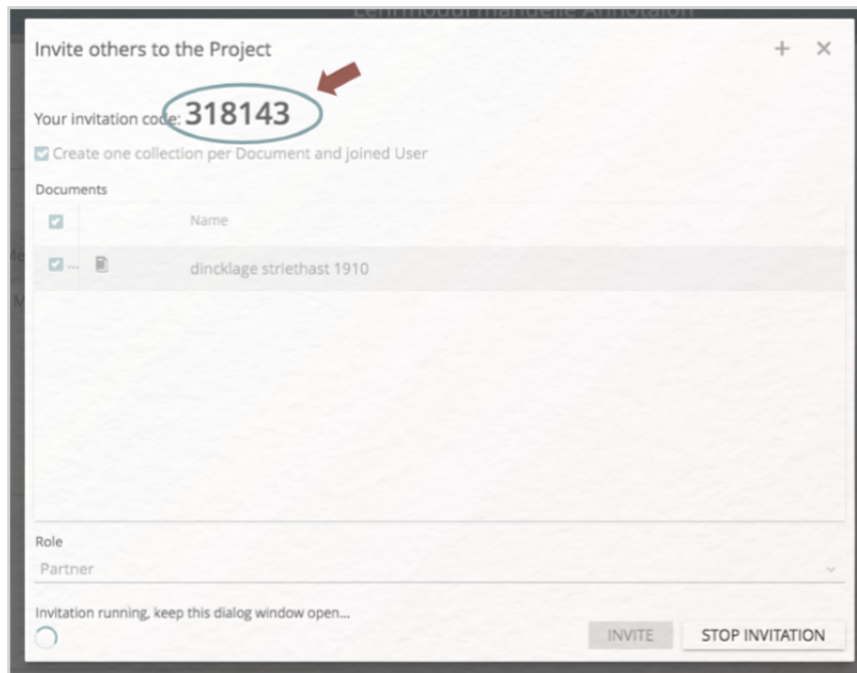


Abb. 3: Fenster mit Invitation-Code

Bitte Sie Ihre Studierenden nun, auf dem Startbildschirm von CATMA auf *Join Projekt* zu gehen. Es öffnet sich ein Fenster, in dem sie den *Invitation-Code* eingeben können. Während CATMA nun für alle neuen Team-Mitglieder und jeden Text im Projekt eine neue *Annotation Collection* anlegt (was einen Moment dauern kann), können Sie mit Ihrer Einführung beginnen.

Die manuelle digitale Annotation ist ein Verfahren, das unmittelbar an eine zentrale Technik der Geisteswissenschaften, ja sogar an eine wichtige Kulturtechnik (Moulin 2010) anknüpft: das Markieren und dichte Beschreiben von Texten bzw. einzelnen Textpassagen (Geertz 1987) zur Bedeutung dichter Beschreibungen für die Wissenschaften. Mithilfe der manuellen Annotation können hypothesengeleitete Untersuchungen durchgeführt werden, die dem hermeneutischen Prinzip folgen. Für ein literaturwissenschaftliches Seminar ist es empfehlenswert, diese Anknüpfungspunkte an traditionelle geisteswissenschaftliche Forschung hervorzuheben, da die Methode und insbesondere CATMA diesen Prinzipien entspricht bzw. ihnen folgend entwickelt wurde. Digitale manuelle Annotation kann auch genutzt werden, um den Zusammenhang zwischen Text als Basis literaturwissenschaftlichen Datenmaterials und Annotations- und später auch Analysedaten zu erläutern. Das eingesetzte Tool kombiniert Annotation mit der Möglichkeit, komplexe Abfragen zu erstellen, zu visualisieren und zu analysieren. Somit verbindet es das in den Literaturwissenschaften sehr häufig eingesetzte *Close Reading* (vgl. *Close Reading*) mit neueren Ansätzen des *Distant Reading* (vgl. *Distant Reading*). *Distant Reading* wird hier im Sinne Morettis (2013) als Analyseverfahren verstanden, bei der statistische Auswertungen von Texten, sowie Paratexte und Metadaten berücksichtigt werden. Im Sinne eines *Scalable Reading*-Verfahrens wird hier während des Analyseprozesses in den Text hinein- und wieder hinausgezoomt.

An dieser Stelle bietet es sich an, auch die Unterschiede und Gemeinsamkeiten von linguistischer und literaturwissenschaftlicher Annotation herauszuarbeiten. In der Linguistik werden meist eindeutige, intersubjektive Annotationen angestrebt und es wurden bereits Standards wie z.B. *Gold Standard* oder *Inter-Annotator-Agreement* ausgearbeitet. In den Literaturwissenschaften sind dagegen gerade Phänomene von Interesse, die uneindeutig sind. Konfligierende Annotationen werden darum nicht unbedingt als Problem betrachtet. Intersubjektive Übertragbarkeit ist nur selten möglich und Annotationen können stark voneinander abweichen. Dieses Phänomen werden die Studierenden während der hier angebotenen Sitzungen selbst ganz aktiv erfahren. An dieser Stelle ist es sinnvoll, eine erste, kritische Diskussion anzuregen. Fragen Sie die Studierenden, welche Vor- und Nachteile sie im kollaborativen Arbeiten sehen, und geben Sie ihnen die Möglichkeit, Rückfragen zu den vorbereiteten theoriebasierten Materialien (Methodeneinträge digitale manuelle Annotation (Horstmann 2024a) und kollaborative Annotation (Jacke 2024b)) zu stellen.

2.3 Erarbeitung

Die Studierenden haben zur Vorbereitung des Seminars bereits die Lerneinheit (Horstmann 2024a) mithilfe der Tutorial-Videos durchgearbeitet (forTEXT (2019); Tutorial-Videos auf Zenodo und Youtube, bzw. fortext.net). Leiten Sie also nun von der Diskussion der Methode zu konkreten technischen Erfahrungen über. Geben Sie den

Studierenden die Möglichkeit, von Schwierigkeiten zu berichten. Wenn Sie in Ihrem Seminar bereits andere digitale Tools eingeführt haben, z. B. mit dem Lehrmodul Netzwerkanalyse (Schumacher 2024b) oder Lehrmodul Topic Modeling (Horstmann 2024b), können Sie die in dieser Sitzung gelehrt Methode auch zu den vorigen in Bezug setzen. Lassen Sie dazu Ihre Studierenden die Methoden vergleichen und kurz schildern, was ihnen bei der Nutzung im Vergleich zu den anderen Tools aufgefallen ist.

Teilen Sie das Seminar nun in drei Gruppen ein. Jede Gruppe sollte aus mindestens drei Studierenden bestehen. Bilden Sie notfalls lieber weniger als zu kleinere Gruppen. Jede der drei Gruppen wird einen zwei Kapitel umfassenden Ausschnitt der Novelle analysieren. Alle Gruppen arbeiten mit den beiden Tag-Unterkategorien Distanz und Dauer. Weisen Sie darauf hin, dass beim Annotieren diese Guidelines (Jacke 2024c) beachtet werden sollten. Wahrscheinlich werden die Studierenden einen Teil der Annotationsaufgaben als Nachbereitung des Seminars mit nach Hause nehmen müssen.

Aufgabe 1: Annotieren Sie Ihren Abschnitt der Novelle *Der Striethast* von Emmy von Dincklage mit den Tagset-Kategorien Dauer und Distanz. Jede*r in Ihrer Gruppe annotiert zunächst einzeln und parallel zu den anderen Gruppenmitgliedern, sodass am Ende unterschiedliche Annotationsvarianten vorhanden sind.

Aufgabe 2: Gehen Sie ins *Project*-Modul und wählen Sie im Drei-Punkte-Menü *Commit all changes*. Gehen Sie anschließend im selben Drei-Punkte-Menü auf *Synchronize with the team*. Gehen Sie zurück ins *Annotate*-Modul und geben Sie in der grauen Lasche an, dass Sie die *Annotation-Collections* Ihrer Gruppenmitglieder angezeigt bekommen möchten. Schauen Sie sich Ihre Annotationen gemeinsam an. Was fällt Ihnen auf?

Zu Beginn der zweiten Sitzung gehen Sie mit dem Seminar die Funktionen des *Analyze*-Moduls durch. Nutzen Sie dafür zunächst die vorinstallierten Standard-Queries (vgl. *Query*) (Wordlist, Wildcard, Taglist). Erstellen Sie eine *Wordcloud* aus der Wordlist, einen Distributionsgrafen mit der Wildcard und eine *KWIC*-Tabelle mit der Taglist. Lassen Sie die Studierenden dann erneut in die Gruppen gehen.

Aufgabe 3: Erstellen Sie aus der *Taglist* für Ihre *Annotation Collection* eine *Tagcloud* (Achtung, dafür müssen Sie die Funktion *group by path* aktivieren). Welche Tags haben Sie besonders häufig verwendet? Was sagt Ihnen das über die Erzählstruktur des von Ihnen analysierten Abschnitts der Novelle *Der Striethast*? Vergleichen Sie Ihre *Tagcloud* mit denen Ihrer Gruppenmitglieder. Was fällt auf? Speichern Sie die *Tagcloud* von jedem Gruppenmitglied ab.

Aufgabe 4: Erstellen Sie einen *Distributionsgrafen*, in dem jeder Ihrer *Tagpaths* abgebildet wird. Gibt es Passagen, in denen ein bestimmter Tag besonders häufig vergeben wurde? Klicken Sie doppelt auf den höchsten Punkt der Verteilung. Es öffnet sich eine *Keyword-in-Context*-Tabelle. Klicken Sie doppelt auf das oberste Keyword, um zurück ins Annotate Modul und an die Stelle genau dieser Annotation zu springen. Lesen Sie die Passage und interpretieren Sie, warum gerade hier diese Tagkategorie vorherrscht. Vergleichen Sie dann auch hier Ihren Grafen mit denen Ihrer Teammitglieder und speichern Sie alle Visualisierungen ab. Sprechen Sie über Ihre Interpretation.

Aufgabe 5: Bereiten Sie eine fünfminütige Zusammenfassung Ihrer Arbeit vor. Wählen Sie eine besonders prägnante Visualisierung oder stellen Sie alle Visualisierungen zu einer Abfrage für all Ihre *Annotation-Collections* nebeneinander. Die Leitfrage Ihrer Zusammenfassung sollte sein: Welche Tagkategorie war besonders relevant für die Novelle *Der Striethast* und was sagt das über die Erzählstruktur aus?

2.4 Sicherung

Lassen Sie nun jede Gruppe kurz präsentieren. Die Studierenden sollen dazu jeweils die ausgewählte Grafik mit dem Beamer an die Wand projizieren. Sie sollten ca. fünf Minuten über die Leitfrage sprechen. Dann können noch Nachfragen aus dem Plenum gestellt werden. Wenn alle Gruppen ihre Ergebnisse vorgestellt haben, gehen Sie zurück ins Plenum und die letzte Phase der Lehreinheit kann beginnen.

2.5 Transfer & Reflexion

Bitten Sie nun die Studierenden, von der zuletzt vorgestellten Kategorie ausgehend die jeweils eigene Annotationsperspektive hinzuzufügen. Versuchen Sie so die erzähltheoretischen Spezifika der Novelle gemeinsam herauszuarbeiten. Wenn die Diskussion nicht von selbst zu dem Punkt kommt, an dem die Studierenden von ihren Erfahrungen mit der kollaborativen Annotation berichten, so führen Sie sie dorthin. Fragen Sie, ob die häufigste Tagkategorie bei allen Gruppenmitgliedern dieselbe war und was der Vergleich der Visualisierungen ergeben hat. Was ist ihnen beim Annotieren aufgefallen? Wie sind sie in der Gruppe zur Interpretation der Novelle im Hinblick auf ihre Annotationskategorie gekommen?

Am Ende fragen Sie Ihre Studierenden, ob sie eine Idee haben, wie sie die digitale Annotation in eine Hausarbeit integrieren könnten. Was wäre eine mögliche Fragestellung, die sie mittels digitaler Annotation beleuchten könnten? Fragen Sie sie auch, inwiefern sie digitale Annotation für ihr weiteres Studium oder ihr späteres Berufsleben nutzen könnten. Könnte diese eher an die traditionellen Forschungsmethoden der Geisteswissenschaften angelehnte Technik sinnvoll mit anderen Methoden verknüpft werden? Weisen Sie Ihre Studierenden auch

darauf hin, dass das verwendete narratologische Tagset durch andere ergänzt oder ersetzt werden kann. Gibt es andere Themenbereiche oder Forschungstraditionen, die sich als Hintergrund zur Erarbeitung eines Tagsets eignen (z.B. Gender, Diskursanalyse, Dekonstruktivismus)? Zum Schluss können Sie noch einmal einen Bogen zum Beginn der Einheit schlagen und verdeutlichen, dass CATMA zwar den Close-Reading-Prozess (und damit eine sehr traditionsreiche literaturwissenschaftliche Forschungsmethode) unterstützt, dass das kollaborative Arbeiten und die quantitativen Auswertungen aber den Weg in Richtung Distant Reading eröffnen – und das ohne den Rückweg in den Text zu verschließen.

3. Lösungen zu den Beispielaufgaben

Die Lösungsvorschläge erhalten Sie auf Zenodo (forTEXT 2020b).

Externe und weiterführende Links

- Verlaufsraster, Beispielpräsentation und Lösungen zum Lehrmodul auf Zenodo: <https://web.archive.org/save/https://zenodo.org/records/12530243> (Letzter Zugriff: 03.07.2024)
- Novelle *Der Striethast* im DTA: https://web.archive.org/save/http://www.deutschestextarchiv.de/book/show/dincklage_striethast_1910 (Letzter Zugriff: 03.07.2024)
- Übersicht über die Rechte der einzelnen Rollen (CATMA): <https://web.archive.org/save/https://catma.de/documentation/roles-and-permissions/> (Letzter Zugriff: 03.07.2024)
- CATMA: <https://web.archive.org/save/https://app.catma.de> (Letzter Zugriff: 03.07.2024)
- Videoreihe „Manuelle und Annotation und Literaturanalyse“:
 - auf fortext.net: <https://web.archive.org/save/https://fortext.net/ressourcen/videos/tutorials/manuelle-annotation-und-literaturanalyse> (Letzter Zugriff: 03.07.2024)
 - auf Youtube: <https://web.archive.org/save/https://youtu.be/d5yxEOQu2Dw> (Letzter Zugriff: 03.07.2024)
 - auf Zenodo: <https://web.archive.org/save/https://doi.org/10.5281/zenodo.10353556> (Letzter Zugriff: 03.07.2024)

Bibliographie

- forTEXT. 2019. Tutorial: CATMA 6 zur manuellen Annotation nutzen. Manuelle Annotation und Literaturanalyse. 28. Oktober. doi: 10.5281/zenodo.10353556, <https://zenodo.org/records/10353556>.
- . 2020a. Tagset Narratologie (discours). Zenodo, 27. Januar. doi: 10.5281/zenodo.105196488, <https://zenodo.org/records/105196488>.
- . 2020b. Digitale Annotation mit CATMA lehren. 23. März. doi: 10.5281/zenodo.10519503, <https://zenodo.org/records/12530243>.
- Geertz, Clifford. 1987. *Dichte Beschreibung*. Frankfurt am Main: Suhrkamp.
- Horstmann, Jan. 2024a. Lerneinheit: Manuelle Annotation mit CATMA. *forTEXT Heft 1*, Nr. 4. Manuelle Annotation (7. August). doi: 10.48694/fortext.3750, <https://fortext.net/routinen/lerneinheiten/manuelle-annotation-mit-catma>.
- . 2024b. Lerneinheit: Topic Modeling mit dem DARIAH Topics Explorer lehren. *forTEXT Heft 1*. Topic Modeling (30. September). doi: 10.48694/fortext.3716, <https://fortext.net/routinen/lehrmodule/topic-modeling-mit-dem-dariah-topics-explorer-lehren>.
- Horstmann, Jan und Alexandra Kern. Ressourcenbeitrag: Deutsches Textarchiv (DTA). *forTEXT Heft 1*. Bibliografie. doi: 10.48694/fortext.3791, <https://fortext.net/ressourcen/textsammlungen/deutsches-textarchiv-dta>.
- Jacke, Janina. 2024b. Methodenbeitrag: Kollaboratives literaturwissenschaftliches Annotieren. *forTEXT Heft 1*, Nr. 4. Manuelle Annotation (7. August). doi: 10.48694/fortext.3749, <https://fortext.net/routinen/methoden/kollaboratives-literaturwissenschaftliches-annotieren>.
- . 2024a. Methodenbeitrag: Manuelle Annotation. *forTEXT Heft 1*, Nr. 4. Manuelle Annotation (7. August). doi: 10.48694/fortext.3748, <https://fortext.net/routinen/methoden/manuelle-annotation>.
- . 2024c. Ressourcenbeitrag: Tagset Narratologie (discours). *forTEXT Heft 1*, Nr. 4. Manuelle Annotation (7. August). doi: 10.48694/fortext.3756, <https://fortext.net/ressourcen/tagsets/tagset-narratologie-discours>.
- Moulin, Claudine. 2010. Am Rande der Blätter. Gebrauchsspuren, Glossen und Annotationen in Handschriften und Büchern aus kulturhistorischer Perspektive. *Autorenbibliotheken, Quarto. Zeitschrift des Schweizerischen Literaturarchivs* 30/31: 19–26.
- Schumacher, Mareike. 2024a. Toolbeitrag: CATMA. *forTEXT Heft 1*, Nr. 4. Manuelle Annotation (7. August). doi: 10.48694/fortext.3761, <https://fortext.net/tools/tools/catma>.
- . 2024b. Lehrmodul: Netzwerkanalyse mit Gephi lehren. *forTEXT Heft 1*, Nr. 6. Netzwerkanalyse (30. August). doi: 10.48694/fortext.3780, <https://fortext.net/routinen/lehrmodule/netzwerkanalyse-mit-gephi-lehren>.

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).
- Distant Reading** Distant Reading ist ein Ansatz aus den digitalen Literaturwissenschaften, bei dem computationale 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.
- KWIC** KWIC steht für *Keyword in Context*. Dabei handelt es sich um eine Darstellungsform, bei welcher die Treffer eines bestimmten Suchbegriffs in ihrem Kontext zeilenweise aufgelistet werden. Die Größe der Kontexte, also die Anzahl der angezeigten Umgebungswörter, kann meist individuell festgelegt werden.
- 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 unbekanntem 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.
- 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**.
- Query** *Query* bedeutet „Abfrage“ oder „Frage“ und bezeichnet eine computergestützte Abfrage zur Analyse eines Textes. Um Datenbestände zu durchsuchen, werden Abfragesprachen eingesetzt, die *Queries* (Anfragen) an den Datenbestand senden. So bilden alle möglichen Queries zusammen die *Query Language* eines Tools.
- 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.

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.

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.