

Ressourcenbeitrag: Tagset Narratologie (discours)

Janina Jacke  ¹

1. Christian-Albrechts-Universität zu Kiel

forTEXT

Thema:	Manuelle Annotation	DOI:	10.48694/fortext.3756
Jahrgang:	1	Ausgabe:	4
Erscheinungsdatum:	2024-08-07	Erstveröffentlichung:	2020-01-27 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.

1. Kurzbeschreibung

Das Tagset „Narratologie (discours)“ ist für die Annotation narrativer Elemente in Texten geeignet. Es enthält grundlegende Kategorien für die Analyse des Diskurses (*discours*, d. h. der Art und Weise der Darstellung) und basiert maßgeblich auf Arbeiten der strukturalistisch orientierten Narratologie. Das Tagset erhalten Sie auf Zenodo (forTEXT 2020).

2. Anwendungsbeispiel

Angenommen, Sie möchten an einem kleinen **Korpus** deutschsprachiger Novellen des 18. Jahrhunderts eine erste explorative Studie zu der Frage durchführen, ob bzw. in welchen Hinsicht sich die Erzählweise in kürzeren narrativen Texten im Laufe des Jahrhunderts verändert hat (z. B. hinsichtlich der zeitlichen Gestaltung oder hinsichtlich der Perspektive, aus der die Geschichte präsentiert wird). Sie können hierfür die im Tagset Narratologie (discours) zur Verfügung gestellten Kategorien für die *discours*-Analyse nutzen, um unterschiedliche Facetten der erzählerischen Darstellung in Ihrem Korpus manuell zu annotieren (Jacke 2024). Diese Annotationen lassen sich dann beispielsweise mithilfe von Queries (vgl. Query) oder geeigneter Visualisierungen (Horstmann und Stange 2024) untersuchen, um erste Muster zu erkennen.

3. Diskussion

Die Narratologie ist eine geisteswissenschaftliche Disziplin, die Prinzipien und Praktiken des Erzählens zum Gegenstand hat (Meister 2014). In diesem Zusammenhang gehört zur narratologischen Arbeit zum einen die Entwicklung theoretischer Modelle für die Beschreibung, Analyse und teilweise Interpretation erzählender Texte, zum anderen die Erforschung konkreter (meist, aber nicht ausschließlich fiktionaler) Erzählungen.

Die Narratologie ist in den 1960er Jahren aus dem russischen Formalismus und dem französischen Strukturalismus hervorgegangen und orientiert sich deswegen in ihrer Ursprungsform an den entsprechenden Forschungsparadigmen: Maßgeblich geprägt durch Gérard Genettes *Discours du récit* (Genette 2007), konzentrieren sich narratologische Arbeiten insbesondere auf die Entwicklung von systematisch organisierten Kategorien zur genauen Beschreibung narrativer Textmerkmale (vor allem der Facetten der Darstellung: Wer erzählt die Geschichte, an wessen Perspektive orientiert sich die Erzählung, wie ist die Präsentation zeitlich organisiert?). Die strukturalistisch ausgerichtete Variante der Narratologie versteht sich aufgrund dieser festgelegten Maßstäbe und Systematiken als wenig kontextabhängiger bzw. wenig interpretativer Zugang zur Texterforschung, der meist intersubjektiv nachvollziehbare Ergebnisse liefert und als Heuristik für die Textinterpretation dienen kann (Kindt und Müller 2003).

Ungefähr seit den 1980er, vermehrt seit den 1990er Jahren entwickelt sich die Narratologie dann zu einer zunehmend diverseren Disziplin bzw. zu einem Konglomerat von Unterdisziplinen. Hinzu kommen dabei im Rahmen sogenannter postklassischer Narratologien neue Untersuchungsgegenstände in Form anderer Medien (Kuhn 2011; Thon 2016) sowie neue Kontexte (Herman 1999). Besonders durch die Kopplung erzähltheoretischer Untersuchungen an bestimmte Kontexte (beispielsweise Leseforschung, Gendertheorie oder Kulturwissenschaft) transformieren sich narratologische Untersuchungen teilweise von beschreibender und analytischer Texterforschung zu einem genuin interpretativen Unterfangen. Gerade die strukturalistisch-formalistische Variante narratologischer Untersuchungen scheint aufgrund ihres Fokus auf Textbeschreibung und ihres systematischen Kategorieninventars prädestiniert für eine computergestützte Umsetzung. Tatsächlich gibt es bereits zahlreiche Projekte, die sich der Computational Narratology (Mani 2013) zuordnen lassen – insbesondere auch solche,

die stark von Close-Reading-Methoden (vgl. **Close Reading**) und manueller Annotation (Jacke 2024) Gebrauch machen (Bögel u. a. 2015; Brunner 2015; Gius 2015; Modrow 2016). Dennoch eignen sich bestimmte digitale Methoden auch zur Unterstützung postklassisch-narratologischer Projekte – so beispielsweise hermeneutische Annotation.

Die Kategorien des hier vorgestellten Tagsets sind weitgehend als Beschreibungskategorien zu verstehen und sind daher an keine bestimmte Interpretationstheorie gekoppelt. Je nach Ambiguität oder Deutungsoffenheit der mithilfe des Tagsets analysierten Texte kann die Annotation von Textelementen allerdings dennoch (bestimmte Formen von) Interpretation erfordern. Die Kategorien für die *discours*-Analyse orientieren sich weitestgehend an Genette (2007) – mit Ausnahme derjenigen zu Erzählebenen, die auf einem Modell von Ryan (1991) basieren. Das Tagset kann entsprechend als Reduktion und starke Vereinfachung der in Gius (2015) bzw. Modrow (2016) entwickelten narratologischen Tagsets verstanden werden.

4. Tagset

Sie können das Tagset Narratologie (discours) auf Zenodo über [diesen Link](#) (forTEXT 2020) als XML-Datei herunterladen, um es dann in geeignete Tools (beispielsweise CATMA (Schumacher 2024)) zu importieren und es dort zu verwenden. Abbildung 1 zeigt die im Tagset enthaltenen Tags in ihrer hierarchischen Struktur sowie die Properties (vgl. **Property**) und Values.

Tagsets	Tags	Properties	Values
Narratologie (Teil 1)	discours		
	Stimme		
	Erzählebenen		
	primäre Ebene	Grenzüberschreitung	illokutionär ontologisch
	sekundäre Ebene	Grenzüberschreitung	illokutionär ontologisch
	tertiäre Ebene	Grenzüberschreitung	illokutionär ontologisch
	Erzählinstanz		
	homodiegetisch		
	heterodiegetisch		
	Modus		
	Distanz/Redewiedergabe		
	geringe Distanz/zitierte Rede		
	mittlere Distanz/transponierte Rede		
	hohe Distanz/erzählte Rede		
	Fokalisierung		
	Nullfokalisierung		
	interne Fokalisierung		
	externe Fokalisierung		
	Zeit		
	Ordnung		
	Analepse		
	Prolepse		
	Dauer		
	zeitraffend		
	zeitdeckend		
	zeitdehnend		
	Frequenz		
	repetitiv		
	iterativ		

Abb. 1: Tagset Narratologie (discours)

5. Richtlinien zur Anwendung

Zunächst ein allgemeiner Hinweis: Diese Richtlinien enthalten nur spezifische Anwendungshinweise für die Tags, die sich auf der untersten Hierarchieebene befinden (also beispielsweise nicht *discours*, *Stimme* oder *Erzähler*, sondern *homodiegetisch* oder *heterodiegetisch*). Hierarchisch höherliegende Kategorien dienen dagegen vor allem der Systematisierung.

Im Folgenden werden die einzelnen Kategorien kurz definiert – für die Tags der untersten Hierarchieebene werden darüber hinaus Hinweise zur Länge der annotierten Passage und zu textuellen Indikatoren angegeben sowie i. d. R. ein Beispiel (Gius und Jacke 2016).

discours: Die hier versammelten Kategorien dienen der Annotation der Art und Weise des Erzählens.

5.1 Stimme

Stimme: Diese *discours*-Kategorien werden für die Analyse der Erzählstimme verwendet.

Erzählebenen: Diese Unterkategorien zu *Stimme* dienen der Analyse von Rahmen- und Binnenerzählungen. Sie eignen sich als Einstieg für eine explorative narratologische Annotation. Erzählebenen segmentieren Erzählungen und haben Einfluss auf die weitere narratologische Analyse (s. u.). Dem in diesem Tagset operationalisierten Ebenenkonzept zufolge unterscheiden sich Erzählebenen entweder illokutionär (d. h. sie werden durch unterschiedliche Erzählinstanzen präsentiert) oder ontologisch (d. h. sie erzählen von verschiedenen Realitätssystemen; Beispiele finden Sie unten).

Erzählebenen-Annotationen werden in der Regel gestapelt: Tritt beispielsweise im Rahmen einer sekundären Ebene eine tertiäre Ebene auf, wird die tertiäre Ebene an der relevanten Textstelle *zusätzlich* annotiert.

- Tag *primäre Ebene*: Hierbei handelt es sich um die grundlegende Erzählebene, die jede Erzählung notwendigerweise aufweist.
 - Länge der annotierten Passage: i. d. R. mehrere Absätze.
 - Indikatoren: Die primäre Erzählebene umfasst normalerweise die gesamte Erzählung.
 - Hinweis: Da diese Annotation i. d. R. die gesamte Erzählung umfasst, empfiehlt es sich oft aus pragmatischen Gründen, sie wegzulassen.
- Tag *sekundäre Ebene*: Mit diesem Tag werden Passagen annotiert, die Erzählebenen enthalten, die direkt in die primäre Erzählebene eingebettet sind. Eine solche Einbettung kann durch eine illokutionäre Grenzüberschreitung stattfinden (d. h. eine erzählte Figur wird zur erzählenden; dies umfasst auch den Fall, in dem eine Erzählinstanz Äußerungen wiedergibt, die sie als erzählte Figur getätigt hat) oder durch eine ontologische Grenzüberschreitung (d. h. wenn in ein anderes Realitätssystem gewechselt wird, z. B. in Traumsequenzen, Wunschvorstellungen oder intratextuellen Fiktionen).
 - Länge der annotierten Passage: i. d. R. Sätze oder Absätze.
 - Indikatoren: Figurenrede (oft eingeleitet durch *verba dicendi* und/oder eingerahmt von Anführungsstrichen) für illokutionäre Grenzüberschreitungen; Konjunktiv oder *verba pensendi* für ontologische Grenzüberschreitungen.
 - Property *Grenzüberschreitung*: Diese Property wird verwendet, um in Annotationen für sekundäre Erzählebenen festzuhalten, durch welche Art von Grenzüberschreitung die eingebettete Ebene entsteht. Mögliche Values sind *illokutionär* und *ontologisch* (s. o.). Es ist auch möglich, beide Values zusammen zu vergeben.
 - Beispiel: „*Ich war krank!*“ sagte Matteo leise.“ (C. F. Hebbel: *Matteo*; illokutionäre Grenzüberschreitung)
- Tag *tertiäre Ebene*: Mit diesem Tag werden Passagen annotiert, die Erzählebenen enthalten, die direkt in sekundäre Erzählebenen eingebettet sind. In allen weiteren Punkten gleichen sie sekundären Erzählebenen.
 - Beispiel: „Am nächsten Morgen erzählte uns Frau von B., daß sie geträumt habe, *der schöne Armenier läge schlafend auf einer Bank [...]*“ (F. G. zu Reventlow: *Das polierte Männchen*; ontologische Grenzüberschreitung)

Erzählinstanz: Diese Kategorie ist ebenfalls der Analyse der Erzählstimme gewidmet. Das hier relevante Kriterium ist der ontologische Status der Erzählinstanz in Relation der von ihr erzählten Welt. Die Tags werden pro Erzählebene vergeben, d. h. es wird in der Regel eine gesamte Erzählebene mit einem Tag annotiert, auch wenn der ontologische Status der Erzählinstanz nicht in jedem Satz eindeutig markiert ist. In eingebetteten Erzählungen (sekundäre Ebene etc.) wird der ontologische Status des jeweiligen Binnenerzählers in Relation zu seiner Erzählung annotiert.

- Tag *homodiegetisch*: Eine Erzählinstanz ist genau dann homodiegetisch, wenn sie Teil der erzählten Welt ist. Dabei ist unwichtig, welche Rolle sie in der erzählten Geschichte spielt.
 - Länge der annotierten Passage: i. d. R. mehrere Absätze (entspricht Länge der Erzählebenen)
 - Indikatoren: Verwendung von Pronomen in erster Person (Singular)
 - Beispiel: „*Da bin ich nun den ganzen Nachmittag in den Straßen herumspaziert, [...]*“ (A. Schnitzler: *Blumen*).
- Tag *heterodiegetisch*: Eine Erzählinstanz ist genau dann heterodiegetisch, wenn sie nicht Teil der erzählten Welt ist.
 - Länge der annotieren Passage: i. d. R. mehrere Absätze (entspricht Länge der Erzählebenen)

- Indikatoren: keine Verwendung von Pronomen in erster Person (Singular)
- Beispiel: „*Mitschüler erzählten als Witz, seine Mutter sei Leichenbändigerin und seine Großmutter Löwenfrau gewesen. [...]*“ (J. Ringelnatz: *Nervosipopel*)

5.2 Modus

Modus: Mit diesen *discours*-Kategorien wird die Perspektive analysiert, aus der die Ereignisse einer Geschichte präsentiert werden.

Distanz/Redewiedergabe: Die hier versammelten Tags dienen der Kategorisierung der Art und Weise, wie Figurenrede wiedergegeben wird. Mit diesen Formen der Redewiedergabe gehen zugleich verschiedene Grade der Mittelbarkeit der Darstellung einher. Die Analyse von Redewiedergabe kann auf der Annotation der Erzählebenen aufbauen, indem erstere auf alle als illokutionäre Grenzüberschreitungen ausgezeichneten Passagen angewandt wird.

- Tag *geringe Distanz/zitierte Rede*: Mit diesem Tag werden alle Passagen ausgezeichnet, in denen Figurenrede wörtlich wiedergegeben wird.
 - Länge der annotierten Passage: i. d. R. Sätze oder Absätze (entspricht Länge der Erzählebenen)
 - Indikatoren: Anführungszeichen, Pronomen in erster Person
 - Beispiel: „*Ihr betrachtet da mein Instrument*“, sagte Albert“ (L. Tieck: *Der Pokal*)
- Tag *mittlere Distanz/transponierte Rede*: Dieser Tag wird für Passagen verwendet, in denen Figurenrede in indirekter Rede wiedergegeben wird.
 - Länge der annotierten Passage: i. d. R. Sätze oder Absätze (entspricht Länge der Erzählebenen)
 - Indikatoren: Konjunktiv, weitgehende Beibehaltung figuraler Ausdrucksweise mit Ausnahme von Personalpronomen und indexikalischen Ausdrücken
 - Beispiel: „*Er hätte jetzt überhaupt viel in der Stadt zu thun*, sagte sie [...]“ (J. Proelß: *Lili*)
- Tag *hohe Distanz/erzählte Rede*: Mit diesem Tag werden Passagen versehen, in denen figurale Äußerungen (stark) zusammengefasst werden.
 - Länge der annotierten Passage: i. d. R. Teilsätze oder Sätze (entspricht Länge der Erzählebenen)
 - Indikatoren: Konjunktiv, Abweichung von figuraler Ausdrucksweise, (starke) Verkürzung
 - Beispiel: „Herr Bulverin, so hieß der Drogist, wußte *nur Gutes* an die alten Daddeldus zu berichten.“ (J. Ringelnatz: *Nervosipopel*)

Fokalisierung: Mithilfe dieser Tags wird festgehalten, an wessen Wahrnehmung bzw. Wissen sich die Darstellung in einer bestimmten Textpassage orientiert.

- Tag *Nullfokalisierung*: Eine Passage gilt genau dann als nullfokalisiert, wenn keine Einschränkung der Wahrnehmung bzw. des Wissens erkennbar ist, d. h. wenn eine allwissende Erzählinstanz anzunehmen ist. Die Erzählinstanz weiß also mehr als eine Figur (bzw. als alle Figuren zusammen).
 - Länge der annotierten Passage: i. d. R. Sätze oder Absätze
 - Indikatoren: Erzählen von Ereignissen, von denen keine der Figuren weiß
 - Beispiel: „[...] Allen war die unerwartete Erscheinung des Fürsten Isidor ein unerklärliches Rätsel, denn niemand wußte, daß Fürst Isidor, durch das Alter, überdem aber noch durch falsches Haar, durch Schminke entstellt und auf diese Weise unerkannt, im Lande hauste, daß er in den letzten Tagen in jenem verfallenen Schloß auf den Tod des Fürsten lauerte.“ (E. T. A. Hoffmann: *Die Doppelgänger*)
- Tag *interne Fokalisierung*: Eine Passage ist intern fokalisiert, wenn sie sich an der Wahrnehmung bzw. dem Wissen einer Figur orientiert – d. h. die Erzählinstanz weiß so viel wie eine Figur.
 - Länge der annotierten Passage: i. d. R. Sätze oder Absätze
 - Indikatoren: Verben der Wahrnehmung, die sich auf eine Figur als Subjekt beschränken
 - Beispiel: „Da sah er die ganze Herrlichkeit Gottes um sich gebreitet. All die wunderbaren in goldenem Violett schimmernden Berge, die Seen mit ihrem halbversteckten Flimmer [...]“ (M. Janitschek: *Poverino*)
- Tag *externe Fokalisierung*: Eine externe Fokalisierung liegt vor, wenn sich die Erzählung auf die unkommentierte Darstellung äußerer Geschehnisse beschränkt. Die Erzählinstanz weiß also weniger als eine Figur.
 - Länge der annotierten Passage: i. d. R. Sätze oder Absätze
 - Indikatoren: keine Verben der Wahrnehmung
 - Beispiel: „Vom großen Dom erscholl das vormittägige Geläute. Über den weiten Platz wandelten in verschiedenen Richtungen Männer und Weiber, Wagen fuhren vorüber und Priester gingen nach ihren Kirchen. [...]“ (L. Tieck: *Der Pokal*)

5.3 Zeit

Zeit: Diese *discours*-Kategorien dienen der Analyse des temporalen Verhältnisses zwischen den erzählten Ereignissen und ihrer Präsentation in der Erzählung.

Ordnung: Unter der zeitlichen Ordnung wird die Reihenfolge der Präsentation der Ereignisse im Verhältnis zur tatsächlichen Reihenfolge der Ereignisse in der erzählten Welt verstanden. (Ontologisch verschiedene Erzählebenen können zueinander nicht im Rahmen von Ordnungsanalysen in Beziehung gesetzt werden.) Ein häufiger Fall ist hierbei die chronologische Darstellung, bei der die Reihenfolge der Präsentation der Reihenfolge der Ereignisse entspricht. Passagen, in denen in die chronologische Erzählung zeitliche Vor- oder Rückgriffe eingeschoben werden, werden anachronisch genannt und als Abweichungen im Text annotiert. Hierbei werden die folgenden zwei Fälle unterschieden:

- *Tag Analepse*: Eine Passage, durch die eine chronologische Erzählung unterbrochen wird, um von früheren Ereignissen zu berichten, wird Analepse genannt.
 - Länge der annotierten Passage: i. d. R. Teilsätze, Sätze oder Absätze
 - Indikatoren: Tempus, das Vorzeitigkeit ausdrückt (z. B. Plusquamperfekt); charakteristische Zeitausdrücke (z. B. „früher“, „vor einiger Zeit“, „drei Jahre zuvor“)
 - Beispiel: „Jetzt sah man, was geschehen war: *der Hansjörg hatte sich am mittleren Gelenk den Zeigefinger der rechten Hand abgeschossen.*“ (B. Auerbach: *Die Kriegspfeife*)
- *Tag Prolepse*: Eine Passage, durch die eine chronologische Erzählung unterbrochen wird, um Ereignisse vorwegzunehmen, die in der Zukunft stattfinden, wird Prolepse genannt.
 - Länge der annotierten Passage: i. d. R. Teilsätze, Sätze oder Absätze
 - Indikatoren: Tempus, das Zukunft ausdrückt (z. B. Futur); charakteristische Zeitausdrücke (z. B. „später“, „nach ein paar Wochen“)
 - Beispiel: „Gleich beim ersten Anblick des Hundes war er von der Zuneigung ergriffen worden, *die dauern sollte bis zu seinem letzten Atemzuge.*“ (M. von Ebner-Eschenbach: *Krambambuli*)

Dauer: Diese Kategorien dienen der Analyse des Erzähltempos, d. h. des Verhältnisses zwischen der Zeitspanne, in der die erzählten Ereignisse in der erzählten Welt stattgefunden hat, und der Zeitspanne, die das Erzählen davon dauert. (Im Falle von Dauer-Analysen, die durch illokutionäre Grenzüberschreitung zustande kommen, sollte im Rahmen von Dauer-Analysen zunächst nur die Dauer der Wiedergabe des figuralen Sprechakts annotiert werden.) In der Regel ist nach einer gründlichen Dauer-Analyse *jede* Textstelle mit einem Dauer-Tag annotiert.

- *Tag zeitraffend*: Wenn Ereignisse zusammengefasst werden, die erzählte Zeit also länger ist als die Erzählzeit, dann wird dies als zeitraffendes Erzählen bezeichnet.
 - Länge der annotierten Passage: i. d. R. Sätze oder Absätze
 - Indikatoren: Zeitausdrücke, die längere Zeitspannen zusammenfassen (z. B. „ein Jahr lang“)
 - Beispiel: „Acht Wochen habe ich in dieser Entlegenheit verlebt“ (G. Trakl: *Traumland. Eine Episode*)
- *Tag zeitdeckend*: Zeitdeckendes Erzählen liegt vor, wenn die Erzählung von Ereignissen genau so lange dauert wie die Ereignisse selbst.
 - Länge der annotierten Passage: i. d. R. Sätze oder Absätze
 - Indikatoren: zitierte Rede
 - Beispiel: „Drei! Nicht wahr?‘ ‚Ja! Erst!!‘ ‚Schön! ... Ist noch Bier da?‘ ‚Ja! Ich glaube.“ (A. Holz und J. Schlaf: *Ein Tod*)
- *Tag zeitdehnend*: Zeitdehnend ist Erzählen dann, wenn die Erzählzeit länger ist als die Ereignisse, von denen berichtet wird.
 - Länge der annotierten Passage: i. d. R. Sätze oder Absätze
 - Indikatoren: deskriptive Einschübe
 - Beispiel: „Ich trat an das Fenster und bemerkte – obgleich wir uns im zweiten Stockwerk befanden – dicht vor den Scheiben ein gewaltiges, störriges und sträubiges Roß, das mit flatternder Mähne in der Luft zu schweben schien.“ (J. von Eichendorff: *Auch ich war in Arkadien*)

Frequenz: Diese Kategorien dienen der Analyse des Verhältnisses der Häufigkeit, mit der Ereignisse in der fiktiven Welt stattgefunden haben, zu der Häufigkeit, mit der von ihnen erzählt wird. Der häufige Fall, dass ein Ereignis so oft erzählt wird, wie es stattfindet, wird nicht annotiert.

- *repetitiv*: Repetitives Erzählen liegt vor, wenn von einem Ereignis, das einmal stattfand, mehrmals berichtet wird.
 - Länge der annotierten Passage: i. d. R. Sätze oder Absätze
 - Beispiel: „Ich halte Sie nämlich für einen Philosophen [...]“; „Doktor Gudehus hält mich für einen Philosophen [...]“ (T. Mann: *Der Tod*)
- *iterativ*: Wenn ein Ereignis (bzw. Ereignisse eines Ereignistyps) mehrmals stattfand, aber nur einmal davon erzählt wird, spricht man von iterativem Erzählen.

- Länge der annotierten Passage: i. d. R. Sätze oder Absätze
- Indikatoren: charakteristische Zeitausdrücke (z. B. „jeden Morgen“, „immer wieder“)
- Beispiel: „Oftmals, wenn meine Gedanken sich wie graue Gewässer vor mir ausbreiten [...] sehe ich etwas wie den Zusammenhang der Dinge [...]“ (T. Mann: *Der Tod*)

Externe und weiterführende Links

- CATMA: <https://web.archive.org/save/https://catma.de> (Letzter Zugriff: 03.07.2024)
- Tagset „Narratologie (discours)“ auf Zenodo: <https://web.archive.org/save/https://zenodo.org/records/10519648> (Letzter Zugriff: 03.07.2024)

Bibliographie

- Bögel, Thomas, Michael Gertz, Evelyn Gius, Janina Jacke, Jan Christoph Meister, Marco Petris und Jannik Strötgen. 2015. Collaborative Text Annotation Meets Machine Learning: heureCLÉA, a Digital Heuristic of Narrative. *DHCommons Journal*, Nr. 1. doi: 10.5281/zenodo.3240591, <http://dhcommons.org/journal/issue-1/collaborative-text-annotation-meets-machine-learning-heurecl%C3%A9a-digital-heuristic> (zugegriffen: 4. November 2015).
- Brunner, Annelen. 2015. *Automatische Erkennung von Redewiedergabe: ein Beitrag zur quantitativen Narratologie*. Berlin, Boston: de Gruyter.
- forTEXT. 2020. Tagset Narratologie (discours). Zenodo, 27. Januar. doi: 10.5281/zenodo.105196488, <https://zenodo.org/records/10519648>.
- Genette, Gérard. 2007. *Discours du récit*. Points Essais 581. Paris: Éd. du Seuil.
- Gius, Evelyn. 2015. *Erzählen über Konflikte: Ein Beitrag zur digitalen Narratologie*. Bd. 46. Narratologia. Berlin; Boston: De Gruyter.
- Gius, Evelyn und Janina Jacke. 2016. Zur Annotation narratologischer Kategorien der Zeit. Guidelines zur Nutzung des CATMA-Tagsets. <http://heureclea.de/wp-content/uploads/2016/11/guidelinesV2.pdf>.
- Herman, David, Hrsg. 1999. *Narratologies: New Perspectives on Narrative Analysis*. Columbus, Ohio: Ohio State University Press.
- Horstmann, Jan und Jan-Erik Stange. 2024. Methodenbeitrag: Textvisualisierung. Hg. von Evelyn Gius. *forTEXT* 1, Nr. 5. Textvisualisierung (7. August). doi: 10.48694/fortext.3772, <https://fortext.net/routinen/methoden/textvisualisierung>.
- Jacke, Janina. 2024. Methodenbeitrag: Manuelle Annotation. Hg. von Evelyn Gius. *forTEXT* 1, Nr. 4. Manuelle Annotation (7. August). doi: 10.48694/fortext.3748, <https://fortext.net/routinen/methoden/manuelle-annotation>.
- Kindt, Tom und Hans-Harald Müller. 2003. Wieviel Interpretation enthalten Beschreibungen? Überlegungen zu einer umstrittenen Unterscheidung am Beispiel der Narratologie. In: *Regeln der Bedeutung. Zur Theorie der Bedeutung literarischer Texte*, hg. von Fotis Jannidis, Gerhard Lauer, Matias Martinez, und Simone Winko, 286–304. Berlin, New York: de Gruyter.
- Kuhn, Markus. 2011. *Filmnarratologie: ein erzähltheoretisches Analysemodell*. Berlin, New York: de Gruyter.
- Mani, Inderjeet. 2013. Computational Narratology. In: *the living handbook of narratology*, hg. von Peter Hühn, Jan Christoph Meister, John Pier, und Wolf Schmid. Hamburg: Hamburg University. <https://www-archiv.fdm.uni-hamburg.de/lhn/node/43.html> (zugegriffen: 15. Januar 2020).
- Meister, Jan Christoph. 2014. Narratology. In: *the living handbook of narratology*, hg. von Peter Hühn, Jan Christoph Meister, John Pier, und Wolf Schmid. Hamburg: Hamburg University. <http://www.lhn.uni-hamburg.de/article/narratology> (zugegriffen: 24. November 2017).
- Modrow, Lena. 2016. *Wie Songs erzählen. Eine computergestützte, intermediale Analyse der Narrativität*. Frankfurt am Main: Peter Lang.
- Ryan, Marie-Laure. 1991. *Possible Worlds, Artificial Intelligence, and Narrative Theory*. Bloomington: Indiana University Press.
- 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>.
- Thon, Jan-Noël. 2016. *Transmedial narratology and contemporary media culture*. Lincoln, London: University of Nebraska Press.

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.

Feature Unter Features können Einzelfunktionen eines Tools verstanden werden, die beispielsweise komplexe Funktionen wie die Visualisierung eines Textes als **Wordcloud** ermöglichen, oder auch kleinere Funktionseinheiten wie den Abgleich einzelner Spracheigenschaften (**Properties**) mit **annotierten** Beispieltexten darstellen.

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.

Korpus Ein Textkorpus ist eine Sammlung von Texten. Korpora (Plural für „das Korpus“) sind typischerweise nach Textsorte, Epoche, Sprache oder Autor*in zusammengestellt.

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 (Textauszeichnung) Die Textauszeichnung (eng. *Markup*) fällt in den Bereich der Daten- bzw. Textverarbeitung, genauer in das Gebiet der Textformatierung, welche durch **Auszeichnungssprachen** wie **XML** implementiert wird. Dabei geht es um die Beschreibung, wie einzelne Elemente eines Textes beispielsweise auf Webseiten grafisch dargestellt werden sollen.

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 of 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**.

Property Property steht für „Eigenschaft“, „Komponente“ oder „Attribut“. In der automatischen **Annotation** dienen konkrete Wortheigenschaften wie Groß- und Kleinschreibung zur Klassifizierung von Wörtern oder Phrasen. Durch die Berücksichtigung solcher Eigenschaften in den **Features** eines Tools kann

maschinelles Lernen bestimmter Phänomene umgesetzt werden. In der manuellen Annotation können als Properties auch Eigenschaften von **Annotationen** benannt werden.

- 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.
- 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.
- TEI** Die *Text Encoding Initiative* (TEI) ist ein Konsortium, das gemeinsam einen Standard für die Darstellung von Texten in digitaler Form entwickelt. Die TEI bietet beispielsweise Standards zur Kodierung von gedruckten Werken und zur Auszeichnung von sprachlichen Informationen in maschinenlesbaren Texten (siehe auch **XML** und **Markup**).
- 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.
- XML** XML steht für *Extensible Markup Language* und ist eine Form von **Markup Language**, die sowohl computer- als auch menschenlesbar und hochgradig anpassbar ist. Dabei werden Textdateien hierarchisch strukturiert dargestellt und Zusatzinformationen i. d. R. in einer anderen Farbe als der eigentliche (schwarz gedruckte) Text dargestellt. Eine standardisierte Form von XML ist das **TEI-XML**.