

Methodenbeitrag: Entwicklung von Kategoriensystemen

Janina Jacke  ¹

Dominik Gerstorfer  ¹

1. Technische Universität Darmstadt

forTEXT

Thema:	Projektkonzeption	DOI:	10.48694/fortext.3805
Jahrgang:	1	Ausgabe:	12
Erscheinungsdatum:	30-11-2024	Erstveröffentlichung:	2021-01-11 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. Definition

Unter der Entwicklung von Kategoriensystemen ist die Erstellung einer terminologischen Ordnungssystematik zur Erfassung eines Gegenstandsbereichs zu verstehen. Diese kann unterschiedliche Formen annehmen, die jeweils unterschiedlichen formalen Vorgaben genügen bzw. unterschiedliche Kriterien zur Entwicklung und Auswahl der Kategorien ansetzen. Ontologien, Taxonomien, Typologien und kontrollierte Vokabulare sind häufig verwendete Typen von Kategoriensystemen, die sich, abhängig vom Anwendungszweck und den beteiligten wissenschaftlichen Disziplinen, hinsichtlich ihrer Definitionen und Abgrenzungen voneinander unterscheiden. In textbasierten geisteswissenschaftlichen Disziplinen wie der Literaturwissenschaft können Kategoriensysteme z. B. der Klassifikation ganzer Werke oder der Kategorisierung einzelner Textphänomene dienen. In den digitalen Geisteswissenschaften finden Kategoriensysteme häufig als Annotationsschemata bzw. **Tagsets** im Rahmen (manueller) **Annotation** Verwendung (Jacke 2024a).

2. Anwendungsbeispiel

Angenommen, Sie möchten untersuchen, wie weibliche Hauptfiguren im zeitgenössischen Roman präsentiert werden. Um eine differenzierte und anschlussfähige Analyse durchzuführen, können Sie Kategorien für die Klassifikation von Textstellen entwickeln, die für die Charakterisierung von Protagonistinnen relevant sind. Hierfür greifen Sie möglicherweise zunächst auf allgemeine Kategorien für die Figurenanalyse zurück, die in der Literaturwissenschaft entwickelt worden sind (Jannidis 2012). Hier finden sich z. B. Kategorien für die Analyse der Art und Weise, wie auf Figuren Bezug genommen wird, sowie für die Analyse von Varianten impliziter und expliziter Charakterisierung. Auf Basis eines solchen Kategoriensystems erstellen Sie in einem geeigneten Annotationsprogramm ein Tagset, das Sie für die Annotation und Analyse ausgewählter zeitgenössischer Romane nutzen können.

Im Rahmen der Analyse könnte sich herausstellen, dass Sie für Ihre spezifische Fragestellung an einigen Stellen gern genauere Differenzierungen vornehmen möchten, als es mit den bereits existierenden Kategorien möglich ist. Sie fügen Ihrem Tagset deswegen an geeigneten Stellen weitere Kategorien hinzu. Bei der Entwicklung neuer Kategorien greifen Sie – neben Beobachtungen am Text – auf verschiedene Theorien zurück, hier beispielsweise auf Gendertheorien oder linguistische Differenzierungen. So könnten Sie es für fruchtbar halten, bei expliziter Charakterisierung einer Protagonistin die Typen von Adjektiven (prädikative, adverbiale und attributive) zu unterscheiden, die für die Beschreibung verwendet werden. Die Auswahl der Kategorien beruht unter anderem auf Ihrer Einschätzung darüber, welche Unterscheidungen im Rahmen einer Textinterpretation relevant sein könnten. Die Kriterien, die die Kategorien Ihres Kategoriensystems definieren, sowie die Beziehungen zwischen den Kategorien dokumentieren Sie sorgfältig.

Durch die Entwicklung eines für Ihre Fragestellung relevanten Kategoriensystems haben Sie einen systematischen Überblick darüber gewonnen, in welcher Form weibliche Hauptfiguren grundsätzlich dargestellt werden können und welchem Verhältnis diese Formen zueinander stehen. Durch die Nutzung eines Kategoriensystems für die Literaturanalyse erhalten Sie pointierte und nachvollziehbare Ergebnisse hinsichtlich der Frage, wie Protagonistinnen im zeitgenössischen Roman dargestellt werden.

3. Literaturwissenschaftliche Tradition

Die eingangs genannten Typen von Kategoriensystemen sind in der Literaturwissenschaft unterschiedlich bekannt bzw. gebräuchlich. So sind beispielsweise sowohl das Konzept als auch der Terminus „Ontologie“ der Literaturwissenschaft weitgehend fremd – im *Reallexikon der deutschen Literaturwissenschaft* (Braungart u. a. 2007) beispielsweise findet sich kein entsprechender Eintrag.

Die Begriffe „Ontologie“ und „Kategorie“ stammen aus der Philosophie und lassen sich bis zu Aristoteles zurückverfolgen. „Ontologie“ bezeichnet die Lehre vom Seienden (*Met.* Δ 7), d. h. die Bestimmung der grundlegendsten Bestandteile der Welt.

Hinweis: Aristoteles wird, wie üblich, nach Bekker-Zählung zitiert. Die entsprechende Paginierung findet sich auch in modernen Ausgaben der *Kategorien* (Aristoteles 2019a) und der *Metaphysik* (Aristoteles 2019b). Das Ziel der philosophischen Ontologie ist es, festzustellen, welche Entitäten es gibt und in welchen Verhältnissen sie zueinander stehen. Um dieses Ziel zu erreichen, entwickelt Aristoteles ein System allgemeinsten Aussagen (*Kat.* 4, 1b25 ff.), welche die Entitäten, deren Eigenschaften und Relationen beschreiben, Aristoteles nennt diese Aussagen „Kategorien“. Aus der philosophischen Tradition der metaphysischen Ontologie heraus hat sich die *formale* Ontologie (Husserl 1929) entwickelt, der es nicht mehr um die grundlegende Beschaffenheit der Wirklichkeit an sich geht, sondern um die Beziehungen zwischen Gegenständen und deren Teilen. Dies geschieht in Analogie zur formalen Logik: So, wie die formale Logik allgemeine Sätze untersucht und nicht etwa spezifische, untersucht die formale Ontologie allgemeine Eigenschaften von Gegenständen (Smith und Mulligan 1983).

Die Bezeichnung „Ontologie“ wird in den 1990er Jahren schließlich von der Informatik übernommen (Studer und Sure-Vetter 2019; Busse u. a. 2014). Dort bezeichnet der Ausdruck formale Repräsentationssysteme, die der Erfassung und Verarbeitung von Wissen in einem bestimmten Gegenstandsbereich dienen. Ontologien finden hier beispielsweise im Bereich der Künstlichen Intelligenz Verwendung: Wissen über einen Gegenstandsbereich wird in Softwaresystemen gespeichert, die dieses Wissen eigenständig zur Problemlösung anwenden können (Gruber 2018). Konstitutiv für eine Ontologie in Abgrenzung zu anderen Kategoriensystemen sind vor allem formale Spezifiziertheit, dies gilt sowohl für das verwendete Vokabular als auch für die unterschiedlichen Elemente von Ontologien, die in einer sogenannten *top level ontology* oder *upper ontology* festgehalten werden (z. B. Objekt, Klasse, Eigenschaft, Relation, Funktion und Axiom), und die Möglichkeit, aufgrund der festgelegten Relationen, zwischen den Kategorien Schlussfolgerungen zu ziehen. Unter „Ontologie“ wird in technischen Zusammenhängen auch oft die jeweilige spezifische Implementation einer Wissensrepräsentation verstanden. Der Ausdruck bezeichnet dann die jeweilige Kombination aus formaler Beschreibungssprache und Schemadefinitionen (zum Beispiel XML und XDS oder RDF und RDFS), die zur Kategorisierung einer bestimmten Domäne verwendet wird (Hesse 2002). Gegliederte Kategoriensysteme sind in literaturwissenschaftlichen Kontexten eher unter den Begriffen „Taxonomie“ oder „Typologie“ bekannt. De facto werden die Bezeichnungen „Taxonomie“ und „Typologie“ in der Literaturwissenschaft oft synonym und unspezifisch für (in der Regel hierarchisch gegliederte) Kategoriensysteme verwendet. Einer möglichen spezifischeren Verwendungsweise zufolge können Taxonomien als Ordnungssysteme verstanden werden, deren Klassen sich auf der Basis empirisch auftretender Ähnlichkeiten und Unterschiede im untersuchten Gegenstandsbereich ergeben. Bei Typologien dagegen werden die einzelnen Kategorien aufgrund typischer Merkmalskombinationen oder durch theoriegeleitete Auswahl der Kriterien gebildet, die eine Kategorie auszeichnen (Bailey 1994, 4–6: 34). Letztlich können alle Kategoriensysteme als spezielle Ontologien konstruiert werden.

Ein Bereich, in dem Klassifikationssysteme eine wichtige Rolle in der Literaturwissenschaft spielen, ist die *Systematisierung literarischer Texte*, beispielsweise im Kontext von Gattungs- und Genretheorie (Lamping 2009; Zymner 2010). Eine Entsprechung in der Linguistik ist die Klassifikation von Texten nach *Textsorte* (Adamzik 2007). In der Bibliotheksklassifikation werden Klassifikationssysteme zur Erfassung und Verwaltung (literarischer) Texte auf eine Weise verwendet, die der typisch informatischen Verwendungsweise recht nahe kommt (Alex, Bee und Junger 2018). Schwierigkeiten stellen sich u. a. oft hinsichtlich der Frage, ob Mehrfachzuordnungen einzelner Texte zu Klassen erlaubt bzw. wie diese konzipiert sein sollte, sowie hinsichtlich der Frage, ob die Zugehörigkeit zu einer Klasse durch mehrere Kriterien bestimmt ist und, wenn ja, ob und wie diese Kriterien in einer Ontologie hierarchisch geordnet sind (Jansen 2008).

Neben der Textklassifikation finden Ontologien in geisteswissenschaftlichen Kontexten auch in der *Textanalyse* Verwendung, insbesondere in der (strukturalistischen) Linguistik (Nickel 1969). In der Literaturwissenschaft wird im Rahmen von Textanalyse vor allem dort mit Kategoriensystemen gearbeitet, wo vornehmlich die *Form* literarischer Texte im Fokus steht, beispielsweise in der formalen Gedichtanalyse oder im Rahmen der Narratologie (Genette u. a. 2010). Hier existieren oft detaillierte hierarchisch gegliederte Begriffssysteme, die der Kategorisierung einzelner Textelemente dienen und so eine genaue Textbeschreibung ermöglichen, die dann wiederum für die Interpretation genutzt werden kann. So ließen sich beispielsweise Genettes Kategorien (20190) für die Analyse der Erzählweise, die Unterkategorien zur Erfassung von Erzählzeit, -modus und -stimme beinhalten, als Typologie bezeichnen.

Werden in der Literaturwissenschaft Kategoriensysteme für die Textanalyse erarbeitet, dann geschieht dies in der Regel im Rahmen der Textanalyse selbst in einem Bottom-up- oder einem Middle-out-Verfahren (Fellmann 2013): Literaturwissenschaftler*innen stoßen auf Einzelphänomene, für deren Einordnung sie zunächst spe-

zifische (oder semi-spezifische) Kategorien entwickeln, um dem System dann darauf aufbauend schrittweise allgemeinere Kategorien (oder teilweise auch weitere Unterdifferenzierungen) hinzufügen. Top-down-Verfahren kommen insbesondere dann ins Spiel, wenn bestehende allgemeine Kategoriensystementwürfe – teilweise auch aus anderen Disziplinen – oder andere Theorien herangezogen werden, um diese im Rahmen der Textanalyse zu verfeinern bzw. zu operationalisieren.

Allgemein ist die Arbeit an und mit Kategoriensystemen zum Zweck der Textanalyse in der traditionellen Literaturwissenschaft eher in formalistisch-strukturalistisch orientierten Teilbereichen verbreitet. Eine theoretische Reflexion der Kriterien, die fruchtbare literaturwissenschaftliche Kategoriensysteme erfüllen sollten, sowie eine methodologische Diskussion des Vorgehens bei der Entwicklung solcher Systeme findet in der nicht-digitalen Literaturwissenschaft ausschließlich in wenigen Arbeiten statt, die sich der analytischen Literaturwissenschaft zuordnen lassen (Strube 1993).

4. Diskussion

Generell bietet die Arbeit an und mit Kategoriensystemen auch in geisteswissenschaftlichen Kontexten eine Reihe von Vorteilen – unabhängig davon, ob sie digital oder nicht-digital durchgeführt wird. So ist es bei der Entwicklung eines Kategoriensystems notwendig, den relevanten Gegenstandsbereich sehr genau zu reflektieren: Welche Arten von Phänomenen existieren und welche Gemeinsamkeiten und Unterschiede bestehen zwischen diesen Phänomenarten? Diese Arbeit kann nicht nur auf einer allgemeinen Ebene interessante systematisierende Aussagen über einen Gegenstandsbereich ermöglichen. Sie hilft auch dabei, Einzelaussagen (z. B. über konkrete literarische Texte) in einen größeren Zusammenhang zu setzen, da beispielsweise deutlich wird, welche alternativen Möglichkeiten nicht umgesetzt wurden bzw. vorliegen.

Wenn Ontologieentwicklung sorgfältig betrieben wird, dann befördert sie in unterschiedlichen Hinsichten ein geschärftes Verständnis des jeweiligen Gegenstandsbereichs: Für identifizierte Kategorien werden Definitionen dokumentiert, die die geisteswissenschaftliche Terminologie explizieren und somit einen pointierten wissenschaftlichen Austausch ermöglichen (Fricke 1977, 255–265). Definitionen oder Erläuterungen zu den Kategorien eines Kategoriensystems können darüber hinaus auch Hinweise zur Operationalisierung der Kategorien enthalten, d. h. Angaben dazu, wie die Zugehörigkeit eines konkreten Gegenstands zu einer Kategorie festgestellt werden kann (Moretti 2013; Reiter, Willand und Gius 2019).

Wenn das Kategoriensystem mit dem Ziel entwickelt werden, *formalisierte* Kategorien für einen Gegenstandsbereich zu erarbeiten, schärft und vereinfacht dies den wissenschaftlichen Austausch in noch stärkerem Maße (Munn und Smith 2008). Auf diese Weise ist es möglich, der Polysemie wissenschaftlicher Termini entgegenzuwirken, indem beispielsweise unterschiedliche Verwendungsweisen eines Ausdrucks gesondert expliziert und in das Begriffssystem eingeordnet werden.

Werden Kategoriensysteme als Tagsets bzw. Annotationsschemata im Rahmen (manueller) Annotation (Jacke 2024a) für die Textanalyse verwendet, so sind damit oft ebenfalls einige Vorteile verbunden – beispielsweise sehr textnahe Arbeiten, eine gute Dokumentation der Analysen, intersubjektiv nachvollziehbare Kategorisierungen und anschlussfähige Ergebnisse. Wenn die Relationen zwischen den einzelnen Kategorien im Rahmen des für die Annotation verwendeten Kategoriensystems außerdem sorgfältig bestimmt werden, dann können auf Basis des Systems und der Annotationen leichter weitere Schlussfolgerungen hinsichtlich der annotierten Textpassagen gezogen werden.

Tatsächlich stellen sich bei der Arbeit an und mit Kategoriensystemen in geisteswissenschaftlichen (insbesondere in literaturwissenschaftlichen) Kontexten allerdings einige Probleme und Herausforderungen. An den wenigen Stellen, an denen in der Literaturwissenschaft Kategoriensysteme entwickelt und verwendet werden, geschieht dies oft ohne Angabe klarer Definitionen und ohne eine genaue Spezifizierung der Beziehungen zwischen den Kategorien. Hier ist zum einen meist unklar, ob die festgelegten Kategorien wechselseitig exklusiv sind (d. h. ein Gegenstand kann nur einer der Kategorien angehören) bzw. ob sie gemeinsam exhaustiv sind (d. h. jeder Gegenstand fällt in mindestens eine der Kategorien). Zum anderen wird in Kategoriensystemen meist ausschließlich mit hierarchischen Strukturen gearbeitet, die eine Subsumtionsbeziehung implizieren (d. h. eine Unterklasse übernimmt alle Merkmale der Oberklasse) – dabei ist jedoch nicht immer klar, ob diese Subsumtionssemantik auch wirklich umgesetzt ist. Dies führt dazu, dass unter Rückgriff auf das Klassifikationssystem nur sehr eingeschränkt weitere Schlüsse (z. B. über annotierte Textpassagen) gezogen werden können. Darüber hinaus wird auch die gemeinsame Arbeit an standardisierten Klassifikationssystemen kaum verfolgt – stattdessen werden solche Systeme meist ohne Allgemeinheitsanspruch für einzelne Anwendungskontexte entwickelt.

Letztlich sind die Probleme im Zusammenhang mit Kategorienentwicklung in geisteswissenschaftlicher Forschung jedoch auf einer noch tieferen Ebene angesiedelt: Da in den einzelnen geisteswissenschaftlichen Disziplinen (wie der Literaturwissenschaft) kaum eine theoretisch-methodologische Auseinandersetzung mit Kategorienbildung und -systemen stattfindet, ist unklar, welche der Gütekriterien für Kategoriensysteme, die für andere Kontexte entwickelt worden sind, sich auf den literaturwissenschaftlichen Bereich übertragen lassen (Strube 1993, 56).

Hier ist erstens zu bemerken, dass die Formalisierung des Gegenstandsbereichs *Literatur* noch nicht so weit

fortgeschritten ist, wie es in vielen Naturwissenschaften für die relevanten Domänen der Fall ist. Deswegen ist es de facto unwahrscheinlich, dass für die Literaturwissenschaft in ihrer Gesamtheit Ontologien entwickelt werden, die gegenseitig exklusive und zusammen exhaustive Kategorien enthalten.

Zweitens existiert in der Literaturwissenschaft eine außergewöhnliche Fülle unterschiedlicher Perspektiven auf und teilweise inkompatibler Theorien über Literatur (Köppe und Winko 2008). Deswegen ist schwer vorstellbar, wie es möglich sein kann, eine konsistente und standardisierte Ontologie für Literatur zu erstellen. Monohierarchische Strukturen (d. h. Strukturen, in der jede Unterklasse nur eine direkte Oberklasse haben kann), wie sie für die Naturwissenschaften oft als vorbildlich erachtet werden (Arp, Smith und Spear 2015), scheinen kein geeignetes System für die Abbildung einer derartigen Vielfalt zu sein.

Drittens scheint Literatur grundsätzlich ein weniger klar einzuordnendes Phänomen zu sein als naturwissenschaftliche Gegenstände – einige Literaturtheorien erlauben deshalb beispielsweise explizit auch widersprüchliche Textanalysen oder Interpretationen (z. B. Dekonstruktion). Zumindest aus der Perspektive solcher Theorien stellt sich die Frage, ob literaturwissenschaftliche Kategoriensysteme beispielsweise auch Inkonsistenzen, Redundanzen oder Zirkelstrukturen enthalten dürfen sollten – dies würde allerdings dazu führen, dass kaum Schlussfolgerungen unter Rückgriff auf solche Kategoriensysteme möglich sind.

Generell ist außerdem noch unklar, für welche Arten literaturwissenschaftlich relevanter Fragestellungen die Entwicklung von Kategoriensystemen möglich und sinnvoll ist. Während Kategorien bisher für die Systematisierung literarischer Texte und für bestimmte Aspekte der Textanalyse genutzt werden, wäre noch zu ergründen, ob bzw. inwiefern und unter welchen Umständen Kategorisierungen auch im Zusammenhang mit Interpretation fruchtbar sein können.

Neben diesen theoretischen Fragen zu Kategoriensystemen in den Geisteswissenschaften gibt es auch praktisch-pragmatische Fragen – hier wird das Thema der digitalen Unterstützung relevant. Während Kategorienentwicklung auch in nicht-digitalen Kontexten betrieben werden kann, ist sie mit digitaler Unterstützung einfacher und für den Einsatz in digitalen Projekten stärker verbreitet. Anstatt Kategoriensysteme mit Stift und Papier zu entwerfen, können dafür auf der einfachsten Ebene Tabellenprogramme wie Excel oder Diagrammprogramme wie draw.io genutzt werden.

Nutzerfreundlicher (bzw. in Eingabe und Darstellung besser angepasst an diese Aufgabe) sind aber meist Annotationstools, von denen viele die freie Erstellung von Kategoriensystemen unterstützen (z. B. CATMA (Schumacher 2024) oder WebAnno (Schumacher und Bläß 2024)). Da außerdem die Entwicklung solcher Systeme in geisteswissenschaftlichen Zusammenhängen meist im Zuge von Annotation stattfindet bzw. dort maßgeblich modifiziert und weiterentwickelt wird (Gius und Jacke 2017), ist es von Vorteil, wenn Annotation und Arbeit an dem Kategoriensystem nahtlos in derselben Umgebung möglich sind. Allerdings könnten einige Organisations- und Bearbeitungsfunktionen für Kategoriensysteme in Annotationstools noch im Hinblick auf Flexibilität optimiert werden, indem Nutzer*innen beispielsweise dabei unterstützt werden, unstrukturierte Beobachtungen zum Text schrittweise in Kategoriensysteme auszuarbeiten (Horstmann und Jacke 2020), oder es ermöglicht wird, die Hierarchie von Kategorien zu verändern oder weitere Typen von Beziehungen hinzuzufügen.

Die komplexesten Möglichkeiten zur digital gestützten Erstellung von Kategoriensystemen in Form von Ontologien ermöglicht das Tool Protégé (Jacke 2024b). Hier können auch umfangreichere Beschreibungen von Kategorien (beispielsweise Definition und Hinweise zur Operationalisierung) eingefügt sowie unterschiedliche Relationen festgelegt werden. Dadurch bietet Protégé auch die Möglichkeit, Schlüsse auf Basis von Ontologien zu ziehen. Protégé ermöglicht selbst nicht die Anwendung von Ontologien im Rahmen von Annotation – Annotationstools, in denen Nutzer*innen die mit Protégé erstellten Ontologien für die Textauszeichnung verwenden können, sind nur eingeschränkt verfügbar (Corcho 2006). Stattdessen wird Protégé vor allem zur Erstellung von Ontologien für Wissensmanagement genutzt.

In der digitalen Literaturwissenschaft lassen sich erste Ansätze einer theoretisch-methodologischen Reflexion von Kategorienentwicklung feststellen. Beispielhaft zu nennen ist hier ein kürzlich erschienener, von Reiter, Willand und Gius (2019) herausgegebener Band, der den Fokus allerdings stärker auf die Erstellung von Operationalisierungshinweisen („Guidelines“) zur Anwendung von Kategoriensystemen in der Textanalyse setzt und nicht auf die Ontologieentwicklung selbst (Reiter, Willand und Gius 2019). Generell scheinen die digitalen Geisteswissenschaften ein Feld zu sein, in dem derartige Reflexionen durch das interdisziplinäre Arbeiten befördert werden, indem unterschiedliche disziplinspezifische Annahmen, Theorien, Methoden und Praxen aufeinandertreffen und ausgehandelt werden müssen.

5. Technische Grundlagen

Technische Kenntnisse im engeren Sinne sind keine notwendige Voraussetzung, um Kategorienentwicklung betreiben und mit Kategorien arbeiten zu können. Grundlegende wissenschaftstheoretische bzw. logische Kenntnisse sind allerdings von Vorteil, um beispielsweise die Grundvoraussetzungen fruchtbarer Definitionen (Gupta 2015; Carnap 1929; Pawłowski 1980) und die logischen Eigenschaften verschiedener Relationen zwischen Kategorien (z. B. Reflexivität, Symmetrie und Transitivität) zu kennen oder um einschätzen zu können, welche Implikationen Inkonsistenzen in Kategoriensystemen bzw. Ontologien haben.

Bei der Arbeit an Kategoriensystemen, die inhaltlich und technisch standardisiert (oder zumindest anschlussfä-

hig) sein sollen, sollten Sie darauf achten, dass die Kategoriensysteme in einem für die intendierte Verwendung passenden Format erstellt werden. Vor diesem Hintergrund ist es sinnvoll, ein für die Textannotation vorgesehenes Kategoriensystem bereits – sofern dies möglich ist – in dem Annotationsprogramm zu erstellen, mit welchem Sie die Annotation vornehmen möchten. Wenn Sie das Kategoriensystem in einem anderen Programm erstellen möchten, dann prüfen Sie vorab, in welchen Formaten Sie es exportieren können und ob das Annotationsprogramm Ihrer Wahl eines dieser Formate für den Datenimport unterstützt.

Wenn Sie eine komplexe Ontologie in Protégé erstellen möchten (das nicht primär auf Annotationsontologien, sondern auf Wissenserfassung und -verarbeitung ausgelegt ist), so sind Kenntnisse der „Web Ontology Language“ (OWL) hilfreich bzw. notwendig (Smith, Welty und McGuinness 2004).

Bibliographie

- Adamzik, Kirsten. 2007. *Textsorten. Reflexionen und Analysen*. Tübingen: Stauffenburg-Verlag.
- Alex, Heidrun, Guido Bee und Ulrike Junger, Hrsg. 2018. *Klassifikationen in Bibliotheken. Theorie – Anwendung – Nutzen*. Berlin (u.a.): de Gruyter Saur.
- Aristoteles. 2019a. *Aristoteles. Philosophische Schriften in sechs Bänden*. Bd. 1. 6 Bde. Hamburg: Felix Meiner.
- . 2019b. *Aristoteles. Philosophische Schriften in sechs Bänden*. Bd. 5. 6 Bde. Hamburg: Felix Meiner.
- Arp, Robert, Barry Smith und Andrew Spear. 2015. *Building ontologies with Basic Formal Ontology*. Cambridge, Mass.: MIT Press.
- Bailey, Kenneth D. 1994. *Typologies and taxonomies: An Introduction to Classification Techniques*. Bd. 102. Quantitative Applications in the Social Sciences. SAGE Publications.
- Braungart, Georg, Harald Fricke, Klaus Grubmüller, Jan-Dirk Müller, Friedrich Vollhardt und Klaus Weimar, Hrsg. 2007. *Reallexikon der deutschen Literaturwissenschaft*. Bd. 2: H–O. Berlin: de Gruyter.
- Busse, Johannes, Bernhard Humm, Christoph Lübbert, Frank Moelter, Anatol Reibold, Matthias Rewald, Veronika Schlüter, Bernhard Seiler, Erwin Tegtmeier und Thomas Zeh. 2014. Was bedeutet eigentlich Ontologie?: Ein Begriff aus der Philosophie im Licht verschiedener Disziplinen. *Informatik-Spektrum* 37, Nr. 4: 286–97.
- Carnap, Rudolf. 1929. *Abriss der Logistik Mit Besonderer Berücksichtigung der Relationstheorie und Ihrer Anwendungen*. Wien: Springer.
- Corcho, Oscar. 2006. Ontology Based Document Annotation: Trends and Open Research Problems. *International Journal of Metadata Semantics and Ontologies* 1, Nr. 1: 47–57. doi: 10.1504/IJMSO.2006.008769..
- Fellmann, Michael. 2013. Semantisches Prozessmanagement und E-Business. http://www.mfellmann.net/content/slides/V06_OWL_SPARQL.pdf (zugegriffen: 18. Dezember 2020).
- Fricke, Harald. 1977. *Die Sprache der Literaturwissenschaft*. München: Beck.
- Genette, Gérard, Andreas Knop, Jochen Vogt und Isabel Kranz. 2010. *Die Erzählung*. 3., durchges. und korrigierte Aufl. UTB Literatur- und Sprachwissenschaft 8083. Paderborn: Fink.
- Gius, Evelyn und Janina Jacke. 2017. The Hermeneutic Profit of Annotation: On Preventing and Fostering Disagreement in Literary Analysis. *International Journal of Humanities and Arts Computing* 11, Nr. 2 (Oktober): 233–254. doi: 10.3366/ijhac.2017.0194, <https://www.eupublishing.com/doi/10.3366/ijhac.2017.0194> (zugegriffen: 4. September 2020).
- Gruber, Tom. 2018. Ontology. In: *Encyclopedia of Database Systems*, hg. von Ling Liu und M. Tamer Özsu, 2574–2576. New York: Springer.
- Gupta, Anil. 2015. Definitions. In: *The Stanford Encyclopedia of Philosophy*, hg. von Edward N. Zalta. <https://plato.stanford.edu/archives/win2019/entries/definitions/> (zugegriffen: 18. Dezember 2020).
- Hesse, Wolfgang. 2002. Ontologie(n). *Informatik Spektrum* 25, Nr. 6: 477–480. doi: 10.1007/s002870200265..
- Horstmann, Jan und Janina Jacke. 2020. Interpretationsspielräume. Undogmatisches Annotieren in CATMA 6. In: *DHd 2020 Spielräume: Digital Humanities zwischen Modellierung und Interpretation. Konferenzabstracts.*, hg. von Christof Schöch, 154–158. Paderborn, 2–6. März.
- Husserl, Edmund. 1929. *Formale und transzendente Logik*. Halle: Niemeyer.
- Jacke, Janina. 2024a. 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>.
- . 2024b. Toolbeitrag: Protégé. Hg. von Evelyn Gius. *forTEXT* 1, Nr. 10. Projektkonzeption (29. November). doi: 10.48694/fortext.3806, <https://fortext.net/tools/tools/protege>.
- Jannidis, Fotis. 2012. Character. 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/character> (zugegriffen: 8. Mai 2018).
- Jansen, Ludger. 2008. Classifications. In: *Applied Ontology: An Introduction*, hg. von Katherine Munn und Barry Smith, 159–172. Frankfurt: Ontos.
- Köppe, Tilmann und Simone Winko. 2008. *Neuere Literaturtheorien. Eine Einführung*. Stuttgart (u.a.): Metzler.
- Lamping, Dieter, Hrsg. 2009. *Handbuch der literarischen Gattungen*. Stuttgart: Alfred Kröner Verlag.
- Moretti, Franco. 2013. „Operationalizing“: Or, the Function of Measurement in Literary Theory. *Stanford Literary Lab Pamphlets* 2. <https://litlab.stanford.edu/LiteraryLabPamphlet6.pdf> (zugegriffen: 18. Dezember 2020).

- Munn, Katherine und Barry Smith. 2008. *Applied Ontology: An Introduction*. Frankfurt: Ontos.
- Nickel, Gerhard. 1969. Geschichte und Leistung des Taxonomischen Strukturalismus. *Zeitschrift für Dialektologie und Linguistik* 36, Nr. 1: 2–18.
- Pawłowski, Tadeusz. 1980. *Begriffsbildung und Definition*. Berlin (u.a.): De Gruyter.
- Reiter, Nils, Marcus Willand und Evelyn Gius. 2019. A Shared Task for the Digital Humanities Chapter 1: Introduction to Annotation, Narrative Levels and Shared Tasks. *Journal of Cultural Analytics*. doi: 10.22148/16.048, <https://doi.org/10.22148/16.048> (zugegriffen: 12. November 2019).
- 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>.
- Schumacher, Mareike und Sandra Bläß. 2024. Toolbeitrag: WebAnno. Hg. von Evelyn Gius. *forTEXT* 1, Nr. 4. Manuelle Annotation (7. August). doi: 10.48694/fortext.3764, <https://fortext.net/tools/tools/webanno>.
- Smith, Barry und Kevin Mulligan. 1983. Framework for Formal Ontology. *Topoi* 2, Nr. 1: 73–85.
- Smith, Michael K., Chris Welty und Deborah L. McGuinness. 2004. OWL Web Ontology Language Guide. *W3C Recommendation*. <https://www.w3.org/TR/owl-guide/> (zugegriffen: 18. Dezember 2020).
- Strube, Werner. 1993. Die Klassifikation literarischer Texte. In: *Analytische Philosophie der Literaturwissenschaft*, 41–66. Paderborn: Schöningh.
- Studer, Rudi und York Sure-Vetter. 2019. Ontologien. In: *Enzyklopädie der Wirtschaftsinformatik. Online-Lexikon*. <https://enzyklopaedie-der-wirtschaftsinformatik.de/lexikon/daten-wissen/Wissensmanagement/Wissensmodellierung/Wissensrepräsentation/Semantisches-Netz/Ontologien> (zugegriffen: 2. April 2020).
- Zymner, Rüdiger, Hrsg. 2010. *Handbuch Gattungstheorie*. Stuttgart (u.a.): Metzler.

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.
- 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.
- 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**.
- 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.