

Platinstandard-Annotation in der digitalen Literaturwissenschaft: Definition, Funktionen und diskursive Argumentvisualisierung als Best-Practice-Beispiel

Jacke, Janina

jacke@ndl-medien.uni-kiel.de

Christian-Albrechts-Universität zu Kiel, Deutschland

ORCID: 0000-0001-7217-3136

Einleitung

Dieser theoretisch-methodologische Beitrag skizziert das Konzept und die Erstellung von Platinstandard-Annotationsdaten im Zusammenhang mit komplexen literaturwissenschaftlichen Phänomenen. Den Überlegungen liegt die Idee zugrunde, dass bei der Erstellung von Annotationsdaten in der digitalen Literaturwissenschaft je nach dem primären Verwendungszweck unterschiedliche Standards angesetzt werden sollten. Im Folgenden werde ich zunächst kurz das traditionelle Konzept der Goldstandard-Daten vorstellen und potenzielle Probleme aufzeigen (Abschnitt 2). Im Anschluss wird das Konzept des Platinstandards eingeführt, das andere und striktere Qualitätskontrollen für Annotationsdaten vorsieht und dadurch hermeneutische literaturwissenschaftliche Belange stärker berücksichtigt (Abschnitt 3). Abschließend wird ein Best-Practice-Beispiel für diesen Ansatz aus dem Projekt CAUTION präsentiert, das sich mit der Annotation des interpretationsabhängigen Phänomens des unzuverlässigen Erzählens beschäftigt (Abschnitt 4).

Goldstandard-Annotation und potenzielle Probleme

Das Konzept der Gold-Standard-Annotationen ist so tief in der Methodik der Computerlinguistik verwurzelt, dass es sogar in Standardlehrbüchern vorausgesetzt wird (vgl. Jurafsky/Martin 2022). Die Idee ist, dass Menschen Zielphänomene in Korpora manuell annotieren, die dann als Ground Truth für das Training oder die Evaluierung von

Computermodellen dienen, die diese Phänomene automatisch erkennen (vgl. Wissler et al., 2014).

Im Kontext der digitalen Literaturwissenschaft wirft dies gegensätzliche Probleme auf.¹ Zum einen ist die Erstellung von ausreichend großen Gold-Standard-Korpora extrem arbeits- und zeitintensiv. Zum anderen kann es je nach Annotationsverfahren und annotierten Phänomenen zu Qualitätsmängeln bei den Annotationen kommen. Eine Möglichkeit, das erste Problem zu adressieren, sind Silberstandard-Annotationen (vgl. Rebholz-Schumann et al., 2019), d. h. kombinierte, automatisch generierte Annotationen, die „provide a level of annotation quality that is better than unchecked output of automated processing, even though it might not reach gold standard“ (Schweitzer et al., 2018).

Das zweite Problem wird durch verstärkte Maßnahmen zur Sicherung der Qualität manueller Annotationen angegangen, z. B. durch die Annotation eines Textes durch mehrere Annotator:innen, um idiosynkratische Entscheidungen zu verhindern, oder durch Iterationsschleifen mit verfeinerten Annotationskategorien/-richtlinien (vgl. Gius und Jacke, 2017; Reiter, 2020). In der Regel werden in solch einem Setting nur einstimmige Annotationen verwendet. Diese Annotationen sind als Goldstandard für die Evaluation oder das Training besser geeignet, da sie die tatsächlichen Zielphänomene adäquater widerspiegeln.

Platinstandard für Annotationen in hermeneutischen Kontexten

Die Annotation komplexer literarischer Phänomene (vgl. Gius, 2019) erfordert jedoch manchmal eine noch strengere Qualitätskontrolle. Allein durch Multi-Annotatoren-Settings und Iterationsschleifen kann nicht sichergestellt werden, dass einstimmige Annotationen fundierten Daten entsprechen und widersprüchliche Entscheidungen tatsächlich unentscheidbare Fälle darstellen. Im Gegensatz zu vielen traditionellen NLP-Aufgaben reicht die Intuition kompetenter Muttersprachler:innen (auch bei Kenntnis der relevanten literaturwissenschaftlichen Terminologie) nicht aus, um intuitiv adäquate Entscheidungen zu treffen, da die Erkennung einiger literarischer Phänomene in einem Text stark auf verschiedenen impliziten Annahmen und Schlussfolgerungen beruht (vgl. Kuhn, 2023).

Gerade in Fällen, in denen das primäre Ziel der Annotation nicht die möglichst zeitnahe Automatisierung der Annotationsaufgabe darstellt, um auf großen Korpora beispielsweise einen historischen Überblick über die (ungefähre) Entwicklung literarischer Phänomene zu erlangen, sondern hermeneutische Fragen im Fokus stehen, kann die digitale Literaturwissenschaft von Platinstandard-Annotationen profitieren. Darunter sind Annotationen zu verstehen, die nicht durchschnittliche bzw. mehrheitliche menschliche Entscheidungen widerspiegeln, sondern *gut begründete*: Annotator:innen sollen in solch einem Kontext Annahmen und Schlussfolgerungen, die zu Annotationen

führen, dokumentieren, systematisieren, diskutieren und reflektieren.

In einigen Forschungsprojekten im Feld der digitalen Literaturwissenschaft sind bereits diskursive Prozesse zur Optimierung von Annotationen implementiert (vgl. z.B. Gius und Jacke, 2017, Andresen et al., 2022, Gödeke et al., 2022). Obwohl die so produzierten Annotationen innerhalb dieser Projekte teilweise als „Goldstandard“ bezeichnet werden (Gödeke et al. 2022), weisen sie durch die Diskussion (u. U. mit anschließender Revision) Elemente des hier vorgeschlagenen Konzepts des Platinstandards auf. Ich möchte indes als Bedingungen für Platinstandard-Annotationen vorschlagen, dass der diskursive Aushandlungs- bzw. Begründungsprozess von Annotationen stärker regelgeleitet vorgenommen und systematisch dokumentiert werden sollte. In diesem Rahmen können dann auch die Kriterien entwickelt und expliziert werden, nach denen entschieden wird, unter welchen Umständen Annotationen bearbeitet werden sollten und in welchen Fällen etwa auch Uneinigkeit zwischen Annotator:innen bestehen bleiben darf.² Durch derartige Kriterien ergibt sich die Möglichkeit einer (fachlich begründeten und vorerst nicht automatisierbaren) Evaluation der Annotationen – denn Inter-Annotator Agreement ist, wie auch Andresen et al. anmerken, bei komplexen und interpretationsabhängigen Annotationsaufgaben nicht immer ein angemessenes Evaluationsmittel (vgl. Andresen et al., 2022).

Im Folgenden soll ein mögliches Vorgehen zu Erstellung von Platinstandard-Annotationen anhand eines konkreten Beispiels aus dem Projekt CAUTION illustriert werden. Die hier vorgestellte Methode der Nutzung von Argumentbäumen hat vor dem Hintergrund der obigen Ausführungen folgende Vorteile: (1) Sie bietet die Möglichkeit einer systematischen Dokumentation von Argumentations- und Aushandlungsprozessen. (2) Einige der Evaluationskriterien liegen bereits zu Beginn offen und können durch die gewählte Dokumentationsform leichter überprüft werden, andere können im Prozess entwickelt bzw. identifiziert werden. (3) Die vorgestellte Methode hat den Vorteil, dass nicht grundsätzlich nur solche Annotationen auf den Prüfstand gestellt werden, bezüglich derer Uneinigkeit unter den Annotator:innen herrscht.

Die im Folgenden exemplarisch erläuterte Methode der Nutzung von Argumentbäumen ist zwar nicht die einzige Möglichkeit, Begründungs- und Aushandlungsprozesse bei der Entwicklung von Platinstandard-Annotationen zu systematisieren und zu dokumentieren – sie lässt sich aber (ggf. mit Modifikationen) generell auf Annotationsvorhaben anwenden, bei denen die Annotationen auf hermeneutischen Prozessen basieren.

Use Case: Argumentbäume als Methode zur Erstellung von Platinstan-

dard-Annotationen unzuverlässigen Erzählens

Im Projekt CAUTION (Computer-aided Analysis of Unreliability and Truth in Fiction – Interconnecting and Operationalizing Narratology) wird das interpretationsabhängige Phänomen des unzuverlässigen Erzählens – genauer: ein Teilkonzept mimetischer Unzuverlässigkeit (vgl. Kindt, 2008) – im Rahmen eines mehrgleisigen Zugangs untersucht, bei dem jeweils mit unterschiedlichem primären Erkenntnisinteresse Silber-, Gold- und Platinstandard-Annotationen erstellt werden (vgl. Jacke, 2023). Die hier untersuchte Form von Unzuverlässigkeit liegt vor, wenn die Erzählfigur in einem fiktionalen Text inkorrekte Behauptungen über die fiktive Welt äußert.

Im Folgenden soll der Fokus auf der Vorstellung des im Projekt umgesetzten Prozederes für die Erstellung der Platinstandard-Annotationen liegen. Da diese allerdings auf (einer literaturwissenschaftlichen Entsprechung von) Goldstandard-Annotationen aufbauen, wird deren Genese ebenfalls kurz erläutert.

Für die Erstellung der Goldstandard-Annotationen wurden zunächst neun deutschsprachige Erzählungen aus der Zeit zwischen dem 19. und dem 21. Jahrhundert³ mithilfe des Programms CATMA auf der Basis gemeinsamer Annotationsrichtlinien annotiert. Dabei haben mindestens drei Annotator:innen denselben Text bearbeitet, wobei Teilsätze, die eine inkorrekte Behauptung über die fiktive Welt enthalten, als „incorrect statement“ und unentscheidbare Teilsätze als „undecided“ ausgezeichnet werden sollten. Rechtfertigung oder Diskussion von Annotationsentscheidungen waren dabei nicht Teil des Prozederes. Um trotz der Interpretationsabhängigkeit des Phänomens (die zum einen in der Kategorie „undecided“, zum anderem im erwartungsgemäß niedrigen Inter-Annotator Agreement zum Ausdruck kommt, vgl. Blessing et al., 2024) einen gemeinsamen Datensatz zu kreieren, der als literaturwissenschaftliche Entsprechung von Goldstandard-Daten verstanden werden kann, wurden die Daten auf eine bestimmte Weise zusammengeführt: Alle annotierten Passagen erhalten einen Dezimalwert, der widerspiegelt, wie oft „incorrect statement“ (1,0) und „undecided“ (0,5) gewählt wurde (Addition der Werte aller Annotator:innen für die Passage geteilt durch Zahl der Annotator:innen). Das Ergebnis stellt einen durchschnittlichen Wert der Sicherheit dar, mit der menschliche Annotator:innen einer Passage vergleichsweise intuitiv unzuverlässiges Erzählen zuschreiben.

Für ein hermeneutisch besonders interessantes Teilkorpus⁴ wird zur Erstellung von Platinstandard-Annotationen eine Iterationsschleife mit einem experimentellen Setting hinzugefügt. Einschätzungen zur mimetischen Unzuverlässigkeit von Erzählinstanzen, insbesondere die Feststellung inkorrektur Äußerungen, basieren auf einer Gesamtinterpretation des Textes im Hinblick auf die fiktive Welt (vgl. Kindt, 2008, 53). Diese Art der Interpretation lässt sich als inhaltspezifisierende (*content-specifying*) Interpretation

bezeichnen (vgl. Folde, 2015). Im Rahmen der ersten intuitiven Annotationsrunde liegen den Annotationsentscheidungen solche Interpretationen in der Regel implizit und unvollständig zugrunde. In der Iterationsschleife zur Erstellung der Platinstandard-Annotationen sollen diese Interpretationen der Annotator:innen nicht nur explizit gemacht, sondern auch in einem kollaborativen Prozess optimiert werden. Diesem Anspruch liegt die Annahme zugrunde, dass (auch inhaltspezifisierende) Interpretationen literarischer Texte zwar legitimerweise variieren können (z.B. aufgrund von textueller Mehrdeutigkeit, unterschiedlicher Fokussierung oder verschiedener interpretationstheoretischer Ansätze), dass Interpretationen aber zugleich argumentativ gut untermauert sein sollten (vgl. Descher und Petraschka, 2018).

Um Interpretationen und Argumentationen gut strukturiert und möglichst vergleichbar zu visualisieren, werden die Annotator:innen gebeten, im Argument-Visualisierungs-Modus des Programms MindMup⁵ individuell einen Argumentbaum zu erstellen, bei dem eine These über die Zuverlässigkeit bzw. Unzuverlässigkeit der Erzählfigur des jeweiligen Textes die Hauptthese bildet,⁶ die dann durch unterschiedliche Argumente bzw. Prämissen gestützt wird.⁷ Ein Beispiel für einen solchen Argumentbaum (hier für den Text „Meine falschen Eltern“ von Katharina Bendixen, vgl. Bendixen, 2012) findet sich in Abb. 1. Der Argumentbaum ist folgendermaßen zu lesen: Jeder Knoten (Konklusion) wird durch die mit ihm verbundenen darunterliegenden Knoten (Prämissen) gestützt. Prämissen können wiederum selbst argumentativer Stützung bedürfen. Sie werden dann ebenfalls zu Konklusionen, die durch darunterliegende Prämissen gestützt werden. Sind mehrere Prämissen gemeinsam Teil eines Arguments, wird dies durch eine horizontale Klammer über den relevanten Prämissen visualisiert. Die Prämissen sind in erster Instanz hauptsächlich Thesen über die fiktive Welt des Textes, in späterer Instanz sind sie dann aber oft anderen Typen zugehörig – einen besonders wichtigen Typ bilden konkrete Textstellen aus dem relevanten literarischen Text.⁸

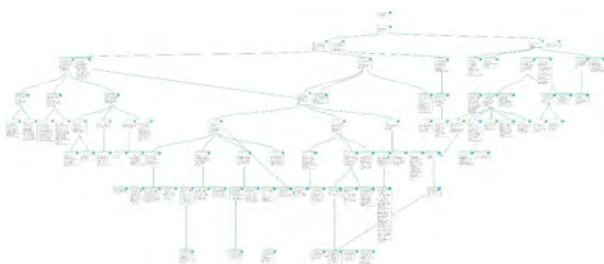


Abbildung 1: Beispiel-Argumentbaum: unzuverlässiges Erzählen in „Meine falschen Eltern“

Nachdem jede Annotatorin einen Argumentbaum erstellt hat, werden die einzelnen Bäume in der Annotator:innengruppe untersucht und diskutiert. Dabei sollen Stellen identifiziert werden, an denen die Argumentationsstruktur nicht korrekt ist (Bilden vertikale Kanten wirklich Prä-

missen-Konklusion-Zusammenhänge ab?), an denen Thesen nicht ausreichend begründet erscheinen⁹ und an denen mögliche Gegenargumente unberücksichtigt bleiben. Zudem soll geprüft werden, ob die Argumentbäume relevante Textmerkmale erklären bzw. relevante Fragen (hier: über die fiktive Welt) erklären kann. Relevante Kriterien sind hier also u.a. plausible Schlussregeln (wobei Regeln für literaturwissenschaftliche Argumentation teilweise im Detail erst noch ausgehandelt werden müssen), Plausibilität von Prämissen (wobei sich die Einschätzung je nach interpretationstheoretischem Ansatz unterscheiden kann), Entkräftigung von Gegenargumenten sowie Kohärenz/Reichweite. Die Annotator:innen haben in den Diskussionen die Möglichkeit, auf die Einwände zu reagieren, und können im Anschluss ihre Argumentbäume überarbeiten, um den Einwänden zu begegnen.

Das Ziel dieses Vorgehens besteht nicht darin, dass die Annotator:innen sich am Ende auf eine inhaltspezifisierende Interpretation, eine Argumentation oder gar auf einen Argumentbaum einigen. Es geht lediglich darum, die impliziten und intuitiven Interpretationen, die den Unzuverlässigkeitsannotationen zugrunde liegen, explizit zu machen, in einem diskursiven Verfahren auf den Prüfstand zu stellen und möglichst gut argumentativ zu stützen. Nach abgeschlossener Arbeit an den Argumentbäumen überarbeiten die Annotator:innen auf dieser argumentativen Basis ihre vormals intuitiven Unzuverlässigkeitsannotationen.¹⁰

Fazit

Auf die dargestellte Weise lassen sich für ein relevantes Teilkorpus Platinstandard-Annotationen erstellen, also gut begründete Annotationen für interpretationsabhängige literarische Phänomene. Diese Daten sind literaturwissenschaftlich belastbare hermeneutische Annotationen, die sowohl reflektierte Informationen über die literarischen Texte enthalten (etwa im Vergleich der Annotationsversionen) als auch theoretische und methodologische Einblicke erlauben (etwa in die Interpretationsabhängigkeit der Annotationen). Zudem werden mit den Argumentbäumen und den Vorarbeiten¹¹ weitere auswertbare Datensätze generiert, die die Möglichkeit bieten, eine Brücke zwischen den Arbeitsfeldern Digitale Literaturwissenschaft und Computational Argumentation (vgl. z.B. Habernal und Gurevych, 2016) zu schlagen, um die Spezifika literaturwissenschaftlichen Argumentierens in den Blick zu rücken.

Aufgrund des umfangreichen Prozederes sind die generierten Datensätze in der Regel zu klein, um unmittelbar ML-Modelle zu trainieren. Wenn allerdings durch gemeinsame Anstrengungen und über mehrere Jahre hinweg ein ausreichend großes Platinstandard-Korpus für ein komplexes Phänomen generiert werden kann, um Modelle zu trainieren, könnte sogar die Aussicht bestehen, parametrisierte automatische Annotationen zu erhalten, die Annotationen relational zu bestimmten Grundannahmen und -argumenten generieren. Bis dahin kann (auch innerhalb eines Projekts) parallel mit verschiedenen Standards gearbeitet werden, so

dass beispielsweise einerseits mittels Silber- oder Goldstandard-Annotationen ein ComputermodeLL als Annäherung an ein literaturwissenschaftliches Phänomen, andererseits mittels Platinstandard-Annotationen ein literaturwissenschaftlich belastbarer Datensatz für besonders interessante Einzeltexte generiert wird.

Fußnoten

1. Probleme mit Goldstandard-Annotation im Bereich des NLP werden in Kovář et al., 2016 diskutiert.
2. Bei Andresen et al. werden als Kriterien „plausibility, consistency and formal correctness of the annotations“ genannt (Andresen et al. 2022), wobei allerdings offenbleibt, wie etwa die Konsistenz auf den Prüfstand gestellt oder wie Plausibilität definiert wird.
3. Zu den Kriterien für die Korpuserstellung siehe Blessing et al., 2024.
4. Das können beispielsweise Texte sein, in denen besonders viele Unzuverlässigkeitsannotationen vorkommen oder in denen sich die Annotator:innen besonders uneinig sind, das IAA also besonders niedrig ist.
5. Im Kontext von CAUTION ist die Wahl auf die intuitiv und kollaborativ nutzbare Software MindMup gefallen (<https://www.mindmup.com/mindmup-gold/>, 22.07.2024). Es ließe sich aber mit anderer Software experimentieren, die durch mehr Codierungsoptionen weitere Analysen der Argumentationsdaten zulässt, etwa mit Argdown (<https://argdown.org/>, 22.07.2024).
6. Das Verhältnis zwischen ‚pauschalen‘ Aussagen über die Zuverlässigkeit von Erzählinstanzen einerseits und der textstellenbasierten Feststellung inkorrektur Aussagen andererseits ist nicht trivial. Eine plausible These wäre, dass Erzählinstanzen dann als (mimetisch) unzuverlässig bezeichnet werden, wenn ihre inkorrektur Aussagen einen Grenzwert erreichen, der sich aus Häufigkeit, Intensität und interpretatorischer Relevanz zusammensetzt (vgl. Jacke, 2020, 180–182). Dieser Zusammenhang wird in CAUTION ebenfalls untersucht, steht aber nicht im Fokus.
7. Vorbild für die Arbeit mit Argumentbäumen im literaturwissenschaftlichen Bereich ist das Projekt ArguLit (vgl. Descher et al., 2023), in dem Argumentbäume allerdings zur Rekonstruktion von Argumentationen in literaturwissenschaftlichen Interpretationstexten genutzt wurden und nicht, wie hier, als heuristisches Werkzeug zur Explizierung und kollaborativen Verbesserung eigener Interpretationen und Annotationen.
8. Als methodische Heranführung an die Erstellung eines Argumentbaums erstellen die Annotatorinnen (zumindest für den ersten analysierten Text) zunächst eine Liste relevanter Fragen über die fiktive Welt sowie mögliche Antworten. Im Anschluss wird ein knapper Text verfasst, der diese Fragen beantwortet und argumentativ stützt. In einem letzten Schritt wird dieser Text in einen Argumentbaum überführt.

9. Im Zusammenhang mit literaturwissenschaftlicher Argumentation sind Argumente oft nicht formal (deduktiv) schlüssig. Was als hinreichend argumentativ gestützt gilt, ist hier also ebenfalls Teil des diskursiven Aushandlungsprozesses.

10. Zum Verfassungszeitpunkt dieses Abstracts läuft der Prozess der Überarbeitung der Annotationen noch – bis März 2025 werden aber für mehrere Texte alle Daten vorliegen, so dass im Vortrag auch die überarbeiteten mit den intuitiven Annotationen verglichen werden können.

11. Siehe Fußnote 8.

Bibliographie

- Andresen, Melanie, Benjamin Krautter, Janis Pagel und Nils Reiter.** 2022. „Who Knows What in German Drama? A Composite Annotation Scheme for Knowledge Transfer. Annotation, Evaluation, and Analysis“. *Journal of Computational Literary Studies* 1 (1). <https://doi.org/10.48694/jcls.107>.
- Bendixen, Katharina.** 2012. „Meine falschen Eltern“. In Dies.: *Gern, wenn du willst*. Poetenladen: Leipzig.
- Blessing, André, Janina Jacke und Jonas Kuhn.** 2024. „Agreement und Kookkurrenz bei unzuverlässigem Erzählen. Ziele, Herausforderungen und erste Ergebnisse aus dem Projekt CAUTION“. In *DH Quo Vadis? DHd 2024 Passau. Konferenzabstracts*. DOI: 10.5281/zenodo.10686564.
- Descher, Stefan, Merten Kröncke und Simone Winko.** 2023. „Wie plausibilisieren Literaturwissenschaftler*innen ihre Interpretationen? Das DFG-Projekt ‚Das Herstellen von Plausibilität in Interpretationstexten. Untersuchungen zur Argumentationspraxis in der Literaturwissenschaft‘ (ArguLit)“. *Sonderausgabe 7 von Textpraxis* (2.2023). <https://www.textpraxis.net/descher-kroencke-winko-wie-plausibilisieren-literaturwissenschaftler-ihre-interpretationen> (zugegriffen: 22. Juli 2024).
- Descher, Stefan und Thomas Petraschka.** 2019. *Argumentieren in der Literaturwissenschaft. Eine Einführung*. Ditzingen: Reclam.
- Folde, Christian.** 2015. „Grounding Interpretation“. *British Journal of Aesthetics* 55 (3), 361–374.
- Gius, Evelyn.** 2019. „Computationelle Textanalysen als fünfdimensionales Problem. Ein Modell zur Beschreibung von Komplexität“. *LitLab Pamphlet* 8, hrsg. v. Thomas Weitin. https://www.digitalhumanitiescooperation.de/wp-content/uploads/2019/12/pamphlet_gius_2.0.pdf (zugegriffen: 22. Juli 2024).
- 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 (2), 233–254.
- Gödeke, Luisa, Florian Barth, Tillmann Döncke, Hanna Varachkina, Anna Mareike Weimer, Benjamin Gittel, Anke Holler und**

Caroline Sporleder. 2022. „Generalisierungen als literarisches Phänomen. Charakterisierung, Annotation und automatische Erkennung“. In *Zeitschrift für digitale Geisteswissenschaften* 7. https://zfdg.de/2022_010 (zugegriffen: 3. Dezember 2024).

Habernal, Ivan und Iryna Gurevych. 2016. „Which argument is more convincing? Analyzing and predicting convincingness of Web arguments using bidirectional LSTM“. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, 1589–1599. <https://aclanthology.org/P16-1150.pdf> (zugegriffen: 22. Juli 2024).

Jacke, Janina. 2020. *Systematik unzuverlässigen Erzählens. Analytische Aufarbeitung und Explikation einer problematischen Kategorie*. Berlin und Boston: de Gruyter.

Jacke, Janina. 2023. „Die (computationelle?) Operationalisierung unzuverlässigen Erzählens. Ein Beitrag zur Theorie und Methodik literaturwissenschaftlichen Interpretierens“. *Sonderausgabe 7 von Textpraxis* (2.2023). <https://www.textpraxis.net/janina-jacke-operationalisierung-unzuverlaessigen-erzaehlens> (zugegriffen: 22. Juli 2024).

Jurafsky, Daniel und James H. Martin. 2022. *Speech and Language Processing. An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. https://web.stanford.edu/~jurafsky/slp3/ed3book_jan122022.pdf (zugegriffen: 22. Juli 2024).

Kindt, Tom. 2008. *Unzuverlässiges Erzählen und literarische Moderne. Eine Untersuchung der Romane von Ernst Weiß*. Tübingen: Niemeyer.

Kovář, Vojtěch, Miloš Jakubíček und Aleš Horák. 2016. „On Evaluation of Natural Language Processing Tasks – Is Gold Standard Evaluation Methodology a Good Solution?“. In *Proceedings of the 8th International Conference on Agents and Artificial Intelligence (ICAART 2016)*, Bd. 2, 540–545. <https://www.scitepress.org/papers/2016/58248/58248.pdf> (zugegriffen: 22. Juli 2024).

Kuhn, Jonas. 2023. „Empirie – Beschreibung – Interpretation. Über den Platz von Computermodellen in den hermeneutisch-historisch orientierten Literaturwissenschaften“. In *Digitale Literaturwissenschaft. DFG Symposium 2017*, hg. von Fotis Jannidis. Springer. 57–95. https://link.springer.com/chapter/10.1007/978-3-476-05886-7_4 (22.07.2024).

Rebholz-Schuhmann, Dietrich, Antonio José Jimeno Yepes, Erik M. van Mulligen, Ning Kang, Jan Kors, David Milward, Peter Corbett, Ekaterina Buyko, Katrin Tomanek, Elena Beisswanger und Udo Hahn. 2010. „The CALBC Silver Standard Corpus for Biomedical Named Entities — A Study in Harmonizing the Contributions from Four Independent Named Entity Taggers“. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, 568–573. <https://aclanthology.org/L10-1609/> (zugegriffen: 22. Juli 2024).

Reiter, Nils. 2020. Anleitung zur Erstellung von Annotationsrichtlinien. *Reflektierte algorithmische Textanalyse*, ed. by Nils Reiter,

Axel Pichler und Jonas Kuhn. De Gruyter, 193–198, <https://www.degruyter.com/document/doi/10.1515/9783110693973-009/html> (zugegriffen: 22. Juli 2024).

Schweitzer, Katrin, Kerstin Eckart, Markus Gärtner, Agnieszka Falenska, Arndt Riester, Ina Rösiger, Antje Schweitzer, Sabrina Stehwen und Jonas Kuhn. 2018. German Radio Interviews: The GRAIN Release of the SFB732 Silver Standard Collection. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*. <https://aclanthology.org/L18-1457/> (zugegriffen: 22. Juli 2024).

Wissler, Lars, Mohammed Almashraee, Dagmar Monett und Adrian Paschke. 2014. „The Gold Standard in Corpus Annotation“. In *Proceedings of 5th IEEE Germany Student Conference*. DOI: 10.13140/2.1.4316.3523.