

Janis Pagel, Nils Reiter, Ina Rösiger und Sarah Schulz

Annotation als flexibel einsetzbare Methode

Zusammenfassung: In der Computerlinguistik (CL) wird Annotation mit dem Ziel eingesetzt, Daten als Grundlage für maschinelle Lernansätze und Automatisierung zu sammeln. Gleichzeitig nutzen Geisteswissenschaftler Annotation in Form von Notizen beim Lesen von Texten. Wir behaupten, dass mit der Entwicklung der Digital Humanities (DH) die Annotation zu einer Methode geworden ist, die als Mittel zur Unterstützung der Interpretation und Entwicklung von Theorien eingesetzt werden kann. In diesem Beitrag zeigen wir, wie diese verschiedenen Annotationsziele in einem einheitlichen Workflow abgebildet werden können. Wir reflektieren die Komponenten dieses Workflows und geben Beispiele, wie Annotation im Rahmen von DH-Projekten einen Mehrwert schaffen kann.



Abstract: In computational linguistics (CL), annotation is used with the goal of compiling data as the basis for machine learning approaches and automation. At the same time, in the Humanities scholars use annotation in the form of note-taking while reading texts. We claim that with the development of Digital Humanities (DH), annotation has become a method that can be utilized as a means to support interpretation and develop theories. In this paper, we show how these different annotation goals can be modeled in a unified workflow. We reflect on the components of this workflow and give examples for how annotation can contribute additional value in the context of DH projects.

1 Einführung

Unter Annotation verstehen wir eine Methode, um Textdaten mit zusätzlichen Daten anzureichern. Diese zusätzlichen Daten sind an mehr oder weniger klar abgegrenzte Textstellen gebunden. Im Sprachgebrauch changiert der Begriff ‚Annotation‘ zwischen einer Bezeichnung für einen Prozess und dessen Ergebnis. In diesem Artikel liegt unser Fokus auf der Methode, und nicht auf den erstellten Annotationen als Datenobjekt oder Analysegegenstand. Wir konzentrieren uns zu-

Anmerkung: Dieser Beitrag ist eine übersetzte und überarbeitete Fassung von: Janis Pagel, Nils Reiter, Ina Rösiger und Sarah Schulz (2018). „A Unified Text Annotation Workflow for Diverse Goals“. In: *Proceedings of the Workshop for Annotation in Digital Humanities (annDH)*. Sofia, Bulgarien, S. 31–36. URL: <http://ceur-ws.org/Vol-2155/pagel.pdf> (besucht am 1. Juni 2020)

Janis Pagel, Nils Reiter, Ina Rösiger, Sarah Schulz, Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart

 Open Access. © 2020 Janis Pagel, Nils Reiter, Ina Rösiger und Sarah Schulz; publiziert von De Gruyter
 Dieses Werk ist lizenziert unter der Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International Lizenz.
<https://doi.org/10.1515/9783110693973-006>

dem auf Annotationsaufgaben, die interpretative oder assoziative Aspekte haben (d. h., sich auf den expliziten oder impliziten Inhalt eines Textes beziehen).¹

Annotationsprojekte in der Computerlinguistik (CL) haben eine große Anzahl an Korpora hervorgebracht, die mit linguistischen Konzepten annotiert sind (z. B. Wortarten oder semantische Rollen). Computerlinguistische Annotationsprojekte legen darüber hinaus Wert auf konsistente und konsensuale Entscheidungen zwischen den Annotator*innen, da die getroffenen Entscheidungen oft als Trainings- oder Testdaten für (überwachte) maschinelle Lernmethoden verwendet werden.

Annotationen in den Geisteswissenschaften folgen nicht dem gleichen Paradigma wie in der Computerlinguistik, in der die Annotationen auch als Methode verwendet werden, um intersubjektive Entscheidungen zu treffen. Vielmehr spielt in den Geisteswissenschaften die Annotation als Hinzufügen von Notizen am Rand oft eine Rolle, um individuelle Interpretationen z. B. eines literarischen Textes für den eigenen Gebrauch zu visualisieren, wenn auch manchmal implizit. Die explizite Darstellung dieses Prozesses hat Vorteile, da explizite Annotationen die Interpretation unterstützen können, indem sie sie klarer, eindeutiger und auch nachvollziehbarer machen. Darüber hinaus könnte eine Zukunftsperspektive für die Geisteswissenschaften ein kollaborativer Prozess der Theorieentwicklung sein. Ein Ansatz zur Erreichung dieses Ziels ist die Integration der Annotationsmethode in die geisteswissenschaftliche Forschung, bei der theoretische Begriffe durch den Annotationsprozess formalisiert und iterativ geschärft werden.

Das vorliegende Kapitel vergleicht die in der CL vorherrschenden Annotationsprozesse mit den in den (digitalen) Geisteswissenschaften verwendeten Prozessen. Wir argumentieren, dass die Annotationsprozesse zwar unterschiedlichen Zielen dienen und unterschiedliche Prioritäten setzen, aber viele Gemeinsamkeiten aufweisen und in ein gemeinsames konzeptionelles Modell integriert werden können. Darüber hinaus argumentieren wir, dass Annotation ein produktives Werkzeug sein kann, um theoretische Begriffe und Definitionen in den Geisteswissenschaften zu präzisieren und verbessern, was eine neue Art der Verwendung von Annotation darstellt.

¹ Im Gegensatz dazu stehen strukturelle Annotationen, die z. B. Formatierungsinformationen oder Metadaten markieren. Auch wenn hier technisch etwas Ähnliches passiert, ist es im Regelfall kein interpretativer Akt und steht (oft) in keinem Zusammenhang zum Textinhalt. Die Grenzen zwischen interpretativen und strukturellen Annotationen sind aber in Einzelfällen schwer zu ziehen.

2 Verschiedene Annotationsziele

Die verschiedenen Annotationsziele, die in der Literatur diskutiert werden, lassen sich grob in vier Bereiche sortieren, die sich teilweise überlappen:

Die **explorative Annotation** bietet zunächst die Möglichkeit, einen Text (oder ein anderes Datenobjekt) strukturiert kennenzulernen, ohne sich vorab auf Kategorien oder Schemata festzulegen. Diese Art der Auszeichnung ist der langjährigen Tradition der Kommentierung in den traditionellen Geisteswissenschaften am nächsten (Bradley 2008), bei der Forscher*innen ihnen wichtig erscheinende Aspekte sowie Ideen, die beim Lesen entstanden sind, am Rand einer Seite notieren. Bradley führt aus:

[T]his kind of annotation, indeed note-taking more generally, provides one of the bases for much scholarly research in the humanities. In this view note-taking fits into the activity of developing a personal interpretation of the materials the reader is interested in. (Bradley 2012, Abs. 11)

Ziel dieser Art von Annotation ist es also, am Ende ein vorläufiges Textwissen samt möglicher Assoziationen zu erhalten, das es den Forscher*innen ermöglicht, eine konkretere Forschungsfrage oder Hypothese zu formulieren. Diese Frage oder Hypothese kann später vor dem Hintergrund einer theoretischen Grundlage bearbeitet werden, während die erste Lektüre ohne spezifische Annahmen oder Fragen erfolgt.

Zweitens zielt die **konzeptualisierende Annotation**² darauf ab, die Definition von theoretischen Begriffen oder prä-theoretischen Beobachtungen, die erklärungsbedürftig sind, zu verbessern. Beide werden oft in Sekundärliteratur beschrieben, aber selten so definiert, dass sie auf neue Texte ‚anwendbar‘ sind. Der Versuch, die Begriffe oder Beobachtungen systematisch in weiteren Texten zu finden, fördert zunächst Definitionslücken oder -schwierigkeiten zutage. Durch das Beheben dieser Lücken oder Schwierigkeiten werden die Definitionen besser und breiter anwendbar. Primäres Mittel hierbei ist es, Fälle von Meinungsverschiedenheiten zwischen verschiedenen Annotator*innen zu identifizieren und die Definitionen zu verfeinern, bis eine ausreichende Übereinstimmung erreicht wird.

Drittens zielt die **erklärende Annotation** darauf ab, eine kondensierte und ggf. formalisierte Darstellung der textlichen Grundlage für eine Interpretationshypothese zu liefern. Während Interpretationshypothesen (z. B. in der Literaturwissenschaft) typischerweise auf Textnachweisen basieren, sind die entsprechen-

² Konzeptualisierende Annotation ist Ziel des *shared tasks SANTA*, der ab Seite 390, in Teil IV des Bandes vorgestellt wird.

den Textstellen nicht explizit markiert und der Argumentationspfad von Textstelle zur Interpretation bleibt implizit. Durch die Erläuterung in Form von Annotationen werden diese Schritte explizit und formalisiert. Diese Annotationen sind nicht auf ein einzelnes Phänomen beschränkt, sondern decken alle Phänomene ab, die für eine Interpretation als relevant erachtet werden. Bei dieser Konstellation besteht das Hauptziel nicht darin, eine einzige ‚wahre‘ Annotation zu erstellen, sondern verschiedene plausible Annotationen, die unterschiedliche Lesarten des Textes darstellen. Erläuternde Annotationen sind grundsätzlich in verschiedenen Formalisierungsgraden denkbar. Im unformalisierten Fall bestehen die Annotationen aus Fußnoten, die Text enthalten, der sich an Menschen richtet. Am anderen Ende des Spektrums sind Annotationen denkbar, in denen maschinenlesbare Informationen geliefert werden.

Viertens zielt die **automatisierungsorientierte Annotation**³ (cf. Hovy und Lavid 2010; Pustejovsky und Stubbs 2012) auf die Zusammenstellung konsistent annotierter Daten, die als Trainings- und Testmaterial für automatische Annotationswerkzeuge verwendet werden. Automatisierte Annotationswerkzeuge benutzen dabei typischerweise Verfahren des überwachten maschinellen Lernens, bei dem Modelle trainiert werden, die den Zusammenhang zwischen Oberflächenmerkmalen und Zielkategorien aus großen Datenmengen erschließen. Die Konsistenz der Annotation ist für die Automatisierung von größter Bedeutung, da Inkonsistenzen die Klassifikationsleistung negativ beeinflussen. Annotationsprojekte, die Trainings- oder Testdaten generieren, legen Wert auf ein hohes *inter-annotator agreement* (IAA), also auf eine hohe Übereinstimmung zwischen den Annotator*innen.

Die vier genannten Anwendungsfälle der Annotationsmethode schließen sich nicht gegenseitig aus. Tatsächlich ist es schwierig, nicht zumindest einige Aspekte der anderen Ziele zu berühren, auch wenn nur ein einziges Ziel momentan im Fokus liegt. Annotation zur Generierung von Trainings- oder Testdaten z. B. legt oft Probleme in den Definitionen und Annotationsrichtlinien offen, die verfeinert werden müssen. Dabei werden Entscheidungen getroffen, die Rückwirkungen in die konzeptuelle Ebene haben, selbst wenn diese nicht intendiert, sondern die Entscheidungen vor allem pragmatisch motiviert sind. Die Entscheidung z. B., zwei Kategorien zu verschmelzen, mag pragmatisch motiviert sein. Sie ist aber Teil der Konzeptarbeit, die innerhalb des Annotationsprojektes stattfindet.

³ Annotation zum Ziel der automatisierten Erkennung des Zielphänomens ist außerdem Diskussionsgegenstand des Beitrages von Pichler und Reiter (2020), S. 43ff. in diesem Band.

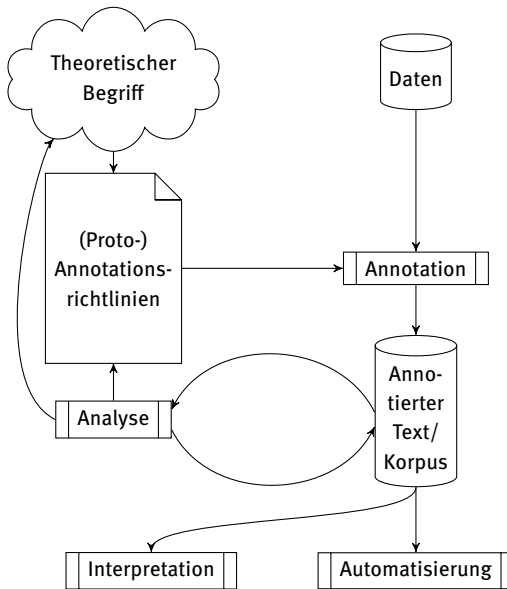


Abb. 1: Annotations-Workflow-Schema. Pfeile zeigen eine (grobe) zeitliche Abfolge an.

3 Ein einheitlicher Annotationsworkflow

Abbildung 1 veranschaulicht ein Modell für einen Annotationsworkflow, der auf die im vorigen Abschnitt genannten Ziele ausgerichtet ist. Es beschreibt sowohl das in der CL vorherrschende Annotationsmodell als auch Annotationsmodelle aus den Geisteswissenschaften und Anwendungsfälle, die neu und spezifisch für die DH sind. Die Bezeichnung als Workflow bedeutet nicht, dass jedes Annotationsprojekt jeden Teilschritt verwenden oder dass der gesamte Workflow innerhalb eines einzelnen Projekts abgeschlossen werden muss. Die in einem Projekt festgelegten Annotationsvorgaben können problemlos im nächsten Projekt fortgesetzt oder ausgearbeitet werden. Je nach Ziel des Projektes werden verschiedene Bereiche betont oder ignoriert. Generell können die verschiedenen Annotationsprozesse auch als Phasen betrachtet werden, die ein Phänomen durchläuft, bis ein intersubjektives Verständnis erreicht worden ist.

Der Ausgangspunkt des Workflows ist ein **theoretischer Begriff**. Wir verwenden hier die Bezeichnung ‚Begriff‘, um eine Vielzahl von Fällen einzubeziehen: Der Begriff kann auf der Grundlage einer vollwertigen Theorie (z. B. Wortarten oder Erzählebenen) beschrieben/vorhergesagt werden, aber er kann auch auf einer Beobachtung in Textdaten beruhen, die erklärt werden muss oder in existie-

render Fachliteratur diskutiert wurde (z. B. Ähnlichkeiten in der Figurendarstellung bei Adaptionen eines literarischen Stückes). Der theoretische Begriff wird in der Abbildung durch eine Wolke dargestellt, um anzuzeigen, dass er oft ‚unscharfe Kanten‘ aufweist, und die Anwendung auf textuelle Daten Interpretationsschritte beinhaltet. Der Begriff kann natürlich, etwa durch Sichtung von Forschungsliteratur, weiter und feiner spezifiziert werden, bevor er in den Annotationszyklus eintritt.

Theoretische Begriffe interagieren auf vielfältige Weise mit **Daten**: Viele Begriffe beruhen auf an Daten gemachten Beobachtungen, selbst wenn sie indirekt oder nur durch den bisherigen wissenschaftlichen Diskurs übermittelt werden. Daher spielt das Zustandekommen der Beobachtung ebenfalls eine Rolle: Eine konkrete Datensammlung kann nie wirklich zufällig ausgewählt werden und setzt daher zumindest die (breite) Beschränkung auf ein Interessengebiet voraus. Die Auswahl der Daten führt zu einer Verzerrung und schränkt den Raum für mögliche Beobachtungen ein. Dazu führen Kanonisierungs- und Standardisierungsprozesse zu einer Verengung des Blickwinkels und machen bestimmte Phänomene grundsätzlich nicht beobachtbar. Dies ist unabhängig vom genauen Status des theoretischen Begriffs. Daher muss der Datenauswahl große Aufmerksamkeit geschenkt werden und Kriterien für die Auswahl müssen für die Nutzer*innen der Sammlung explizit festgelegt werden, um die Forschung transparent und nachvollziehbar zu machen.

Die eigentliche Annotation basiert immer auf **Annotationsrichtlinien**, selbst wenn diese nicht expliziert sind (und es vielleicht auch nie werden). Wenn ein theoretisches Konzept zum ersten Mal annotiert wird, können die Richtlinien zunächst nur eine Fixierung auf ein bestimmtes theoretisches Werk (für die Narratologie z. B. Genette 1980) oder einen Teil davon (z. B. narrative Ebenen) sein. Iterationen im Annotations-Workflow können zu immer aufwendigeren Annotationvorgaben führen, die vom theoretischen Konzept abweichen können. Für das Alltagsgeschäft der Annotation dienen Richtlinien als Vermittler zwischen theoretischen Begriffen und der eigentlichen Annotationspraxis. Im Idealfall können so auch Nicht-Expert*innen (z. B. studentische Hilfskräfte oder Crowdsourcing-Arbeiter*innen) die Annotationen durchführen. Annotationsrichtlinien sollten zwar so generisch wie möglich gehalten werden, durch die Art ihres Zustandekommens sind sie jedoch oft ‚optimiert‘ für bestimmte Texte oder Korpora. Wenn theoretische Konzepte für Nicht-Expert*innen aufgeschlüsselt werden, werden sie in Bezug auf das zu kommentierende Korpus beschrieben; schwierige, aber irrelevante Aspekte können völlig ignoriert werden.

Der eigentliche **Annotationsprozess** besteht dann darin, Texte zu lesen, Textteile zu markieren/auszuwählen und mit den in den Richtlinien definierten Kategorien zu verknüpfen. Manchmal werden zusätzliche Merkmale einer

Instanz des theoretischen Begriffs annotiert. Abhängig von den Zielen der Annotation können Annotationen parallel durchgeführt werden, d. h. mehrere Annotator*innen bearbeiten den gleichen Text parallel. Dies ermöglicht den direkten Vergleich der Annotationen, um mögliche Mängel der Annotationsrichtlinien aufzudecken. Ein weiterer Parameter im Annotationsprozess ist die Annotationseinheit: Einige Annotationen funktionieren wortweise (d. h., jedes Wort ist annotiert), andere basieren auf Sätzen oder Phrasen. Um (nicht-triviale) linguistische Einheiten zuverlässig als Basiseinheit zu verwenden, ist es sinnvoll, diese manuell oder automatisch zu erzeugen und als Vorannotation bereitzustellen oder klare Kriterien für diese ebenfalls in den Richtlinien zu verankern. Auch wenn die Annotationen (je nach Ziel) grundsätzlich auf Papier erfolgen können, können computergestützte Annotationswerkzeuge den Annotationsprozess unterstützen, indem sie Annotationskandidaten vorschlagen oder die Nachnutzung der annotierten Daten ermöglichen.

Das unmittelbare Ergebnis des Annotationsprozesses ist ein **annotiertes Korpus**, das auf verschiedene Arten genutzt werden kann. Eine naheliegende Art der **Analyse** ist es, bestimmte Hypothesen oder Annahmen mit den neu erstellten Daten zu vergleichen. Diese Art der Analyse fördert ein besseres Verständnis der Theorie, z. B. in Form von präziseren theoretischen Begriffen. Die Analyse von konkreten Daten kann auch dazu führen, dass Belege für oder gegen bestimmte theoretisch motivierte Hypothesen gefunden werden. Diese Ergebnisse können dann genutzt werden, um die zugrundeliegende Theorie zu verfeinern. Eine andere Art von Analysen basiert auf den Meinungsverschiedenheiten zwischen mehreren Annotator*innen. Das Hauptziel dieser Art von Analyse ist es, sicherzustellen, dass i) die Annotationsrichtlinien ausreichend genau und klar definiert sind und ii) sie von den Annotator*innen gelesen, verstanden und befolgt wurden. Ein zielführendes Vorgehen dabei besteht darin, diejenigen Annotationen manuell zu überprüfen, bei denen die Annotator*innen verschiedener Meinung waren, d. h., unterschiedliche Annotationsentscheidungen getroffen haben. Dies kann durch die Annotator*innen selbst oder durch die ‚Vorgesetzten‘ erfolgen. Quantitativ kann die Menge an Meinungsverschiedenheiten als *inter-annotator agreement* ausgedrückt werden, das typischerweise in der Dokumentation zu einer Korpusveröffentlichung enthalten ist. Während die Messung von IAA eine recht lange Tradition hat (frühe Veröffentlichungen dazu stammen aus den 1960ern, wie etwa Cohen 1960), ist die Diskussion darüber, wie sich IAA genau quantifizieren lässt, noch nicht abgeschlossen (Mathet et al. 2015). Unterschiedliche Aufgaben, die auf unterschiedlichen Einheiten basieren erfordern zudem unterschiedliche Metriken für das *inter-annotator agreement*. Die quantitative Messung von IAA ist beim Vergleich verschiedener Annotationsrichtlinien oder annotierter Korpora zentral und das gemessene IAA kann auch als Obergrenze für die Maschinenleis-

tung dienen. Wenn das Ziel der Annotation darin besteht, theoretische Konzepte zu entwickeln, ist die Überprüfung der tatsächlichen Meinungsverschiedenheiten der Annotator*innen aufschlussreicher als die Zählung derselben. Gius und Jacke (2017) schlagen vor, Meinungsverschiedenheiten in vier Kategorien einzuteilen, basierend auf ihren Ursachen: i) Annotationsfehler, ii) Schwächen in den Annotationsrichtlinien, iii) abweichende Vorannahmen und iv) ‚echte‘ Mehrdeutigkeiten. Annotationsfehler basieren auf Nachlässigkeiten und können sofort behoben werden, die Meinungsverschiedenheiten der Kategorien ii) und iii) erfordern eine Anpassung der Annotationsrichtlinien. Wenn Meinungsverschiedenheiten der Kategorie iv) nicht durch Berücksichtigung eines zusätzlichen Kontextes gelöst werden können, bleiben sie als zwei Lesarten und Annotationsmöglichkeiten im Korpus erhalten.

Sobald ein annotiertes Korpus verfügbar ist, sind zwei verschiedene Folgeschritte möglich: Interpretation und Automatisierung. Die **Interpretation** eines Textes auf der Grundlage von Annotationen führt zu zusätzlichen Lesarten, die auf konkreten Datenpunkten beruhen, was letztlich auch zu einer intersubjektiveren Interpretation des Textes beitragen kann. Wir werden hier nicht im Detail auf den **Automatisierungsprozess** eingehen, aber er erfordert typischerweise annotierte Daten (siehe in diesem Band den Beitrag von Pichler und Reiter (2020) für eine abstrakte Beschreibung und den von Klinger et al. (2020) für ein konkretes Fallbeispiel).

Eine Annahme in der CL ist, dass die Annotationen eindeutig sind, dass also alle Meinungsverschiedenheiten gelöst wurden. Wie mit Unstimmigkeiten, die auch durch korrekte Anwendung der Annotationsregeln nicht eindeutig entschieden werden können in Bezug auf die Automatisierung umgegangen werden kann, ist noch nicht abschließend geklärt. Gius und Jacke (2017) schlagen unterschiedlich parametrisierte Modelle für die automatische Vorhersage vor, zumindest für die Uneinigkeitskategorie (iii). So kann beispielsweise die Annotation einer bestimmten Kategorie eine Entscheidung über eine grundlegendere verwandte Kategorie erfordern. In einem Tool zur automatischen Erkennung eines bestimmten Begriffs kann dieser Parameter manuell eingestellt werden, um einen bestimmten Messwert zu erzwingen. Gius und Jacke lassen jedoch die Frage offen, wie dies bei Meinungsverschiedenheiten, die sich aus einer gültigen Mehrdeutigkeit des Textes ergeben, realisiert werden kann.

Der robuste Umgang statischer oder quantitativer Methoden mit echt mehrdeutigen Daten bleibt also ein Desideratum, gerade auch in Hinblick auf Daten aus den Geistes- und Sozialwissenschaften.

4 Beispielhafte Annotationsprojekte

Wir besprechen nun mehrere DH-Projekte, um die verschiedenen Ziele von Annotation zu veranschaulichen und verschiedene Wege aufzuzeigen, die Projekte mit unserem Annotationsworkflow einschlagen können.

4.1 Explorative Annotation

Ein Beispiel für explorative Annotation ist das Anfertigen von Notizen. McCarty (2020) beschreibt seinen *note-taking*-Workflow, der aus dem Sammeln von Notizen in digitaler Form, deren Ausdrucken auf kleinen Kärtchen und anschließend der manuellen Sortierung besteht.

Ein frühes Projekt, das ein solches Vorgehen in einem Tool implementierte, ist das Pliny-Projekt (Bradley 2008). Pliny ist eine 2009 veröffentlichte Software, um neue Möglichkeiten der Annotation in der digitalen Welt zu erforschen. Es soll den traditionellen geisteswissenschaftlichen Workflow (Bradley 2008) unterstützen, indem es den Prozess des *note-taking* und die Aufzeichnung erster Reaktionen auf einen Text mit dem Ziel einer nachfolgenden Phase, in der eine Forschungsfrage entwickelt wird, ermöglicht. Die Entwickler*innen geben das Beispiel einer Webseite⁴, auf der die Benutzer*innen Beobachtungen notieren, die sie während des Besuchs der Seite machen. In unserem Workflow entspricht diese Phase der Annotation einer prä-theoretischen Stufe, in der Daten die Annotation auslösen. Dies kann in einem nächsten Schritt möglicherweise zur Analyse des annotierten Textes führen, was wiederum zu Annotierungsrichtlinien führen kann. Obwohl die Verantwortlichen behaupten, dass Pliny den ‚traditionellen Weg‘ des Notierens in die digitale Welt verlagere, scheint Pliny in der DH-Wissenschaftswelt keine Akzeptanz zu finden: Es gibt – wenn überhaupt – nur wenige Projekte, die das Tool nutzen (und es dokumentiert haben). Dies könnte aber auch ein Hinweis auf eine wenig entwickelte Tradition der geisteswissenschaftlichen Methodendiskussion sein, die zu einem Mangel an Publikationen über den Annotationsprozess innerhalb bestimmter Projekte führt.

Ein neueres Projekt zur Unterstützung von explorativen Annotationen ist das 3DH-Projekt⁵, bei dem zunächst Prototypen entwickelt werden (Kleymann et al. 2018). Eine Möglichkeit besteht darin, Textstellen tentativ gruppieren zu können, ohne sie direkt benennen zu müssen. Dies wiederum soll in einer intuitiv

⁴ Aus den „Proceedings of the Old Bailey“: <http://www.oldbaileyonline.org>

⁵ <http://threedh.net/3dh/>

benutzbaren und durchdachten Oberfläche auch technischen Laien möglich sein, wodurch ein Arbeitsschritt *vor* einer kategoriengeleiteten Annotation unterstützt würde.

Daneben wird eine explorative Annotation auch in Version 6 des Annotationswerkzeuges CATMA⁶ unterstützt: Durch simples Hervorheben von Textstellen, sowie durch eine freie Kommentarfunktion kann explorative Annotation leicht umgesetzt werden (Horstmann und Jacke 2020). Durch das iterative Ineinandergreifen mehrerer Annotationsmodi (Hervorhebung, Kommentierung, Kategorisierung) kann zudem ein Teil des in Abbildung 1 gezeigten Workflows direkt innerhalb eines Tools umgesetzt werden.

Für explorative Annotation ist das *inter-annotator agreement* nicht wichtig, da sie in erster Linie dem Ziel dient, ein Verständnis für wichtige Konzepte und mögliche Ansatzpunkte für ein Forschungsprojekt zu entwickeln.

4.2 Konzeptualisierende Annotation

Als Beispiel für die Konzeptualisierung von Annotation wollen wir zunächst Moretti (2013) nennen. Er beschreibt die Abweichung von der Definition des ‚character-space‘ von Woloch (2003). Die Operationalisierung dieser Literaturtheorie durch Annäherung an den Textraum, den eine Figur einnimmt, durch Kenntlichmachung dessen, wie viele Wörter diese in einem dramatischen Text spricht, stärkt die zugrunde liegende Theorie und führt „back from theories, through data, to the empirical world.“ (Moretti 2013, S. 4). Moretti hält dies für entscheidend für literarische Theorien, weil es einige Konzepte „actual‘ in the strong sense of the word“ mache (Moretti 2013, S. 4). In unserem Arbeitsablauf konzentriert sich dieses Projekt stark auf die Formalisierung eines theoretischen Begriffs, also die Übersetzung vom Konzept des Zeichenraums in den Raum des eigentlichen Textabschnitts. Die Annotation selbst ist trivial, aber der annotierte Text wird dann als Grundlage für die Interpretation verwendet.

Ein gründlicherer Versuch, die Annotation zur Entwicklung theoretischer Konzepte zu nutzen, wurde von Bögel et al. (2015) unternommen. Ziel des Projekts heureCLÉA ist es, zeitbezogene narrative Phänomene in literarischen Texten zu annotieren. Die veröffentlichten Richtlinien⁷ sind bereits spezifischer als die zugrundeliegende Theorie, da sie den Umgang mit z. B. hypothetischen Prolepsen definieren. Dieser Prozess der Verfeinerung des theoretischen Begriffs durch

⁶ <https://catma.de>

⁷ <http://heureclea.de/wp-content/uploads/2016/11/guidelinesV2.pdf>

Annotation kann auch als *shared task* durchgeführt werden (siehe auch Willand et al. 2020 in diesem Band, für einen *shared task* mit Fokus auf Erzählebenen).

Potenziell wird die systematische Konfrontation der Theorie mit Textbeispielen zu Implikationen für diese Theorie führen. Für diese Art von Annotation bildet das *inter-annotator agreement* eine Grundlage, um vordefinierte theoretische Konzepte intersubjektiv zu diskutieren. So ist IAA eine Metrik, die Informationen darüber liefern kann, wie spezifiziert eine Theorie ist und in welchem Maße sie die Definition von Indikatoren zur intersubjektiven Überprüfung ermöglicht.

Ein weiteres Beispiel für die konzeptualisierende Annotation ist die Koreferenzannotation. Die Annotation von Koreferenzen ist in der Computerlinguistik gut etabliert und wird durch bereits bestehende theoretische Begriffe und Richtlinien unterstützt (Pradhan et al. 2007; Dipper und Zinsmeister 2009; Riester und Baumann 2017). Die Anwendung dieser Richtlinien auf ‚neue‘ Textsorten zeigt jedoch, dass sie weiter verbessert werden müssen. Im Rahmen des QuaDrama-Projektes⁸ wurden dramatische Texte mit Koreferenzketten annotiert. Dabei ergaben sich neue Fragen, die in Weiterentwicklungen der Annotationsrichtlinien mündeten (Rösiger et al. 2018). Das Projekt entspricht dem Workflow wie folgt: Bestehende Annotationsrichtlinien wurden übernommen und der Annotationsprozess eingeleitet. Nach der Annotation der ersten Texte wurde eine Analyse durchgeführt und die Richtlinien wurden an die Daten und spezifischen Probleme angepasst. Weitere Texte wurden dann mit der neuen Version der Richtlinien annotiert. Bei der Anpassung der Richtlinien handelt es sich um einen Konzeptualisierungsschritt, da die neuen Richtlinien neue Erkenntnisse widerspiegeln, die aus der Betrachtung konkreter Koreferenzphänomene gewonnen wurden.

Ein Beispiel, bei dem die theoretische Vorarbeit eine größere Rolle spielt, ist die Annotation von Gender-Stereotypen im m*w-Projekt (Schumacher und Flüh 2020)⁹. Ausgehend von theoretischen Unterscheidungen von Butler (zwischen Genus, Gender Identity und Gender Performanz) wurde weitere Literatur herangezogen, um zu annotierbaren Kategorien zu gelangen, bevor die eigentliche Annotation startete (die dann ihrerseits wieder zu kategoriellen Verschiebungen führte).

4.3 Erklärende Annotation

Ein Beispiel für ein erklärendes Annotationsprojekt in einem frühen Stadium ist die in Nantke und Schlupkothén 2018 vorgestellte Arbeit. Die Autor*innen kon-

⁸ <https://quadrama.github.io>

⁹ <https://msternchenw.de>

zentrieren sich auf die Kommentierung von intertextuellen Referenzen, um mögliche Interpretationen eines Textes zu formalisieren. Nur eine Teilmenge der vorgeschlagenen Formalisierungen sind tatsächlich textuelle Annotationen im engeren Sinne – andere sind Beziehungen zwischen textuellen Annotationen oder zwischen textuellen Annotationen und (digitalen Darstellungen von) historischen Kontexten. Auf technischer Ebene werden sowohl die Annotationen als auch die Beziehungen durch *semantic-web*-Technologien dargestellt. Wichtig ist, dass diese Annotationen kein einzelnes Phänomen, sondern eine große Anzahl an ‚grundlegenden Annotationen‘ zu verschiedenen Phänomenen abdecken. Angesichts der Komplexität dieser Annotationen erscheint ein groß angelegtes Annotationsprojekt schwierig zu realisieren – solche Annotationen werden hauptsächlich für einen einzigen Text erstellt. Da zudem keine festen Kategorien existieren (müssen), ist ein *inter-annotator agreement* schwer zu berechnen.

Als zweites Beispiel sei hier das Projekt TEASys¹⁰ genannt. Ziel des Projektes ist es, erklärende Annotationen systematisch zur Unterstützung des Leseverständnisses zu erproben. Erklärungsbedürftigen Stellen englischer Literatur werden dabei (u. a. in Seminaren, also im Unterrichtskontext) Erklärungen beigelegt, die auf verschiedenen Ebenen strukturiert sind. So können u. a. historische Verwendungen von Wörtern oder Formulierungen, intertextuelle Verweise oder Interpretationsansätze ergänzt werden (Bauer und Zirker 2017).

In Bezug auf den in Abbildung 1 dargestellten Arbeitsablauf verwenden erklärende Annotationen theoretische Begriffe zur grundlegenden Bestandsaufnahme von Textnachweisen (möglichst unter Verwendung von Annotationsrichtlinien), ohne darauf abzuzielen, diese zu verbessern. Stattdessen folgen diese Projekte dem rechten Pfad im Workflow, was zu einem annotierten Text führt, gefolgt von einer Interpretation oder einer Begründung der Interpretation mithilfe der Annotationen.

4.4 Automatisierungsorientierte Annotation

Die letzte Art der Annotation, die wir diskutieren wollen, ist die automatisierungsorientierte Annotation, die in der Computerlinguistik weit verbreitet ist. Dabei besteht das Ziel der Annotation darin, Daten bereitzustellen, um statistische Zusammenhänge zwischen Oberflächenmerkmalen und den jeweiligen Zielkategorien zu ermitteln. Die annotierten Daten fungieren dann zum einen als Trainings- und

¹⁰ <http://www.annotation.es.uni-tuebingen.de>

zum anderen als Testdaten, also für die Evaluation der erstellten automatischen Erkenner.

Ein prominentes Beispiel für Annotationen, die als Input für einen vollautomatischen Ansatz verwendet werden, ist die Annotation von Wortarten. Part-of-Speech-Tagging (PoS-Tagging) ist eine konzeptionell klare Aufgabe, was sich im hohen *inter-annotator agreement* zeigt, das für diese Aufgabe erreicht wird. Das kürzlich veröffentlichte GRAIN-Korpus (Schweitzer et al. 2018) beispielsweise enthält Annotationen von drei Annotator*innen für deutsche Radiointerviews, die recht komplexe und spontane Rede umfassen. In ihrer Arbeit geben die Autoren ein *pairwise* Cohen's κ (Cohen 1960) von 0,97 an, das allgemein als fast perfekte Übereinstimmung gilt. Die Tatsache, dass die Annotation vom Menschen konsequent durchgeführt werden kann, ist eine notwendige Voraussetzung für die Entwicklung von automatischen Werkzeugen. In der Folge war das PoS-Tagging eine der ersten CL-Aufgaben, bei denen die Leistung von automatischen Werkzeugen mit einer Genauigkeit von über 97 % ein zufriedenstellendes Niveau erreicht hat (cf. Manning 2011) und heute zumindest für Standardtexte als eine fast gelöste Aufgabe gilt.

Das PoS-Tagging wurde auch auf Texte aus der DH-Domäne angewendet, z. B. historische Texte, bei denen die Leistung von Standardwerkzeugen natürlich nicht zufriedenstellend ist. Schulz und Kuhn (2016) konnten jedoch zeigen, dass für Mittelhochdeutsch bereits eine geringe Menge an annotierten Daten (z. B. rund 200 Sätze) zu tragfähigen Ergebnissen von automatischen Systemen führen kann.

Ein weiteres Beispiel für Annotation, die letztlich der Automatisierung dient, ist in dem Beitrag von Klinger et al. (2020), in diesem Band ab S. 238, dargestellt. Dabei ist das Ziel, Emotionen zu kategorisieren, die von literarischen Figuren empfunden werden. Dazu wurde ein Korpus annotiert und anschließend als Trainingsmaterial für ein maschinelles Lernverfahren verwendet. Auftretende Hürden bei der Annotation wurden gelöst, indem die Annotationsrichtlinien erweitert wurden.

5 Diskussion und Schlussfolgerungen

In diesem Beitrag zeigen wir, dass Annotation nicht nur als Mittel zur Erstellung von Trainingsmaterial für maschinelle Lernansätze dienen kann. Eine als Forschungsmethode verstandene Annotation kann zur Entwicklung eines fokussierten Verständnisses relevanter Konzepte, die in Texten zu finden sind, sowie als Instrument zur Spezifikation und Verifikation theoretischer oder prätheore-

tischer Konzepte dienen. Dies ist besonders fruchtbar für Disziplinen wie die Literaturwissenschaft, in denen Konzepte im wissenschaftlichen Diskurs oft implizite Vorannahmen mitbringen und daher unterspezifiziert wirken, was eine intersubjektive Anwendung erschwert.

Im Allgemeinen kann die Kommentierung von nicht-standardisierten (aus Sicht der Computerlinguistik) Texten dazu beitragen, neue Phänomene aufzudecken, die eine Anpassung oder Erweiterung von Annahmen erfordern. So muss beispielsweise die computerlinguistische Annahme, dass es eine *ground truth* – eine einzige korrekte Annotation – gibt, für literarische Textkonzepte möglicherweise ausgeweitet werden, da das Lesen und Interpretieren eines Textes unterschiedliche und dennoch korrekte Lesarten zulässt. Es bleibt eine Herausforderung für die Methoden des maschinellen Lernens, wie mit diesen ‚echten‘ Unklarheiten in Bezug auf das Training und die Bewertung automatischer Systeme umgegangen werden kann.

Eine weitere Überlegung, die diese verschiedenen Arten von Annotationen auslösen, ist die Wahl des Annotationswerkzeugs: Annotationswerkzeuge, die in der Computerlinguistik entwickelt wurden (z. B. WebAnno, Yimam et al. 2013, oder MMAX2, Müller und Strube 2006), treffen selbstverständlich die dort verwendeten Annahmen. So beinhalten sie typischerweise eine Methode zum Vergleichen von Annotationen und die Annotationskategorien und -richtlinien werden im Voraus definiert.

Explorative und auch erklärende Annotation haben andere Anforderungen: Das Annotationswerkzeug CATMA 6 unterstützt die Annotation freier Kommentare oder simpler Hervorhebungen, das im Rahmen des 3DH-Projekts (Kleymann et al. 2018) entwickelte Tool ermöglicht die Markierung beliebiger Textbereiche und Interaktion mit ihnen (z. B. Gruppierung und/oder Visualisierung). Das Annotationssystem TEASys verwendet Kategorien, aber Kategorien für Freitext-Kommentare.

In jedem Fall ist das Verhältnis zwischen der vom Tool angebotenen Funktionalität und dem Ziel des Annotationsprozesses noch ein wenig erforschter Bereich. Leider scheint es kaum Publikationen zu geben, in denen die Verwendung der Annotationsmethodik als Mittel zur Erforschung neuer Texte oder zur Schärfung von Forschungsfragen dokumentiert wird, während Beschreibungen und Diskussionen von automatisierungsorientierten Annotationen häufig zu finden sind.

Zusammenfassend haben wir einen Workflow für Annotationen beschrieben, die in den DH durchgeführt werden. Der Workflow soll so offen und flexibel wie möglich sein, um den verschiedenen möglichen Perspektiven und Bereichen, die in den DH zusammenkommen, Rechnung zu tragen, während er sich gleichzeitig auf Schritte konzentriert und solche erfordert, die von allen Annotationsunterneh-

mungen geteilt werden sollten. Wir definieren vier Hauptziele, die die verschiedenen Bereiche von DH verfolgen könnten: Explorative, konzeptionelle, erklärende und automatisierungsorientierte Ziele. Wir besprechen den Zweck und die Unterschiede der einzelnen Ziele auf allgemeiner Ebene, gefolgt von einer Übersicht konkreter Projekte in den DH mit einem dieser Ziele. Diese Übersicht zeigt auch den Einsatz des Workflows in verschiedenen Situationen und unterstreicht seine Flexibilität. Wir glauben, dass unser Workflow für alle Arten von DH-Zielen allgemein anwendbar ist und hoffen, dass in Zukunft weitere Projekte Annotation nutzen werden, um alte Fragen der Geisteswissenschaften aus einer neuen Perspektive zu betrachten.

Literatur

- Bauer, Matthias und Angelika Zirker (2017). „Explanatory Annotation of Literary Texts and the Reader: Seven Types of Problems“. In: *International Journal of Humanities and Arts Computing* 11.2, S. 212–232. doi: 10.3366/ijhac.2017.0193.
- Bögel, Thomas, Michael Gertz, Evelyn Gius, Janina Jacke, Jan Christoph Meister, Marco Petris und Jannik Strötgen (2015). „Collaborative Text Annotation Meets Machine Learning: heure-CLÉA, a Digital Heuristic of Narrative“. In: *DHCommons* 1. doi: 10.5281/zenodo.3240591.
- Bradley, John (2008). „Thinking about interpretation: Pliny and scholarship in the humanities“. In: *Literary and Linguistic Computing* 23.3, S. 263–279. doi: 10.1093/lc/fqn021.
- Bradley, John (2012). „Towards a Richer Sense of Digital Annotation: Moving Beyond a ‚Media‘ Orientation of the Annotation of Digital Objects“. In: *Digital Humanities Quarterly* 6.2. URL: <http://www.digitalhumanities.org/dhq/vol/6/2/000121/000121.html> (besucht am 1. Juni 2020).
- Cohen, Jacob (1960). „A Coefficient of Agreement for Nominal Scales“. In: *Educational and Psychological Measurement* 20.1, S. 37–46.
- Dipper, Stefanie und Heike Zinsmeister (2009). „Annotating Discourse Anaphora“. In: *Proceedings of the Third Linguistic Annotation Workshop (LAW III)*. ACL-IJCNLP. Singapore, S. 166–169.
- Genette, Gérard (1980). *Narrative Discourse – An Essay in Method*. Übers. von Jane E. Lewin. Ithaca, New York: Cornell University Press.
- Gius, Evelyn und Janina Jacke (2017). „The Hermeneutic Profit of Annotation: On Preventing and Fostering Disagreement in Literary Analysis“. In: *International Journal of Humanities and Arts Computing* 11.2, S. 233–254. doi: 10.3366/ijhac.2017.0194.
- Horstmann, Jan und Janina Jacke (2020). „Interpretationsspielräume. Undogmatisches Annotieren literarischer Texte in CATMA 6“. In: *Dhd 2020 Digital Humanities: Spielräume. Conference abstracts*. Paderborn, S. 154–158.
- Hovy, Eduard und Julia Lavid (2010). „Towards a ‚Science‘ of Corpus Annotation: A New Methodological Challenge for Corpus Linguistics“. In: *International Journal of Translation Studies* 22.1, S. 13–36.

- Kleymann, Rabea, Jan Christoph Meister und Jan-Erik Stange (2018). „Perspektiven kritischer Interfaces für die Digital Humanities im 3DH-Projekt“. In: *Abstracts der DHd: Kritik der digitalen Vernunft*. Köln: Digital Humanities im deutschsprachigen Raum e.V., S. 279–284.
- Klinger, Roman, Evgeny Kim und Sebastian Padó (2020). „Emotion Analysis for Literary Studies“. In: *Reflektierte Algorithmische Textanalyse*. Hrsg. von Nils Reiter, Axel Pichler und Jonas Kuhn. Berlin: De Gruyter, S. 238–268.
- Manning, Christopher D. (2011). „Part-of-Speech Tagging from 97% to 100%: Is It Time for Some Linguistics?“ In: *Computational Linguistics and Intelligent Text Processing*. Lecture Notes in Computer Science. Berlin/Heidelberg: Springer, S. 171–189. doi: 10.1007/978-3-642-19400-9_14.
- Mathet, Yann, Antoine Widlöcher und Jean-Philippe Métivier (2015). „The Unified and Holistic Method Gamma (γ) for Inter-Annotator Agreement Measure and Alignment“. In: *Computational Linguistics* 41.3, S. 437–479. doi: 10.1162/COLI_a_00227.
- McCarty, Willard (2020). „Making and studying notes: Towards a cognitive ecology of annotation“. In: *Annotations in Scholarly Editions and Research*. Hrsg. von Julia Nantke und Frederik Schlupkothen. Berlin: De Gruyter, S. 269–295.
- Moretti, Franco (2013). „Operationalizing‘: or, the function of measurement in modern literary theory“. Pamphlet 6. Stanford Literary Lab. URL: <http://litlab.stanford.edu/LiteraryLabPamphlet6.pdf> (besucht am 1. Juni 2020).
- Müller, Christoph und Michael Strube (2006). „Multi-level annotation of linguistic data with MMAX2“. In: *Corpus Technology and Language Pedagogy: New Resources, New Tools, New Methods*. Hrsg. von Sabine Braun, Kurt Kohn und Joybrato Mukherjee. New York: P. Lang, S. 197–214.
- Nantke, Julia und Frederik Schlupkothen (2018). „Zwischen Polysemie und Formalisierung: Mehrstufige Modellierung komplexer intertextueller Relationen als Annäherung an ein ‚literarisches‘ Semantic Web“. In: *Abstracts der DHd: Kritik der digitalen Vernunft*. Köln: Digital Humanities im deutschsprachigen Raum e.V., S. 345–349.
- Pagel, Janis, Nils Reiter, Ina Rösiger und Sarah Schulz (2018). „A Unified Text Annotation Workflow for Diverse Goals“. In: *Proceedings of the Workshop for Annotation in Digital Humanities (annDH)*. Sofia, Bulgarien, S. 31–36. URL: <http://ceur-ws.org/Vol-2155/pagel.pdf> (besucht am 1. Juni 2020).
- Pichler, Axel und Nils Reiter (2020). „Reflektierte Textanalyse“. In: *Reflektierte Algorithmische Textanalyse*. Hrsg. von Nils Reiter, Axel Pichler und Jonas Kuhn. Berlin: De Gruyter, S. 43–59.
- Pradhan, Sameer S., Lance Ramshaw, Ralph Weischedel, Jessica Macbride und Linnea Micciulla (2007). „Unrestricted coreference: Identifying entities and events“. In: *ICSC '07: Proceedings of the International Conference on Semantic Computing*, S. 446–453. doi: 10.1109/ICSC.2007.106.
- Pustejovsky, James und Amber Stubbs (2012). *Natural Language Annotation for Machine Learning: A Guide to Corpus-Building for Applications*. Sebastopol, Boston, Farnham: O'Reilly Media.
- Riester, Arndt und Stefan Baumann (2017). „The RefLex Scheme – Annotation Guidelines“. Sin-SpeC. Working papers of the SFB 732 14. University of Stuttgart.
- Rösiger, Ina, Sarah Schulz und Nils Reiter (2018). „Towards Coreference for Literary Text: Analyzing Domain-Specific Phenomena“. In: *Proceedings of the Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*. Santa Fe, USA, S. 129–138.

- Schulz, Sarah und Jonas Kuhn (2016). „Learning from Within? Comparing PoS Tagging Approaches for Historical Text.“ In: *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*. Portorož, Slowenien: European Language Resources Association (ELRA), S. 4316–4322.
- Schumacher, Mareike und Marie Flüh (2020). „m*w Figurengender zwischen Stereotypisierung und literarischen und theoretischen Spielräumen Genderstereotype und -bewertungen in der Literatur des 19. Jahrhunderts“. In: *Dhd 2020 Digital Humanities: Spielräume. Conference abstracts*. Paderborn, S. 162–166.
- Schweitzer, Katrin, Kerstin Eckart, Markus Gärtner, Agnieszka Faleńska, Arndt Riester, Ina Rösiger, Antje Schweitzer, Sabrina Stehwien 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)*. Miyazaki, Japan: European Language Resources Association (ELRA), S. 2887–2895.
- Willand, Marcus, Evelyn Gius und Nils Reiter (2020). „SANTA: Idee und Durchführung“. In: *Reflektierte Algorithmische Textanalyse*. Hrsg. von Nils Reiter, Axel Pichler und Jonas Kuhn. Berlin: De Gruyter, S. 391–422.
- Woloch, Alex (2003). *The One Vs. the Many: Minor Characters and the Space of the Protagonist in the Novel*. Princeton, New Jersey: Princeton University Press.
- Yimam, Seid Muhie, Iryna Gurevych, Richard Eckart de Castilho und Chris Biemann (2013). „WebAnno: A Flexible, Web-based and Visually Supported System for Distributed Annotations“. In: *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics: System Demonstrations*. Association for Computational Linguistics, S. 1–6. URL: <http://www.aclweb.org/anthology/P13-4001> (besucht am 1. Juni 2020).

