

Artikel aus:  
Zeitschrift für digitale Geisteswissenschaften

Titel:  
Generalisierungen als literarisches Phänomen. Charakterisierung, Annotation und automatische Erkennung

---

Autor/in:  
Luisa Gödeke

Kontakt: [luisa.goedeke@uni-goettingen.de](mailto:luisa.goedeke@uni-goettingen.de)  
Institution:  
GND: [1228354081](#) ORCID: [0000-0003-3230-8993](#)

---

Autor/in:  
Florian Barth

Kontakt: [barth@sub.uni-goettingen.de](mailto:barth@sub.uni-goettingen.de)  
Institution:  
GND: [1228353344](#) ORCID: [0000-0003-3408-7311](#)

---

Autor/in:  
Tillmann Dönicke

Kontakt: [tillmann.doenicke@uni-goettingen.de](mailto:tillmann.doenicke@uni-goettingen.de)  
Institution:  
GND: [122835300X](#) ORCID: [0000-0001-5844-9167](#)

---

Autor/in:  
Hanna Varachkina

Kontakt: [hanna.varachkina@stud.uni-goettingen.de](mailto:hanna.varachkina@stud.uni-goettingen.de)  
Institution:  
GND: [1269647261](#) ORCID: [0000-0002-5832-3983](#)

---

Autor/in:  
Anna Mareike Weimer

Kontakt: [annamareike.hofmann@uni-goettingen.de](mailto:annamareike.hofmann@uni-goettingen.de)  
Institution:  
GND: [1228354707](#) ORCID: [0000-0002-2352-4793](#)

---

Autor/in:  
Benjamin Gittel

Kontakt: [benjamin.gittel@uni-goettingen.de](mailto:benjamin.gittel@uni-goettingen.de)  
Institution:  
GND: [1041237715](#) ORCID: [0000-0002-1855-0049](#)

---

Autor/in:  
Anke Holler

Kontakt: [anke.holler@phil.uni-goettingen.de](mailto:anke.holler@phil.uni-goettingen.de)  
Institution:  
GND: [142630578](#) ORCID: [0000-0003-0566-1150](#)

---

Autor/in:  
Caroline Sporleder

Kontakt: [csporled@cs.uni-goettingen.de](mailto:csporled@cs.uni-goettingen.de)  
Institution:  
GND: [142630578](#) ORCID: [0000-0003-0566-1150](#)

---

DOI des Artikels:  
[10.17175/2022\\_010](#)

Nachweis im OPAC der Herzog August Bibliothek:  
[1817977911](#)

Erstveröffentlichung:  
24.11.2022

Lizenz:

Sofern nicht anders angegeben



Medienlizenzen:  
Medienrechte liegen bei den Autor\*innen

Letzte Überprüfung aller Verweise:  
20.10.2022

GND-Verschlagwortung:

[Computerlinguistik](#) | [Erzähltheorie](#) | [Quantifizierung \(Linguistik\)](#) | [Literaturwissenschaft](#) |

Zitierweise:

Luisa Gödeke, Florian Barth, Tillmann Dönicke, Hanna Varachkina, Anna Mareike Weimer, Benjamin Gittel, Anke Holler: Generalisierungen als literarisches Phänomen. Charakterisierung, Annotation und automatische Erkennung. In: Zeitschrift für digitale Geisteswissenschaften. Wolfenbüttel 2022. PDF Format ohne Paginierung. Als text/html abrufbar unter DOI: [10.17175/2022\\_010](https://doi.org/10.17175/2022_010).

Luisa Gödeke, Florian Barth, Tillmann Dönicke, Hanna  
Varachkina, Anna Mareike Weimer, Benjamin Gittel, Anke Holler

## Generalisierungen als literarisches Phänomen. Charakterisierung, Annotation und automatische Erkennung

---

### Abstracts

Generalisierungen in Erzähltexten dienen typischerweise nicht oder nicht ausschließlich dazu, die erzählte Welt aufzubauen, sondern um Informationen über die Erzählinstanz, den Sinn des Erzählten oder die reale Welt bereitzustellen. Der Aufsatz leistet einen Beitrag zur linguistisch-literaturwissenschaftlichen Beschreibung von Generalisierungen, entwickelt ein Tagset für ihre Identifikation und Klassifikation und stellt die Ergebnisse ihrer kollaborativen Annotation in einem diachronen Korpus (1616–1930) dar. Die schließlich vorgestellten regelbasierten und statistischen Tagger zur automatischen Erkennung von Generalisierungen gewährleisten den Zugang zu einer breiten Basis von Textbeispielen und ermöglichen eine empirisch gesättigte Analyse der Funktionen von Generalisierungen und der assoziierten narratologischen Phänomene.

Generalisations in narrative texts typically do not or not exclusively serve to devise the narrated world, but provide information about the narrative instance, the meaning of what is told or the real world. The paper contributes to the linguistic-literary description of generalisations, develops a tagset for their identification and classification, and presents the results of their collaborative annotation in a diachronic corpus (1616–1930). Finally, the paper presents a rule-based and a statistical tagger for the automatic recognition of generalisations that allow users to access a variety of examples of the phenomenon and can be used in the analysis of the functions of generalisations and the associated narratological phenomena.

## 1. Einleitung

Das wohl bekannteste Beispiel für Generalisierungen in der literaturwissenschaftlichen Forschung ist der erste Satz aus Tolstoj's *Anna Karenina*, an dem eine Vielzahl aktueller Forschungspositionen erläutert werden können:

(1) Alle glücklichen Familien sind einander ähnlich; aber jede unglückliche Familie ist auf ihre besondere Art unglücklich.<sup>1</sup>

Martínez und Scheffel analysieren dieses Beispiel in ihrer Diskussion des *theoretischen Satzes* als eine »kommentierende Stellungnahme des Erzählers über die Welt überhaupt« und unterscheiden sie von mimetischen Sätzen, die »Informationen über die konkrete Beschaffenheit und das Geschehen in der erzählten Welt« vermitteln.<sup>2</sup> Bei genauerer Betrachtung fällt allerdings auf, dass diese vermeintlich strikte Trennung der Funktionen dem Beispiel nicht gerecht wird. Durch die Generalisierung entsteht augenscheinlich eine Distanz zur erzählten Welt,<sup>3</sup> wodurch der Satz als Bemerkung über die nicht-fiktionale Welt interpretiert werden kann – also als eine Behauptung über die Beschaffenheit von Familien im Allgemeinen.<sup>4</sup> Der Satz kann aber zusätzlich als Vorhersage über die Geschehnisse in der fiktionalen Welt interpretiert werden und somit als Leitfaden zum Verständnis des Werkes an sich dienen. In (1) kann die Generalisierung demnach auf eine explizite Erzählinstanz (*overt narrator*) hinweisen und zugleich als Indikator für sogenannte *nicht-fiktionale-Rede* (*non-fictional speech*)<sup>5</sup> gelten.<sup>6</sup> Vertreter\*innen einer Theorie der Literatur als soziale Institution betrachten generalisierende Aussagen wie (1) als *thematic statements*, also als Hinweise auf das Thema des betreffenden Werks.<sup>7</sup> In der Debatte um das Verhältnis von Wissen und Literatur werden Generalisierungen als Beispiele für sogenannte *explicit truths* oder *explizite Propositionen* herangezogen, die möglicherweise eine besondere Rolle für den kognitiven Wert fiktionaler Werke spielen.<sup>8</sup>

Beispiel (1) illustriert demnach, dass Generalisierungen im Erzähltext offenbar Aussagen darstellen, die typischerweise nicht oder nicht nur dem Aufbau der erzählten Welt dienen, sondern diverse weitere Funktionen erfüllen können. Nichtsdestoweniger spielen sie in der Narratologie nur eine untergeordnete Rolle. Häufig ist umgangssprachlich von *Verallgemeinerungen* die Rede; eine empirische Untersuchung von Generalisierungen hinsichtlich ihrer werkimmanenten und textexternen Funktionen blieb bislang aus. So beruhen die bis dato in der Forschung formulierten Hypothesen zur Funktion von Generalisierungen meist auf wenigen ausgewählten Beispielsätzen kanonisierter Werke. Diachrone Perspektiven, etwa hinsichtlich einzelner Autor\*innen, Epochen, literarischer Gattungen oder Strömungen, bleiben unausgeschöpft.

---

<sup>1</sup> Tolstoj 2012, S. 7.

<sup>2</sup> Scheffel / Martínez 2016, S. 104f.

<sup>3</sup> Nünning 1989, S. 50f.

<sup>4</sup> Vgl. die Diskussion in Gittel 2015, S. 148–168.

<sup>5</sup> Vgl. Searle 1975, S. 331f.; Konrad 2014, S. 424–474; Konrad 2017, S. 57.

<sup>6</sup> Chatman 1978, S. 243.

<sup>7</sup> Vgl. Lamarque / Olsen 1994, S. 328–331; Lahn / Meister 2013, S. 176f.

<sup>8</sup> Vgl. Petraschka 2014, S. 99–106; Vesper 2014; in anderem Vokabular Mikkonen 2009.

Im Mittelpunkt des vorliegenden Aufsatzes stehen zwei Generalisierungstagger: ein regelbasierter und ein statistischer Tagger. Die Entwicklung der Tagger soll dazu dienen, einerseits Generalisierungen formal (d. h. linguistisch) zu untersuchen. Andererseits sollen die mit Generalisierungen assoziierten Funktionen und Effekte im Erzähltext analysiert werden können, insbesondere vor dem Hintergrund diachroner narratologischer Fragestellungen. Dementsprechend wird mit einem Korpus deutschsprachiger fiktionaler Erzähltexte gearbeitet, die zwischen 1616 und 1930 veröffentlicht wurden. Der **Abschnitt 2** wird eingangs in die mit Generalisierungen assoziierten linguistischen Phänomene Quantifikation und Generizität und ihre computergestützte Operationalisierung einführen. Ein besonderes Augenmerk wird darauf liegen, die für uns zentrale Herausforderung für die Annotation von Generalisierungen zu bewältigen: Quantifikation und Generizität stellen einerseits klassische Gegenstände der semantischen Theoriebildung dar und sollen auf eine Weise annotiert werden, die eine linguistische Analyse der Annotation ermöglichen. Dennoch müssen annotierbare Indikatoren an der Satzoberfläche erkennbar sein, was insbesondere bei generischen Ausdrücken komplex ist. Diese Herausforderung ergibt sich aus unserer Engführung linguistischer und narratologischer Perspektiven auf das Thema Generalisierungen. Unser Ziel ist es, die Tagger insbesondere dafür zu nutzen, die Schnittstelle zwischen Formen und Funktionen von Generalisierungen im Erzähltext auszuleuchten. In **Abschnitt 3** schildern wir unser Annotationsvorgehen, das die in der Forschungsliteratur vorhandenen Tagsets zur Annotation von Quantifikations- bzw. Generizitätsphänomenen um ein intuitives und theorieübergreifendes Vorgehen für deutschsprachige fiktionale Texte ergänzt. Zudem wird das Annotationskorpus beschrieben sowie Ergebnisse des *Inter-Annotator-Agreements* dargestellt. Es folgt eine Erläuterung zur Implementierung der Tagger in **Abschnitt 4** und ihre Evaluation in **Abschnitt 5**.

## 2. Generalisierungen in Theorie und Operationalisierung

Generalisierungen können auf verschiedene Weise im Satz sprachlich markiert sein. Typische Indikatoren sind allquantifizierte und generische Ausdrücke, die zu einer Äußerung naturgemäß generalisierende Bedeutungsanteile beitragen können. Quantifikation und Generizität stellen dabei klassische Gegenstände der semantischen Theoriebildung dar, weswegen bereits zahlreiche detaillierte und differenzierte Analysen und Modelle zu ihrer adäquaten Beschreibung vorliegen. Daraus ergibt sich die zentrale Herausforderung für die Annotation und die automatische Erkennung von Generalisierungen in Erzähltexten: Einerseits müssen annotierbare Indikatoren für Generalisierungen an der Satzoberfläche identifiziert werden können, andererseits muss jedwedes Annotationsschema auch theoretisch plausibel mit den etablierten semantischen Theorieansätzen kompatibel sein.

Im Folgenden werden zunächst kurz die hier vorausgesetzten Annahmen der theoretischen Semantik zur Generizität und Quantifikation eingeführt. Daran anschließend werden relevante computerlinguistische Vorschläge zur Annotation generischer Ausdrücke diskutiert. Dabei wird auch deutlich werden, dass die vorgeschlagene Herangehensweise nur begrenzt auf Erzähltexte übertragbar ist.

### 2.1 Generizität und Quantifikation

Mit generischen Aussagen ist es möglich, über eine Klasse von Dingen oder Individuen gleichzeitig etwas auszusagen, entweder durch auf diese *Art* oder *Gattung* referierende Nominalphrasen (NPn) wie in Beispiel (2) oder auf Satzebene – in diesem Fall ist die Rede von charakterisierenden oder generischen Sätzen.<sup>9</sup> Diese teilen sich in entweder lexikalisch-charakterisierende Aussagen über Fakten (vgl. (3)) oder habituell-charakterisierende Aussagen über wiederkehrende Ereignisse auf (vgl. (4)). Beide Formen können auch im selben Satz auftreten (vgl. (3)).

(2) Löwen fressen Fleisch. (NP ›Löwen‹ referiert auf Gattung)

(3) Ein Löwe hat eine Mähne.<sup>10</sup> (lexikalisch-charakterisierend, kombiniert mit generisch verwendeter NP)

(4) John trinkt nach dem Essen ein Bier. (*habitueller Satz*)

<sup>9</sup> Krifka et al. 1995, S. 2.

<sup>10</sup> Ein wichtiger Unterschied zwischen Ein Löwe hat eine Mähne und Jeder Löwe hat eine Mähne besteht darin, dass generische Aussagen immer Ausnahmen erlauben, allquantifizierte Aussagen hingegen nicht. Aus linguistischer Perspektive ist es daher nicht angemessen, beiden Formen dieselben Wahrheitswerte zuzuordnen. Ergebnisse aus der Psycholinguistik stellen diese Annahme allerdings infrage, vgl. Leslie et al. 2011.

Allen diesen generischen Aussagen ist gemein, dass eindeutige Indikatoren für ihre generalisierende Lesart an der Satzoberfläche fehlen. Aus diesem Grund sind die aufgeführten Aussagen in der Regel ambig: Die indefinite Plural-NP Löwen in (2) ist eine Form, die zwar häufig generische Lesarten auslöst, aber auch spezifisch verwendet werden kann:

(5) Die Schließanlage im Zoo ist kaputt. Es sind Löwen ausgebrochen. (spezifische Verwendung)

Ebenso verhält es sich mit Massennomen im Singular. Nur in (6), nicht aber in (7) wird Gold generisch verwendet:

(6) Gold glänzt.

(7) Die Bank wurde ausgeraubt und es wurde Gold gestohlen.

Auch definite und indefinite NPn im Singular können generisch oder spezifisch interpretiert werden. Oft ist der Kontext der ausschlaggebende Faktor dafür, ob eine generische Lesart vorliegt, vgl. (8)–(10):

(8) Ein Löwe hat eine Mähne. (Indefinite NP, löst standardmäßig eine generalisierende Lesart aus)

(9) Der Löwe hat eine Mähne. (Definite NP, generalisierende Lesart je nach Kontext möglich)

(10) Ein Vater erklärt seinem Kind Eigenschaften von Wildtieren. Er sagt: Der Elefant hat einen Rüssel. Der Löwe hat eine Mähne. (Kontext triggert generalisierende Lesart)

(11) Ein Vater und sein Kind stehen im Zoo vor dem Löwenkäfig. Der Vater fragt das Kind, welches der Tiere eine Mähne hat. Das Kind zeigt auf den männlichen, ausgewachsenen Löwen und sagt: Der Löwe hat eine Mähne. (Kontext triggert eine spezifische Lesart)

Diese Beispiele illustrieren die hohe Kontextabhängigkeit generischer Interpretationen; anhand der sprachlichen Formen einer NP allein lässt sich noch nicht sicher bestimmen, ob diese NP generisch interpretiert wird oder nicht. Die Theorie der generalisierten Quantoren vorausgesetzt, nehmen wir an, dass charakterisierende und quantifizierte Sätze dieselbe dreiteilige Struktur aufweisen, die aus einem Quantor (Q), einem Restriktor (R) und einem nuklearen Skopus (S) besteht,<sup>11</sup> wobei jeder Bestandteil einen bestimmten eigenen Bedeutungsbeitrag leistet. Der Quantor allerdings muss in natürlichsprachlichen Sätzen nicht zwingend an der Satzoberfläche ablesbar sein. Dies ist insbesondere bei generischen Sätzen der Fall. Wir erläutern diese Struktur an einem Beispiel:

(12) Jede Wissenschaftlerin ist klug.

Ein Quantor kann als eine Relation zwischen zwei Mengen aufgefasst werden. Demnach drückt jede in (12) die Relation zwischen der Menge der Wissenschaftlerinnen und der Menge der klugen Personen aus. Dabei legt er fest, dass die Menge der Wissenschaftlerinnen vollständig in der Menge der klugen Personen enthalten ist, also eine Teilmenge der klugen Personen bildet. Die NP Wissenschaftlerin stellt den Restriktor des Quantors jede dar und beschreibt die Teilmenge, über die etwas ausgesagt wird: die Menge aller Wissenschaftlerinnen. Das Prädikat ist klug bildet den (nuklearen) Skopus des Quantors, wodurch die Eigenschaft, die dieser Menge zugeschrieben wird, erfasst ist. In Beispielen wie (12) sind Q, R und S noch recht leicht oberflächennah zu erkennen, bei Satzmaterial aus Erzähltexten wird sie ungleich komplexer, wie in Abschnitt 2.3 gezeigt wird.

Sowohl mit quantifizierten als auch mit generischen Äußerungen können Generalisierungen ausgedrückt werden. Das mitunter größere Interesse, auch in der computerlinguistischen Forschung, liegt auf generischen Aussagen, da diese nicht durch Indikatoren an der Satzoberfläche gekennzeichnet und damit schwerer formal zu beschreiben und automatisch zu erkennen

---

<sup>11</sup> Vgl. Lewis 1975; Barwise / Cooper 1981; Heim 1982; Benthem 1983; Krifka et al. 1995, S. 25; Kamp 2002.

sind. In der Narratologie wiederum ist nicht klar, welche Formen von Generalisierungen prominenter in fiktionalen Texten sind oder welche Formen mit welchen Funktionen im Text eher korrelieren. Eine Einschränkung auf den einen oder den anderen linguistischen Forschungsschwerpunkt ist daher zum jetzigen Zeitpunkt nicht sinnvoll.

## 2.2 Operationalisierungsansätze von Generizität und Quantifikation im Überblick

Es existiert eine beträchtliche Anzahl an Annotationsschemata, die dazu dienen sollen, quantifizierte oder generische Ausdrücke zu markieren. Ein Großteil dieser Vorschläge bezieht sich ausschließlich auf generische NPn<sup>12</sup> und arbeitet entsprechend mit einer limitierten Anzahl an Tags. So annotieren Friedrich et al. 2015, ob das Subjekt eines Satzes generisch auf eine *Art* oder *Klasse* referiert, und wenn ja, ob der umliegende Satz eine charakterisierende oder spezifische Aussage über diese Klasse trifft. Entsprechend dieser Unterteilung nutzen sie ein dreiteiliges Tagset:

Tagset von Friedrich et al.:

- GEN; gen: generischer Satz mit generischem Subjekt
- NON-GEN; non gen: nicht-generischer Satz mit nicht-generischem Subjekt
- NON-Gen; gen: nicht generischer Satz mit generischem Subjekt

Durch die Fokussierung auf generische Subjekte werden Sätze mit generischen Objekten, aber nicht-generischen Subjekten aus der Erhebung ausgeschlossen, obwohl auch in diesen Generalisierungen vorliegen können, wie Beispiel (13) illustriert:

(13) John mag Kaffee.

Friedrich / Pinkal widmen sich der automatischen Erkennung von habituellen Sätzen anhand des lexikalischen Aspekts des Vollverbs.<sup>13</sup> Dieser Ansatz ist strikt linguistisch motiviert: Der Aspekt des Vollverbs eines Satzes (dynamisch oder statisch) bestimmt den Aspekt des Satzes. Hat das Vollverb eines Satzes dynamischen Aspekt, kann dieser entweder auf habituelle oder episodische Sätze hindeuten. Statische Verben hingegen führen üblicherweise zu statischen Sätzen. Je nach Aspekt verändert sich die Funktion eines Satzes im (narrativen) Diskurs, so werden generische Sätze üblicherweise mit Erzählpausen assoziiert.<sup>14</sup>

Schließlich stellen Friedrich et al. 2016 auch ein integriertes Vorgehen vor, um generische NPn und habituelle Ausdrücke simultan zu annotieren und automatisch zu erkennen.<sup>15</sup> Das Ziel ihres Ansatzes ist die automatische Erkennung sogenannter *situation entity types* (SE types). Darunter fallen neben Ereignissen und Fakten auch sogenannte *generative states*, die einerseits generische Äußerungen und andererseits habituelle Äußerungen sein können. Bei der Annotation auf Satzebene wird der *main referent*, üblicherweise das Subjekt, als generisch oder nicht-generisch klassifiziert. Zudem wird das Vollverb hinsichtlich des lexikalischen Aspekts (dynamisch oder statisch) annotiert. Je nach Ergebnis werden dann die Tags *generic sentence*, *generalizing sentence*, *state* oder *event* vergeben.<sup>16</sup> Die Problematik, Sätze mit nicht-generischen Subjekten aber generischen Objekten (vgl. (13)) durch die Annotation auszuschließen, bleibt hier zwar bestehen, kann aber durch die Klasse *state* aufgefangen werden.

Insgesamt ist die automatische Erkennung von Generalisierungen also nicht nur mit der automatischen Erkennung von generischen NPn oder Habitualität, sondern durch die Erkennung von (lexikalischem) Aspekt auch mit der Identifikation von temporalen Strukturen in Erzähltexten verknüpft. Nach unserer Kenntnis ist bis heute noch kein Versuch unternommen worden, Generalisierungen spezifisch in deutschsprachigen fiktionalen Erzähltexten zu erkennen. Wie sich in diesem Abschnitt herausgestellt hat, ist der Großteil der vorhandenen Ansätze computerlinguistisch ausgerichtet und bezieht sich meist auf faktuale Textsorten wie Zeitungstexte, (Online-) Lexikoneinträge und dergleichen. Außerdem wird überwiegend mit englischsprachigen Texten gearbeitet. Eine direkte Übertragung dieser Ansätze auf deutschsprachige fiktionale Texte ist nicht ohne Weiteres möglich – die Gründe hierfür werden im folgenden Abschnitt diskutiert.

<sup>12</sup> Vgl. Mitchell et al. 2003; Walker et al. 2006; Reiter / Frank 2010; Friedrich et al. 2015.

<sup>13</sup> Vgl. Friedrich / Pinkal 2015.

<sup>14</sup> Vgl. Carlson 2006.

<sup>15</sup> Vgl. Friedrich et al. 2016.

<sup>16</sup> Friedrich et al. 2015, S. 41.

## 2.3 Operationalisierung von Generalisierungen in fiktionalen Erzähltexten

Die im vorangegangenen Abschnitt erläuterten Tagsets sind computerlinguistisch motiviert und daher auf bestimmte semantische oder syntaktische Formen begrenzt. Da für uns nicht von vornherein klar ist, welche Formen von Generalisierungen für narratologische Fragestellungen relevant sind, ist ein solches ausschließlich an der Form orientiertes Vorgehen im Fall fiktionaler Erzähltexte ungeeignet. Es bestünde die Gefahr, einen Teil generalisierender Aussagen aus der Erhebung von vornherein auszuschließen und damit das Forschungsziel zu verfehlen. Hinzu kommt die Schwierigkeit, dass die hier vorgestellten Tagger auch auf älteren Sprachstufen bzw. nicht-kanonischen Ausdrucksweisen operieren können müssen, wenn auch diachrone narratologische Forschungsfragen beantwortet werden sollen.

Die drei zentralen Herausforderungen für die Annotation von Generalisierungen lassen sich wie folgt umreißen: Die erste Schwierigkeit besteht darin, dass der Tagger Generalisierungen in Texten erkennen können muss, die ab 1600 veröffentlicht wurden. Speziell in den älteren Texten treten nicht nur typische lexikalische Ausdrücke und syntaktische Konstruktionen der jeweiligen Sprachstufen auf, sondern darüber hinaus lässt sich auch eine höhere Zahl an komplexen, teils mehrfach rekursiv eingebetteten Sätzen beobachten. Das ist nicht nur der Struktur des Deutschen geschuldet, sondern auch ein Charakteristikum fiktionaler Erzähltexte. In diesen Texten im Detail Restriktor, Skopus und Quantor zu identifizieren und adäquat zu annotieren, ist eine anspruchsvolle und zeitintensive Aufgabe, die auch ein hohes Maß an (linguistischem) Training für die Annotator\*innen beansprucht. Dies sei anhand des Beispiels (14) kurz illustriert:

(14) Wenn Luciane, meine Tochter, die für die Welt geboren ist, sich dort für die Welt bildet, [...]; wenn sie durch Freiheit des Betragens, Anmut im Tanze, schickliche Bequemlichkeit des Gesprächs sich vor allen auszeichnet und durch ein angebornes herrschendes Wesen sich zur Königin des kleinen Kreises macht, wenn die Vorsteherin dieser Anstalt sie als kleine Gottheit ansieht, die nun erst unter ihren Händen recht gedeiht, die ihr Ehre machen, Zutrauen erwerben und einen Zufluß von andern jungen Personen verschaffen wird, wenn [...]: so ist dagegen, was sie schließlich von Ottilien erwähnt, nur immer Entschuldigung auf Entschuldigung [...].<sup>17</sup>

Die durch wenn eingeleiteten konditionalen Nebensätze in Beispiel (14) fungieren augenscheinlich als Restriktoren für den durch so eingeleiteten Nebensatz und damit für nur eine Skopus-Einheit. Es ist allerdings unklar, ob hier quantifizierte Aussagen aneinandergereiht werden oder ob die einzelnen Restriktoren eine Einheit bilden.

Zweitens ist es in der jetzigen Untersuchungsphase zu früh, schon im Vorhinein abschließend festzulegen, welche Formen von Generalisierungen für die narratologischen Fragestellungen im Zuge einer quantitativen Datenerhebung relevant würden. Vorrangig scheint vielmehr zu sein, Indikatoren für Generalisierungen an der Satzoberfläche aufzuspüren und einen Umgang für die Fälle zu finden, in denen mehrere konkurrierende Marker zugleich auftreten. So kommt es in fiktionalen Texten gehäuft vor, dass syntaktische Strukturen mehrere Lesarten auslösen:

(15) Wer ein Übel los sein will, der weiß immer, was er will.<sup>18</sup>

Hier liegt einerseits eine Generalisierung über Personen vor, die ein Übel loswerden wollen und andererseits über die Situationen, in denen sie das Übel loswerden wollen, nämlich immer. Welche der beiden Generalisierungen über die andere regiert, ist ambig und die Auflösung dieser Ambiguität ist für die automatische Erkennung nicht von zentralem Interesse. Eine linguistisch exakte Annotation wäre in solchen Fällen zeitaufwändig und fehleranfällig.

Drittens ist, wie bereits erläutert wurde, nicht nur der Überfluss von Markern an der Satzoberfläche eine Herausforderung, sondern auch die koverte Markierung von Generalisierungen etwa durch generische NPn.

(16) Die Landleute haben die rechten Kenntnisse; ihre Mitteilungen aber sind konfus und nicht ehrlich. Die Studierten aus der Stadt und von den Akademien sind wohl klar und ordentlich, aber es fehlt ihnen an der unmittelbaren Einsicht in die Sache.<sup>19</sup>

In (16) wird eine Aussage über die Gruppe von Landleuten und die Gruppe von Studierten gemacht. Diesen beiden Gruppen werden zugleich prototypische Eigenschaften zugeschrieben. Ein overter Quantor findet sich aber weder beim Subjekt noch beim Prädikat – nichtsdestotrotz ist eine generalisierende Lesart möglich.

---

<sup>17</sup> Goethe 2012, S. 250.

<sup>18</sup> Goethe 2012, S. 255.

<sup>19</sup> Goethe 2012, S. 244.

Kurz zusammengefasst muss ein Tagset, das Generalisierungen in fiktionalen Texten umfassend annotierbar macht und der Implementierung eines Taggers vorausgeht, folgenden Adäquatheitsbedingungen genügen. Einerseits muss die linguistische Komplexität der unterschiedlichen Auftretensformen im Tagset abgebildet werden können. Dabei sollten aber nicht allzu formale Kategorien angesetzt werden, welche die Komplexität der Annotation in fiktionalen Texten unverhältnismäßig erhöhten und auch dem Anspruch, sämtliche Generalisierungen in Texten erkennen zu können, nicht gerecht würden. Andererseits sollte das Tagset jedoch derart auf linguistische Parameter zurückführbar sein, dass es auch einer qualitativen Datenbetrachtung standhält.

### 3. Annotation

Nachdem im vorherigen Abschnitt die Hürden der Annotation von Generalisierungen in fiktionalen Erzähltexten beschrieben wurden, wird nun das Tagset und der Annotationsprozess geschildert. Der Abschnitt endet mit dem berechneten Inter-Annotator-Agreement und einigen Erläuterungen zur Erstellung der Goldstandards, auf deren Basis die Tagger implementiert werden.

#### 3.1 Tagset und Anwendungsregeln

Das Tagset erfasst mit dem Tag GI (für *Generalisierende Interpretation*), ob eine Passage eine Generalisierung ausdrückt, und mit diversen Subtags, welche Art der semantischen Quantifikation (universell, existenziell oder vage) vorliegt (vgl. Tabelle 1). Eine Passage besteht hierbei aus mindestens einem, möglicherweise mehreren aufeinanderfolgenden Teilsätzen. Weil Teilsätze somit die kleinste annotierbare Einheit darstellen, sprechen wir im Folgenden von einer Annotation auf Teilsatzebene. Die Annotation auf Teilsatzebene ermöglicht die Gratwanderung zwischen einer aus linguistischer Perspektive verwertbaren Annotation und einer Konzentration auf Oberflächenmerkmale, die für die Implementierung des Taggers zielführend ist. Auf diese Weise werden quantifizierte Aussagen in ihrem breiten Spektrum erfasst und gleichzeitig zeitökonomisch annotiert.

Subtag	Typ der Quantifikation	Natürlichsprachliches Beispiel
ALL	universelle Quantifikation	alle, immer, jede*r
MEIST	Mehrheitsquantifikation	meistens, am häufigsten
EXIST	existenzielle Quantifikation	es gibt x, es existieren y
DIV	vage Quantifikation	manchmal, teilweise, gewöhnlich, oft, etc.
BARE	keine der zuvor genannten + koverte Quantifikation	generische NPn (z. B. Gold, Löwen)
NEG	eine der oben genannten + Negation	Es gibt kein Gold, das braun ist.

Tab. 1: Tagset MONACO. [Gödeke et al. 2022 nach Barth et al. 2021]

Das Tagset ist wie folgt aufgebaut: Die Subtags ALL, MEIST und EXIST korrespondieren mit Quantoren aus der formalen Logik, die an der Satzoberfläche mit einer Reihe von Lexemen ausgedrückt werden können.<sup>20</sup> Beim Subtag ALL handelt es sich meist um Lexeme wie immer oder jede\*r, jedes:

(17) Gedenken wir nur des Kalks, [der zu allen Säuren eine große Neigung, eine entschiedene Vereinigungslust äußert]GI ALL!<sup>21</sup>

Beim Subtag MEIST kommen Lexeme wie meist oder meistens vor, die sich auf mehr als die Hälfte der bezeichneten Menge beziehen:

(18) [...]; [aber es sind meistens unbewußte Erinnerungen glücklicher und unglücklicher Folgen, die wir an eigenen oder fremden Handlungen erlebt haben]GI MEIST.<sup>22</sup>

<sup>20</sup> Siehe Dönicke et al. 2021 für eine ausführlichere formal-semantische Betrachtung des Tagsets.

<sup>21</sup> Vgl. Goethe 2012, S. 272.

<sup>22</sup> Vgl. Goethe 2012, S. 247.



Das Subtag EXIST umfasst generalisierende Passagen, die sich auf die Existenz der Individuen einer Klasse beziehen und durch Prädikate der Existenz wie *geben* und *existieren* markiert werden.<sup>23</sup>

(19) [Aber es gibt dergleichen noch, es muß dergleichen geben oder doch wieder geben]GI EXIST.<sup>24</sup>

Das Subtag DIV fängt alle vagen Quantoren auf, insbesondere also Quantifikationen, die durch Adverbien wie *manchmal*, *häufig*, *gelegentlich*, *selten* ausgelöst werden. Diese unterscheiden sich von den anderen natürlichsprachlichen Quantoren dadurch, dass ihre Wahrheitsbedingungen nicht präzise beschreibbar sind.

(20) [Unsere vortreffliche Vorsteherin läßt mich gewöhnlich die Briefe lesen, in welchen sie Beobachtungen über ihre Zöglinge den Eltern und Vorgesetzten mitteilt]GI DIV.<sup>25</sup>

(21) »Es betrifft unsern Freund, den Hauptmann,« antwortete Eduard. »Du kennst die traurige Lage, [in die er, wie so mancher andere, ohne sein Verschulden gesetzt ist]GI DIV.«<sup>26</sup>

Zusätzlich kann das Tagset nicht-overt Quantifikation mit dem Subtag BARE erfassen. Dieser kann etwa im Fall von generischen Subjekten oder generischen Objekten gesetzt werden, also dann, wenn keine Quantifikation an der Satzoberfläche identifizierbar ist.

(22) [Die Landleute haben die rechten Kenntnisse]GI BARE; [ihre Mitteilungen aber sind konfus]GI BARE [und nicht ehrlich]GI NEG. [Die Studierten aus der Stadt und von den Akademien sind wohl klar und ordentlich, aber es fehlt an der unmittelbaren Einsicht in die Sache]GI BARE.<sup>27</sup>

(23) [Die Männer denken mehr auf das Einzelne, auf das Gegenwärtige]GI BARE;<sup>28</sup>

Das Subtag NEG umfasst sämtliche generalisierte und negierte Aussagen, unabhängig von der syntaktischen Struktur. Die syntaktische Struktur kann deshalb interessant sein, weil mit unterschiedlichen syntaktischen Positionen von Negationen verschiedene Skopen von Negationen einhergehen. Daher bestimmt die Syntax, was genau in einem Satz negiert wird; der Quantor oder (Teile vom) Restriktor oder Skopus. Da diese informationsstrukturellen Unterscheidungen für die automatische Erkennung von generalisierenden Aussagen nicht relevant sind, wird jede generalisierende Aussage, die eine Negation enthält, mit dem Subtag NEG annotiert. In diesem Fall überschreibt NEG alle möglichen anderen Annotationen.

(24) [Unanfechtbare Wahrheiten gibt es überhaupt nicht]GI NEG<sup>29</sup>

Zudem gelten bestimmte Regeln zum Annotationsverfahren, die Mehrfachannotationen und ambige Fälle betreffen. So kann es vorkommen, dass in derselben Passage auf verschiedene Weisen generalisiert wird. Im folgenden Beispiel wird sowohl *u#ber* *Zeitra#ume* (immer) als auch existenziell *u#ber* *Zank* und *Streit* generalisiert. Da hier konkurrierende Generalisierungsarten vorliegen, müssen zwei Subtags vergeben werden.

(25) [[Es gibt doch bloß immer Zank und Streit]GI ALL]GI EXIST<sup>30</sup>

Eine wichtige Ausnahme bildet hier das Subtag BARE. BARE wird ausschließlich in den Fällen vergeben, in denen keine anderen Subtags vergeben werden können:

<sup>23</sup> Das Subtag EXIST ist ausschließlich explizit existenziellen Aussagen vorbehalten. Wir gehen also etwa davon aus, dass eine indefinite Nominalphrase wie *Ein Löwe* keine existenzielle Quantifikation auslöst, sondern eine generische NP ist – insofern eine generalisierende Lesart vorliegt. Entsprechend würde *Ein Einhorn* existiert mit dem Tag EXIST annotiert werden.

<sup>24</sup> Vgl. Fontane 2012, S. 166.

<sup>25</sup> Vgl. Goethe 2012, S. 263.

<sup>26</sup> Vgl. Goethe 2012, S. 242.

<sup>27</sup> Vgl. Goethe 2012, S. 244.

<sup>28</sup> Vgl. Goethe 2012, S. 244.

<sup>29</sup> Vgl. Fontane 2012, S. 9.

<sup>30</sup> Vgl. Fontane 2012, S. 125.

(26) [Der Heilige Vater liebt seine Untertanen gleichmäßig]GI BARE. (fikt. Bsp.)

In (27) wird allerdings sowohl u#ber Pa#pste (Der Heilige Vater) als auch u#ber alle Untertanen generalisiert. Deshalb wird hier nur ALL gesetzt.

(27) [Der Heilige Vater liebt alle seine Untertanen gleichma#ßig]GI ALL<sup>31</sup>

Finden sich zwei Generalisierungen in der gleichen Passage, die beide mit demselben Tag annotiert werden wu#rden, wird nicht doppelt annotiert. Im folgenden Beispiel fallen Jeder Heilige Vater und alle seine Untertanen beide in die Kategorie ALL:

(28) [Jeder Heilige Vater liebt alle seine Untertanen gleichma#ßig]GI ALL (fikt. Bsp.)

Neben den Subtags stellt das Tagset die Möglichkeit bereit, eine Passage als ambig zu markieren. Ist in einem Beispiel wie (9) nicht eindeutig zu bestimmen, ob die Aussage generalisierend oder spezifisch gemeint ist, kann dieses Subtag gesetzt werden.

Wie eingangs erwähnt, ermöglicht die Annotationspraxis auf Teilsatzebene die Einbeziehung aller Quantifikationsformen in natürlicher Sprache und ermöglicht zugleich ein zügiges Annotationsprozedere. Das dargestellte Vorgehen unterscheidet sich aber auch in wesentlichen Punkten von bereits bestehenden Annotationspraktiken in der Forschung zu Generizität und Quantifikation, die in Abschnitt 2.2 erläutert worden sind. Der wesentlichste Unterschied besteht darin, dass kein linguistisches Training für die Annotierenden nötig ist, etwa Restriktor, Skopus und Quantor auseinanderzuhalten. Zur Illustration der verschiedenen Vorgehensweisen sei hier ein Beispiel aus Goethes *Die Wahlverwandtschaften* anhand des Annotationsschemas von Friedrich et al. und unserem Konzept vergleichend analysiert.<sup>32</sup> Das Schema von Friedrich et al. (2016) bietet sich für einen Vergleich insbesondere deshalb an, weil ebenfalls primär intuitiv auf Satzebene annotiert wird und dadurch die verschiedenen Formen von Generizität erfasst werden, nämlich generische NPn und habituelle Äußerungen.

(29) [Das Bewusstsein ist keine hinlängliche Waffe, ja, manchmal eine gefährliche für den, der sie führt]GENERIC SENTENCE.<sup>33</sup>

Der main referent des Satzes, hier die Subjekt-NP Das Bewusstsein, würde als generische NP identifiziert werden. Das Verb sein hat statischen lexikalischen Aspekt, entsprechend würde der gesamte Satz als Generic Sentence klassifiziert werden.<sup>34</sup> Die Negation verändert hier nicht den situation entity type des Satzes, weil sie sich nur auf die Subjekt-NP bezieht.<sup>35</sup> Das Annotationsprozedere ist damit elegant und in diesem Satz niedrigschwellig anwendbar. Insbesondere in älteren fiktionalen Texten kommt aber die Problematik hinzu, dass Verben hinsichtlich ihres Aspekts klassifiziert werden müssten, die im aktuellen Sprachgebrauch nicht mehr vorkommen. Es ist daher mitunter schwierig, den lexikalischen Aspekt abzuleiten, der einen ausschlaggebenden Hinweis für die Klassifikation als generisch gibt. Zudem zeigt das Temporaladverb manchmal im zweiten Teilsatz die Habitualität an. Diese Information geht durch die Art der Annotation verloren. Generalisierungen kommen in vielfältigen Formen vor, daher ist es eine Priorität für uns, verschiedene Lesarten durch die Annotation und später durch den Tagger erfassen zu können. Vergleichend sei nun gezeigt, welche Tags mit unserem Tagset vergeben werden:

(30) [Das Bewusstsein ist keine hinlängliche Waffe]NEG, [ja, manchmal eine gefährliche für den, der sie führt]DIV.<sup>36</sup>

Der erste Teilsatz besteht aus einer definiten generischen NP und einem diese NP charakterisierenden Prädikat. Würde keine Negation in diesem Teilsatz stehen, qualifizierte sich dieser Teilsatz als BARE; da aber eine Negation vorliegt und Negation alle anderen Subtags überschreibt, wird das Subtag NEG vergeben. Der zweite Teilsatz wird mit dem Subtag Divers (DIV) versehen, weil das Adverb manchmal eine habituelle Generalisierung über eine vage Anzahl an Situationen ausdrückt.

<sup>31</sup> Vgl. Ernst 2012, S. 240.

<sup>32</sup> Vgl. Friedrich et al. 2016.

<sup>33</sup> Vgl. Goethe 2012, S. 247.

<sup>34</sup> Vgl. Friedrich et al. 2016, S. 41.

<sup>35</sup> Vgl. Friedrich et al. 2016, S. 36.

<sup>36</sup> Vgl. von Goethe 2012, S. 247.

### 3.2 Korpus, Workflow und Agreement

Wie in der Einleitung geschildert, soll der Generalisierungstagger in allen Texten von 1600 bis 1950 Generalisierungen automatisch erkennen können. Entsprechend muss das Tagset auf Texten erprobt werden, die diesen Zeitraum und seine verschiedenen stilistischen und sprachlichen Eigenheiten repräsentieren. Das Gesamtkorpus besteht aus ausschließlich originär deutschsprachiger Erzählliteratur, von dem eine Teilmenge an Texten als Annotationskorpus genutzt wird. Jede\*r Autor\*in im Annotationskorpus ist durch einen Text vertreten. Im Korpus enthalten sind Romane, Novellen, Erzählungen und epische Versdichtungen mit sowohl Ich- (homodiegetischen) als auch Er- (heterodiegetischen) Erzählinstanzen. Die genaue Auswahl kann Tabelle 2 entnommen werden.

Um der Repräsentationsfunktion für das Gesamtkorpus gerecht zu werden, werden die Annotationstexte zunächst nicht komplett annotiert, sondern auf die ersten 200 Sätze beschränkt. Der Nachteil unserer Auswahl ist zunächst, dass ohne den Rückgriff auf eine automatische Erkennung eine Betrachtung generalisierender Passagen werkimmanent erst nach fertiger Entwicklung der Tagger möglich sein wird. Es überwiegen aber die Vorteile – so sind die Annotationen nicht auf eine kleinere Menge von Texten beschränkt, erfassen dadurch mehrere Autor\*innen und demzufolge eine größere Breite literarischer Epochen und Genres. Außerdem ist es so möglich, unterschiedliche Entwicklungsstadien der Sprache, von der Frühen Neuzeit an, zu berücksichtigen.

Wir annotieren Generalisierungen mittels kollaborativer Annotation im webbasierten Tool **CATMA 6**.<sup>37</sup> Das Best-Practice-Modell von Gius / Jacke dient als Orientierung bei der Besprechung und Qualitätsprüfung der erstellten Annotationen.<sup>38</sup> Die bereits annotierten Texte sind unter dem Titel MONACO<sup>39</sup> in einem GitLab-Repository publiziert.

Text	Jahr	Goldstandard	Annotator*in	Gattung	Typ der Erzählinstanz	Kappa (Tokenebene)	Kappa (Teilsatzebene)
Andreae: Die chymische Hochzeit	1616	nein	{‘a_112’, ‘a_107’}	Roman	homodiegetisch	0,4883	0,5376
von Zesen: Adriatische Rosemund	1645	nein	{‘a_111’, ‘a_110’}	Roman	heterodiegetisch	0,6017	0,5988
Grimmelshausen: Der abenteuerliche Simplicissimus	1668	nein	{‘a_111’, ‘a_107’}	Roman	homodiegetisch	0,4652	0,6046
Lohenstein: Großmüthiger Feldherr Arminius oder Herrmann	1689	nein	{‘a_108’, ‘a_109’}	Roman	heterodiegetisch	0,8062	0,7906
Schnabel: Die Insel Felsenburg	1731	nein	{‘a_111’, ‘a_112’}	Roman	homodiegetisch	0,2169	0,1946
Gellert: Das Leben der schwedischen Gräfin von G	1748	ja	{‘a_109’, ‘a_108’}	Roman	homodiegetisch	0,8663	0,8724
Wieland: Geschichte des Agathon	1766, 1767	ja	{‘a_109’, ‘a_112’, ‘a_108’, ‘a_111’, ‘a_110’, ‘a_107’}	Roman	homodiegetisch	0,7763	0,7812
LaRoche: Geschichte des Fräuleins von Sternheim	1771	ja	{‘a_109’, ‘a_108’}	Roman	homodiegetisch	0,8435	0,8352

Tab. 2: Korpus MONACO [Gödeke et al. 2022 nach Barth et al. 2021]

<sup>37</sup> Vgl. Gius et al. 2022.

<sup>38</sup> Vgl. Gius / Jacke 2016.

<sup>39</sup> Vgl. Barth et al. 2021.

Novalis: Die Lehrlinge zu Sais	1802	ja	{'a_112', 'a_111'}	Roman	homodiegetisch	0,4045	0,4108
Kleist: Michael Kohlhaas	1808	nein	{'a_109', 'a_110'}	Erzählung	heterodiegetisch	0,6376	0,5889
Goethe: Die Wahlverwandtschaften	1809	ja	{'a_111', 'a_108'}	Roman	heterodiegetisch	0,579	0,5688
Goethe: Die Wahlverwandtschaften	1809	ja	{'a_112', 'a_107'}	Roman	heterodiegetisch	0,7396	0,7948
Hoffmann: Der Sandmann	1816	ja	{'a_112', 'a_107'}	Erzählung	homodiegetisch	0,82	0,8316
Dahn: Kampf um Rom	1876	ja	{'a_112', 'a_107'}	Roman	heterodiegetisch	0,4173	0,5278
May: Winnetou II	1893	nein	{'a_107', 'a_r_112'}	Roman	homodiegetisch	0,4099	0,3502
Fontane: Der Stechlin	1895	ja	{'a_109', 'a_112'}	Roman	heterodiegetisch	0,9118	0,8784
Kafka: Der Bau	1923–1924	ja	{'a_111', 'a_110'}	Erzählung	homodiegetisch	0,672	0,6816
Musil: Der Mann ohne Eigenschaften	1930	ja	{'a_111', 'a_108'}	Roman	heterodiegetisch	0,5029	0,4655
Makro-Durchschnitt						0,6214	0,6293

Tab. 2: Korpus MONACO [Gödeke et al. 2022 nach Barth et al. 2021]

Bisher wurden mit dem Tagset 17 Texte aus dem Annotationskorpus annotiert, die in der Regel jeweils von zwei Annotierenden zu bearbeiten sind. Für jede Annotierendenkombination berechnen wir Fleiss' Kappa<sup>40</sup> basierend auf übereinstimmenden Tokens bzw. auf übereinstimmenden Teilsätzen. Bei überlappenden Annotationen wird ein Token bzw. ein Teilsatz nur dann als Übereinstimmung gewertet, wenn beide Annotierende die überlappenden Tags gesetzt haben.

Im Mittel über alle Texte wird ein überzeugendes Agreement von 0,62 (Tokenebene) bzw. 0,63 (Teilsatzebene) erreicht. Die deutlichen Schwankungen (z. B. Schnabel *Die Insel Felsenburg* mit 0,22 bzw. Fontane *Der Stechlin* mit 0,91, vgl. Tabelle 2) sind präsumtiv neben den individuellen Fähigkeiten der Annotierenden auch auf die unterschiedliche Komplexität der literarischen Texte sowie auf die zugrundeliegende Sprachstufe zurückzuführen.

Bisher wurden für zehn der Texte Goldstandards auf Basis der Erstannotationen erstellt. Hierbei gehen zwei Adjudikator\*innen (aus der Gruppe der promovierenden Autor\*innen dieses Aufsatzes) erneut die Erstannotationen durch, diskutieren die annotierten Textstellen und erstellen eine Expert\*innenannotation, welche dann als Goldstandard genutzt werden kann. Neben der Beseitigung von Unachtsamkeitsfehlern in der Erstannotation hat die Adjudikation auch die Aufgabe, prävalente Lesarten von mehrdeutigen oder vagen Aussagen festzustellen (oder, in Einzelfällen, Aussagen als linguistisch ambig zu markieren, falls keine prävalente Lesart festzustellen ist).

## 4. Automatische Erkennung

Im folgenden Abschnitt werden die in den regelbasierten Tagger eingebundenen Merkmale erläutert, auf denen später der statistische Tagger trainiert wurde.

### 4.1 Merkmale

Das wichtigste Merkmal von Generalisierungen sind Quantoren und quantifizierende Ausdrücke wie jede\*r, alle, immer und dergleichen. Diese bestimmen die Klasse, in die Teilsätze mit den entsprechenden Quantoren fallen, beispielsweise in die Klasse ALL:

<sup>40</sup> Vgl. Fleiss 1971.

(31) [Jede Sache will gelernt sein]GI ALL<sup>41</sup>

(32) [Ebenso pünktlich liefen darauf auch jedesmal kurze Antwortschreiben ein]GI ALL, die dem lieben Freund und geschätzten Gelehrten dankten.<sup>42</sup>

Wir organisieren die entsprechenden Quantoren und quantifizierenden Ausdrücke in einem Wörterbuch mit Lemma-Subtag-Paaren (z. B. jede : ALL). Entsprechend können ganze Teilsätze, in denen sie vorkommen, dem entsprechenden Subtag zugeordnet werden. Eine Ausnahme davon bilden Phrasen, die mit deiktischen Ausdrücken auf konkrete Objekte, Personen, Zeiten oder Orte referieren, weil in diesen Fällen keine Generalisierungen vorliegen können. Um quantifizierte Ausdrücke mit einer deiktischen Referenz ausschließen zu können, wurde eine Liste mit deiktischen Markern erstellt.

Wie bereits oben erläutert wurde, sind wir mit dem Problem der koverten Quantifikation konfrontiert, die mit dem Subtag BARE gekennzeichnet ist. Einerseits bieten sich NPn im Plural für den Ausdruck von Generalisierungen an:

(33) [Regentropfen vereinigen sich gern zu Strömen]GI BARE<sup>43</sup>

Andererseits können auch definite NPn Generalisierungen markieren, sowohl im Plural als auch im Singular:

(34) [Der Räuber kann sehr leicht mein Opfer werden, und ein süß schmeckendes]GI BARE<sup>44</sup>

Und auch der unbestimmte Artikel mit Nomen im Singular kann generalisierend verwendet werden.

(35) [Aber ein Narr steckt sie [tausend Mark] in den Strumpf]GI BARE, sagen die Wirklichkeitsmenschen, [und ein Tüchtiger schafft etwas mit ihnen]GI BARE;<sup>45</sup>

Insgesamt sind Artikel, sowohl bestimmte als auch unbestimmte, aber wenig verlässlich. Endgültig kann nur der Kontext disambiguieren, ob eine definite oder indefinite NP generalisierend oder spezifisch verwendet wird. Diese Unterscheidung ist in einem regelbasierten System nicht umsetzbar. Wir haben festgestellt, dass die Suche nach NPn im Plural ohne Artikel am wenigsten Fehler verursacht.

Verbalphrasen (VPn) können auch generalisierend sein. In diesem Fall handelt es sich meistens um habituelle Äußerungen, die durch Adverbien wie täglich, oft oder überall markiert werden. Außerdem gibt es Verben, die eine generalisierende Bedeutung haben. Das sind zum einen Verben der Existenz, wie existieren und geben:

(36) [Aber es gibt leidenschaftliche Räuber]GI EXIST<sup>46</sup>

Und zum anderen Verben, die über Situationen generalisieren, wie pflegen zu:

(37) [Dann pflegen besonders friedliche Zeiten zu kommen]GI BARE<sup>47</sup>

Eine weitere für Generalisierungen relevante Kategorie des Verbs ist *gnomisches Präsens*, also ein Präsens ohne zeitliche Bedeutung. Unter der (simplifizierenden) Annahme, dass das Präsens in fiktionalen Texten meistens keine Handlung ausdrückt, kann der Tagger nach einzelnen Teilsätzen in dieser Zeitform suchen. Eine konkrete Ausnahme ist die direkte Rede, die zwar oft im Präsens steht, aber nicht ohne Weiteres als Merkmal für gnomisches Präsens gewertet werden kann. Ebenso verhält es sich mit dem *historischen* oder *szenischen Präsens*, also einem Präsens, das aus stilistischen Gründen für vergangene Geschehnisse verwendet wird, und typischerweise mehrere Teilsätze im Präsens aufeinander folgen lässt:

---

<sup>41</sup> Vgl. Fontane 2012, S. 119.

<sup>42</sup> Vgl. Musil 1970, Kapitel 3.

<sup>43</sup> Vgl. Goethe 2012, S. 270.

<sup>44</sup> Vgl. Kafka 1990.

<sup>45</sup> Vgl. Musil 1970, Kapitel 4.

<sup>46</sup> Vgl. Kafka 1990.

<sup>47</sup> Vgl. Kafka 1990.

(38) Von außen ist eigentlich nur ein großes Loch sichtbar, dieses führt aber in Wirklichkeit nirgends hin, schon nach ein paar Schritten stößt man auf natürliches festes Gestein.<sup>48</sup>

Darüber hinaus können auch bestimmte syntaktische Strukturen auf Generalisierungen hindeuten. Komplexe Sätze der Formen wenn ... dann oder wer ... der sind in der Regel generalisierend. Solche Strukturen können einfach in einem regelbasierten Tagger implementiert werden.

(39) [Wenn es aber Wirklichkeitssinn gibt, und niemand wird bezweifeln, dass er seine Daseinsberechtigung hat, dann muss es auch etwas geben, das man Möglichkeitssinn nennen kann]GI EXIST<sup>49</sup>

(40) [Wer unter solchen Umständen was Besseres sagen will, sagt immer was Schlechteres]GI ALL<sup>50</sup>

Im Folgenden wird beschrieben, wie die beschriebenen Merkmale in den beiden Taggern eingesetzt werden.

## 4.2 Modelle und Algorithmen

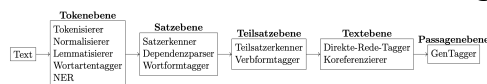


Abb. 1: Textverarbeitungskomponenten für verschiedene sprachliche Ebenen: Ein Text durchläuft die einzelnen Komponenten von links nach rechts und von oben nach unten. [Gödeke et al. 2022]

Für die Implementierung der Textverarbeitungskomponenten wurde die Open-Source-Bibliothek **spaCy**<sup>51</sup> benutzt. Dem Generalisierungstagger (GenTagger) selbst sind eine Reihe von Komponenten vorgeschaltet, um linguistische Merkmale zu extrahieren (vgl. Abbildung 1). Da sich im Korpus teilweise Texte mit zeitspezifischer Schreibvariation befinden, ist die erste Komponente ein auf dem **Deutschen Textarchiv**<sup>52</sup> trainierter Normalisierer. Es folgen Lemmatisierer, Wortartentagger, Eigennamenerkenner (NER) und Satzwerker, welche bereits im vortrainierten spaCy-Modell für das Deutsche enthalten sind. Als Nächstes folgen Abhängigkeitsparser, Wortformtagger (morphologische Analyse), Teilsatzerkenner und Verbformtagger (Teilsatzanalyse) aus Dönicke.<sup>53</sup> Als Letztes kommen ein Direkte-Rede-Tagger, der Text zwischen öffnenden und schließenden Anführungszeichen erkennt, sowie ein Koreferenzierer, der erkennt, wenn zwei verschiedene sprachliche Ausdrücke sich auf dieselbe Entität beziehen; bei diesem handelt es sich um eine Reimplementierung des Algorithmus aus Krug et al., der so erweitert wurde, dass nicht nur Figurenentitäten, sondern alle Eigennamen, NPn und Pronomina in die Koreferenzresolution einbezogen werden.<sup>54 55</sup>

### 4.2.1 Regelbasierter GenTagger

Da zu Beginn der Arbeit weder Trainings- noch Testdaten vorlagen, wurde zunächst ein regelbasierter GenTagger entwickelt. Als Entwicklungsdatensatz dienten hierzu die ersten vier Kapitel aus Goethes *Die Wahlverwandtschaften*. Der Algorithmus wird in Abbildung 2 skizziert.

<sup>48</sup> Vgl. Kafka 1990.

<sup>49</sup> Vgl. Musil 1970, Kapitel 4.

<sup>50</sup> Vgl. Fontane 2012, S. 17.

<sup>51</sup> Vgl. Honnibal / Johnson 2015.

<sup>52</sup> Vgl. Deutsches Textarchiv 2022.

<sup>53</sup> Vgl. Dönicke 2020.

<sup>54</sup> Vgl. Krug et al. 2015.

<sup>55</sup> Wir wurden im Review darauf hingewiesen, dass Fehler bei der automatischen Erkennung von Generalisierung auch auf Fehler in Vorverarbeitungsschritten zurückzuführen sein können. In welchem Ausmaß Fehler aus den einzelnen Pipeline-Komponenten propagiert werden, kann an dieser Stelle nicht quantitativ eingeschätzt werden; jedoch wurden der Teilsatzerkenner und der Verbformtagger – zwei Komponenten, die relativ am Ende in der Pipeline eingebunden sind –, bereits in Dönicke 2020 in der Literaturdomäne getestet. Dort erreichten sie Performanzen (F-Maß) von 81% für die Erkennung von Teilsätzen sowie 93% für Tempus, 79% für Modus, 94% für Genus Verbi und 80% für Modalität.

```

1 Passagen := {}
2 für jeden Teilsatz in Dokument tue
3   Subtags := {}
4   für jedes Token in Teilsatz tue
5     wenn Token.Lemma in QuantorenDict dann
6       Subtag := QuantorenDict[Token.Lemma]
7       Restriktoren := {nächste.übergeordnete.NP.oder.VP(Token)}
8     sonst wenn Token.Lemma = existieren oder (Token.Lemma = geben
9       und es in Kinder(Token)) dann
10      Subtag := EXIST
11      Restriktoren := direkt.untergeordnete.NPs(Token)
12    sonst wenn ist.generische.NP.oder.VP(Token) dann
13      Subtag := BARE
14      Restriktoren := {Token}
15    für jeden Restriktor in Restriktoren tue
16      wenn nicht enthält.deiktische.Koreferenz(Restriktor) dann
17        Subtags := Subtags ∪ {Subtag}
18    wenn Anzahl(Subtags) > 0 dann
19      Passagen := Passagen ∪ {Passage({Teilsatz}, Subtags)}
20 Passagen := verbinde.Teilsätze.zu.Passagen(Passagen)
21 für jede Passage in Passagen tue
22   wenn enthält.Negation(Passage) dann
23     Passage.Subtags := {NEG}
24   wenn Anzahl(Passage.Subtags) > 1 dann
25     Passage.Subtags := Passage.Subtags \ {BARE}

```

Abb. 2: Regelbasierter GenTagger. [Gödeke et al. 2022]

Zunächst werden Generalisierungen auf Teilsatzebene gesucht (2–18), indem nach Quantor-Restriktor-Kombinationen gesucht wird. Dabei werden drei mögliche Fälle unterschieden:

Wenn ein overter Quantor im Teilsatz vorkommt, ist der Restriktor die nächste übergeordnete NP oder VP (5–7). Lemma-Subtag-Paare (z. B. immer : ALL) werden dabei in einem manuell erstellten Wörterbuch gespeichert, wie in Abschnitt 4.1 beschrieben wurde.

Wenn eine Form von existieren oder es gibt im Teilsatz vorkommt, kommen als Restriktor alle direkt untergeordneten NPn in Frage. In diesem Fall ist das Subtag EXIST (8–10).

Wenn eine generische NP oder VP im Teilsatz vorkommt, hat diese keinen overten Quantor und das Subtag ist BARE (11–13). Indikatoren für generische NPn und VPn werden durch eine Vielzahl von Regeln beschrieben, die auf die zuvor extrahierten und oben beschriebenen linguistischen Merkmale zurückgreifen.

Für einen gefundenen Restriktor wird dann überprüft, ob er oder eine mit ihm koreferente NP einen deiktischen Ausdruck enthält (14–16). In diesem Fall referiert er mit hoher Wahrscheinlichkeit auf eine spezifische Entität und kommt somit nicht für eine Generalisierung in Frage. Andernfalls wird das entsprechende Subtag für den Teilsatz vorgemerkt. Wurde mindestens ein Subtag für den aktuellen Teilsatz gespeichert, wird eine Passage instanziiert, die aus genau diesem Teilsatz und den zugehörigen Subtags besteht (17–18).

Als Nächstes werden bestimmte Teilsatztypen zu Passagen zusammengeführt. In diesem Schritt werden Relativ- und Konditionalsätze mit ihren übergeordneten Teilsätzen verbunden und ihre Subtags vereinigt. In einem letzten Schritt wird erneut über alle Passagen iteriert (20–24). Kommt ein Negationsmarker vor, werden die Subtags für diese Passage auf NEG reduziert. Kommen mehr als ein Subtag in der Passage vor, wird BARE ggf. entfernt. So wird sichergestellt, dass NEG und BARE nicht in Kombination mit anderen Subtags auftreten.

## 4.2.2 Statistischer GenTagger

Nach *Die Wahlverwandtschaften* wurden bisher die ersten 200 Sätze von 9 weiteren Texten mit Goldstandards annotiert, was es ermöglicht, einen statistischen GenTagger zu trainieren und zu testen. Als Testdaten verwenden wir Wielands *Geschichte des Agathon*,<sup>56</sup> in welchem wir Generalisierungen von allen sechs statt der üblichen zwei Annotierenden annotieren ließen.

Für den statistischen GenTagger lässt sich ein Klassifikationsproblem auf Teilsatzebene definieren. Obwohl sich generalisierende Passagen überlappen können und somit theoretisch mehrere Subtags pro Teilsatz möglich sind, kommt dieser Fall in der Praxis kaum vor, weshalb wir uns auf eine einfache 7-Klassen-Klassifikation (eine Klasse pro Subtag plus eine Klasse für nicht-generalisierende Teilsätze) beschränken. Tabelle 3 zeigt, dass die Verteilung der Klassen zudem extrem unausgeglich ist. Einige wenige generalisierende Passagen sind linguistisch (d. h. syntaktisch, semantisch, pragmatisch etc.) ambig und können daher nicht eindeutig annotiert werden. Solche Fälle wurden im Goldstandard markiert und sowohl beim Training als auch beim Testen der GenTagger ausgeschlossen.

<sup>56</sup> Wieland 2012.

	Texte	Passagen	Teilsätze						
nicht-GI	BARE	NEG	ALL	DIV	EXIST	MEIST			
Trainingsdaten	8	492	3316	421	197	175	94	23	5
Entwicklungsdaten	1	191	1262	206	81	76	25	5	2
Testdaten	1	45	794	37	40	27	15	3	3

Tab. 3: Größe der Trainings-, Entwicklungs- und Testdaten.  
[Gödeke et al. 2022]

Da der regelbasierte Algorithmus einem Entscheidungsbaum ähnelt, wurde für die automatische Klassifikation ebenfalls ein Entscheidungsbaum<sup>57</sup> verwendet. Um einer Überanpassung an die Trainingsdaten entgegenzuwirken, wurde die maximale Baumtiefe auf 15 und die minimale Anzahl an Daten pro Blattknoten auf 2 gesetzt. Diese Werte haben sich auf den Entwicklungsdaten als gut erwiesen. Im Training wurde außerdem allen Klassen das gleiche Gewicht gegeben.

Wir verwenden die gleichen Basismerkmale wie für den regelbasierten Tagger (vgl. Tabelle 4). Somit besitzt der Entscheidungsbaum die gleichen Grundlagen wie wir um Regeln zu formen. Ähnliche Merkmale wurden bereits von Reiter / Frank für die Erkennung generischer NPn extrahiert.<sup>58</sup> Zuerst werden Merkmale von allen NPn und VPn im Satz extrahiert. Diese werden zusätzlich zu neuen Merkmalen, z. B. »Subtag des Quantors : ALL, Numerus des NP-Kopfs : Plural, Artikel in NP? : nein«, kombiniert, die aus dem Subtag des Quantors und bis zu zwei weiteren Merkmalen bestehen. Das Maximum von drei Merkmalen hat sich auf den Entwicklungsdaten als optimal erwiesen. Auf Teilsatzebene werden hauptsächlich grammatikalische Eigenschaften wie z. B. Tempus extrahiert, um somit Phänomene wie das gnomische Präsens abzudecken. Da eine Passage aus mehreren Teilsätzen bestehen kann, wir aber nur Teilsätze klassifizieren, werden alle Merkmale nicht nur vom zu klassifizierenden Teilsatz extrahiert, sondern auch von seinen beiden Nachbarteilsätzen. Darüber hinaus werden noch einige Merkmale extrahiert, welche die Verknüpfung der benachbarten Teilsätze zum gefragten Teilsatz kennzeichnen, z. B. ob sie im gleichen Satz stehen. Es wurden außerdem eine Reihe von lexikalischen Merkmalen getestet, z. B. N-Gramme für die Teilsätze oder die Lemmata von Quantor und Restriktor; allerdings führten alle diese Erweiterungen zu einer niedrigeren Performanz auf den Entwicklungsdaten.

Ebene	Merkmale
NPn im Teilsatz	Tag des Quantors oder BARE, Wortart des Quantors oder BARE, Wortart des NP-Kopfs, Dependenzrelation des NP-Kopfs, Numerus des NP-Kopfs, Artikel in NP?, Zahlwort in NP?, spezifisch?
VPn im Teilsatz	Tag des Quantors oder BARE, Wortart des Quantors oder BARE, Wortart des VP-Kopfs, Häufigkeitsadverb in VP?, vager Mehrwortausdruck in VP?, spezifisch?
Teilsatz	Tempus, Aspekt, Modus, Genus verbi, Fintheit, pflegen zu in Teilsatz?, direkte Rede?, Negation in Teilsatz?, Konditionalsatz?, Relativsatz?
benachbarte Teilsätze	gleicher Satz?, direkt untergeordnet?, direkt übergeordnet?

Tab. 4: Merkmale für den statistischen GenTagger; Fragezeichen markieren Boole'sche Merkmale. [Gödeke et al. 2022]

<sup>57</sup> Vgl. Breiman et al. 1984.

<sup>58</sup> Vgl. Reiter / Frank 2010.



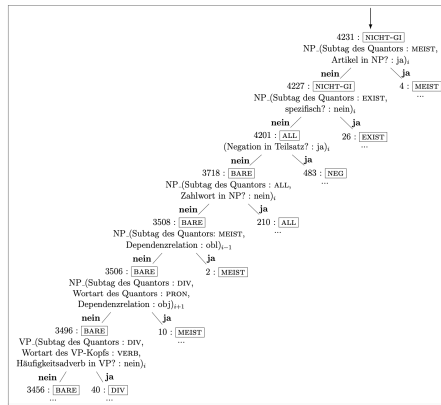


Abb. 3: Die wichtigsten Merkmale des Entscheidungsbaum-GenTaggers. [Gödeke et al. 2022]

Eine Inspektion des automatisch erstellten Entscheidungsbaums (vgl. Abbildung 3<sup>59</sup>) zeigt, dass zunächst Teilsätze mit den Merkmalen für MEIST, EXIST, NEG und ALL abgefragt werden, bevor nach Merkmalen für DIV, BARE und nicht-GI gesucht wird.

Der zweite statistische GenTagger ist ein Zufallswald mit 100 Entscheidungsbäumen.<sup>60</sup> Da die statistischen GenTagger Teilsätze taggen, aber keine Passagen bestimmen, werden im Anschluss direkt aufeinanderfolgende Teilsätze mit dem gleichen Subtag zu Passagen verbunden. Dies dient lediglich der Weiterverarbeitung und hat keinen Einfluss auf die Evaluation, da die GenTagger ebenfalls auf Teilsatzebene evaluiert werden.

## 5. Evaluation und Diskussion

Als Evaluationsmaße verwenden wir Precision (P), Recall (R) und F-Maß (F), welche für ein einzelnes Subtag  $y$  wie folgt definiert sind:

$$P(y) = \frac{| \{ \text{Teilsätze annotiert mit } y \} \cap \{ \text{Teilsätze getaggt mit } y \} |}{| \{ \text{Teilsätze getaggt mit } y \} |}$$

$$R(y) = \frac{| \{ \text{Teilsätze annotiert mit } y \} \cap \{ \text{Teilsätze getaggt mit } y \} |}{| \{ \text{Teilsätze annotiert mit } y \} |}$$

$$F(y) = \frac{2 \cdot P(y) \cdot R(y)}{P(y) + R(y)}$$

Abb. 4: Berechnung von Precision (P), Recall (R) und F-Maß (F) für ein einzelnes Subtag y. [Gödeke et al. 2022]

Die Precision gibt also an, wie viele der mit  $y$  getaggtten Teilsätze auch im Goldstandard mit  $y$  annotiert sind. Der Recall wiederum gibt an, wie viele der im Goldstandard mit  $y$  annotierten Teilsätze auch mit  $y$  getaggt wurden. Das F-Maß ist das harmonische Mittel der beiden.

Aus anwendungstechnischer Sicht besitzt die Precision einen höheren Stellenwert als der Recall: Wenn nicht alle generalisierenden Passagen in einem Text ausgegeben werden (niedriger Recall), gehen lediglich interessante Fälle verloren. Wenn hingegen viele nicht-generalisierende Passagen ausgegeben werden (niedrige Precision), müssen die ausgegebenen Passagen manuell nachgefiltert werden. Ein hoher Recall ist natürlich ebenfalls wünschenswert. Bei der Entwicklung der GenTagger wurden daher hohe Werte für Precision und F-Maß auf den Entwicklungsdaten angestrebt.

Im Falle mehrerer Subtags Y kann der Mikro-Durchschnitt der einzelnen Maße wie folgt gebildet werden:

<sup>59</sup> Die Knoten lesen sich wie folgt: In der ersten Zeile stehen Anzahl und häufigste Klasse der Trainingsdaten (d. h. Teilsätze), die den Knoten erreichen (da wir jeder Klasse die gleiche Gewichtung geben, wird die Häufigkeit relativ zur Klassengröße berechnet); darunter steht die abgefragte Merkmalskombination, wobei der Index angibt, ob sie sich auf den zu klassifizierenden Teilsatz (i), seinen Vorgänger (i-1) oder seinen Nachfolger (i+1) bezieht. Auslassungspunkte markieren gestützte Stellen im Baum.

<sup>60</sup> Vgl. Breiman 2001.

$$\text{MikroP}(Y) = \frac{\sum_{y \in Y} |\{ \text{Teilsätze annotiert mit } y \} \cap \{ \text{Teilsätze getaggt mit } y \}|}{\sum_{y \in Y} |\{ \text{Teilsätze getaggt mit } y \}|}$$

$$\text{MikroR}(Y) = \frac{\sum_{y \in Y} |\{ \text{Teilsätze annotiert mit } y \} \cap \{ \text{Teilsätze getaggt mit } y \}|}{\sum_{y \in Y} |\{ \text{Teilsätze annotiert mit } y \}|}$$

$$\text{MikroF}(Y) = \frac{2 \cdot \text{MikroP}(Y) \cdot \text{MikroR}(Y)}{\text{MikroP}(Y) + \text{MikroR}(Y)}$$

Abb. 5: Berechnung des Mikro-Durchschnitts von Precision (P), Recall (R) und F-Maß (F). [Gödeke et al. 2022]

Die Mikro-Precision gibt also an, wie viele der getaggtten Teilsätze korrekt getaggt wurden, während der Mikro-Recall angibt, wie viele der im Goldstandard annotierten Teilsätze korrekt getaggt wurden.

Da die Subtags unterschiedlich häufig vorkommen, berechnen wir außerdem den Makro-Durchschnitt. Im Gegensatz zum Mikro-Durchschnitt, welcher jedem Teilsatz die gleiche Gewichtung gibt, gibt der Makro-Durchschnitt jedem Subtag die gleiche Gewichtung:

$$\text{MakroP}(Y) = \frac{1}{|Y|} \cdot \sum_{y \in Y} P(y)$$

$$\text{MakroR}(Y) = \frac{1}{|Y|} \cdot \sum_{y \in Y} R(y)$$

$$\text{MakroF}(Y) = \frac{1}{|Y|} \cdot \sum_{y \in Y} F(y)$$

Abb. 6: Berechnung des Makro-Durchschnitts von Precision (P), Recall (R) und F-Maß (F). [Gödeke et al. 2022]

Methode	Entwicklungsdaten	Testdaten				
Mikro-P	Mikro-R	Mikro-F	Mikro-P	Mikro-R	Mikro-F	
Regelbasiert	0,42	0,26	0,32	0,21	0,26	0,23
Entscheidungsbaum	0,16	0,62	0,25	0,07	0,48	0,12
Zufallswald	0,26	0,40	0,31	0,14	0,33	0,20

Tab. 5: Mikro-Durchschnitts-Werte der GenTagger auf Entwicklungs- und Testdaten. [Gödeke et al. 2022]

Um einen ersten vergleichenden Überblick über die drei GenTagger zu bekommen, seien zunächst die Mikro-Durchschnitte in Tabelle 5 betrachtet. Sowohl auf den Entwicklungsdaten als auch auf den Testdaten erzielt der regelbasierte GenTagger höhere Werte für Precision und F-Maß als die statistischen GenTagger. Der Zufallswald-GenTagger zieht zwar fast mit dem regelbasierten GenTagger gleich, was das F-Maß betrifft, kann dies aber auf einen höheren Recall, nicht auf eine höhere Precision zurückführen. Wie zu erwarten, erzielt der Entscheidungsbaum niedrigere Werte für Precision und F-Maß als der Zufallswald, weshalb Ersterer im Folgenden nicht weiter betrachtet wird.

	Regelbasiert	Zufallswald				
	P	R	F	P	R	F
ALL	0,26	0,48	0,34	0,27	0,40	0,33
MEIST	0,00	0,00	0,00	0,50	0,33	0,40
EXIST	0,75	1,00	0,86	1,00	1,00	1,00
DIV	0,09	0,13	0,10	0,31	0,33	0,32
BARE	0,18	0,32	0,23	0,08	0,46	0,14
NEG	0,30	0,75	0,12	0,14	0,10	0,12

Tab. 6: Tag-spezifische Makro-Durchschnittswerte der GenTagger auf den Testdaten. [Gödeke et al. 2022]

Makro-	0,26	0,33	0,28	0,39	0,44	0,39
--------	------	------	------	------	------	------

Tab. 6: Tag-spezifische Makro-Durchschnittswerte der GenTagger auf den Testdaten.  
[Gödeke et al. 2022]

Tabelle 6 schlüsselt die Evaluation nach den einzelnen Subtags auf, wodurch individuelle Stärken und Schwächen der beiden GenTagger sichtbar werden. Im Makro-Durchschnitt erzielt der Zufallswald um ca. 10% höhere Werte für Precision, Recall und F-Maß als der regelbasierte GenTagger. Für ALL erzielen beide GenTagger eine Precision von ca. 27% und ein F-Maß von ca. 33%. Für MEIST und EXIST gibt es in den Testdaten nur jeweils zwei Passagen:

(41) Und da es mehr als zu gewiß ist, [daß der größte Teil derjenigen, welche die große Welt ausmachen, wie Hippias denkt, oder doch nach seinen Grundsätzen handelt]GI MEIST.<sup>61</sup>

(42) Niemals hatte ihn ein weibliches Aug erblickt, ohne die Schuld ihres Geschlechts zu bezahlen, [welches die Natur für die Schönheit so empfindlich gemacht zu haben scheint, daß diese einzige Eigenschaft den meisten unter ihnen die Abwesenheit aller übrigen verbirgt]GI MEIST.<sup>62</sup>

(43) [Ohne Zweifel gibt es wichtigere als derjenige, auf den seine Wahl gefallen ist]GI EXIST.<sup>63</sup>

(44) Aus diesem Grunde kann er ganz zuverlässig versichern, [daß Agathon und die meisten übrigen Personen, [welche in seine Geschichte eingeflochten sind]GI EXIST]GI MEIST, wirkliche Personen sind, [dergleichen es von je her viele gegeben hat, und in dieser Stunde noch gibt]GI EXIST.<sup>64</sup>

Der regelbasierte GenTagger vergibt das Subtag MEIST nur zweimal, und zwar fälschlicherweise für die beiden Teilsätze ab daß in (44): Bei den meisten übrigen Personen handelt es sich um eine spezifische Gruppe. Somit sind Precision und F-Maß 0. Der statistische GenTagger vergibt MEIST ebenfalls zweimal, einmal korrekterweise in (42) und einmal fälschlicherweise für einen Teilsatz, in dem am meisten adverbial, nicht quantifizierend gebraucht wird. Somit ist die Precision 50%. Keiner der beiden GenTagger erkennt (41) als MEIST, da die Formulierung der größte Teil für eine Mehrheitsquantifikation nicht im Quantorenwörterbuch enthalten ist, auf welche beide GenTagger zugreifen. (Diese Liste wird zwar nach wie vor im Zuge der Annotation erweitert, aber natürlich wurden die Testdaten nicht dafür herangezogen.) Die EXIST-Fälle erkennt der statistische GenTagger korrekt, während der regelbasierte GenTagger einen zusätzlichen Teilsatz mit EXIST versieht, ebenfalls in (44). Dabei handelt es sich um einen Folgefehler beim Verbinden von Teilsätzen zu Passagen, der wiederum durch einen Fehler beim Parsen des Satzes verursacht wird. Diese Art von Fehler kommt zwar auch bei den anderen Subtags vor, sorgt bei nur vier getaggten Teilsätzen aber gleich für einen Precision-Verlust von 25%.

Für DIV erzielt der Zufallswald ca. 20% höhere Werte für Precision und F-Maß als der regelbasierte GenTagger. Wir vermuten die Ursache dafür darin, dass vage Quantifikation in einer deutlich größeren Formvielfalt auftritt als All-, Mehrheits- und Existenzquantifikation. Der regelbasierte GenTagger verwendet dennoch nur eine Liste vager Quantoren und den Spezifitätstest, wohingegen der Zufallswald-GenTagger viel genauere Regeln erlernen kann. Für BARE verhält es sich umgekehrt – hier erzielt der Zufallswald ca. 10% schlechtere Werte für Precision und F-Maß. Da wir uns bei BARE nicht auf overt Quantoren verlassen konnten, wurde sehr viel Aufwand in die manuellen Regeln investiert, mit denen die automatisch erlernten Regeln des Zufallswaldes (noch) nicht mithalten können.<sup>65</sup>

Obwohl NEG häufig in den Trainingsdaten vorkommt, erzielen beide GenTagger für dieses Subtag den niedrigsten Recall (abgesehen von MEIST für den regelbasierten Tagger), d. h. es werden kaum Teilsätze ausgegeben, die im Goldstandard mit NEG annotiert wurden. Zum Teil kann man dies darauf zurückführen, dass im Testtext einige ungewöhnlich lange NEG-Passagen vorkommen (die längste besteht aus 18 Teilsätzen), die GenTagger aber nur Merkmale von direkt benachbarten Teilsätzen mit

<sup>61</sup> Wieland 2012, S. 377.

<sup>62</sup> Wieland 2012, S. 385.

<sup>63</sup> Wieland 2012, S. 375.

<sup>64</sup> Wieland 2012, S. 375.

<sup>65</sup> Es sei an dieser Stelle angemerkt, dass der statistische GenTagger durchaus hätte dieselben Regeln erlernen können, da ihm alle dafür notwendigen Merkmale zur Verfügung standen.

einbeziehen. Befindet sich der Negationsmarker weiter vom aktuellen Teilsatz entfernt, wird dieser auch nicht mehr für die Klassifikation herangezogen, und es ist unwahrscheinlich, dass das Subtag NEG vergeben wird. Da BARE und NEG die häufigsten Subtags sind, haben diese Klassen auch den größten Einfluss auf die Mikro-Durchschnitts-Werte.

Abschließend zeigt Abbildung 7 den Lerngraphen des Zufallswald-GenTaggers auf den Testdaten für eine wachsende Menge an Trainingstexten. Eine mögliche Interpretation des Graphen ist, dass das volle Potenzial des statistischen GenTagger noch nicht erreicht ist und die Performanz allein durch die Zugabe weiterer Trainingstexte gesteigert werden kann.

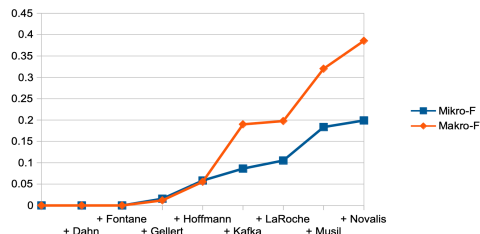


Abb. 7: Performanz des Zufallswald-GenTaggers auf den Testdaten bei schrittweiser Hinzunahme von Trainingstexten. Autorennamen stehen als Kürzel für die einzelnen Werke. [Gödeke et al. 2022]

## 6. Fazit und Ausblick

Das zentrale Ziel dieses Beitrags war es, einen Tagger für die automatische Erkennung von Generalisierungen in fiktionalen Erzähltexten vorzustellen. Für die Taggerimplementierung haben wir ein eigenes Tagset entwickelt, das alle Formen natürlichsprachlicher Quantifikation (allquantifiziert, vage, existenziell) auf Teilsatzebene annotierbar macht. Auf der Basis manuell erstellter Goldstandards wurden ein regelbasierter und ein statistischer Tagger entwickelt, die als zusätzliche Komponenten in die Textverarbeitungspipeline von spaCy eingebunden werden können.<sup>66</sup> Der statistische Zufallswald-Tagger erzielt in fast allen Kategorien die besseren Ergebnisse. Eine Ausnahme bilden Generalisierungen ohne overte Quantifikation, die der regelbasierte Tagger mit einer höheren Performanz klassifiziert. Da Generalisierungen ohne overte Quantifikation zahlenmäßig am häufigsten vorkommen, erreicht der regelbasierte Tagger ein höheres Mikro-F-Maß (24% vs. 20%), wohingegen der statistische Tagger ein höheres Makro-F-Maß erreicht (39% vs. 28%). Allerdings ist beim statistischen Tagger eine kontinuierlich steigende Lernkurve zu erkennen, was nahelegt, dass seine Performanz mit der verarbeiteten Textmenge steigt.

Der Tagger trägt dazu bei, Generalisierungen in fiktionalen Erzähltexten als komplexes literarisches Phänomen erkennbar werden zu lassen. Davon profitiert insbesondere die Narratologie, insofern Generalisierungen Indikatoren für Phänomene wie *Metanarration*, einen *overt narrator*, *nicht-fiktionale Rede* und *explicit truths* sein können. Es werden diachrone Analysen ermöglicht, indem autor\*innen-, strömungs- oder gattungsübergreifend Generalisierungen in Texten bezüglich ihrer Funktion empirisch untersucht werden können. Zudem wird eine Analyse von Generalisierungen auf verschiedenen Textebenen möglich, da sie sowohl im Erzähler- als auch im Figurentext auftreten können (vgl. (16), Abschnitt 2.3). Der Tagger bereichert jedoch nicht nur die Literaturwissenschaft, sondern auch die theoretische Linguistik, deren Modelle und Theorien zur Generizität und Quantifikation durch die Sprachdaten aus literarischen Erzähltexten vor neue Herausforderungen gestellt werden. Aber auch auf der Textebene kann die Linguistik Erkenntnisse gewinnen, da generische Aussagen typischerweise mit dem statischen Aspekt assoziiert werden.<sup>67</sup> Inwiefern Erzählpausen im narrativen Text mit statischem lexikalischem Aspekt einhergehen, ist eine spannende offene Forschungsfrage. Die zentrale kommende Aufgabe wird es sein, mehr Daten für den statistischen Tagger zu generieren, um die Performanz des Taggers zu verbessern. Auch soll überprüft werden, inwiefern neuronale Methoden eine noch höhere Trefferquote erzielen können.

## Danksagungen

Die VolkswagenStiftung und die Deutsche Forschungsgemeinschaft (424264086) haben die Arbeit an diesem Aufsatz ermöglicht, dafür bedanken wir uns herzlich. Wir danken zudem unseren studentischen Hilfskräften für ihre Annotationsarbeit: Friederike Altmann, Jan P. Lau, Jonas Lipski, Evelyn Ovsjannikov, Noreen Scheffel, Ruben M. Van Wijk und Marina L. Wurzbacher. Ebenfalls danken wir zwei anonymen Gutachter\*innen für ihre Kommentare und Hinweise, von denen der vorliegende Aufsatz profitiert hat.

<sup>66</sup> Die Tagger sind verfügbar in [MONAPipe](#) (Release v2.0); vgl. Barth et al. 2022.

<sup>67</sup> Vgl. Carlson 2006.

## Bibliographische Angaben

- Florian Barth / Tillmann Döncke / Benjamin Gittel / Luisa Gödecke / Anna Mareike Weimer / Anke Holler / Caroline Sporleder / Hanna Varachkina: MONACO: Modes of Narration and Attribution Corpus. 2021. [\[online\]](#)
- Florian Barth / Tillmann Döncke / Hanna Varachkina / Caroline Sporleder: MONAPipe: Modes of Narration and Attribution Pipeline for German Computational Literary Studies and Language Analysis in spaCy. In: Proceedings of the 18th Conference on Natural Language Processing (KONVENS: Konferenz zur Verarbeitung natürlicher Sprache/Conference on Natural Language Processing, Potsdam, 12.–15.09.2022). [\[online\]](#)
- Jon Barwise / Robin Cooper: Generalized Quantifiers and Natural Language. In: *Linguistics and Philosophy* 4 (1981), H. 2, S. 159–219. [\[Nachweis im GVK\]](#)
- Johan van Benthem: The Logic of Natural Language. In: *Philosophical Books* 24 (1983), H. 2, S. 99–102. [\[Nachweis im GVK\]](#)
- Leo Breiman / Jerome H. Friedman / Richard A. Olshen / Charles J. Stone: Classification and Regression Trees. New York u. a. 1984. [\[Nachweis im GVK\]](#)
- Leo Breiman: Random Forests. In: *Machine Learning* 45 (2001). DOI: [10.1023/A:1010933404324](#) [\[Nachweis im GVK\]](#)
- Greg N. Carlson: Generics, Habituals, Iteratives. In: *Encyclopedia of Language & Linguistics*. Hg. von Keith Brown. 2. Auflage. Amsterdam u. a. 2006. [\[Nachweis im GVK\]](#)
- Seymour Chatman: Story and Discourse. Narrative Structure in Fiction and Film. New York, NY 1978. [\[Nachweis im GVK\]](#)
- Deutsches Textarchiv. Grundlage für ein Referenzkorpus der neuhochdeutschen Sprache. Hg. von Berlin-Brandenburgische Akademie der Wissenschaften. Berlin 2022. [\[online\]](#)
- Tillmann Döncke: Clause-Level Tense, Mood, Voice and Modality Tagging for German. Düsseldorf 2020. DOI: [10.18653/v1/2020.tlt-1.1](#)
- Tillmann Döncke / Luisa Gödecke / Hanna Varachkina: Annotating Quantified Phenomena in Complex Sentence Structures Using the Example of Generalising Statements in Literary Texts. In: Proceedings of the 17th Joint ACL - ISO Workshop on Interoperable Semantic Annotation. Hg. von Harry Bunt. (ISA 17, Groningen, 16.–17.06.2021) Stroudsburg 2021, S. 20–32. [\[online\]](#)
- Paul Ernst: Die Uhr. In: TextGrid Repository. Digitale Bibliothek. Göttingen 2012. Handle: [11858/00-1734-0000-0002-A2A0-8](#)
- Joseph L. Fleiss: Measuring nominal scale agreement among many raters. In: *Psychological Bulletin* 76 (1971), H. 5, S. 378–382. [\[Nachweis im GVK\]](#)
- Theodor Fontane: Der Stechlin. In: TextGrid Repository. Digitale Bibliothek. Göttingen 2012. Handle: [11858/00-1734-0000-0002-AECF-D](#)
- Annamarie Friedrich / Manfred Pinkal: Automatic recognition of habituals: a three-way classification of clausal aspect. In: Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. Hg. von Lluís Màrquez / Chris Callison-Burch / Jian Su (EMNLP, Lisbon, 17.–21.09.2015). Red Hook, NY 2015. DOI: [10.18653/v1/D15-1294](#) [\[Nachweis im GVK\]](#)
- Annamarie Friedrich / Alexis Palmer / Melissa Peate Sørensen / Manfred Pinkal: Annotating genericity: a survey, a scheme, and a corpus. In: Proceedings of the 9th Linguistic Annotation Workshop. Hg. von Adam Meyers / Ines Rehbein / Heike Zinsmeister (LAW 9, Denver, CO, 31.05.–05.06.2015). Red Hook, NY 2015, S. 21–30. DOI: [10.3115/v1/W15-1603](#) [\[Nachweis im GVK\]](#)
- Annamarie Friedrich / Alexis Palmer / Manfred Pinkal: Situation entity types: automatic classification of clause-level aspect. In: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Hg. von Katrin Erk / Noah A. Smith (ACL 54, Berlin, 07.–12.08.2016). Stroudsburg, PA 2016, S. 1757–1768. DOI: [10.18653/v1/P16-1166](#)
- Benjamin Gittel: Essayismus als Fiktionalisierung von unsicheres Wissen prozessierender Reflexion. In: *Scientia Poetica* 19 (2015), H. 1, S. 136–171. [\[Nachweis im GVK\]](#)
- Evelyn Gius / Janina Jacke: Zur Annotation narratologischer Kategorien der Zeit. Guidelines zur Nutzung des CATMA-Tagsets. Version 2.0. Hamburg 2016. PDF. [\[online\]](#)
- Evelyn Gius / Jan Christoph Meister / Malte Meister / Marco Petris / Christian Bruck / Janina Jacke / Mareike Schumacher / Dominik Gerstorfer / Marie Flüh / Jan Horstmann: CATMA 6. Version 6.5. 2022. In: Zenodo. DOI: [10.5281/zenodo.1470118](#)
- Johann Wolfgang von Goethe: Die Wahlverwandschaften. In: TextGrid Repository. Digitale Bibliothek. Göttingen 2012. Handle: [11858/00-1734-0000-0006-6A93-D](#)
- Irene Roswitha Heim: The Semantics of Definite and Indefinite Noun Phrases. Ann Arbor, Dissertation, University of Massachusetts 1982. [\[Nachweis im GVK\]](#)
- Matthew Honnibal / Mark Johnson: An Improved Non-monotonic Transition System for Dependency Parsing. In: Conference on Empirical Methods in Natural Language Processing. Hg. von Association for Computational Linguistics. (EMNLP, Lissabon, 17.–21.09.2015). Red Hook, NY 2015, S. 1373–1378. DOI: [10.18653/v1/D15-1162](#) [\[Nachweis im GVK\]](#)
- Franz Kafka: Der Bau. In: Projekt Gutenberg. Hamburg. Basierend auf der Druckausgabe Frankfurt / Main 1990. [\[online\]](#)
- Hans Kamp: A Theory of Truth and Semantic Representation. In: *Formal Semantics. The Essential Readings*. Hg. von Paul Portner / Barbara H. Partee. Oxford 2002, S. 189–222. [\[Nachweis im GVK\]](#)
- Eva-Maria Konrad: Signposts of Factuality: On Genuine Assertions in Fictional Literature. In: *Art and Belief*. Hg. von Ema Sullivan-Bissett / Helen Bradley / Paul Noordhof. Oxford 2017, S. 42–62. (= Mind Association Occasional Series) [\[Nachweis im GVK\]](#)
- Eva-Maria Konrad: Dimensionen der Fiktionalität. Analyse eines Grundbegriffs der Literaturwissenschaft. Münster 2014. [\[Nachweis im GVK\]](#)
- Manfred Krifka / Francis Jeffry Pelletier / Gregory N. Carlson / Alice ter Meulen / Godehard Link / Gennaro Chierchia: Genericity: An Introduction. In: *The generic book*. Hg. von Greg N. Carlson / Francis Jeffry Pelletier. Chicago u. a. 1995, S. 1–124. [\[Nachweis im GVK\]](#)
- Markus Krug / Frank Puppe / Fotis Jannidis / Luisa Macharowsky / Isabella Reger / Lukas Weimar: Rule-based Coreference Resolution in German Historic Novels. In: Proceedings of the Fourth Workshop on Computational Linguistics for Literature. Hg. von Anna Feldman / Anna Kazantseva / Stan Szpakowicz / Corina Koolen (NAACL-HLT 4, Denver, 04.06.2015). Red Hook, NY 2015, S. 98–104. DOI: [10.3115/v1/W15-0711](#) [\[Nachweis im GVK\]](#)
- Silke Lahn / Jan Christoph Meister: Einführung in die Erzähltextanalyse. 2., aktualisierte Auflage. Stuttgart u. a. 2013. [\[Nachweis im GVK\]](#)
- Peter Lamarque / Stein H. Olsen: Truth, Fiction, and Literature. A Philosophical Perspective. Oxford 1994. [\[Nachweis im GVK\]](#)
- Sarah-Jane Leslie / Sangeet Khemlani / Sam Glucksberg: All Ducks Lay Eggs: The Generic Overgeneralization Effect. In: *Journal of Memory and Language* 65 (2011), H. 1, S. 15–31. [\[Nachweis im GVK\]](#)
- David K. Lewis: Adverbs of Quantification. In: *Formal Semantics of Natural Language*. Hg. von Edward L. Keenan. Cambridge, UK 1975, S. 178–188. [\[Nachweis im GVK\]](#)
- Jukka Mikkonen: Assertions in Literary Fiction. In: *Minerva* 13 (2009), S. 144–180. [\[online\]](#)
- Alexis Mitchell / Stephanie Strassel / Mark Przybicki / JK Davis / George R. Doddington / Ralph Grishman / Adam Meyers / Ada Brunstein / Lisa Ferro / Beth Sundheim: ACE-2 Version 1.0. LDC2003T11, Philadelphia 2003. DOI: [10.35111/kcqq-v224](#)
- Robert Musil: Der Mann ohne Eigenschaften. Erstes Buch. Hg. von Adolf Frisé. In: Projekt Gutenberg. Hamburg. Basierend auf der Druckausgabe 1970. [\[online\]](#)
- Ansgar Nünning: Grundzüge eines kommunikationstheoretischen Modells der erzählerischen Vermittlung. Trier 1989. [\[Nachweis im GVK\]](#)
- Thomas Petraschka: Interpretation und Rationalität. Billigkeitsprinzipien in der philologischen Hermeneutik. Berlin 2014 (= *Historia Hermeneutica / Series Studia*, 11) [\[Nachweis im GVK\]](#)

Nils Reiter / Anette Frank: Identifying Generic Noun Phrases. In: Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics. Hg. von Jan Hajič / Sandra Carberry / Stephan Clark / Joakim Nivre (ACL 48, Uppsala, 11.–16.07.2010). Red Hook, NY 2010, S. 40–49. [[online](#)] [[Nachweis im GVK](#)]

Michael Scheffel / Matías Martínez: Einführung in die Erzähltheorie. 10., überarbeitete Auflage. München 2016. [[Nachweis im GVK](#)]

John R. Searle: The Logical Status of Fictional Discourse. In: New Literary History 6 (1975), H. 2, S. 319–332. [[Nachweis im GVK](#)]

Lev Nikolaevič Tolstoj: Anna Karenina. In: TextGrid Repository. Digitale Bibliothek. Göttingen 2012. Handle: [11858/00-1734-0000-0005-57A8-B](#)

Achim Vesper: Literatur und Aussagen über Allgemeines. In: Wahrheit, Wissen und Erkenntnis in der Literatur. Philosophische Beiträge. Hg. von Christoph Demmerling / Ingrid Vendrell Ferran. Berlin 2014, S. 181–196. [[Nachweis im GVK](#)]

Christopher Walker / Stephanie Strassel / Julie Medero / Kazuaki Maeda: ACE 2005 Multilingual Training Corpus LDC2006T06. Philadelphia 2006. DOI: [10.35111/mwxc-vh88](#)

Christoph Martin Wieland: Geschichte des Agathon. In: TextGrid Repository. Digitale Bibliothek. 2012. Handle: [11858/00-1734-0000-0005-A68F-A](#)

## Abbildungs- und Tabellenverzeichnis

Tab. 1: Tagset MONACO. [Gödeke et al. 2022 nach Barth et al. 2021]

Tab. 2: Korpus MONACO. [Gödeke et al. 2022 nach Barth et al. 2021]

Abb. 1: Textverarbeitungskomponenten für verschiedene sprachliche Ebenen: Ein Text durchläuft die einzelnen Komponenten von links nach rechts und von oben nach unten. [Gödeke et al. 2022]

Abb. 2: Regelbasierter GenTagger. [Gödeke et al. 2022]

Tab. 3: Größe der Trainings-, Entwicklungs- und Testdaten. [Gödeke et al. 2022]

Tab. 4: Merkmale für den statistischen GenTagger; Fragezeichen markieren Boole'sche Merkmale. [Gödeke et al. 2022]

Abb. 3: Die wichtigsten Merkmale des Entscheidungsbaum-GenTaggers. [Gödeke et al. 2022]

Abb. 4: Berechnung von Precision (P), Recall (R) und F-Maß (F) für ein einzelnes Subtag y. [Gödeke et al. 2022]

Abb. 5: Berechnung des Mikro-Durchschnitts von Precision (P), Recall (R) und F-Maß (F). [Gödeke et al. 2022]

Abb. 6: Berechnung des Makro-Durchschnitts von Precision (P), Recall (R) und F-Maß (F). [Gödeke et al. 2022]

Tab. 5: Mikro-Durchschnittswerte der GenTagger auf Entwicklungs- und Testdaten. [Gödeke et al. 2022]

Tab. 6: Tag-spezifische Makro-Durchschnittswerte der GenTagger auf den Testdaten. [Gödeke et al. 2022]

Abb. 7: Performanz des Zufallswald-GenTaggers auf den Testdaten bei schrittweiser Hinzunahme von Trainingstexten. Autorennamen stehen als Kürzel für die einzelnen Werke. [Gödeke et al. 2022]