

# Konzept und Klassifikation literarischer Raumentitäten<sup>1</sup>

Florian Barth<sup>2</sup>

**Abstract:** Innerhalb der “Computational Narratology” werden literaturwissenschaftlich-textuelle Konzepte formalisiert, um sie anschließend mit algorithmischen und maschinellen Verfahren automatisch identifizieren zu können. Dieser Prozess, der eine enge Verknüpfung theoretischer Grundlagen und computationeller Umsetzung erfordert, wird in diesem Beitrag am Beispiel narratologischer Raumkategorien vorgestellt. Eine Pilotannotation demonstriert konzeptuelle Grundlagen der Kategorien und deren intersubjektives Verständnis anhand des Inter-Annotator-Agreements. Anschließend werden Features zur Erkennung jener Ortsreferenzen entwickelt, an denen die diegetische Handlung eines literarischen Textes angesiedelt ist, und prototypische Machine-Learning-Modelle zur Klassifikation präsentiert.

**Keywords:** Computational Literary Studies; Natural Language Processing; Machine Learning; Spatial Analysis; Annotation

## 1 Einordnung

Unter dem Begriff der “Computational Humanities” lassen sich jene Methoden innerhalb der Digital Humanities fassen, die einen stärkeren Fokus auf die Entwicklung numerischer und formaler Modelle legen, beispielsweise durch die Verwendung von Machine Learning, statistischen Verfahren oder algorithmischer Sprachverarbeitung [Bu20; Ro19]. Als spezifischer Unterbegriff hat sich innerhalb der digitalen Literaturwissenschaft in jüngster Zeit das Schlagwort “Computational Literary Studies” durchgesetzt, worunter man die Anwendung obiger Verfahren für genuin literaturwissenschaftliche Fragestellungen versteht. Bei Erzähltexten wird insbesondere die Formalisierung narratologischer Konzepte forciert, etwa im Rahmen von Shared-Tasks [GRW19], sowie eine anschließende maschinelle Erkennung angestrebt, z. B. von narrativen Szenen [Gi19] oder dem diegetischen Status der Erzählinstanz und ihrer Perspektive (point of view; [EF16]).

In diesem Kontext widmet sich der vorliegende Beitrag der Formalisierung und Klassifikation narratologischer Kategorien des Raums nach Dennerlein [De09] und Piatti [Pi08]. In der bisherigen literaturwissenschaftlichen Forschung wurde die Kategorie des Raums insbesondere vor dem Hintergrund des *Spatial Turn* ausführlich besprochen [So89] und innerhalb der Narratologie konzeptualisiert – dies erfolgte jedoch vor allem theoretisch,

---

<sup>1</sup> Der Beitrag basiert auf der Masterarbeit des Autors, die von Gabriel Viehhauser, Nils Reiter und Roman Klinger unterstützt wurde, und denen hierfür ein besonderer Dank gewidmet ist.

<sup>2</sup> Universität Göttingen, Institut für Informatik / Niedersächsische Staats- und Universitätsbibliothek Göttingen (SUB), Abteilung Forschung & Entwicklung, florian.barth@uni-goettingen.de

ohne eine Anwendung des Konzepts bei längeren Textpassagen oder einer Prüfung auf Formalisierbarkeit. Deshalb wird in diesem Beitrag eine Pilotannotation vorgestellt, welche die Kategorie der literarischen RAUMENTITÄT (SPATIAL ENTITY) anhand einheitlicher Richtlinien erfasst und mittels Inter-Annotator-Agreement evaluiert.

Anschließend erfolgt die Klassifikation der RAUMENTITÄTEN in handlungsrelevante SETTINGS und lediglich ERWÄHNT RÄUME (MENTIONS) – eine Differenzierung, die einen zentralen Aspekt der narratologischen Auseinandersetzung mit Raum darstellt [De09; Pi08]. Wir bezeichnen diese Unterscheidung im Folgenden auch als *Klassifikation der Handlungsrelevanz*, für deren Automatisierung linguistisch-textuelle Features und prototypische Machine-Learning-Verfahren präsentiert werden. Eine Schwierigkeit dieser Klassifikation liegt in der Einbettung der Kategorie des SETTINGS in übergeordnete Textstrukturen, zu denen insbesondere *narrative Ebenen* gehören, von denen die Klassifikation und bereits die formale Konzeptualisierung von SETTINGS abhängt. Beim vorliegenden SETTINGS-Begriff wurde daher auf eine definitorische Unabhängigkeit von Erzählebenen geachtet und die Textauswahl für die Annotation derart getroffen, dass lokale Features nicht von einer narrativen Ebene beeinflusst werden (siehe Abschnitt 2.2).

## 2 Konzept und Pilotannotation

### 2.1 RAUMENTITÄTEN

Literarische RAUMENTITÄTEN (SPATIAL ENTITIES) umfassen i.) Toponyme (konkrete geographische Raumnamen wie “Frankreich” oder “Berlin”), ii.) fiktionale und faktuale Eigennamen (“Eiffelturm”, “Schicksalsberg”) sowie iii.) Gattungsbezeichnungen (“Problemviertel”, “Speisekammer”) [De09].<sup>3</sup> Grundsätzlich erfassen wir RAUMENTITÄTEN als ganze Nominalphrase (NP) und, sofern eine Präposition voransteht, als Präpositionalphrase (PP; Bsp. 1.b, 1.d). NPs bzw. PPs, die sich auf unterschiedliche RAUMENTITÄTEN beziehen, können sich zudem überlappen oder umschließen, und werden dann verschachtelt annotiert (1.a).

1. a Passepartout befand sich also, nachdem halb zwölf vorüber war, allein [**im Hause** [**der Savile Row**]<sub>SPATIAL ENTITY</sub>]<sub>SPATIAL ENTITY</sub>. Sogleich machte er sich daran, [**es**]<sub>SPATIAL ENTITY</sub> [[**vom Keller**] [**bis zum Speicher**]<sub>SPATIAL ENTITY</sub>]<sub>SPATIAL ENTITY</sub> zu besichtigen. (Verne: Reise um die Erde in 80 Tagen)
- b Ein junger Mann saß zusammengesunken [**neben der Gaslaterne**]<sub>SPATIAL ENTITY</sub> [**auf dem Bordstein**]<sub>SPATIAL ENTITY</sub>. (Dos Passos: Manhattan Transfer)

<sup>3</sup> Im Natural Language Processing werden lediglich Toponyme von bestehenden Named-Entity-Klassifikatoren als Geographical Entity (GEO) oder Geopolitical Entity (GPE) erkannt. Für die englische Sprache gibt es z. B. den Named-Entity-Classifer von Finkel et al. [FGM05] und fürs Deutsche jenen von Faruqui; Padó [FP10].

- c Felicie lief [zu Homais]<sub>SPATIAL ENTITY</sub>, der es aller Welt ausposaunte.<sup>4</sup> (Flaubert: Madame Bovary)
- d Als ich [auf dem Wege]<sub>SPATIAL ENTITY</sub> hinunter [zum Mittagessen]<sub>SPATIAL ENTITY</sub> [an dem Zimmer]<sub>SPATIAL ENTITY</sub> vorüberging, sah ich [durch die geöffnete Thür]<sub>SPATIAL ENTITY</sub>. (Brontë: Jane Eyre)

Als RAUMENTITÄTEN gelten zudem iv.) Objekte, die von Figuren betretbar sind (“Flugzeuge”, “Autos”, aber auch “Schränke”, “Kisten”),<sup>5</sup> v.) Distanzen zwischen zwei RAUMENTITÄTEN (Bsp. 1.a), vi.) Deiktika (die Adverbien “hier”, “da” und “dort”) sowie vi.) unspezifische Konkreta (“außen”/“innen”, “Heimat”, “Zuhause”) [De09]. Die zuvor genannten Formen lassen sich allein anhand des Nomens innerhalb einer RAUMENTITÄT bzw. des Adverbs (bei Deiktika und unspezifischen Konkreta) bestimmen. Andere Formen sind abhängig vom Kontextverständnis einer Textpassage, etwa bei vi.) Objekten, gegenüber denen sich Figuren verorten (Bsp. 1.b) [De09], vii.) Figurennennungen, die auf einen Ort referieren (Bsp. 1.c), oder bei viii.) Ereignissen, die innerhalb einer NP bzw. PP ausgedrückt werden und gleichzeitig eine Ortsangabe darstellen (“zum Mittagessen” in Bsp. 1.d). Darüber hinaus wurden in der Pilotannotation Pronomen und Konjunktionen als RAUMENTITÄTEN berücksichtigt, wenn sie mit einer zuvor (Katapher) oder danach genannten RAUMENTITÄT (Anapher) koreferent sind (“es” in Bsp. 1.a) [Re13].

## 2.2 Klassifikation der Handlungsrelevanz

Bei jeder RAUMENTITÄT soll entschieden werden, ob dort die Handlung der Erzählung aktiv stattfindet, wodurch sich die Entität zu einem SETTING qualifiziert. Ein SETTING zeichnet sich aus durch 1.) die Präsenz und aktive Handlung von Figuren verbunden mit einer *Jetzt-Zeitlichkeit* (Bsp. 2.a) [De09; Pi08], oder durch 2.) eine besondere Ereignishaftigkeit bei Vorgängen ohne direkte Figurenbeteiligung (2.b) [De09].<sup>6</sup>

2. a In diesem Augenblicke klopfte es an die Türe [des kleinen Salons, worin sich Phileas Fogg aufhielt.]<sub>SETTING</sub>.<sup>7</sup> (Verne: Reise)
- b Um halb ein Uhr hielt [der Zug]<sub>SETTING</sub> [auf der Station Benares]<sub>SETTING</sub>. (Verne: Reise)

<sup>4</sup> Dennoch wird “Homais” in diesem Beispiel auch als eigenständige Figuren-Entität betrachtet (neben der Funktion als Raumreferenz).

<sup>5</sup> In der Raumnarratologie wird zwischen punktuellen Orten und weiterreichenden Räumen unterschieden [De09], die in unserem Fall beide innerhalb des Konzepts der RAUMENTITÄT angesiedelt sind.

<sup>6</sup> Bestehende informatische oder computerlinguistische Ansätze wie das Konzept des *ISO-Space* [PMV11] erfassen zwar u. a. literaturwissenschaftlich relevante Aspekte der Raumdarstellung sowie die Ereignishaftigkeit eines Raumes im Zusammenhang einer Bewegung [Pu15]. Jedoch wird kein Bezug zur Narration bzw. dem Plot eines Textes hergestellt, und zudem bestehen grundlegende Diskrepanzen zu den literaturwissenschaftlichen Kategorien, z. B. hinsichtlich der Beurteilung von Figuren als räumliche Entitäten.

<sup>7</sup> Relativsätze werden mitannotiert, wenn sie direkt an die NP bzw. PP angeschlossen sind.

Als ERWÄHNTER RAUM (MENTION) werden dagegen RAUMENTITÄTEN im Kontext von Reflexionen oder Beschreibungen (3.a) sowie repetitiv erzählten (3.b) oder zukünftig bevorstehenden Ereignissen (3.c) annotiert.

3. a [Das nicht eben prachtvolle Haus [in Savile Row]<sub>MENTION</sub>]<sub>MENTION</sub> empfahl sich durch größte Bequemlichkeit. (Verne: Reise)
- b Er machte regelmäßig die Fahrten [[von Brindisi]<sub>MENTION</sub> [nach Bombay]<sub>MENTION</sub>]<sub>MENTION</sub> [durch den Suez-Canal]<sub>MENTION</sub>. (Verne: Reise)
- c In einem Monat wollte sie mit ihrem Bruder [mit dem Auto]<sub>MENTION</sub> [nach New York]<sub>MENTION</sub> fahren. (Kerouac: Unterwegs)

Wie eingangs dargelegt, beeinflusst die jeweilige narrative Ebene die Klassifikation aller darin befindlichen SETTINGS: Auf einer typischen zweiten Erzählebene – einer Geschichte innerhalb einer Geschichte, die z. B. von einer intradiegetischen Figur erzählt wird [De09; Ge88] – entstehen neue SETTINGS, obwohl diese im genannten Beispiel lediglich von einer Figur der ersten Erzählebene “erwähnt” werden. Zum Konzept einer Erzählebene gehören nach Ryan ebenfalls Passagen mit einem abweichenden ontologischen Status, z. B. Traumsequenzen oder eingeschobene Textartefakte wie Briefe [Ry91], die Dennerlein als modale oder mediale Komponente erfasst, und die darin befindlichen RAUMENTITÄTEN als ERWÄHNTE RÄUME klassifiziert [De09]. Im vorliegenden SETTING-Konzept gehen wir hingegen auch bei ontologisch abweichenden Erzählebenen davon aus, dass dort neue SETTINGS erkannt werden können, sodass eine Unabhängigkeit von der Kategorie der narrativen Ebene gewährleistet ist (siehe dazu: [Ba19]).

### 2.3 Annotation

Die Pilotannotation wurde von zwei Annotatorinnen im Annotationstool *WebAnno* durchgeführt [Yi13],<sup>8</sup> und basiert auf fünf Texten aus dem *Corpus of German Novels (DROC; [Kr17])*, in welchem Annotations-Layer für direkte Rede sowie Figuren und deren Koreferenz vorliegen.<sup>9</sup>

Für das Agreement auf Token-Ebene wurde *Cohens Kappa* verwendet, welches sich für den Vergleich von zwei Annotationen eignet [Co60]. Da die Berechnung von Kappa auf dem beobachteten Agreement  $P(A)$  sowie dem erwarteten Agreement  $P(E)$  basiert, kann jedes Token nur einmalig in den Vergleich einbezogen werden. Bei überlappenden Annotationen

<sup>8</sup> Eine Annotation wurde vom Autor selbst durchgeführt und eine zweite von einer wissenschaftlichen Hilfskraft des Instituts für Literaturwissenschaft (Abteilung Digital Humanities) der Universität Stuttgart. Die Annotationen, die in *WebAnno* intern mit Apache UIMA realisiert sind [Sc04], wurden ins Format *XML Metadata Interchange (XMI)* exportiert. XMI ist ein OMG-Standard und geeignet Objekt-Graphen wie die *Common Analysis Structure (CAS)* von UIMA zu repräsentieren [Ap10]. Im Falle der vorliegenden Annotationen eignet sich XMI als Stand-Off-Format, um komplexe, mehrfach überlappende Annotationen abzubilden (vgl. Tabelle 1).

<sup>9</sup> Konkret handelt es sich um Romanauszüge aus Theodor Fontanes *Cécile* sowie *Quitt*, Gustave Flauberts *Madame Bovary*, Lew Tolstois *Anna Karenina* und Émile Zolas *Nana*. Alle Textausschnitte beinhalten lediglich eine narrative Ebene.

	In	einer	Viertelstunde	mußte	der	von	Galveston	nach	dem	Norden	führende	Zug	da sein.
1					M	M	M	M	M	M	M	M	
2						M	M	M	M	M			
3						M	M						
4								M	M	M			

Tab. 1: Mehrfach überlappende Annotationen von ERWÄHNTEN RÄUMEN (MENTIONS)

besteht dadurch die Schwierigkeit, dass für ein Token (z. B. “Galveston” in Tabelle 1) mehrere Annotationen einer Annotatorin bestehen, für Kappa pro Token allerdings nur eine Annotations-Ebene berücksichtigt werden kann. Hier wurde daher der Ansatz gewählt, Überlappungen zu verodern: Aus den maximalen Spans überlappender Annotationen wurde bei jeder Annotatorin zufällig eine Annotation mitsamt aller Tokens herausgegriffen. Insbesondere bei mehrfachen Überlappungen gehen dadurch jedoch Informationen verloren.

Der  $F_1$ -Score funktioniert hingegen auf Entitäten-Ebene: Übereinstimmende Entitäten werden als True Positives (TP) erfasst; eine fehlende Übereinstimmung zwischen den Annotatorinnen erzeugt jeweils ein False Positive (FP) und ein False Negative (FN). Wir verwenden einen *Exact-F-Score*, bei dem nur die exakte Übereinstimmung aller Tokens einer RAUMENTITÄT als TP gewertet wird, und einen *Fuzzy-F-Score*, bei dem die Übereinstimmung von mindestens einem Token für ein TP ausreichend ist. Insgesamt kann festgehalten werden, dass eine Entitäten-basierte Agreement-Berechnung mittels  $F_1$ -Score bei überlappenden und aus mehreren Token bestehenden RAUMENTITÄTEN ein präziseres Agreement ermöglicht, da eine konkrete Zuordnung der von den Annotatorinnen gemeinten Entitäten vorgenommen und bewertet wird.<sup>10</sup>

	F-Fuzzy	F-Exact	Cohens Kappa
<b>Erkennung der RAUMENTITÄTEN</b>			
F-Macro / Kappa (Mittelwert aus Einzeltexten)	0.86	0.79	0.79
F-Micro / Kappa (alle Tokens)	0.86	0.72	0.78
<b>Erkennung der RAUMENTITÄTEN &amp; SETTING/MENTION-Klassifikation</b>			
F-Macro / Kappa (Mittelwert aus Einzeltexten)	0.73	0.64	0.72
F-Micro / Kappa (alle Tokens)	0.72	0.61	0.71

Tab. 2: Agreement für RAUMENTITÄTEN sowie RAUMENTITÄTEN inkl. Klassifikation der Handlungsrelevanz

Das Agreement wird zunächst für die Erkennung der RAUMENTITÄTEN berechnet sowie in einem zweiten Schritt für die RAUMENTITÄTEN inklusive der Klassifikation der Handlungsrelevanz. Bei letzterer Variante wird ein TP bzw. eine Übereinstimmung bei Kappa nur akzeptiert, wenn auch die Klassifikation der Handlungsrelevanz für eine RAUMENTITÄT zwischen beiden Annotatorinnen übereinstimmt. Tabelle 2 zeigt die Evaluations-Ergebnisse, bei welchen die Werte als Macro- und Micro-F-Score sowie in zwei Varianten für Kappa angegeben sind. Analog zum Macro-F-Score, bei dem für alle Einzeltexte zunächst

<sup>10</sup> Beim Alignment komplexer überlappender Annotationen wurden zuerst die Entitäten mit dem jeweils niedrigsten Start-Index und der längsten Token-Folge verglichen.

Precision und Recall berechnet und aus den gemittelten Werten (Macro-Precision sowie Macro-Recall) der F-Score gebildet wird [SL09], berechnen wir ein gemitteltes Kappa aus den Kappa-Werten der Einzeltexte. Damit finden unterschiedlich lange Einzeltexte die gleiche Berücksichtigung im Endergebnis. Beim Micro-F-Score, ermittelt aus Precision und Recall über alle TP, FP und FN der Einzeltexte, haben längere Texte einen stärkeren Einfluss auf das Ergebnis – ergänzend dazu wurde ein Kappa für alle Token-Vergleiche in allen Texten ermittelt. Das Agreement beim Macro-F-Score weist mit 0.79 (Exakt) bzw. 0.86 (Fuzzy) eine hohe Übereinstimmung auf, wie auch die gemittelten Kappa-Werte der Einzeltexte mit 0.79, während der Micro-F-Score und der Kappa-Wert basierend auf allen Tokens etwas darunter liegen.<sup>11</sup> Insbesondere der Fuzzy-F-Score und die Kappa-Werte belegen somit ein konstant hohes Agreement für die RAUMENTITÄTEN. Wird zusätzlich die Handlungsrelevanz in das Agreement miteinbezogen, sinken die Werte im Mittel über Micro- und Macro-F-Score um 0.13 und bei Kappa um 0.7. Somit herrscht auch bei der Klassifikation der Handlungsrelevanz ein hohes Agreement.

### 3 Feature-Engineering für die Klassifikation der Handlungsrelevanz

Als Features für die Klassifikation der Handlungsrelevanz dienen linguistisch-textuelle Marker, die Indikatoren für SETTINGS bzw. ERWÄHNTE RÄUME darstellen, und deren Extraktion sowie Funktionalität im Folgenden dargelegt wird. Wir stellen zunächst Features in gestufter Satzumgebung vor (Verbkategorien, temporale Marker, Zeitformen), bei denen der inhaltliche Kontext von einem oder mehreren Teilsätzen Berücksichtigung findet. Dem schließen sich weitere einstufige Features an, die sich auf die jeweilige RAUMENTITÄT bzw. den kompletten umgebenden Satz beziehen (Direkte Rede, Figurenreferenzen, Präpositionen).

#### 3.1 Features in gestufter Satzumgebung

Zur Funktionalität einiger Features ist eine Eingrenzung auf den lokalen Teilsatz notwendig, der die RAUMENTITÄT umgibt. Beispielsweise können innerhalb eines Satzes verschiedene Klassifikationen getroffen werden (Bsp. 4.), die durch derartige Features anhand der lokalen Umgebung unterscheidbar sind.

4. Endlich hielt er es nicht mehr aus, und da er vermutete, sie sei [nach Rouen]<sub>MENTION</sub> gefahren, ging er ihr [auf der Landstraße]<sub>SETTING</sub> eine halbe Wegstunde weit entgegen. (Flaubert: Madame Bovary)

Zur Verarbeitung dieser Fälle werden die oben genannten Features (Verbkategorien, temporale Marker, Zeitformen) basierend auf dem Abhängigkeitsbaum nach Satzumgebungen

<sup>11</sup> Die niedrigeren Werte für den Micro-F-Score und das Kappa für alle Tokens resultieren aus der stärkeren Gewichtung von Fontanes *Quitt*, welches aufgrund zahlreicher überlappender Entitäten das geringste Agreement aller Einzeltexte aufweist, jedoch die umfangreichste Textprobe darstellt.

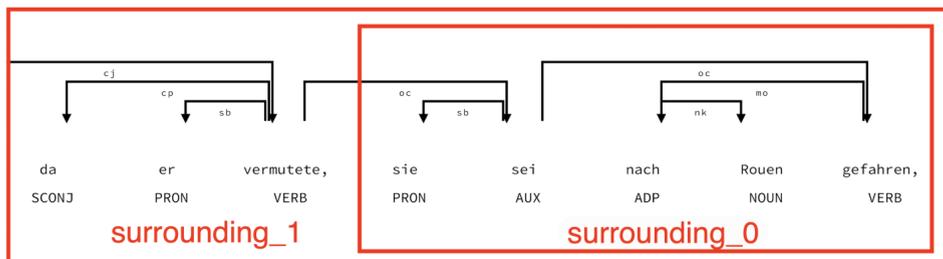


Abb. 1: Satzumgebungen “surrounding\_0” und “surrounding\_1” für Beispiel 4.

gestaffelt. Für die RAUMENTITÄT “nach Rouen” aus Bsp. 4. wird der unmittelbare Teilsatz extrahiert (“surrounding\_0” in Abbildung 1) sowie die im Dependenzbaum nächsthöhere Phrase (“surrounding\_1”).<sup>12</sup> Die Klassifikation des ERWÄHNTEN RAUMES ist in diesem Fall maßgeblich durch die Verbkategorien bestimmt: Der dem ERWÄHNTEN RAUM übergeordnete Teilsatz drückt eine Vermutung aus (Verb “vermuten” in surrounding\_1, Abbildung 1), an die sich ein Konjunktiv (“sei gefahren”) in der lokalen Satzumgebung der RAUMENTITÄT (“surrounding\_0”) anschließt. Wir sprechen von “Umgebungen”, weil bei der im Dependenzbaum nächsthöheren Phrase auch alle untergeordneten Teilsätze abgesehen von den bereits eruierten Umgebungen mit einbezogen werden. So besteht die zweite äußere Umgebung (“surrounding\_2”) für den ERWÄHNTEN RAUM in Beispiel 4. nicht nur aus dem direkt übergeordneten Satz zu “surrounding\_1” (mit dem Kopf der Phrase “hielt”), sondern auch aus dem angeschlossene Hauptsatz (“surrounding\_2”: “Endlich hielt er es nicht mehr aus, und [...] ging er ihr auf der Landstraße eine halbe Wegstunde weit entgegen.”; vgl. auch Tabelle 3). Dies dient der Informationsgewinnung für eine optimale Klassifikation der entsprechenden Entität. Tabelle 3 zeigt für beide RAUMENTITÄTEN aus Beispiel 4. die jeweiligen Umgebungen sowie die extrahierten Verben. Während der ERWÄHNTEN RAUM durch eine Vermutung gekennzeichnet ist, steht das SETTING (“auf der Landstraße”) im Kontext einer aktiven Handlung verknüpft mit dem Verb “gehen”.<sup>13</sup>

Zur Vereinheitlichung der Verben dient eine Kategorisierung anhand der *Levin-Klassen* [Le95], die im lexikalisch-semantischen Netz *GermaNet* Verwendung finden und dort ausgelesen wurden [Ge18]. Jedem Verb wird eine von 15 Klassen zugeordnet, z. B. ist für “vermuten” (Bsp. 4.) die Klasse “Kognition” definiert, was als Indikator für einen ERWÄHNTEN RAUM gelten kann.<sup>14</sup>

Temporale Marker können ein eindeutiges Indiz für die Jetzt-Zeitlichkeit bzw. den vergangenen, zukünftigen oder repetitiven Charakter von Ereignissen im Kontext der Raumdarstellung liefern und werden ebenfalls in gestufter Satzumgebung betrachtet. Wir verwenden den

<sup>12</sup> Das Parsing und die Navigation im Dependenz-Baum wurde mit der Python-Library *spaCy* realisiert [HM17].

<sup>13</sup> Bereits Dennerlein [De09] hebt mit Verweis auf Herman [He04] die besondere Bedeutung von Verben bei der Bestimmung von Aktionen bzw. Ereignissen hervor.

<sup>14</sup> Einigen Verben werden in *GermaNet* mehreren Klassen zugeordnet – in diesem Fall wurde die Mehrheitsklasse als Kategorie gewählt bzw. bei mehreren Klassen eine Kategorie aus den höchstgenannten Klassen gebildet.

Raumentität	surrounding_0	verb_0	surrounding_1	verb_1	surrounding_2	verb_2
nach Rouen (mention)	[sie, sei, nach, Rouen, gefahren]	sein	[da, er, vermutete, .]	vermuten	[Endlich, hielt, er, es, nicht, mehr, aus, ,, und, ,, ging, er, ihr, auf, der, Landstraße, eine, halbe, Wegstunde, weit, entgegen, .]	halten
auf der Landstraße (setting)	[,, ging, er, ihr, auf, der, Landstraße, eine, halbe, Wegstunde, weit, entgegen]	gehen	[Endlich, hielt, er, es, nicht, mehr, aus, ,, und, da, er, vermutete, ,, sie, sei, nach, Rouen, gefahren, .]	halten		

Tab. 3: Gestufte Satzumgebungen und extrahierte Verben für Beispiel 4.

Temponym-Tagger *HeidelTime* [Ku16], der vier verschiedene Typen von Zeitmarkern identifiziert: DURATION (“in einem Monat”, Bsp. 3.c auf S. 4); TIME (“um halb ein Uhr”, Bsp. 2.b); DATE (“September 1828”); SET (“regelmäßig”, Bsp. 3.b).

Das Feature der Zeitform liefert im Idealfall direkte Informationen darüber, ob auf gegenwärtige, vergangene oder zukünftige Ereignisse referiert wird. Die beiden rot dargestellten Teilsätze in Bsp. 5. beinhalten Futur I und referieren 1.) auf die Abfahrt eines Dampfbootes am Datum des 25. und 2.) auf dessen später in der Zukunft liegende Ankunft in Kalkutta. Alle RAUMENTITÄTEN innerhalb dieser Satzteile können anhand der Zeitform klar als ERWÄHNTE RÄUME klassifiziert werden. Nur der grün dargestellte Teilsatz im Präsens beschreibt die Jetzt-Zeitlichkeit. Bei der Integration der Zeitform als Feature greifen wir auf eine Tense-Cluster-Detection von Bögel et al. [BSG14] zurück.

5. » ... **Am 25. zu Mittag wird** [ein Dampfboot]<sub>MENTION</sub> [[von Kalkutta]<sub>MENTION</sub> [nach Hongkong]<sub>MENTION</sub>]<sub>MENTION</sub>; **abgehen. Jetzt haben wir erst den 22., und werden noch zeitig** [zu Kalkutta]<sub>MENTION</sub> **entreffen.**« (Verne: Reise)

### 3.2 Einstufige Features

Beispiel 5. stellt einen Idealfall dar, weil darin direkte Rede vorliegt – Erzähltexte sind in der Regel jedoch im epischen Präteritum verfasst (Bsp. 1.a–1.d und 2.a–2.b). Daher wird für jede RAUMENTITÄT geprüft, ob sich diese in einer im DROC-Korpus annotierten direkten Rede befindet. Wir verwenden das DROC-Korpus als Goldstandard, anstatt auf automatische Tagger zurückzugreifen (z. B. von Tu et al. [TKB19]), um die Fehlerquellen für die Klassifikation der Handlungsrelevanz zu reduzieren.

Die Präsenz einer literarischen Figur ist ein zentrales Kriterium der SETTING-Definition [De09; Pi08]. Als Feature nutzen wir hierfür die numerische Anzahl von Figurenreferenzen pro Satz aus den annotierten Daten des DROC-Korpus. Darin werden die Figurenreferenzen in “Core” (alle konkreten Figurennamen) und “Non-Core” (auf Figurennamen referierende Personal- und Reflexivpronomen) unterschieden. Für das Feature wurde eine Kombination

aus beiden Gruppen generiert, weil damit eine größere Menge an Figurenreferenzen als bei üblichen Named-Entity-Klassifikatoren einbezogen wird.

Anhand der Präposition ist in einigen Fällen die Klassifikation der Handlungsrelevanz ableitbar, z. B. liefert die Präposition in einem Satz wie “Er ist [in Stuttgart]<sub>SETTING</sub>” ein starkes Indiz für ein *SETTING*, ohne dass dafür eine Abhängigkeitsrelation zum Subjekt hergestellt werden muss. Eine Modifikation des Verbs sowie der Präposition verändert die Klassifikation dagegen hin zu einem *ERWÄHNTEN RAUM* (“Er geht [nach Stuttgart]<sub>MENTION</sub>”). Für das Präpositions-Feature berücksichtigen wir deshalb jede individuelle Präposition als kategoriale Ausprägung, die wir direkt aus den annotierten Präpositionalphrasen extrahieren.

#### 4 Machine-Learning-Modelle zur Klassifikation von *SETTINGS* und *ERWÄHNTEN RÄUMEN*

Die Pilotannotation umfasst 510 annotierte *RAUMENTITÄTEN* (263 *SETTINGS* und 247 *MENTIONS*), von denen wir eine *Most Frequent Class Baseline* (Mehrheits-Baseline) mit einer Accuracy von 0.52 ableiten, gegenüber der die folgenden mit der Python-Library *scikit-learn* realisierten Machine-Learning-Modelle verglichen werden [Pe11]. Tabelle 4 zeigt die gemittelten Evaluationsmaße bei einer 10-fold Cross-Validation<sup>15</sup> und zwei genutzten Feature-Sets. Bei allen Features außer der Figurenanzahl pro Satz handelt es sich um kategoriale Variablen, die wir als One-Hot-Vektoren kodieren, um die optimale Funktionalität der Machine-Learning-Algorithmen zu gewährleisten. Feature-Set 1 verwendet alle Merkmale in maximaler Ausprägung (alle Satzumgebungen für Verbkategorien, temporale Marker und Zeitformen; 48 verschiedene Präpositionen; Figurenreferenzen inkl. Personal- und Reflexivpronomen in Core- und Non-Core-Ausprägung; direkte Rede) und Set 2 stellt eine reduzierte Variante mit eingeschränkten Satzumgebungen und Präpositionen dar.<sup>16</sup>

Die höchsten  $F_1$ -Score-Werte werden mit einem Random Forest mit unbegrenzter Tiefe und 10 übereinanderliegenden Bäumen erreicht. Bei Entscheidungsbäumen besteht aufgrund der begrenzten Trainingsdaten jedoch das Risiko eines Overfittings, wofür die zusätzlich angegebene Accuracy auf den Trainingsdaten ein Indiz sein kann (0.95 bei unbegrenzter Tiefe im Random Forest bzw. 0.96 beim Decision Tree mit unbegrenzter Tiefe). Um ein Overfitting vorzubeugen, begrenzen wir die Tiefe des Random Forest bzw. der Decision Trees auf 4 bzw. 8, woraufhin sich die Accuracy-Werte der Test- und Trainingsdaten deutlich annähern.<sup>17</sup> Entscheidungsbäume bringen insgesamt den Vorteil mit sich, die Klassifikation

<sup>15</sup> Zusätzlich wurden vor Bildung der Folds alle Instanzen randomisiert durchmischt, wodurch einseitige Kategorie-Verteilungen ausgeglichen werden.

<sup>16</sup> Für die Verbkategorien und Zeitformen werden zwei Umgebungen (*surrounding\_0+1*) und für temporale Marker die direkte Umgebung (*surrounding\_0*) verwendet. Statt der vollen 48 Präpositionen werden nur die Vektoren für “nach” und “zu” genutzt. Während Feature-Set 1 nach Konvertierung aller kategorialen Features in One-Hot-Vektoren insgesamt 122 Einzel-Features enthält, sind es bei Feature-Set 2 nur noch 48 Feature-Vektoren.

<sup>17</sup> Für die Entscheidungsbäume werden nur Ergebnisse aus dem reduzierten Feature-Set 2 angegeben, weil bei vollem Feature-Umfang die Diskrepanz zwischen Test- und Trainings-Accuracy noch größer ist.

	Feature-Set	Modell-spezifische Parameter	Accuracy (Acc.-Train)	$F_1$ -micro	$F_1$ -macro	$F_1$ -weighted-macro
Decision Tree	2	Tiefe: unbegrenzt	0.76 (0.96)	0.76	0.76	0.76
		Tiefe: 4	0.74 (0.78)	0.74	0.73	0.73
Random Forest	2	<b>Tiefe: unbegrenzt; Bäume: 10</b>	<b>0.78 (0.95)</b>	<b>0.78</b>	<b>0.78</b>	<b>0.78</b>
		<b>Tiefe: 8; Bäume: 10</b>	<b>0.77 (0.84)</b>	<b>0.77</b>	<b>0.76</b>	<b>0.76</b>
Naïve Bayes (Multinomial)	1	-	0.74 (0.79)	0.74	0.73	0.74
Naïve Bayes (Bernoulli)	1x	-	0.75 (0.8)	0.75	0.75	0.75
	2x	-	<b>0.76 (0.77)</b>	<b>0.76</b>	<b>0.75</b>	<b>0.76</b>
KNN	1	neighbors: 5	0.73 (0.8)	0.73	0.73	0.73
	2	neighbors: 10	0.7 (0.75)	0.7	0.7	0.7
Support Vector Machine	1	C=0.1	0.74 (0.81)	0.74	0.73	0.73
	2	C=0.1	0.7 (0.77)	0.7	0.69	0.7
Logistic Regression	1	<b>C=0.1</b>	<b>0.75 (0.81)</b>	<b>0.75</b>	<b>0.74</b>	<b>0.75</b>
	2	<b>C=0.1</b>	<b>0.75 (0.77)</b>	<b>0.75</b>	<b>0.74</b>	<b>0.74</b>

Tab. 4: Machine-Learning-Modelle im Vergleich (gemittelte Evaluationsergebnisse bei einer 10-fold Cross-Validation; Feature-Set 1x und 2x ohne Figurenreferenzen)

direkt anhand der einzelnen Features nachvollziehen zu können:<sup>18</sup> Beispielsweise wird der erste, auf dem Gini-Gain basierende Split der Daten anhand des Features der direkten Rede durchgeführt, da in Passagen ohne direkte Rede deutlich mehr ERWÄHNT E RÄUME auftreten.

Lineare Klassifikations-Algorithmen wie die Support Vector Machine oder die logistische Regression liefern ebenfalls solide Ergebnisse, bei denen durch eine stärkere Regularisierung (C=0.1) die Gefahr eines Overfittings zusätzlich reduziert wurde.<sup>19</sup> Die logistische Regression sowie auch das Bernoulli-Modell des Naïve Bayes erzielen im reduzierten Feature-Set 2 Ergebnisse, die fast an die Werte des Random Forest heranreichen,<sup>20</sup> wobei sich die Modelle weniger stark an die Trainingsdaten anpassen. Für eine genauere Beurteilung der Modellqualität wäre jedoch ein größeres Test-Datenset wünschenswert. Die  $F_1$ -Score-Werte (Micro, Macro, Weighted-Macro) bewegen sich bei allen Modellen im jeweiligen Optimum zwischen 0.73 und 0.78, womit die Mehrheits-Baseline deutlich geschlagen wird. Beim  $F_1$ -Weighted-Macro wird zusätzlich die Verteilung der SETTINGS/MENTIONS einbezogen, sodass kürzere Texte weniger Gewicht erhalten, und der  $F_1$ -Weighted-Macro oft mit dem  $F_1$ -Micro übereinstimmt. Insgesamt demonstrieren die gezeigten Machine-Learning-Verfahren

<sup>18</sup> Bei den Entscheidungsbäumen kommt eine optimierte Version des CART-Algorithmus (Classification and Regression Trees) zum Einsatz, bei der One-Hot-Vektoren vorausgesetzt sind. Beim Random Forest wird für das Sampling der Daten die Bagging-Methode (bootstrap aggregating) verwendet.

<sup>19</sup> Für KNN, Support Vector Machine und die logistische Regression wurden die Features auf die Einheit der Standardabweichung vom Mittelwert normiert.

<sup>20</sup> Für die Verwendung des Bernoulli-Modells des Naïve Bayes, welches binäre Daten erfordert, wurde bei den Feature-Sets 1x und 2x das numerische Feature der Figurenreferenzen entfernt, sodass nur noch One-Hot-Vektoren verarbeitet werden.

die Funktionalität der automatischen Klassifikation der Handlungsrelevanz sowie der dafür verwendeten Features.

## 5 Fazit

Anhand der vorliegenden Evaluation der Annotation von RAUMENTITÄTEN sowie ihrer Klassifikation wurde die Formalisierbarkeit der raumnarratologischen Kategorien demonstriert. Dabei liefert das Agreement der RAUMENTITÄTEN die Basis zur Entwicklung eines Entity-Classifiers, der den spezifischen Ansprüchen literaturwissenschaftlicher Raumkonfiguration gerecht werden kann. In einer zweiten Annotationsrunde sollte das Agreement insbesondere bei Objekten, gegenüber denen sich Figuren verorten (vi. auf S. 3), oder bei RAUMENTITÄTEN, die ein konkretes Ereignis auf Phrasenebene beinhalten (viii.), optimiert werden.

Im Hinblick auf die Unterscheidung von SETTINGS und ERWÄHNTEN RÄUMEN zeigt das Feature-Engineering, wie eine inhaltlich komplexe Klassifikations-Aufgabe anhand lokaler (teil-)satzbasierter Merkmale erfolgreich automatisierbar ist, wenn genügend Wissen über die entsprechende Domäne eingebracht wird. Bei einer Weiterentwicklung des Klassifikations-Modells ist zunächst eine umfassendere Annotation zur Gewinnung hinreichender Trainingsdaten erstrebenswert. Für ein unabhängiges Modell sollten zudem jene Features, die bisher auf Gold-Annotationen des DROC-Korpus beruhen (direkte Rede, Named-Entities), durch bestehende oder eigene Classifier ersetzt werden. Ferner gilt es neben syntaktischen und satzübergreifenden Merkmalen (z. B. Koreferenzketten) insbesondere die Abhängigkeit und Wechselwirkung mit übergeordneten textuell-narratologischen Kategorien (Erzählebenen, narrative Szenen) in die weitere Konzeption des Feature-Designs einzubinden. Angesichts der komplexen, auf fachwissenschaftlichen Konzepten basierenden Features wäre dann auch zu prüfen, wie ein derartig trainierter Classifier gegenüber Deep-Learning-Modellen abschneidet, die weitgehend ohne spezifisches Feature-Engineering auskommen.

## Literatur

- [Ap10] Apache: Development Community, “UIMA Tutorial and Developers’ Guides”, The Apache Software Foundation, 2010, Stand: 23. 06. 2019.
- [Ba19] Barth, F.: Annotation Guideline No. 5: Annotation Guidelines for Narrative Levels and Narrative Acts. *Journal of Cultural Analytics*, S. 11201, 2019.
- [BSG14] Bögel, T.; Strötgen, J.; Gertz, M.: Computational Narratology: Extracting Tense Clusters from Narrative Texts. In: *LREC*. Bd. 14, S. 950–955, 2014.
- [Bu20] Burghardt, M.: Theorie und Digital Humanities – Eine Bestandsaufnahme. In: *Digital Humanities Theorie*. 2020.

- [Co60] Cohen, J.: A coefficient of agreement for nominal scales. *Educational and psychological measurement* 20/1, S. 37–46, 1960.
- [De09] Dennerlein, K.: *Narratologie des Raumes*. De Gruyter, 2009.
- [EF16] Eisenberg, J.; Finlayson, M.: Automatic Identification of Narrative Diegesis and Point of View. In: *Proceedings of the 2nd Workshop on Computing News Storylines (CNS 2016)*. S. 36–46, 2016.
- [FGM05] Finkel, J. R.; Grenager, T.; Manning, C.: Incorporating non-local information into information extraction systems by gibbs sampling. In: *Proceedings of the 43rd annual meeting on association for computational linguistics*. Association for Computational Linguistics, S. 363–370, 2005.
- [FP10] Faruqui, M.; Padó, S.: Training and Evaluating a German Named Entity Recognizer with Semantic Generalization. In: *KONVENS*. S. 129–133, 2010.
- [Ge18] GermaNet: GermaNet – An Introduction, 2018, URL: <http://www.sfs.uni-tuebingen.de/GermaNet/>, Stand: 20.03.2018.
- [Ge88] Genette, G.: *Narrative Discourse Revisited*, translated by Jane E. Lewin, 1988.
- [Gi19] Gius, E.; Jannidis, F.; Krug, M.; Zehe, A.; Hotho, A.; Puppe, F.; Krebs, J.; Reiter, N.; Wiedmer, N.; Konle, L.: Detection of Scenes in Fiction. In: *DH 2019 Book of Abstracts*. 2019.
- [GRW19] Gius, E.; Reiter, N.; Willand, M., Hrsg.: *Cultural Analytics. A Shared Task for the Digital Humanities*, Nov. 2019.
- [He04] Herman, D.: *Story logic: Problems and possibilities of narrative*. U of Nebraska Press, 2004.
- [HM17] Honnibal, M.; Montani, I.: *spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing.*, 2017.
- [Kr17] Krug, M.: *DROC-Release*, 2017, URL: <https://gitlab2.informatik.uni-wuerzburg.de/kallimachos/DROC-Release>, Stand: 26.02.2018.
- [Ku16] Kuzey, E.; Strötgen, J.; Setty, V.; Weikum, G.: Temponym tagging: Temporal scopes for textual phrases. In: *Proceedings of the 25th International Conference Companion on World Wide Web*. S. 841–842, 2016.
- [Le95] Levin, B.: *English verb classes and alternations. A preliminary Investigation* 1/, 1995.
- [Pe11] Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V. et al.: *Scikit-learn: Machine learning in Python*. *the Journal of machine Learning research* 12/, S. 2825–2830, 2011.
- [Pi08] Piatti, B.: *Die Geographie der Literatur: Schauplätze, Handlungsräume, Raumphantasien*. Wallstein, Göttingen, 2008.

- [PMV11] Pustejovsky, J.; Moszkowicz, J. L.; Verhagen, M.: Using ISO-Space for Annotating Spatial Information. In: COSIT 2011: 10th International Conference on Spatial Information Theory. 2011.
- [Pu15] Pustejovsky, J.; Kordjamshidi, P.; Moens, M.-F.; Levine, A.; Dworman, S.; Yocum, Z.: SemEval-2015 Task 8: SpaceEval. In: Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015). Association for Computational Linguistics, S. 884–894, 2015.
- [Re13] Reznicek, M.: Linguistische Annotation von Nichtstandardvarietäten – Guidelines und "Best Practices". Annotation Koreferenz. 2013.
- [Ro19] Roth, C.: Digital, digitized, and numerical humanities. *Digital Scholarship in the Humanities* 34/3, S. 616–632, 2019.
- [Ry91] Ryan, M.-L.: Possible worlds, artificial intelligence, and narrative theory. Indiana University Press, 1991.
- [Sc04] Schor, M.: An Effective, Java-Friendly Interface for the Unstructured Management Architecture (UIMA) Common Analysis System, Techn. Ber. IBM RC23176, IBM T. J. Watson Research Center, 2004.
- [SL09] Sokolova, M.; Lapalme, G.: A systematic analysis of performance measures for classification tasks. *Information processing & management* 45/4, S. 427–437, 2009.
- [So89] Soja, E. W.: Postmodern geographies: The reassertion of space in critical social theory. Verso, 1989.
- [TKB19] Tu, N. D. T.; Krug, M.; Brunner, A.: Automatic recognition of direct speech without quotation marks. A rule-based approach. In (Sahle, P., Hrsg.). *Digital Humanities: multimedial & multimodal.*, Konferenzabstracts, Frankfurt am Main, S. 87–89, 2019.
- [Yi13] Yimam, S. M.; Gurevych, I.; de Castilho, R. E.; Biemann, C.: 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. S. 1–6, 2013.