
Lokale Grammatiken zur Beschreibung von lokativen Sätzen und ihre Anwendung im Information Retrieval

Sebastian Nagel



München 2008

Lokale Grammatiken zur Beschreibung von lokativen Sätzen und ihre Anwendung im Information Retrieval

Sebastian Nagel

Inaugural-Dissertation
zur Erlangung des Doktorgrades der Philosophie
an der Ludwig-Maximilians-Universität München

vorgelegt von
Sebastian Nagel
aus München

München, den 26. März 2008

Erstgutachter: Prof. Dr. Franz Guenther
Zweitgutachter: Prof. Dr. Klaus Schulz
Tag der mündlichen Prüfung: 21. Juli 2008

Danksagung

Der erste Dank gilt meinem Doktorvater Prof. Dr. Franz Guenther für die Geduld und sein Wohlwollen während der Entstehungszeit. Von ihm sind viele Ideen und Anregungen, in Seminaren oder im persönlichen Gespräch geäußert, in den Text eingeflossen, ohne dass sie eigens gekennzeichnet wurden. Vor allem aber möchte ich ihm für die Freiheit danken, auch eigene Wege zu gehen, nicht ohne mich durch kritische Kommentare wieder zurückzuholen und in andere Richtung und auf neue Gebiete zu schicken.

Dem Koreferenten, Prof. Dr. Klaus Schulz, spreche ich meinen Dank aus, nicht nur für das Lesen der Arbeit, sondern auch für das nicht selbstverständliche Interesse daran. Für Gespräche und Diskussionen nicht nur auf dem Weg zum Mittagessen danke ich ihm und meinen Mit-Promovanden Uli Reffle, Annette Gotscharek, Levin Brunner, Eduardo Torres-Schumann, Andreas Hauser, sowie Clemens Marschner, vor allem als Organisator der Lunch-Talks, auf dem ich meine Arbeit erstmals einem größeren und kritischen Publikum vorstellen durfte.

Meinen Kollegen Michaela Geierhos, Olivier Blanc und Yeong Su Lee gebührt Dank für das geteilte Interesse an Lexika und Grammatiken, und Robert Zangenfeind für die vielen Diskussion über Mel'čuk, Dependenz, Idiome, Kollokationen und die Bedeutung so vieler Lexeme.

Dank geht auch an die Kollegen Stefan Langer, Daniel Schnorbusch, Max Hadersbeck, Hans Leiß für die eine oder andere Tasse Kaffee – nicht ohne Diskussion über wissenschaftliche und sonstige Fragen des Lebens. Nicht unerwähnt bleiben darf schließlich der Systemadministrator am CIS, Thomas Schäfer, ohne den der Laden nicht läuft.

Bedanken möchte ich mich auch bei den Studenten meiner Seminare, für ihre Aufmerksamkeit und ihre kritischen Fragen, und schließlich dem Institut CIS selbst für die angenehme Atmosphäre und die offenen Türen.

Danke auch an Felix, Rune, Alexandra, Mascha, Liu aus der schönen Zeit zu Beginn, als die Promotion noch Studium, und nicht Arbeit war.

Der größte Dank gilt schließlich meiner Familie, Kindern und Frau, und meinen Eltern, denen ich das Studium, auch sonst einiges zu verdanken habe.

Diese Arbeit wurde zeitweise von der Fast Search & Transfer ASA durch ein großzügiges Stipendium gefördert und hätte ohne diese Finanzierung sicher nicht zu dieser Form gefunden.

Inhaltsverzeichnis

Danksagung	v
Einleitung	1
1 Lokale Grammatiken	3
1.1 Eigenschaften und Anwendungen lokaler Grammatiken	5
1.2 Lexikongrammatik	11
1.3 Repräsentation und Implementierung lokaler Grammatiken	17
2 Lokative Prädikate	21
2.1 Sprache und Raum: semantische Aspekte	21
2.1.1 Räumliche Relationen	21
2.1.2 Bewegung im Raum	24
2.2 Lokativa im Inhalt⇔Text-Modell	26
2.2.1 Lokative lexikalische Funktionen	27
2.3 Lokative Prädikate: Kriterien und Realisierung	30
2.3.1 Die Suche nach dem Prädikat	30
2.3.2 Kriterien für lokative Prädikate	31
2.3.3 Abgrenzung zwischen (lokativen) Komplementen und Supplementen	32
2.3.4 Bestimmung lokativer Komplemente	35
2.3.5 Realisierung und Synonymie der Prädikate	36
2.4 Räumliche Präpositionen	37
2.5 Lokative Verben	45
2.5.1 Im Rahmen der Lexikongrammatik des LADL	45
2.5.1.1 Französische Verben	45
2.5.1.1.1 Kriterien und Verfahren zur Klassifikation	45
2.5.1.1.2 Klassen französischer lokativer Verben	50
2.5.1.2 Deutsche Verben	53
2.5.1.3 Garrigues (1993,1995)	58
2.5.1.4 Constant (2003b)	61
2.5.2 Lokative Verben außerhalb der Lexikongrammatik	62
2.5.2.1 Alternationen und Klassifikation	62
2.5.2.2 Valenzorientierte Beschreibung	64
2.5.2.3 Kognitive, konzeptuelle und formale Semantik	65
2.5.2.4 Fictive motion: statische Lesarten von Bewegungsverben	66
2.5.3 Zwei Fallbeispiele: <i>wohnen</i> und <i>liegen</i>	67
2.5.3.1 <i>wohnen</i>	67
2.5.3.2 <i>liegen</i>	72
2.6 Lokative (quasi-)prädikative Nomina	73
2.7 Adjektive	75
2.8 Adverbien	76

3	Lokative Argumente	79
3.1	Einführung	80
3.1.1	Was sind Lokativa? – Versuch einer Definition	80
3.1.1.1	Beziehung zu hyperonymen und Abgrenzung gegenüber kohyponymen Klassen	81
3.1.2	Untergliederung der Lokativa	82
3.1.3	Appropriate nouns	85
3.1.4	Reguläre Polysemie	86
3.1.5	Zusammenfassung: Beschreibung und Klassifikation der Lokativa	91
3.1.6	Ein Lexikon lokativer Nomina und der Klassenbezeichner von Toponymen	92
3.2	Geographische (Eigen)Namen	93
3.2.1	Strukturelle Eigenschaften von Toponymen	97
3.2.1.1	Interne Struktur	97
3.2.1.2	Syntaktische Eigenschaften	99
3.2.2	Exonyme	100
3.3	Ein Lexikon geographischer Eigennamen	102
3.3.1	Quellen für ein Toponymlexikon	104
3.3.2	Kodierte Eigenschaften	107
3.3.2.1	Lemma und Grundform	109
3.3.2.2	Genus	110
3.3.2.3	Pluralia tantum	111
3.3.2.4	Artikelgebrauch bei Toponymen	111
3.3.2.5	Toponyme mit obligatorischem Klassenbezeichner	113
3.3.2.6	Flexion	113
3.3.2.7	Abgekürzte Formen	116
3.3.2.8	Kürzel, Ländercodes	118
3.3.2.9	Nicknames, Epitheta, Un-Orte	118
3.3.2.10	Koordinationsellipsen: <i>Ober-, Mittel- und Unterfranken</i>	119
3.3.2.11	Derivationen: Adjektive und Bezeichnungen für Bewohner	120
3.3.2.12	Fugenformen	125
3.4	Erweiterung des Lexikons zur Ontologie	127
3.4.1	Die sprachliche Ebene	127
3.4.1.1	Klassen geographischer Eigennamen	129
3.4.2	Geographische Entitäten	130
3.4.2.1	Klassen geographischer Entitäten	131
3.4.2.1.1	Überblick über Klassifikationssysteme	133
3.4.2.2	Relationen zwischen geographischen Entitäten	136
3.4.3	Zeitliche Dimension: historische Namen und Orte	138
3.5	Phrasengrammatiken zur Erkennung lokativer Argumente	138
3.5.1	Nominalphrasen	140
3.5.1.1	Adjektivmodifikatoren	141
3.5.1.2	Toponyme	142
3.5.1.3	Klassenbezeichner und Toponym	144
3.5.1.4	Relativsätze	146
3.5.1.5	Paraphrasen einer NP	146
3.5.2	Präpositionalphrasen	147
3.5.3	Ortsadverbien	147
3.5.4	Lokalsätze	148
3.5.5	Adressen	149

3.6	Kombination von Ortsangaben	149
3.6.1	Quasikoordinierte statische Ortsangaben	150
3.6.2	Pfadangaben	151
4	Lokative Sätze	153
4.1	Lokale Grammatiken und deutscher Satzbau	153
4.2	Adverbialsupplemente (Inserts)	154
4.3	Kriterienkatalog zur Klassifikation lokativer elementarer Sätze	156
4.4	Präpositionen	157
4.5	Verben	158
4.5.1	Bewegungsverben	158
4.5.1.1	Figure (bewegtes Objekt) als Subjekt	159
4.5.1.2	Figure als syntaktisches Objekt	163
4.5.1.3	Subjektlose Sätze	163
4.5.2	Transportverben und ‘local causatives’	163
4.5.3	Verben mit Infinitivergänzung	163
4.5.4	Mensch – Ort statisch	164
4.5.5	Relationen zwischen Orten (Verben)	165
4.6	Nomina	167
4.6.1	Nomina der Fortbewegung und des Aufenthalts	168
4.6.2	Relation Mensch – Ort	169
4.6.3	Relationen zwischen Orten	170
4.7	Adverbien und Adjektive	170
5	Lokativa im Information Retrieval	175
5.1	Named Entity Recognition	175
5.1.1	Anteil der (geographischen) Eigennamen in (deutschen) Texten	176
5.1.2	Strategien zur Erkennung von Eigennamen	177
5.1.2.1	Listenbasierte Erkennung: Gazetteers	177
5.1.2.2	Interner und externer Kontext	179
5.1.2.3	Die Diskursebene: One-Sense-per-Document und Learn-Apply-Forget	181
5.1.3	Ansätze und Systemarchitekturen	181
5.1.3.1	Statistische Systeme	182
5.1.3.2	Regelbasierte Systeme	184
5.1.3.2.1	NER mit lokalen Grammatiken	184
5.1.4	Evaluation von Systemen zur NER	186
5.1.4.1	Precision, Recall, F-Wert	186
5.1.4.2	Wann ist ein System gut?	188
5.1.5	Lokative Named Entities	188
5.1.5.1	Annotierung und Klassifikation	189
5.1.5.1.1	Probleme der Annotierung	189
5.1.5.1.2	Eingebettete (lokative) Entitäten	190
5.1.5.1.3	Standards zur Annotierung	191
5.1.5.1.3.1	MUC-6 und Nachfolger	192
5.1.5.1.3.2	Auflösung von Metonymien	194
5.1.5.1.3.3	Tiefe Annotation: TEI und Spatial ML	195
5.1.5.1.3.4	Geographical Markup	197
5.1.5.2	Erkennung geographischer Entitäten in deutschen Texten	198

5.1.5.2.1	Überblick über implementierte Systeme	199
5.1.5.2.2	Deutsche Korpora mit NE-Annotationen	204
5.1.5.3	Ein System basierend auf lokalen Grammatiken	205
5.1.5.3.1	Die lokalen Grammatiken des Systems	208
5.1.5.3.1.1	Erkennung unbekannter Toponyme	208
5.1.5.3.1.2	Positive Kontexte	208
5.1.5.3.1.3	Negative Kontexte	210
5.1.5.3.2	Performance und Evaluation des Systems	211
5.2	Geokodierung	215
5.2.1	Räumliche Definition von Toponymen	219
5.3	Automatische Extraktion lexikalischer Ressourcen	219
5.3.1	NER zur Lexikonerweiterung	219
5.3.2	Automatische Extraktion von Relationen zwischen Toponymen	221
5.4	Uservavigation und geographische Queries	221
5.4.1	Uservavigation	222
5.4.1.1	Visualisierung auf Landkarten	222
5.4.1.2	Benutzerführung über Ontologien	222
5.4.2	Lokalisierung von Texten	222
5.4.3	Geographische Queries	224
5.5	Räumliche Repräsentation \Leftrightarrow sprachliche Beschreibung	226
Zusammenfassung		227
Summary		231
Literaturverzeichnis		235
Tabellenverzeichnis		289
Abbildungsverzeichnis		291
Index		293
Anhang		299
A	Beispiele zu den Lexika	299
A.1	Toponyme	299
A.1.1	Zur Orthographie der Toponyme	299
A.1.2	Beispiele für das Lexikon deutscher Toponyme	300
A.1.2.1	Syntax eines Wörterbucheintrags	300
A.1.2.2	Grundformenlexikon	303
A.1.2.3	Vollformenlexikon	303
A.1.3	Regeln und Pattern zur Abkürzung	309
A.1.3.1	Liste der Pattern und Ersetzungsregeln	311
A.1.3.2	Beispiele für die sequentielle Anwendung mehrerer Regeln	316
A.1.4	Derivationsuffixe für Adjektive und Bewohner	317
A.1.5	Taxonomien geographischer Entitäten	332
A.1.5.1	Lingustisch orientierte Klassifikationssysteme	332

A.1.5.2	Geographisch orientierte Taxonomien	340
A.2	Lokative Nomina und Klassenbezeichner der Toponyme	353
A.3	Lokative Adjektive und Adverbien	353
B	Lokale Grammatiken	355
B.1	Komplexe Adverbien und Präpositionen	355
B.2	Erkennung unbekannter Toponyme	358
B.3	Antigrammatiken	360
C	Tabellen der Lexikongrammatik	361
C.1	Präpositionen (Adpositionen)	361
C.2	Verben	394
C.3	Nomina	414
C.4	Adverbien	418
D	Lokativa im Erklärend-Kombinatorischen Wörterbuch	421
D.1	Skizzenhafte Einträge deutscher Lokativa	424
D.1.1	Wortfeld <i>wohnen</i>	426
D.1.2	Wortfeld <i>Siedlung</i>	431
D.1.3	Wortfeld <i>Straße, Weg</i>	433
E	Korpora	437
E.1	Named Entities	437
E.1.1	Bewertung des CONLL-2003-SHARED-TASK-Korpus	437
E.2	Allgemeine Korpora	441

Einleitung

Der sprachliche Ausdruck räumlicher Relationen und von Bewegung im Raum war in den letzten Jahrzehnten Gegenstand der linguistischen Forschung.¹ In der Computerlinguistik waren es zunächst die Bereiche maschinelle Übersetzung² und Künstliche Intelligenz,³ die sich mit der Thematik Sprache – Raum auseinandersetzten. In den letzten Jahren hat sich unter dem Begriff ‚raumbezogenes IR‘ (GIR – Geographic Information Retrieval) eine Unterdisziplin des IR zu etablieren begonnen, die sich die Verbesserung der Qualität des IR für raumbezogene Anfragen (Queries) als Ziel gesetzt hat.⁴ Dies soll einerseits durch die Integration (ungenauer) sprachlicher und (exakter) geographischer Daten, zum anderen durch Verwendung von Ontologien, die Relationen zwischen geographischen Entitäten formal kodieren, ermöglicht werden. Geographische Daten sind im Zusammenhang mit Navigationssystemen zunehmend vorhanden,⁵ aber raumbezogene Texte stehen nicht zurück.⁶ Sprache war, ist und wird das primäre Zeichensystem zum Austausch von (räumlicher) Information zwischen Menschen bleiben. Landkarten sind ein sekundäres Medium, und kein Ersatz, sondern lediglich Ergänzung zur Sprache. Der Zugriff auf räumliche Information über Namen und sprachlich kodierte Relationen ist schneller und natürlicher, aber auch weniger präzise. Ortsbezogene Dienste über Telefonie sind auf Sprache als Medium angewiesen. Das Ziel muss deshalb die Verknüpfung der geometrisch-geographischen und sprachlichen Raumrepräsentationen sein, um die Vorteile von beiden zu vereinen.

Die vorliegende Arbeit konzentriert sich auf die sprachliche Dimension. Wie werden räumliche Relationen und Bewegung im Raum sprachlich kodiert? Durch welche Wortarten können lokative Prädikate realisiert sein? In welchen Formen stehen die Argumente des Prädikats und wie werden sie an das Prädikat angeschlossen? Und schließlich – welche Ausdrücke sind äquivalent, oder Paraphrasen voneinander? Sprecher können Sätze wie *Max wohnt in Essen* und *Max ist in Essen wohnhaft* als Paraphrasen erkennen und bilden. Die Paraphrasierung basiert vor allem auf den Phänomenen Transformation (syntaktische Umformung), „Synonymie“ der Prädikate (siehe das Beispiel *Max wohnt*) und Variationsmöglichkeiten in der Realisierung und im Anschluss der Argumente (*der französische Präsident* vs. *der Präsident Frankreichs*). Alle drei Phänomene wurden bei der Beschreibung lokativer Ausdrücke und Sätze berücksichtigt.

Als Mittel und Methode zur Beschreibung werden zwei von Maurice Gross begründete Ansätze gewählt, die als ‚Lexikongrammatik‘ und ‚Lokale Grammatiken‘ bekannt sind. Hier werden

-
- 1 Siehe Levinson (2003), Levinson & Wilkins (2006), Matsumoto & Slobin (2002) für weitere Literatur.
 - 2 Z.B. Maillat (2003a,b), Trujillo (1995a), Amores & Mora (1998).
 - 3 Zur Übersicht: André, Herzog & Rist (1989), Habel et al. (2003) oder unter dem Stichwort ‚spatial reasoning‘ in Literaturdatenbanken und Internet-Suchmaschinen.
 - 4 Vgl. die GIR-Workshops der Konferenzreihe SIGIR, GEOCLEF, das Projekt SPIRIT, oder ganz allgemein Dienste unter dem Stichwort GIS (Geographic Information System).
 - 5 Z.B. OPENSTREETMAP, GEOCODER.US, aber auch kommerzielle Anbieter wie TeleAtlas/TomTom und Navteq/Nokia.
 - 6 Traditionell Reiseberichte, Reise- und Hotelführer, und neue Gattungen wie WIKIPEDIA, WIKIVOYAGE.ORG und City-Blogs.

sie genutzt, um lokative Ausdrücke aller Art zu beschreiben – angefangen von Bezeichnungen von Orten über Ortsadverbiale bis hin zu vollständigen lokativen Sätzen. Letztere sind definiert als einfache Sätze, die räumliche Relationen oder eine Bewegung im Raum ausdrücken.⁷ Da die Grammatiken auf die Anwendbarkeit im IR ausgerichtet sind, ergibt sich die Forderung, dass mindestens eines der Argumente des Prädikats als Toponym realisierbar sein soll.

Die ersten beiden Kapitel der Arbeit vermitteln einen Überblick über die theoretischen Grundlagen: Kap. 1 über Lokale Grammatik und Lexikongrammatik, die ja als Beschreibungsrahmen und Plattform zur Grammatikimplementierung gewählt wurden, Kap. 2 über Ansätze und Probleme bei der linguistischen Beschreibung von Lokativa.

Der zweite, praktische Teil enthält zunächst einen ausführlichen Kommentar zu den erarbeiteten Ressourcen – die Lexika und lokalen Grammatiken zur Erkennung von lokativen Argumenten (Kap. 3) und die Lexikongrammatiken für alle Arten lokativer Prädikate und Sätze (Kap. 4). Der Schwerpunkt liegt dabei mit auf der Verarbeitung von Teilsätzen, Chunks und der Behandlung von Einschüben – einfache Sätze sind in journalistischen Texten kaum zu beobachten. Die Behandlung komplexer Sätze fällt nicht in den Rahmen dieser Arbeit. Viel Raum erhält auch die Realisierung der Argumente, insbesondere in Form von Toponymen, die einerseits hochinformativ sind, andererseits in vielfacher morphologischer Gestalt als Aktanten (Argumente) realisiert werden können, wodurch eine Paraphrase bereits auf Ebene der Aktanten möglich ist.

Den Abschluss bildet ein Kapitel über Lokativa im IR bzw. GIR (Kap. 5), das einen Bogen spannt von der Toponymerkennung über Geokodierung, Lexikonaufbau, Extraktion von Relationen, bis hin zur Lokalisierung von Texten und der Verarbeitung von lokativen Queries. Der Schwerpunkt dieses Kapitels liegt dabei auf Erkennung von Toponymen im Text (Named Entity Recognition). Vorgestellt wird ein System zur Toponymerkennung in deutschen Texten, bei dem die erstellten Grammatiken und Lexika die zentrale Komponente und Resource bilden.

7 Die Definition eines lokativen Verbs unterscheidet sich damit von der der französischen Lexikongrammatiker, vgl. die Diskussion in Kap. 2.5.1.1 und Kap. 2.3.2 für detaillierte Kriterien.

1 Lokale Grammatiken

Lokale Grammatiken versuchen Sprache (zunächst) nicht global als Menge aller Sätze zu beschreiben. Sie beginnen vielmehr mit der Beschreibung lokaler syntaktischer Phänomene, d.h. Phänomene, die in unmittelbarer Nachbarschaft eines Wortes oder einer Wortklasse auftreten:

[...] grammarians and theoreticians have always attempted to describe the general features of sentences. [...] But beyond these generalities lies an extremely rigid set of dependencies between individual words, which is huge in size; it has been accumulated over the millenia by language users, piece by piece, in micro areas such as those we began to analyze here.
(Gross 1997: 325)

[Local grammars] consist of the description of local constraints, namely restrictions on the surrounding sequences of a given set of words.
(Mohri 1994: 1)

[...] the global nature of language results from the interaction of a multiplicity of local finite-state schemes
(Gross 1997: 330)

Typische Beispiele sind Grammatiken für Eigennamen (Named Entities), Fachtermini oder Adverbiale der Zeit (vgl. Abb. 1.1 auf der folgenden Seite, 1.3 auf S. 7 und B.11 auf S. 360). Allen gemeinsam sind gravierende lexikalische Einschränkungen in der internen Syntax oder im unmittelbaren Kontext, oft verbunden mit Nicht-Kompositionalität in der Bedeutung. Lokale Grammatiken stehen damit im Gegensatz zur traditionellen (generativen) Phrasenstrukturgrammatik, die die Syntax einer Sprache allein auf der Basis einiger weniger Wortklassen wie Determinativ, Adjektiv, Nomen, Verb, Präposition und anhand einiger Typen von Phrasen (NP, PP, VP) zu beschreiben versucht.

Die Lexikalisierung der Syntax, d.h. syntaktische Eigenschaften als lexikalisch gebunden zu betrachten, ist auch in anderen Grammatiktheorien und Sprachmodellen zu finden, insbesondere in der Dependenz- bzw. Valenzgrammatik oder dem Inhalt \Leftrightarrow Text-Modell. Sie ist zentrales Prinzip der Lexikongrammatik, einem weiteren von Maurice Gross begonnen und ungleich umfangreicheren Projekt (siehe Kap. 1.2). Mit beiden Projekten versucht Maurice Gross seiner Auffassung von Linguistik als empirischer Wissenschaft, die sich auf jederzeit reproduzierbare und quantifizierbare Fakten stützt, gerecht zu werden – eine Auffassung, die aus der Auseinandersetzung mit Chomskys generativer Grammatik folgt:

Neither generative nor traditional grammars contain any notion of enumeration of linguistic instances. Not a single attempt to enumerate regular and exceptional cases has ever been made by a syntactician.
(Gross 1979: 865)

[...] empirical evaluation of linguistic facts must be reproducible, a condition met in all hard sciences, but practically unknown to linguistics
(Gross 1999: 236)

Accumulating data is obviously not an aim itself. But in all natural sciences it is a fundamental activity, a necessary condition for evaluating the generality of phenomena. Such a concept of generality or of importance of facts is totally absent from generative grammar, where sentences acquire significance only with respect to formalism. In generative grammar, a linguistic example appears to be significant only if it allows one to choose between competing theories. In the last few years, the consequences of this view have become caricatural. Linguists, now

well-trained in formalistic manipulations, know how to invent new theories at will, and don't restrain themselves. Under these conditions, well-worn sentences are sufficient for all theoretical discussions: any motivation for looking into new phenomena has been lost. In this way, generative grammar has settled into a state of closed-circuit functioning. (Gross 1979: 866)

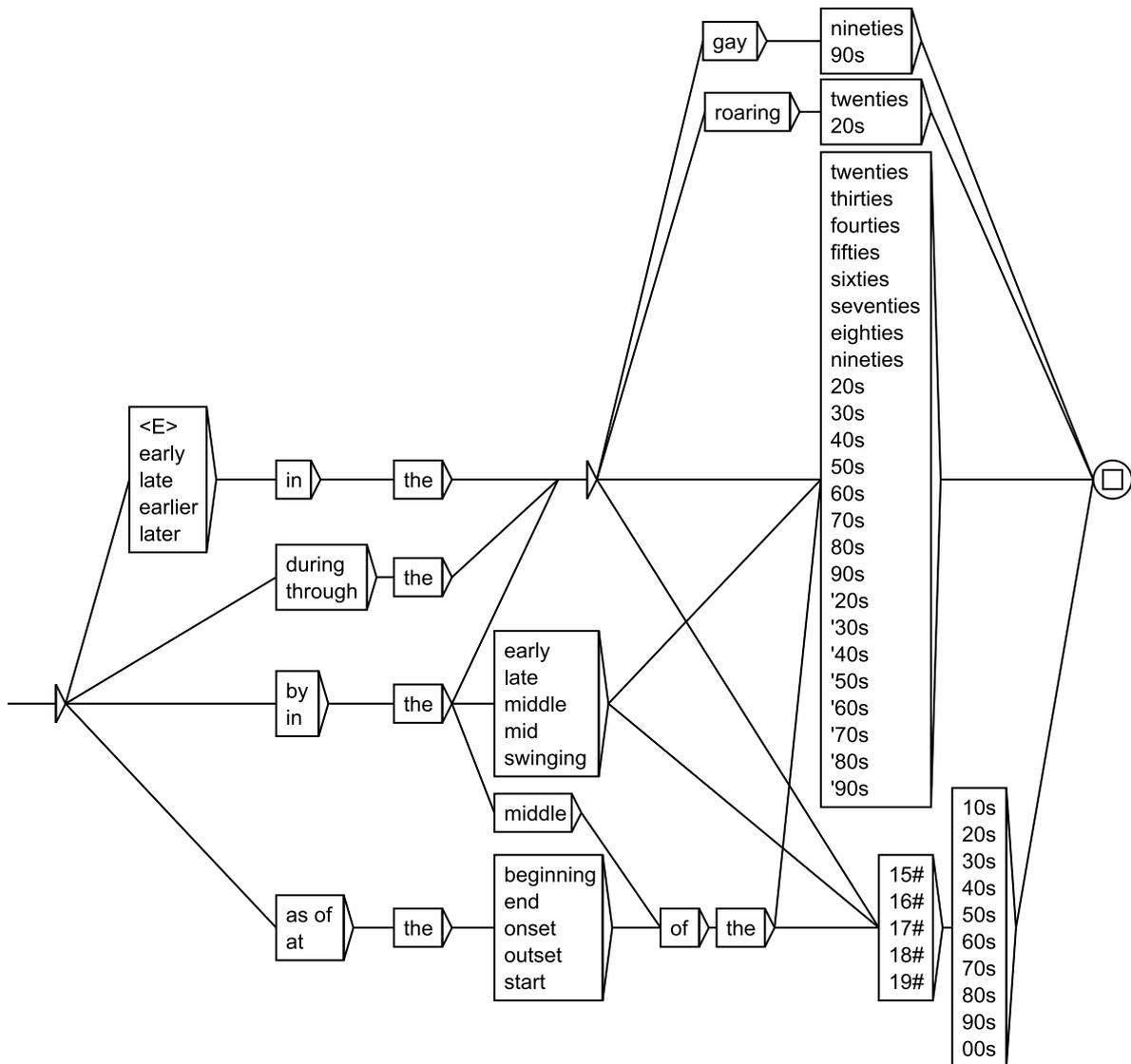


Abbildung 1.1: Lokale Grammatik zur Beschreibung englischer Zeitadverbien des Types „rounded dates“ (vgl. Gross 1993: 31; Gross 1995b: 293)

Lokale Grammatiken werden üblicherweise als Graphen repräsentiert, einer intuitiven, benutzerfreundlichen Form, die anderen, äquivalenten Repräsentationsformen wie regulären Ausdrücken oder Ersetzungsregeln ($NP \rightarrow DET N$) deutlich überlegen ist. So ist auch dem Nichtlinguisten nach kurzer Betrachtung von Abb. 1.1 verständlich, was diese Grammatik beschreiben will.⁸ Für die Anwendung bei der Analyse (und Synthese werden lokale Grammatiken in endliche Automaten, einen effizienten Formalismus zum Erkennen von Ausdrücken in Texten, überführt. Auf Details der Repräsentation und Implementierung lokaler Grammatiken wird in Kap. 1.3 eingegangen.

8 Die Bedeutung von einigen Metasymbolen (hier <E> und #) ausgenommen.

1.1 Eigenschaften und Anwendungen lokaler Grammatiken

Lokale Grammatiken wurden von Maurice Gross (1993: 26)⁹ zur Repräsentation von Kollokationen und festen Redewendungen „in a natural way“ vorgeschlagen. Sie sind damit eine Alternative zu Lexikoneinträgen und haben diesen gegenüber den Vorteil, dass wenige Graphen mit verschiedenen, verknüpften Pfaden viele variierende Einträge repräsentieren können. So beschreibt die lokale Grammatik in Abb. 1.1 zusammengefasst 2252 literarische verschiedene Zeitadverbien. Diese auf den ersten Blick hohe Zahl erhält man, wenn alle Kombinationen aus allen (Teil-)Pfaden aufaddiert werden. Die Variationen ergeben sich aus zwei gegensätzlichen Phänomenen: (1.) regulären, paradigmatischen Ersetzungen. So kann anstelle von *twenties* auch *thirties* etc., also ein Wort aus der gleichen Klasse – hier nicht im Sinne der tradierten Wortarten – treten. (2.) irregulären Einschüben und lexikalisierten Ausnahmen (Idiosynkrasien). Beide, Regularitäten wie Irregularitäten, werden mit dem gleichen Formalismus beschrieben. Für die Irregularitäten wird ein vollständig neuer Pfad im Graphen hinzugefügt, für die Regularitäten ist es dagegen ausreichend, eine bereits bestehende Box um ein weiteres Element zu ergänzen.

Graphen wie in Abb. 1.1 beschreiben ‚funktionale Synonyme‘. Alle von dieser Grammatik beschriebenen Ausdrücke lassen sich in vermutlich jedem englischen Satz durcheinander ersetzen, ohne den Satz ungrammatisch bzw. unsinnig zu machen, auch wenn sich die Ausdrücke semantisch unterscheiden (*in the twenties* \neq *in the thirties*). Das unterscheidet die in lokalen Grammatiken beschriebenen Wortklassen von den traditionellen Wortarten. Ein Nomen im Satz lässt sich eben nicht durch jedes beliebige Nomen austauschen. Auch hier steht Gross im Gegensatz zu Chomsky: „Grammar is best formulated as a self-contained study independent of semantics.“ (Chomsky 1957: 106)

Ein Graph beschreibt also idealerweise (alle) Elemente mit gleicher syntaktischer und semantischer Funktion. Dass die von einem Graphen beschriebenen Einheiten auch semantisch homogen sind, macht lokale Grammatiken geeignet zur semantischen Annotierung (semantisches Tagging), zur Paraphrase und Übersetzung. Mit geringem Aufwand lässt sich der Graph in Abb. 1.1 in einen ‚Transduktor‘ umschreiben, der den erkannten Text Box für Box in eine andere Sprache oder eine kanonische Form übersetzt.

Transduktoren erzeugen für eine erkannte Zeichenfolge (Eingabe) eine neue Zeichenfolge (Ausgabe). So lässt sich die Grammatik aus Abb. 1.1 zu einem Transduktor erweitern, der für die Eingaben *in the twenties* und *during the 1920s* z.B. eine kanonische Form *in 1920s* ausgibt. Die assoziierten Ausgaben bedingen dabei eine Trennung bedeutungsverschiedener und die Alinierung bedeutungsgleicher, aber variabler Bestandteile (Teilpfade). So gelangt

9 Der Begriff ‚local grammar‘ bzw. ‚grammaire locale‘ erscheint erstmals bei Gross (1993), und weiter bei Roche (1993b), Garrigues (1993) und Mohri (1994). Die Idee stark variierende sprachliche Einheiten durch Automaten zu repräsentieren wurde von Gross schon 1989 vorgeschlagen. Er bezieht sich hier auf orthographische Varianten (auch von Mehrwortlexemen), Derivationen, Sequenzen französischer präverbaler klitischer Pronomina und idiomatischer Wendungen. Mit dem Problem, wie komplexe Einheiten und ihre Varianten im Lexikon repräsentiert werden können, hat sich Gross mindestens seit 1986 beschäftigt: „But compound terms raise a problem of representation. The unit of representation in a linear lexicon is roughly the word as defined by its written form, that is, a sequence of letters separated from neighboring sequences by boundary blanks. As a consequence, compound words cannot be directly put into a dictionary the way simple words are. An identification procedure is needed for their occurrences in texts, and this procedure will make use of the various simple parts of the compound utterance.“ (Gross 1986: 1)

Senellart (1998a) auf Basis von Transduktoren zu einer operationalisierbaren Definition von **Synonymie**: Zwei sprachliche Äußerungen sind dann synonym (in Bezug auf einen gg. Transduktor), wenn für beide die gleiche Ausgabe folgt. *In the twenties* und *during the 1920s* wären demnach Synonyme oder **Paraphrasen**.

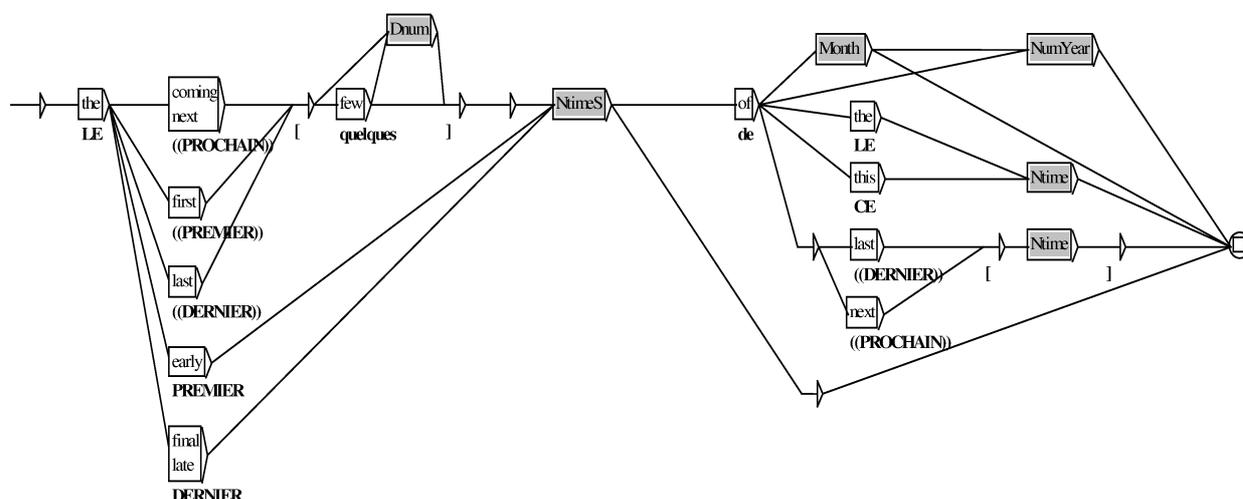


Abbildung 1.2: Übersetzung von Zeitadverbien aus dem Englischen ins Französische mit einem Transduktor. Die grau hinterlegten Boxen stehen für Subgraphenaufrufe. Im Anschluss müssen noch die Wortstellung (durch eckige Klammern angedeutet) verändert, und einzelne Formen (großgeschrieben) nach allgemeinen Kongruenz- und morphologischen Regeln angepasst werden. (aus Fairon & Senellart 1999)

Auf gleiche Weise lassen sich Transduktoren zur **Übersetzung** verwenden. So übersetzt der Transduktor in Abb. 1.2 die englische Wendung *the last few days* in *LE [DERNIER] [quelques] jours*. Bei der Übersetzung müssen freilich noch Wortstellung und Formen nach allgemeinen sprachlichen Regeln angepasst werden, um die korrekte Wendung *les quelques derniers jours* zu erhalten. Die eckigen Klammern und die Großschreibung einiger Wörter markieren die entsprechenden Positionen, bewerkstelligen lässt sich die Anpassung ebenfalls durch Transduktoren. Die Übersetzung erfolgt dann durch eine ‚Kaskade‘ von Transduktoren.

Allerdings erreicht eine direkte Übersetzung mittels lokaler Grammatiken und Transduktoren wegen genereller Regeln (Wortstellung, Kongruenz, Koordination) in der Zielsprache schnell eine Komplexität, die die Graphen unübersichtlich macht und erfordert, dass viele der Pfade einzeln beschrieben werden. Senellart, Dienes & Váradi (2001) schlagen zur Anpassung des maschinellen Übersetzungssystems SYSTRAN an die Fachsprache einer begrenzten Domäne (vgl. Senellart et al. 2001) deshalb vor, lokale Grammatiken nicht unmittelbar in Form von Transduktoren zur maschinellen Übersetzung zu nutzen, sondern nur mittelbar zur **Alinierung** von mehr oder minder festen Wendungen, z.B. Fachtermini.

Zunächst werden dazu typische Wortverbindungen und Konstruktionen mittels lokaler Grammatiken beschrieben. So entsteht eine Bibliothek an Graphen. Die Graphen sind entweder ‚lexikalisch‘ – sie beschreiben ein einzelnes Lexem, oder ‚kontextuell‘, da sie ein Wort/Lexem in seinem unmittelbarem syntaktischen Kontext beschreiben. Die Graphen werden dann (manuell) von der Quellsprache in die Zielsprache übertragen. Abb. 1.3 und 1.4 zeigen ein solches Graphenpaar. Aus einem alinierten mehrsprachigen Korpus (translation memory) werden anschließend die von den Graphen erkannten Sequenzen in ein „alignment dictionary“ extrahiert.

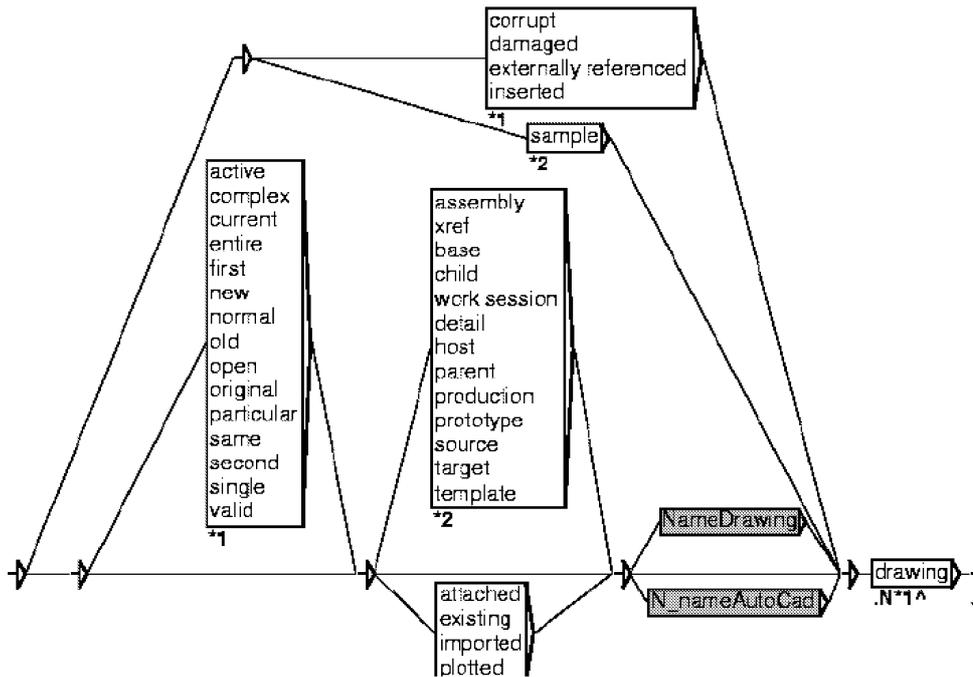


Abbildung 1.3: Repräsentation von Fachtermini durch lokale Grammatiken (aus Senellart et al. 2001)

Dieses wird schließlich in Kombination mit den Graphen während des Übersetzungsprozesses verwendet. In dem von Senellart, Dienes & Váradi beschriebenen Projekt (Adaption an Fachsprache CAD-Software) decken etwa hundert, in fünf Ebenen organisierte Graphen 65% eines Textes ab und beschreiben – ohne alle Subgraph-Expansionen und ohne Rekursionen – 32 000 verschiedene Pfade. Im traditionellen Beschreibungsfomalismus von SYSTRAN entsprechen dem etwa 2400 Transfer-Regeln. Lokale Grammatiken sind also effizient in der Beschreibung von Sprache.

Ein weiteres Anwendungsgebiet lokaler Grammatiken ist die **Informationsextraktion**, z.B. Senellart (1998a,b), Friburger (2002, 2006), Friburger & Maurel (2001a,b, 2004), Mallchok (2005), Traboulsi (2004), Sætre (2006), Watrin (2003, 2006), Bekavac & Tadić (2007), Bsiri, Geierhos & Ringlsetter (2008), vgl. auch Kap. 5.1.3.2.1 und 5.1.5.3 in dieser Arbeit. Die Vorteile sind dabei die graphische Konstruktion komplexer Pattern, die Wiederverwendbarkeit von in Subgraphen abgelegten Pattern und der einfache Zugriff auf Lexika und Gazetteers aus den Grammatiken heraus.

Auch als Grammatiken von Sätzen, Teilsätzen oder Phrasen finden lokale Grammatiken Verwendung. Zum einen als „natural generalization“ (Gross 1997: 352) und Implementierung von **Lexikogrammatiken** (siehe Kap. 1.2), zum anderen zur partiellen Analyse oder ‚Chunk-Parsing‘ (im Gegensatz zur vollständigen syntaktischen Analyse), also der nicht-rekursiven Analyse von Phrasen durch endliche Automaten. Durch die **Kaskadierung** von solchen Automaten bzw. Transduktoren (finite-state cascades) ist schrittweise eine vollständige syntaktische Analyse möglich.¹⁰

¹⁰ Vgl. Joshi (1996), Joshi & Hopely (1999), Abney (1991, 1996a,b), Grefenstette (1996), Blanc, Constant & Watrin (2007a,b) zu den Themen Chunking und Kaskadierung.

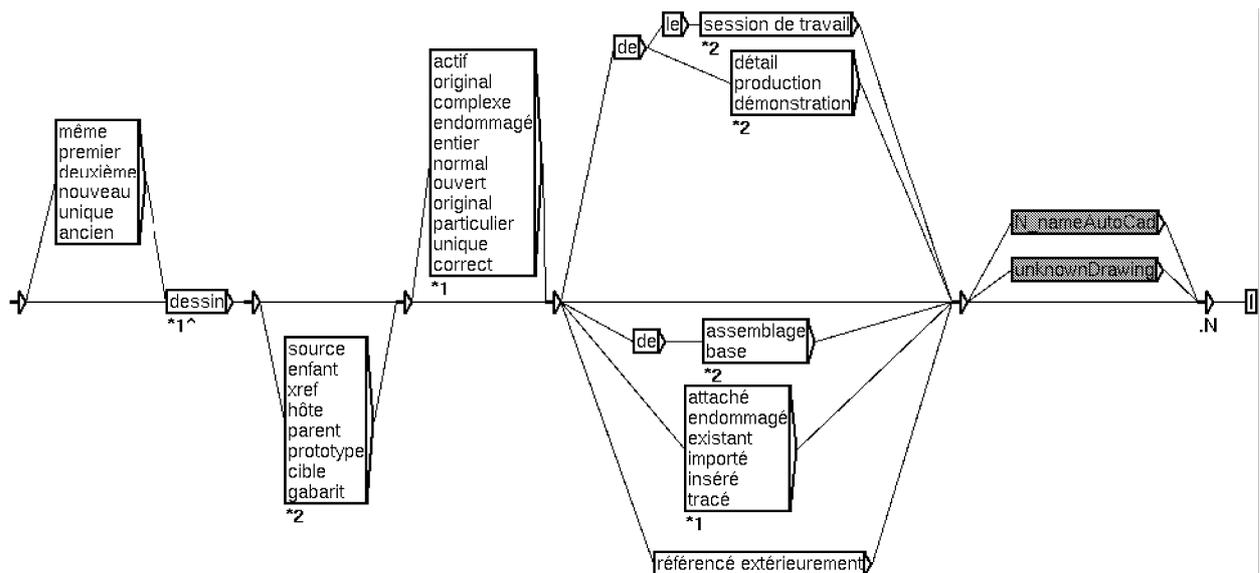


Abbildung 1.4: Repräsentation von Fachtermini durch lokale Grammatiken: das französische Äquivalent zu Abb. 1.3. Teilpfade sind mit Indizes zur Alinierung versehen. (aus Senellart et al. 2001)

Wörter (Wortformen) sind oft ambigue, wenn sie außerhalb ihres Kontextes betrachtet werden: „Questions of ambiguity arise immediately when words are isolated from any context.“ (Gross 2002a: 40). Der Forderung, dass die Kontexte deshalb mit ins Lexikon aufgenommen werden müssen, kommen lokale Grammatiken und auch Lexikogrammatiken ohnehin nach.¹¹ Die Idee, Kontexte, die Ambiguitäten auflösen können, als Muster zu beschreiben, ist bereits in den 50er und 60er Jahren belegt, vgl. das Projekt Uniparse (Joshi 1996, Joshi & Hopely 1999) und Apresjan (1966: 245–6) mit weiteren Literaturhinweisen, und Oliva et al. (2000), Petkevič (2001) Hinrichs & Trushkina (2002, 2004) für aktuelle Anwendungen. Die **Disambiguierung** von Wortarten und morphosyntaktischen Merkmalen ist der wichtigste Schritt für das **Tagging** und kann die syntaktische Analyse vereinfachen. Regeln zur Disambiguierung, wie z.B. für das Französische: „*le* ist kein Artikel (sondern ein Pronomen), wenn eine finite Verbform folgt“, lassen sich auch als lokale Grammatiken oder Automaten repräsentieren. Ebenso lassen sich auch die Ambiguitäten im Satz als alternative Pfade durch einen Automaten (ein sog. **Textautomat**, siehe Abb. 1.5) darstellen. Die Eliminierung von einzelnen Pfaden und damit die Disambiguierung wurde von Mohri (1994), Roche (1992, 1996, 1997), Roche & Schabes (1996, 1997a) als Operation auf zwei Automaten, dem Textautomaten und einem Automaten, der eine Regel beschreibt, formuliert und als System ELAG („Elimination of lexical ambiguities by grammars“) implementiert (Laporte 2001, Laporte & Monceaux 1998-1999; für Anwendungen vgl. Cunqueiro 2005, Krstev & Vitas 2007).

Weitere Anwendungsgebiete lokaler Grammatiken sind schließlich die **Textsegmentierung**, insbesondere die **Satzsegmentierung**, und die Beschreibung von **Flexion** und **Wortbildung**, vgl. Gross (vgl. 1989: 39–44), Savary (2000, 2005), das Kapitel „Lexical Analysis“ im NOOJ-MANUAL, und Literatur zur Finite-State- und Two-Level-Morphology (z.B. Koskenniemi 1983, Karttunen & Beesley 2001, Arppe 2005).

¹¹ Tagging vor der Anwendung der lokalen oder Lexikogrammatiken ist deshalb nicht unbedingt notwendig, wengleich Tagging die Ergebnisse sowohl in der Precision als auch im Recall geringfügig (um 3,5% bzw. 8%) verbessert (Fairon, Paumier & Watrin 2005).

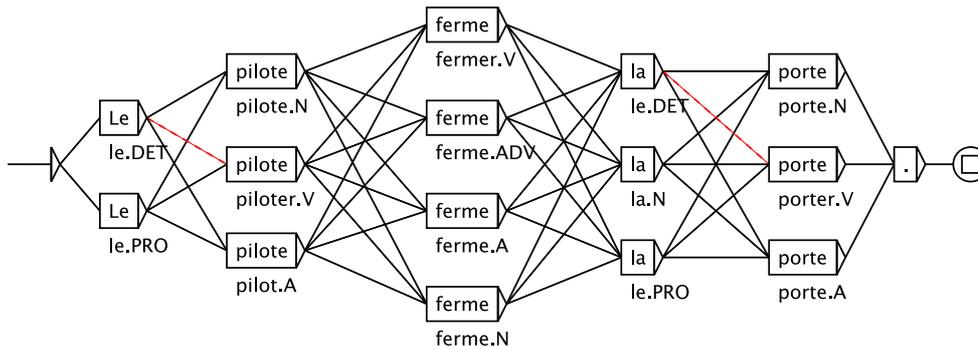


Abbildung 1.5: Textautomat, der die möglichen lexikalischen Analysen des franz. Satzes *Le pilote ferme la porte* (aus Apresjan 1966: 245–6) repräsentiert. Durch die Anwendung der Regel „le/la ist kein Artikel/Determinativ, wenn eine finite Verbform folgt“ (gestrichelte Linien) lässt sich die Anzahl der möglichen Pfade durch den Automaten von 216 auf 160 reduzieren.

Lokale Grammatiken als Methode der Grammatikentwicklung

Eine der ersten methodologischen Fragen, bevor man mit dem Schreiben einer Grammatik beginnt, ist die Wahl einer geeigneten **Entdeckungsmethode** für Grammatiken. Die moderne Linguistik hat die Frage nach einer heuristischen Methode, die es ermöglicht, die zu beschreibenden Phänomene und ihre Realisierungen zu ermitteln, weitgehend ignoriert:

Discovery procedures. It is an essential feature of the method in grammatical studies of the kind inaugurated by Chomsky and pursued here, that no attention is paid to problems of discovering grammatical rules. No straightforward method is presented or implied for arriving at grammatical rules from a corpus of data. It is, in fact, not even assumed that it is possible to develop a rigorous method for establishing grammars, although this possibility is not excluded either. On the whole, linguists are considered to arrive at grammatical descriptions by trial and error, guess-work, on the basis of their previously acquired knowledge of grammatical analysis, etc. The discovery of rules is rather a question of insight than of strict method; it is an art rather than a technique. (Seuren 1969: 5)

Erst in jüngster Zeit zeichnet sich unter dem Stichwort ‚corpus-driven linguistics‘ eine Änderung ab (vgl. u.a. Tognini-Bonelli 2001, Hunston & Francis 2000). Auch Maurice Gross hat die Lexikogrammatiken per Introspektion entwickelt, was sicher vor allem durch das Fehlen ausreichend großer Korpora zum damaligen Zeitpunkt bedingt war. Zur Herleitung von lokalen Grammatiken schlägt er in Gross (1999) ein Verfahren vor, das die Intuition des Linguisten mit Beispielen und Vorschlägen aus einem Korpus kombiniert und gleichzeitig ein Korpus zur Verifizierung der Grammatik heranzieht. Dieses ‚bootstrapping‘ genannte Verfahren wurde von Senellart (1998a,b) exemplifiziert. Es enthält folgende Schritte:

1. beginne mit einer für die zu schreibende Grammatik typischen Struktur – da lokale Grammatiken üblicherweise lexikalisiert sind, meist einem einzelnen Wort oder Lexem
2. erstelle dafür die Konkordanz auf einem geeigneten Korpus
3. durchsuche den linken und rechten Kontext in der Konkordanz nach Elementen, die zu der durch die Grammatik beschriebenen Einheit gehören
4. ergänze die Grammatik um die fehlenden Strukturen
5. erstelle eine Konkordanz der resultierenden Grammatik

6. wiederhole die Schritte 3 bis 5 solange, bis sich die Grammatik nicht mehr vervollständigen lässt

Charakterisiert ist dieses Verfahren durch den Wechsel zwischen Konkordanzanalyse und Vervollständigung der Grammatik. Nach wie vor ist für beide Schritte die Intuition und Urteilskraft des Linguisten notwendig. In ihrer (teilweisen) Automatisierung durch automatische Klassifikation und Clustering der Kontexte steckt das Potenzial des Bootstrap-Ansatzes.

Selten ist ein Korpus jedoch so groß, dass es definitiv alle zu beschreibenden Wendungen enthält. In der Folge ist eine Grammatik auch selten vollständig. Die Vervollständigung einer lokalen Grammatik geschieht durch das Hinzufügen neuer Pfade und gestaltet sich damit verglichen mit der Wartung von in anderen Formalismen geschriebenen Grammatiken recht einfach. Insbesondere, da lokale Grammatiken durch die graphische Repräsentation „leicht zu lesen“ sind. Das macht lokale Grammatiken auch geeignet für den Austausch. Grammatiken natürlicher Sprachen sind zu komplex, um von einer Person erstellt werden zu können. Die Zusammenarbeit erfordert ein geeignetes **Austauschformat**:

Grammars should not be individual pieces of work. Their construction is sufficiently explicit to allow specialists other than the author to use and modify graphs. (Gross 1997: 347)

Published dictionaries are mostly the work of one author. It is impossible to merge two dictionaries or two grammars in order to improve or enlarge them. Each author has a personal point of view, difficult to spell out, which makes it impossible to reproduce the method(s) he used in the descriptions. Electronic dictionaries and grammars must be of a size and a precision that cannot be reached by the methods used for published works. Vocabularies and grammars have to be divided so that specialists can work independently on parts which will have to be merged coherently into a single system. Achieving this goal involves methodological, theoretical and practical constraints which result in a set of standards. (Gross 1999: 235–6)

Lokale Grammatiken werden in der Regel nicht als ein einziger Graph, sondern als Sammlung von Graphen geschrieben. Einzelne Graphen können von anderen Graphen als „Subgraph“ eingebunden werden. Das macht die Grammatik übersichtlicher und erlaubt einzelne Strukturen als „Module“ auch in andere Grammatiken einzubinden und damit „wiederzuverwenden“. Diese **Technik zur Modularisierung** wird ermöglicht durch eine Erweiterung des endlichen Automaten zum Recursive Transition Network (RTN, siehe Kap. 1.3). Sie ermöglicht auch eine effiziente Verwaltung von Grammatiken. Abhängigkeiten zwischen Graphen beschreiben dabei Abhängigkeiten zwischen Teilgrammatiken bzw. Grammatikmodulen (vgl. Constant 2004, 2007, GRAALWEB und Abb. B.8).

Permutationen

Lokale Grammatiken (und endliche Automaten) können zwar die Substitutionen und Elision von Wörtern effizient beschreiben, aber für ein drittes syntaktisches Phänomen, die Permutation, sind lokale Grammatiken kein adäquates Beschreibungsmittel:

[...] the formalism of automata is not well adapted to the description of sentences that differ by a permutation of some of their parts (Gross 1995b: 294)

Finite automata [...] cannot well represent pairs of strings that differ by a permutation. In other terms, the two substrings uv and vu of the strings $AuvB$ and $AvuB$ have to be considered as totally distinct, hence represented by two different paths with common factors A and B . (Gross 1997: 349)

In a more general way, we face a broad limitation: permutation rules cannot be handled in a natural way by finite-state grammars. (Gross 1993: 34)

Gross (1993, 1995b) schlägt deshalb ein zweistufiges System vor. Endliche Automaten beschreiben die Bildung von elementaren Sätzen und Phrasen, Transformationsregeln modifizieren die ursprünglichen Graphen, insbesondere was die Wortstellung betrifft. Trotzdem stellen Permutationen nach wie vor eines der großen Probleme bei der Arbeit mit lokalen Grammatiken dar (vgl. Kap. 4.1), denn eine wirkliche Lösung des Problems existiert bislang nicht.¹²

1.2 Lexikongrammatik

[...] und ich brauchte nur ein Schema tabellarisch auszubilden, wonach man die einzelnen Erfahrungen folgereicht aufreihen und den wunderbaren Lebensgang solcher Geschöpfe deutlich überschauen konnte. (Goethe, Zur Morphologie. Band 1 Heft 1, 1817)

Ich glaube, [...] dass in der gegenwärtigen Zeit der Standpunkt der Wissenschaft, von welcher wir sprechen, der des Sammelns ist. Entfernte Zeiten werden aus dem Stoffe etwas bauen, das wir noch nicht kennen. Das Sammeln geht der Wissenschaft immer voraus; das ist nicht merkwürdig; denn das Sammeln muß ja vor der Wissenschaft sein; aber das ist merkwürdig, dass der Drang des Sammelns in die Geister kommt, wenn eine Wissenschaft erscheinen soll, wenn sie auch noch nicht wissen, was diese Wissenschaft enthalten wird. Es geht gleichsam der Reiz der Ahnung in die Herzen, wozu etwas da sein könnte, und wozu es Gott bestellt haben möge. Aber selbst ohne diesen Reiz hat das Sammeln etwas sehr Einnehmendes. (Adalbert Stifter, Nachsommer)

Die von Maurice Gross initiierten Lexikongrammatiken beschreiben transformationelle und distributionelle Eigenschaften elementarer Sätze. Die beschriebenen Eigenschaften werden in Tabellen dargestellt. Die seit den siebziger Jahren für das Französische gesammelten Daten ermöglichen in bisher unerreichtem Umfang quantitative, und damit empirische Aussagen über die syntaktischen Eigenschaften von elementaren Sätzen bzw. ihrer Prädikate.

In mindestens fünf Punkten stützt sich die Lexikongrammatik auf Ansätze von Zellig Harris, oder zeigt zumindest deutliche Überschneidungen:

I. Die Lexikongrammatik (und die Arbeit von Gross insgesamt, vgl. S. 3) ist „empiriegeleitet“ (Geyken 2001: 1), mit Parallelen zur distributionellen Analyse von Zellig Harris. Sie versucht nicht nur auf empirischem Weg zu Aussagen über die Sprache zu kommen, Empirie spielt auch bei der Ermittlung der zu beschreibenden Fakten eine wichtige Rolle. Wichtigster Fakt, und als solcher experimentell, ist die Bewertung der Grammatikalität/Akzeptabilität eines (elementaren) Satzes. Die Bewertung kann nur binär¹³ erfolgen, d.h. ein Satz ist entweder grammatisch (akzeptabel) oder nicht, wobei sich Gross (1975: 25) für einen großzügigen Akzeptabilitätsbegriff entscheidet. Die Bewertung erfolgt im Satz, nicht in der Phrase, wie folgendes Beispiel von Laporte (vgl. Harris 1991: 359) zeigt:

- (1.1) a. *a quick cup of tea*
b. *I drank a quick cup of tea*

12 Silberztein (NOOJ-MANUAL) schlägt einen Ausweg zumindest für die Analyse vor, indem er eine Pseudoausgabe des Transduktors ($\langle \text{ONCE} \rangle$ oder $\langle \text{ONCE-ID} \rangle$) vorsieht. So lässt sich beschreiben, dass bestimmte Elemente eines Satzes nur einmal auftreten können: z.B. nur ein Subjekt, ein direktes Objekt, ein Zeitadverb, usw. Eine echte Lösung des Problems der Permutationen wird damit allerdings nicht erreicht, da die Beschreibung von Restriktionen in der Abfolge der Elemente damit ausgeschlossen wird.

13 Vgl. kritisch zu einem binären Grammatizitätsbegriff Sampson (2001: 165–79).

c. * *I broke a quick cup of tea*

Obwohl die Lexikongrammatik eine empirische Beschreibung anstrebt, erfolgt die Beschreibung der Einheiten nicht anhand von Korpusbelegen. Das hat zwei Ursachen: (1.) Korpora waren (und sind!) zu klein, um alle beschriebenen distributionellen und transformationellen Eigenschaften in hinlänglich großer Zahl beobachten zu können. (2.) Elementare Sätze (s.u.) sind in Korpora selten und die Rückführung komplexer zu einfachen Sätzen stellt ein nicht triviales Problem der Sprachverarbeitung dar.

II. Ein weiteres Kennzeichen der Arbeiten von Harris wie von Gross ist der Minimalismus im formalen Apparat. Nur wenige Metasymbole finden Verwendung, desgleichen nur wenige Wortklassen, Phrasentypen usw. Motiviert ist dies durch die Erkenntnis von Harris, dass keine Metasprache außerhalb der Sprache selbst möglich ist (vgl. Gross & Schützenberger 1972, Gross 2002b):

The metalanguage is in the language. (Harris 1968:17)

A crucial methodological contribution of the metalanguage is the following: since it is impossible to define the elementary entities and constraints of a language by recourse to its metalanguage (since the metalanguage is itself constructed from those entities by means of those constraints), it follows that the structure of language can be found only from the non-equiprobability of combination of parts. This means that the description of a language is the description of contributory departures from equiprobability, and the last statement of such contributions (constraints) that is adequate to describe the sentences and discourses of the language is the most revealing. (Harris 2002: 8–9)

Zu den wichtigsten und unverzichtbaren metasprachlichen Begriffen der Lexikongrammatik gehören: der Satz, das Prädikat und die Notation eines Satzes als $N_0 V W$.

III. Die zu beschreibenden Einheiten der Lexikongrammatik sind Prädikate oder Operatoren. Wörter können danach klassifiziert werden, ob sie ein oder mehrere Wörter in ihrem Kontext benötigen oder nicht. Die Klasse N enthält Wörter, die keine anderen Wörter im Kontext benötigen (z.B. *John*). Operatoren (O) werden Wörter genannt, die die Präsenz weiterer Wörter erfordern. Sie lassen sich weiter nach der Anzahl und der Klassenzugehörigkeit der von ihnen abhängigen Wörter klassifizieren. So enthält die Klasse O_N alle Wörter, die genau ein Wort der Klasse N fordern (*sleeps* in *John sleeps*), Klasse O_{NN} genau zwei N (*wears* in *John wears boots*), usw. Ein Wort der Klasse O_{OO} (*because*) erfordert dagegen die Präsenz von zwei Wörtern der Klasse O : *John stumbles because John wears boots*. (Harris 1991: 54–9, „Operator Grammar“ in EN.WIKIPEDIA.ORG) Zwischen den Wörtern bestehen also Abhängigkeiten oder ‚Dependenzen‘.¹⁴ Nun bestehen aber Einschränkungen bezüglich der Wahrscheinlichkeit (‚likelihood‘) des gemeinsamen Auftretens eines Operators mit einzelnen Elementen seiner Argumentklassen wie auch für die Argumente untereinander:

14 Harris definiert diese Dependenzrelation unabhängig von der durch Tesnière (1959) begründeten Dependenzgrammatik als rein distributionelle Eigenschaft eines Wortes:

We define the dependence as follows: If A is a simple word, and b, \dots, e is an ordered set of classes of simple words, then A is said to depend on (or, require) b, \dots, e if and only if for every sentence in the base, if A is in the sentence then there occurs in the sentence a sequence of simple words B, \dots, E which are respectively members of b, \dots, e . Within the given sentence, A may then be said to depend on the word sequence B, \dots, E . If in the given sentence there is no other word G such that A depends on G and G depends on the given occurrence of B, \dots, E , then A depends immediately on that occurrence of B, \dots, E . A is then called the operator on B, \dots, E , which in turn is called the argument of A in the sentence. (Harris 1991: 55)

Hence, for a given O_n , different N can have different likelihoods of appearing as its argument; equivalently, the O_n word has different likelihoods of occurring on different N words. [...] For example, *sleep* has reasonable likelihood of appearing on *John, the child, the dog*, but lower likelihood of appearing on *the city* (as in *The city sleeps*), or on *trees* (in *Trees have to sleep each winter*, which is not really metaphoric) or *earth* (*The earth sleeps under a blanket of snow*, and lower yet on *window, chaos*. If an operator has two or more positions in its argument (e.g. subject and object), its likelihoods of occurring on the individual words in one position may be independent of its likelihoods on the words of the other position; [...] For example, the O_{nm} operator *see* has greater likelihood of occurring on *John, dogs, fish* than on *space, water, the blind* [...] in its first argument position, and more or less independently it has greater likelihood on *John, dogs, fish, water, the blind* than on *space, vacuum* in the second position of the NN argument. In contrast, the likelihood of occurrence of the O_{nm} operator *eat* can best be described in respect to pairs of individual words in its NN argument, e.g. with greater likelihoods on *children/cereal, goat/paper, virus/host*, [...] *car/gasoline, Venus fly trap/insects*, than on *children/gasoline, virus/paper, Venus fly trap/cereal*.

(Harris 1991: 61–2)

[...] the likelihood of an operator can be affected not only by its immediate argument but also by some distinguished word further down in the dependency chain—e.g. an argument of its argument to some depth. As an example of one such situation, *melt* is not a likely operator on *house* (as in *The house melted*) but it has some likelihood on *house made of snow*. In this kind of situation, a word with modifier (the latter being reduced from a conjoined sentence) can carry a different selection, which may approximate the selection of a different word, in this case *igloo*, as in *The igloo melted, The house made of snow melted*.

(Harris 1991: 64)

In der Lexikongrammatik entsprechen Harris' ‚likelihood‘ die semantischen Restriktionen bzgl. der Realisierung der Argumente. Im Gegensatz zur Formulierung von Harris (Zitat Harris 1991 Operator *eat*) wird dabei angenommen, dass die Argumente voneinander unabhängig sind. Andernfalls verlieren Modelle wie die Lexikongrammatik, die Prädikat-Argument-Strukturen als zentralen Bestandteil haben, wesentlich an Anziehungskraft (vgl. Laporte 2004a: xiv, Le Pesant 2000: §3).

IV. Eine der von Harris entworfenen Grammatikformen ist die sog. ‚String Grammar‘ oder ‚String Analysis‘:¹⁵

String analysis characterizes the sentences of a language as follows: Each sentence consists of one elementary sentence (its center), plus zero or more elementary adjuncts, i.e. word-sequences of particular structure which are not themselves sentences and which are adjoined immediately to the right or to the left of an elementary sentence or adjunct, or of a stated segment of an elementary sentence or adjunct, or of any one of these with adjuncts adjoined to it. An elementary sentence or adjunct is a string of words, the words (or particular sequences of them) being its successive segments. Each word is assigned (on the basis of its affixes or its position in elementary sentences and adjuncts) to one or more word-categories (rarely, word-sequence categories). Hence, we can replace each word or string by the symbol of its category, thus obtaining a string of category-symbols (called a string formula) as a representation of the word-string in question.

(Harris 1962: 9)

Und genau durch eine solche Folge von Kategoriesymbolen (z.B. $N_0 V N_1 Prép N_2$) klassifiziert die Lexikongrammatik eine Menge von Operatoren und fasst sie in einer Tabelle zusammen. Gemeinsam ist ihnen eben die strukturelle Gleichheit ihrer elementaren Sätze. Johnson (2002: ix) bezeichnet die Lexikongrammatik deshalb als „Lexikalisierung“ der Harrisschen ‚String

15 Betont werden muss, dass Harris die verschiedenen Analysen/Grammatiken (Operator-, String-, Konstituentenstruktur- und Transformationsgrammatik) als gleichberechtigte komplementäre und nicht rivalisierende Ansätze ansieht (Harris 1962: 19–20).

Grammar“: „for each string a table is constructed in which the columns are the word types of the string and the rows are words of the language“.

In der Lexikongrammatik ist der elementare Satz die wichtigste deskriptive Einheit:

The theory of lexicon-grammar is founded on the following axiom: The linguistic unit of meaning is the elementary sentence. [...] A consequence of our axiom is that words are not the elementary units of meaning; [...]

(Gross 1994: 213)

Der Vorteil der Methode, Sätze als die elementaren Bedeutungseinheiten zu verwenden, besteht darin, dass es leichter ist die Bedeutung eines Satzes als die eines Wortes zu bestimmen: „The process of determining what is the elementary sentence and what are adjuncts and to what these adjoin [...] is not hard to grasp“ (Harris 1962: 9–10, vgl. Laporte 2004a: xvi–xvii). Eine Konsequenz ist aber, dass die gleiche Form, ist sie in verschiedenen elementaren Sätzen zu beobachten, auch mehrfach ohne Verbindung in verschiedenen Tabellen der Lexikongrammatik erscheint. Eine semantische Beschreibung, die die Einträge verknüpfen könnte, ist nicht Ziel der Lexikongrammatik.

V. Mittels Transformationen lassen sich elementare Sätze in andere Formen überführen oder in andere Sätze einbetten: „the great bulk of sentences of a language are derived by transformation—to a large extent by transformation—from other sentences, their sources.“ (Harris 1988: 8) Die Transformation ist eine Äquivalenzrelation zwischen Sätzen, die reflexiv (jeder Satz ist über die Identitätstransformation ein Transformant seiner selbst), symmetrisch und transitiv ist (Harris 1964). Zu unterscheiden ist zwischen unären und binären Transformationen (Gross 1993: 26, 1995b, Harris 1962: 15): unäre ändern die Form (nicht aber die Bedeutung) eines (elementaren) Satzes, binäre Transformationen kombinieren zwei (elementare) Sätze zu einem komplexen Satz.

Die wichtigsten Typen von Transformationen sind (vgl. Harris 1964, 1965):

- Permutationen, also die Umstellung von Teilen eines (elementaren) Satzes ohne den syntaktischen Charakter der umgestellten Teile zu ändern (Harris 1972: 116–7)

(1.2) a. *I like this.*
b. = *This I like.*

- Inkremente, Insertionen und Adjunkte: „Local and sentential inserts and adverbial inserts, which do not affect the syntactic status (i.e. the subjectability to transformations) of the sentence parts to which they are adjoined.“ (Harris 1972: 133)

(1.3) a. *He should go.*
b. = *He **in particular** should go.*

- Operatoren auf elementaren Sätzen ($N V \Omega$): „Operators Y , U , W on verb and on sentence, which introduce a new verb, with the original V or sentence being deformed as object or subject of it.“ (Harris 1972: 133)

– auf dem Verb V :

(1.4) a. *He studies eclipses.*
b. = *He is studying eclipses.*

– auf Teilsätzen $V \Omega$

- (1.4) c. = *He is a student of eclipses.*
 d. = *He makes studies of eclipses.*

– auf dem gesamten elementaren Satz

- (1.4) e. = *We know that he studies eclipses.*

- Konnexionen von zwei elementaren Sätzen: „Connectives, which head a sentence and may require a deformation of it, after which the connective-headed sentence is inserted into another sentence by the first operation above.“ (Harris 1972: 133)

- (1.5) a. *The man talked. The man drove.*
 b. = *The man talked and the man drove.*
 c. = *The man talked and drove.*
 d. = *The man who drove talked.*

- Reduktionen, Tilgungen oder Ellipsen (vgl. Harris 1972: 134–42, 1991: 79–83) „The zeroing of redundant material. [This transformation] drops words from a sentence, but only words whose presence can be reconstructed from the environment.“ (Harris 1972: 133) „The reductions neither add nor remove information, and are thus information-preserving.“ (Harris 1991: 80) Im Satz getilgt werden kann vielfältiges Material (vgl. Harris 1991: 79–83), z.B. nach der Konjunktion von Sätzen verdoppelte Elemente (*John plays violin and Mary plays piano* → *John plays violin and Mary piano*) oder (Indefinit)Pronomina (*The place has been taken by someone* → *The place has been taken*). Die wichtigste der von Harris beschriebene Reduktionen sind aber die sog. ‚appropriate words‘ (siehe auch Kap. 3.1.3):

The first type of redundancy removal operates in an insert or a secondary K , i.e. a $C K$, or a K that is under an operator. To consider the actual forms, it will be helpful first to define ‘appropriate X ’, X_{ap} ; X here ranges over the relation-expressing categories V [verb], P [preposition], the operator W , and perhaps classifier-nouns N_{cl} . The X_{ap} of a particular word in a structure is the member (or members) of X which is the main co-occurrent of that word in that structure, for the given subject matter. [...]

In many circumstances [...] the X_{ap} can be eliminated; other members of X cannot. There is no loss of information, for the absence of the X which is required in the $A X B$ structure (whose presence is evidenced by the remaining $A B$), points to the X_{ap} which is determined (up to local synonymy) by the individual words of the $A_i B_i$. Thus from *violin-prodigy* we generally reconstruct *violin-playing prodigy*, and from *violin-merchant* we generally reconstruct *violin-selling merchant*. In any case, the grammatical reality of X_{ap} lies in the fact that it and not other X can be zeroed in this way (or that X_{ap} is the only X that occurs in the given position). (Harris 1972: 134–5)

Neben den Satztransformation spielt die oben beschriebene Nominalisierungstransformation (1.4c) und (1.4d) in der Lexikogrammatik eine besondere Rolle. Diese ist als Relation zwischen Sätzen, und nicht als Relation zwischen einem Satz (1.4a) und einer Nominalphrase definiert. Die Sätze, die Ergebnis der Nominalisierungstransformation sind, benötigen in vielen Sprachen aber ebenfalls ein finites Verb – *make* in (1.4d), das ‚Stützverb‘ (‚support verb‘, ‚ V_{sup} ‘, siehe auch S. 29) genannt wird:

[...] support verbs are verbs which do not present selectional restrictions with respect to their subject and complement. Instead, a selectional relation holds between subject and complement. (Gross 1994: 237)

Support and operator verbs are quite different from ordinary verbs. They do not select their subjects and objects in the way ordinary verbs do; [...] Second, the complement structure of support and operator verbs is different. (Gross 1996: 250)

Stützverben lassen also die Argumentstruktur der Nominalisierung unverändert. Die ebenfalls erwähnten Operatorkonstruktionen (V_{op}) führen zusätzlich zur Argumentstruktur der Nominalisierung ein neues (kausatives oder agentives) Argument ein:

- (1.6) a. *Max got into trouble.* ‚Max geriet in Schwierigkeiten.‘ V_{sup}
 b. *This remark got Max into trouble.* ‚Diese Bemerkung brachte Max in Schwierigkeiten.‘ V_{op} (Gross 1994: 240)

Die Organisation der Lexikogrammtiken, d.h. die Verteilung auf und Beschreibung der Prädikate in separaten Tabellen, basiert auf einer Reihe von Ordnungsprinzipien (siehe Leclère 1990, 2002, 2003, 2005, Gross 1996):

0. Wortart und Typ des Prädikats (Verb, Nomen, Idiom, usw.), mit Verben als der am besten beschriebenen Kategorie
1. Anzahl und Form der Komplemente, d.h. die syntaktische Distributionsklasse des Prädikats basierend auf dem Konzept des einfachen Satzes. Die Notation erfolgt schematisch als Folge von Categoriesymbolen, z.B. $N_0 V N_1 Prép N_2$, wobei die Aktanten einen Index beginnend mit 0 erhalten, während das Prädikat und fixe Elemente des elementaren Satzes ohne Index bleiben. Sechs solcher Basis-Schemata wurden für Sätze mit einem Verb als Prädikat beobachtet:

$N_0 V$
 $N_0 V N_1$
 $N_0 V Prép N_1$
 $N_0 V N_1 Prép N_2$
 $N_0 V Prép N_1 Prép N_2$
 $N_0 V N_1 Prép N_2 Prép N_3$ (Leclère 2002: 33)

Das bedeutet, dass kein Prädikat mehr als vier Aktanten (ein Subjekt und drei Objekte) hat: „we observed that ordinary verbs had practically never three objects.“ (Gross 1996: 254) Dies steht natürlich im Gegensatz zur Methode Mel’čuks, die Semantik eines Prädikats zur Bestimmung von dessen Aktanten heranzuziehen (siehe S. 26). Die Lexikogrammatiker ignorieren solche Überlegungen, obwohl sie elementare Sätze auch als semantische Einheiten betrachten: „elementary sentences of the form subject – verb – essential complements, are considered as elementary units for syntactic as well as semantic composition“ (Gross 1996: 245)

2. nächstes definitorisches Kriterium ist der Inhalt eines bzw. der Komplemente, und zwar als

- a) satzwertiges Komplement (sentential complement)

(1.7) *Paul estime que Luc a raison* ‚Paul denkt, dass Luc Recht hat‘ ($N_0 V Qu P$)

- b) nicht-satzwertiges Komplement / Nominalphrase

(1.8) *Luc a abimé le livre* ‚Luc hat das Buch beschädigt‘ ($N_0 V N_1$)

- c) oder als fixer, idiomatischer (frozen) Bestandteil

(1.9) *Sa dernière heure est arrivée* ‚Sein letztes Stündchen hat geschlagen‘ ($C_0 V$)

3. als weitere Ordnungsprinzipien, damit die Klassen/Tabellen nicht zu groß und heterogen werden, können noch folgende distributionelle Kriterien herangezogen werden:

- a) die Form/Instantiierung der Präposition: $Prép =: \grave{a}$, $Prép =: de$, $Prép =: Loc$ (räumliche Präposition)
- b) Selektionsrestriktionen bzgl. Komplementen: $N =: N_{hum}$ (Mensch), $N =: N_{concret}$ (Konkretum), $N =: N_{loc}$ (Ort, lokatives Nomen)
- c) semantische Interpretation des Prädikats/Verbs: $V =: V_{mvt}$ (Bewegungsverb), usw.

Die letzte Gruppe von Ordnungsprinzipien überschneidet sich mit den in den Tabellen beschriebenen Eigenschaften, d.h. (a) den semantischen Distributionsklassen der Komplemente und (b) den transformationellen Eigenschaften. Die definitorischen Ordnungsprinzipien und die sonstigen in den Tabellen beschriebenen Eigenschaften unterscheiden sich nur in ihrer Priorität. Würden alle Tabellen zu einer einzigen Super-Tabelle zusammengefasst, erscheinen auch die definitorischen Eigenschaften als Spalten dieser Super-Tabelle.

Implementiert und zur Analyse eines Textes angewendet werden Lexikogrammatiken als lokale Grammatiken. Das Verfahren, einen Automaten mit den Einheiten und Eigenschaften einer Lexikogrammatik zu instantiieren, wurde erstmals von Emmanuel Roche (1993a,b) realisiert. Der gleiche Formalismus erlaubt es, lokale Grammatiken, die Konstruktionen mit lokalen Abhängigkeiten wie etwa Zeitadverbialen beschreiben, problemlos mit Lexikogrammatiken zu kombinieren. Die Technik der Instantiierung der Lexikogrammatiken mittels Schablonengraphen ist anhand von Beispielen beschrieben bei Paumier (2002a,b, 2004) und Watrin (2006: 91–3). Sie ist bis heute Thema der Forschung, (z.B. Constant 2003a, 2003b: 169–77, Blanc & Constant 2005).

Maurice Gross begann 1975 mit der Klassifikation französischer Verben im Rahmen der Lexikogrammatik. Mittlerweile wurden von ihm und anderen 12 000 Verben nach 400 Kriterien klassifiziert (Gross 1994: 224). Diese gewaltige empirische Arbeit (siehe auch das Motto dieses Kapitels) hat zumindest ein bemerkenswertes Ergebnis gebracht (Gross 1994, 1996): Vergleicht man die Zeilen der Tabelle nach den 400 Kriterien (Spalten), so gibt es kaum Verben, die die gleichen Eigenschaften realisieren: „no two verbs have the same set of syntactic properties.“ (Gross 1996: 254)

1.3 Repräsentation und Implementierung lokaler Grammatiken

Lokale Grammatiken und endliche Automaten¹⁶ waren bei Gross (1989, 1993) gleichbedeutend (siehe Fußnote 9 auf S. 5). Max Silberstein hat sich dennoch bei der ersten Implementierung einer Engine zur Entwicklung und Anwendung lokaler Grammatiken (INTEX) für Recursive Transition Networks (RTN, Winograd 1983), also ein Äquivalent zu kontextfreien Sprachen entschieden. Vor allem dem Benutzer und Grammatikschreiber spart es Arbeit, wenn er sich wiederholende Strukturen als Subgrammatiken abspeichern kann. Allerdings setzt INTEX die Kontextfreiheit nicht vollständig um. Beim Kompilieren wird eine finite-state-Annäherung der Grammatik erzeugt, indem Selbstrekursionen ab der Tiefe 10 gekappt

16 Für eine formale Definition und Eigenschaften von endlichen Automaten und Transduktoren siehe Roche & Schabes (1996, 1997b).

werden (Silberztein 1998-1999: 20).¹⁷ Silberztein (1998-1999: 22) begründet dies damit, dass „les imbrications de plus de 3 niveaux sont impossibles en français“.¹⁸

Die Diskussion, welcher der drei Typen formaler Sprachen zur Beschreibung natürlicher Sprachen angemessen oder ausreichend ist,¹⁹ ist ebenso alt wie die Unterscheidung von (formalen) Sprachen in reguläre, kontext-freie und kontext-sensitive (Chomsky 1956). Begründet ist sie mit teils fragwürdigen Beispielen wie *The rat the cat the dog chased killed ate the malt* („surely confusing and improbable but it is perfectly grammatical and has a clear and unambiguous meaning“, Chomsky & Miller 1963: 286–287, vgl. Fußnote 18). Sie verstellt den Blick auf die Tatsache, dass reguläre „Sprachen“ als Mittel ausreichend sind, Abermillionen von Sätzen zu beschreiben, wie Gross (1975: 17–8) mit einer Hochrechnung gezeigt hat. Selbst wenn eine natürliche Sprache kontextfrei oder kontextsensitiv wäre, könne es sinnvoll sein, wie Mohri & Sproat (2006) argumentieren, wesentliche Teile der Sprache mit den Mitteln einer regulären Sprache zu beschreiben.

Viel wichtiger als die theoretische Adäquatheit einer Grammatikrepräsentation ist, ob die Form der Repräsentation bei der Entwicklung einer Grammatik hilfreich ist. RTNs sind hier ein Kompromiss: einerseits die intuitive, leicht erfassbare graphische Form des endlichen Automaten, andererseits die Möglichkeit der Modularisierung, wenn Komplexität und mangelnde Bildschirmgröße dies erfordern, oder mehrfach auftretende ähnliche Strukturen dies nahelegen.²⁰

Aus der Perspektive der Anwendung lokaler Grammatiken bei der Analyse (Parsing) haben RTNs jedoch den Nachteil, die Grammatik bzw. den Automaten indeterministisch²¹ werden zu lassen. Die Grammatik durch Einbettung aller Subgraphen in einen Automaten zu überführen ist häufig wegen der Komplexität und der daraus resultierenden Größe des Automaten nicht möglich (vgl. Paumier 2002a, Marschner 2007 auch für Lösungsvorschläge). Allerdings sind lokale Grammatiken ohnehin potentiell indeterministisch: neben den Subgraphen sind auch die Metasymbole (s.u.) und Transduktionen die Ursache.

17 UNITEX lässt Selbstrekursion zwar theoretisch unbeschränkt zu, in der Praxis wird sie jedoch durch ein Makro auf (derzeit) max. 200 Rekursionen beschränkt, um den Stack des Parsers (Locate) nicht zu groß werden zu lassen.

18 Vgl. Gross (1995b: 291–2) zu rekursiv eingebetteten Relativsätzen, die um das Sprachverstehen nicht zu beeinträchtigen auf Rekursionstiefe 3 beschränkt und als Phänomen einzigartig seien. Ebenso eine Korpusstudie (Karlsson 2005) zur rekursiven Einbettung von Phrasen gleichen Typs („self-embedding“):

I have spotted some 200 genuine examples from antiquity to 2005 of nested center-embeddings below depth one. The basic constraint on nested center-embedding in written language is two (with a marginal cline to three), in spoken language one. There are further specific restrictions on which types of clauses may be nested. The practical limit of final embedding (right-branching) is five. Unlimited final embedding is possible only for relative clauses in restricted types of discourse (cumulative folktales). Repeated initial embedding (left-branching) of clauses below depth two does not exist.

Außerdem verändern sich die Restriktionen, denen die eingebetteten Sätze unterworfen sind, je tiefer die Einbettung ist (zit. nach Yli-Jyrä 2005: 17). Die Regeln für verschieden tief eingebettete Sätze gleichen Typs in der Grammatik würden sich damit unterscheiden, was bedeuten würde, dass man überhaupt nicht von „self-embedding“ sprechen kann.

19 Pullum (1991: 131–8) gibt einen forschungsgeschichtlichen Überblick.

20 Vgl. die Anmerkung von Gross (1997: 349) zu „elegance or beauty“ einer lokalen Grammatik.

21 Siehe Roche & Schabes (1996, 1997b) zu einer Definition von (In)Determinismus bei endlichen Automaten.

Metaausdrücke sind ein sinnvoller Bestandteil in lokalen Grammatiken: sie erleichtern die Abstraktion und Generalisierung ähnlich regulären Ausdrücken (z.B. <MAJ> = durchgehend großgeschriebenes Wort, <NB> = Ziffernfolge) und sind notwendig, um die Grammatik mit dem Lexikon zu verknüpfen (<Haus.N:geN> = Genitiv Sg. von *Haus*, <.ADJ:gm> = beliebiges Adjektiv im Gen. Pl.). Lexika im DELA-Format (**D**ictionnaire **E**lectronique du **LADL**) wurden am LADL²² parallel zu lokalen und Lexikogrammatiken entwickelt (Courtois 1990, Silberstein 1990, Courtois 2004) und als Automaten implementiert (Revuz 1991, Clemenceau 1993, Roche 1993c). INTEX/NOOJ, UNITEX und Outilex (Blanc & Constant 2006, Blanc, Constant & Laporte 2006, Blanc 2006) als aktuelle Implementierungen einer Entwicklungsumgebung für lokale Grammatiken kombinieren denn auch Lexikonverwaltung und -Lookup, Grapheneditor, Korpusprozessor und Konkordanzer, und ermöglichen so Entwicklung, Test und Anwendung von lokalen Grammatiken.

22 LADL = Laboratoire d'automatique documentaire et linguistique, Wirkungsstätte von Maurice Gross.

2 Lokative Prädikate

2.1 Sprache und Raum: semantische Aspekte

Dieses Kapitel behandelt stichwortartig wichtige semantische Aspekte, also Aspekte, die bei der Interpretation räumlicher Relationen und Bewegung im Raum unabdingbar sind. Auf Beschreibungsansätze und semantische Modelle wird jeweils knapp Bezug genommen.

2.1.1 Räumliche Relationen

Im einfachsten und gleichwohl nicht trivialen Fall drückt ein lokatives Prädikat die **statische Position** eines Objekts im Raum aus. Solche Prädikate – Zwarts (2003: 40) nennt sie ‚place terms‘ –, sind quer durch alle Wortarten zu finden, z.B. *He lives in the vicinity of a small town* (place noun) oder *How far do you live from town?* (place adjective). Typischerweise erfolgt die Lokalisierung des Objektes (‚figure‘) in Relation zu einem oder mehreren Bezugspunkten oder -objekten (engl. ‚ground‘, ‚landmark‘, ‚reference object‘).

Als **Topologie** wird eine koordinatenfreie Geometrie bezeichnet, in der metrische Eigenschaften (Streckenlängen, Winkel, Krümmung) ohne Belang sind. Nur die gegenseitige Lage und Anordnung der Objekte im Raum ist relevant. Als mathematische Disziplin ist die Topologie der Graphentheorie verwandt. Geometrisch ungleiche Objekte, können topologisch gleich sein. Topologische Elemente sind: null-dimensionale (Punkte/Knoten), eindimensionale (Linien/Kanten), zweidimensionale (Flächen/Maschen), dreidimensionale (Körper). ‚Topologie‘ beinhaltet nur die Konzepte Inklusion und Kontakt (Nähe oder Umgebung), im weiteren Sinn auch mereologische Relationen (‚Teil von‘). Topologische Relationen können formallogisch definiert werden,²³ für eine Übertragung topologischer Relationen auf ‚unscharfe Regionen‘ (‚fuzzy regions‘) vgl. Zhan (1998).

In der Linguistik spricht man von ‚Topologie‘ typischerweise im Zusammenhang mit Präpositionen wie *in*, *auf*, *bei*, *nahe*, *zwischen* – der linguistische Begriff ist also weiter als der mathematische. Sprachlich ausgedrückt werden topologische Relationen aber nicht nur durch Präpositionen, sondern auch durch Kasus (z.B. im Finnischen), Adpositionen, Nomina und Adverbien (vgl. Levinson 2003: 72, 99).

Eine von der topologischen Raumrepräsentationen deutlich verschiedene Strategie zur Wiedergabe räumlicher Relationen ist, ein **Bezugs-** oder **Koordinatensystem** (engl. ‚frame of reference‘) um den Bezugspunkt (Ground) aufzuspannen, das dann die Lokalisierung eines Objektes (Figure) über die Angabe eines Winkels ermöglicht. Natürliche Sprachen kennen drei Arten von Bezugssystemen:

(2.1) *Freising liegt nördlich von München.* (absolut)

²³ Vgl. z.B. RCC Randell, Cui & Cohn (1992). Für formale Definitionen topologischer Relationen siehe Gambarotto & Muller (2003: 150-2).

(2.2) *Vor dem Haus liegt der Garten.* (intrinsisch)

(2.3) *Vor dem Baum liegt ein Ball.* (relativ)

Ein **absolutes Bezugssystem** wird um das Bezugsobjekt (Ground) aufgespannt und an einem Richtungskoeffizienten (engl. ‚slope‘) ausgerichtet. Dieser kann sprachabhängig z.B. durch die dominante (saisonale) Windrichtung, das Gefälle des Reliefs, die Lage von Bergen, die Fließrichtung von Gewässern, die Ausrichtung von Küstenlinien, oder die Position von Gestirnen definiert sein (siehe Levinson 2003: 48–50,90–2).

Ein **intrinsisches Bezugssystem** hat seinen Ursprung im Zentrum des Bezugsobjekts (Ground = G) und wird nach intrinsischen Achsen von G ausgerichtet. Die intrinsische Achse kann durch die Schwerkraft, die Ausrichtung des Wahrnehmungsapparats, die kanonische Bewegung, oder die Lage von Körperteilen definiert sein (siehe Levinson 2003: 42–3,77–81). Miller & Johnson-Laird (1976: 403) entwerfen einen Algorithmus, der die Ausrichtung intrinsischer Achsen für einige Objekttypen voraussagt.

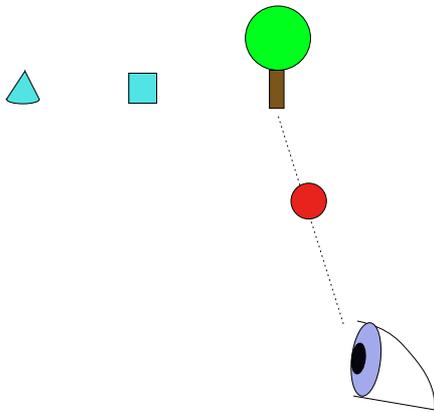


Abbildung 2.1: Relatives räumliches Bezugssystem: *der Ball ist vor dem Baum, der Würfel links vom Baum*

Relatoren, die auf einem **relativen Bezugssystem** basieren, bezeichnen im Gegensatz zu absoluten und intrinsischen Relatoren dreistellige Relationen. Zu Figure und Ground kommt der Betrachterstandpunkt (Viewpoint = V) hinzu, der die Ausrichtung des Bezugssystems bestimmt (siehe Levinson 2003: 47,84–9). Wird V verändert, ändert sich auch die Gesamtkonstellation. Würde der Betrachter z.B. in Abb. 2.1 seine Position mit dem Würfel tauschen, wäre der Ball nicht mehr *vor*, sondern *rechts vom Baum* lokalisiert. Relative Relatoren (*vor*, *hinter*, *links*, *rechts*), die im Deutschen wie vielen anderen Sprachen homonym zu intrinsischen Relatoren sind, können auch dann verwendet werden, wenn das Bezugsobjekt (Ground) über keine intrinsische Ausrichtung, also kein inhärent vorhandenes ‚vorne‘–‚hinten‘ verfügt. Darin unterscheiden sich z.B. Bäume/Bälle von Autos/Menschen.

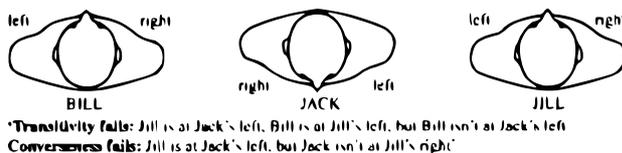


Abbildung 2.2: Intrinsische Relationen sind nicht transitiv und es gibt keine Konversen (aus Levinson 2003: 51)

Im Hinblick auf Relationeneigenschaften unterscheiden sich die drei Typen von Relatoren/Bezugssystemen, siehe Tab. 2.1. Während z.B. relative Relatoren transitiv sind (in Abb. 2.1 steht der Würfel links vom Baum und der Kegel links vom Würfel, damit gilt: der Kegel steht links vom Baum), und Konversen möglich sind (wenn der Kegel *links* vom Würfel steht, steht der Würfel *rechts* vom Kegel), gilt das für intrinsische Relatoren nicht (siehe Abb. 2.2).

	intrinsisch	absolut	relativ
Relation	binär: R(F,G)	binär: R(F,G)	ternär: R(F,G,V)
Transitivität	nein	ja	ja (falls V konstant)
Konversen	nein	ja	ja (falls V konstant)

Tabelle 2.1: Eigenschaften intrinsischer, absoluter und relativer Relatoren

Ein Versuch, die Semantik nicht-topologischer Relationen zu formalisieren (aber ohne Bezug auf die drei Typen von Bezugssystemen), ist die **Vector Space Semantics**. Etabliert von O’Keefe (1996), Zwarts (1997), Zwarts & Winter (2000) zur semantischen Beschreibung von räumlichen Präpositionen, integriert die ‚Vector Space Semantics‘ statische Position und Bewegung in ein Modell. Beide werden als zwei Verwendungsweisen (statisch und direktional) ein und desselben Vektors beschrieben. Dieser beschreibt Lage oder Bewegung eines zu lokalisierenden Objekts (Figure) relativ zu einem Referenzobjekt (Ground). Die Repräsentation durch Vektoren ist vor allem durch die Möglichkeit motiviert, der Präposition Modifikatoren voranzustellen, z.B. *the plow, which lay **ten yards** behind the tractor* (Zwarts 2003: 42). Während sich ein Vektor in der Länge modifizieren lässt, und damit auch die Repräsentation der Lage des Pfluges, ist dies in topologischen Raummodellen, die nur mit Punkten und Regionen, aber ohne Richtungsvektoren operieren, nur schwer realisierbar. In der Vektoralgebra existieren dagegen bereits ausreichende Operationen auf Vektoren (Addition, Skalarprodukt, Inversion, Rotation). Die Vektorrepräsentation lässt sich außerdem auch auf andere räumliche Größen wie Ausrichtung, Gestalt (s.u.) übertragen.

Einige weitere Aspekte in Bezug auf Position im Raum und räumliche Relationen seien hier nur kurz erwähnt:

- **Deixis** stellt eine besondere Art dar, den Bezugspunkt (Ground) zu lokalisieren, nämlich in Bezug auf den Sprecher- oder Hörerstandpunkt (Levinson 2003: 67). Deiktische Ausdrücke basieren auf einem einfachen radialen System, indem Koordinaten und Winkel keine Rolle spielen.
- **Toponymie**: der einfachste Fall zur Angabe einer räumlichen Relation ist die Verwendung eines Ortsnamens. Während Toponyme im Deutschen klar in das topologische System eingebettet sind (*X ist in München*), werden Toponyme in manchen Sprachen ohne räumliche Präposition oder lokativen Kasus verwendet (*X ist München*), was eine Abgrenzung von der Topologie nötig macht (Levinson 2003: 66).
- Zahlreiche Prädikate beschreiben räumliche Eigenschaften eines Objektes und nicht eine räumliche Relation zwischen zwei oder mehr Objekten:
 - **Größe** und **Dimension** eines Objekts: *the height of a tower, to shorten a rope, to grow* (Zwarts 2003: 40). Die These von Bierwisch & Lang (1989), Lang (1989), Dimensionsadjekte bilden den Ausgangspunkt von Raumkonzepten, weist Levinson (2003: 6) strikt zurück als „recent linguistic theorizing about the nature of naïve spatial thought, where it has been supposed that dimensional expressions in language might form the heart of spatial cognition“.
 - **Ausrichtung/Orientierung** eines Objekts oder deren Änderung:
to stand upright, he kicked it over, two rotations, linksherum, rechtsherum, (sich) wenden, zielen, zeigen, schauen, vgl. Zwarts (2003: 40) und Schmidtke et al. (2003)

- **Form/Gestalt** eines Objekts: *gerade, biegen/krümmen, Kugel, Würfel*
- räumliche Untergliederung eines Objekts: *top, backside, corner, centre, bump, notch*, vgl. Zwarts (2003: 40–1)

2.1.2 Bewegung im Raum

Motion is to location like a film is to a picture. It basically it is the change of spatial arrangement over time. (Kracht 2008b)

Die Bewegung im Raum bedeutet, dass ein Objekt (Figure) nacheinander mehrere Bezugsobjekte (Ground) erreicht bzw. passiert, und zwar einen Ausgangs- und Endpunkt (,source‘/ ,origin‘ und ,goal‘/ ,destination‘), und mehrere (,relative‘) Wegpunkte entlang eines Pfades (,path‘, ,trajectory‘)²⁴. Diese Teilpunkte des Pfades können sprachlich kodiert sein, z.B. wie im Finnischen durch Kasusgrammeme. Von einigen Linguisten wird eine Kategorie Modus (Kracht 2002, 2006, 2008a,b) oder Orientierung II (Mel’čuk 1993-2000: II,48–50, 1997-2001: II,58–60) postuliert, die die Richtung und Phase einer Bewegung in Bezug auf ein Bezugsobjekt beschreibt, und zwar mit folgenden Merkmalen:

1. Essiv (static): keine Bewegung: *John is in the store*; oder Lokalisierung der gesamten Bewegung: *John is walking inside the store*
2. Lativ/Illativ (cofinal): der Pfad endet im/am Bezugsobjekt: *John is going into the store*
3. Elativ (coinitial): der Pfad beginnt im/am Bezugsobjekt und führt vom Bezugsobjekt weg: *John is going out of the store*
4. Prolativ (transitory): Pfad passiert (erreicht und verlässt) Bezugspunkt: *John fährt (von Berlin) über Prag (nach Wien)*
5. Approximativ: Objekt nähert sich dem Bezugspunkt ohne ihn notwendig zu erreichen: *John is going towards the store*
6. Rezessiv: Bewegung weg vom Bezugsobjekt: *John is going away from the store*

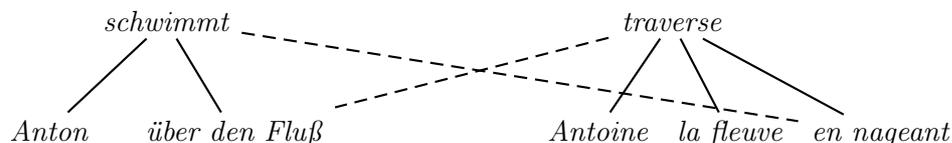
Hinzu kommen aber noch weitere Teilaspekte einer Bewegung, nämlich

- eine temporale/aspektuelle Komponente (Dauer, Phase der Bewegung),
- die Art der Bewegung (z.B. *schlendern* vs. *rennen*),
- das Medium (z.B. Boden, Wasser, Luft),
- das Fortbewegungsmittel, oder
- eine deiktische Komponente (*come* vs. *go*).

Verschiedene Bewegungsverbene fokussieren dabei bestimmte Teilaspekte, z.B. *ankommen* die Phase, *krablen* Art und Medium, *fliegen* das Medium und *segeln* das Fortbewegungsmittel. Die sprachliche Kodierung aller Einzelaspekte erfolgt dabei selten durch ein einziges Wort, sondern durch eine Kombination von Ausdrücken. Dass ein und dieselbe Bedeutung dabei in deutlich verschiedenen syntaktischen Strukturen resultieren kann, formulierte Tesnière in prägnanter Weise:

24 Siehe Habel (1990), Eschenbach et al. (2000), Zwarts (2005), Weisgerber (2005) zum Konzept des Pfades, und Kap. 3.6.2 zur Realisierung einer Grammatik von Pfadangaben.

Daß deutsche Verben eher die Fortbewegungsart spezifizieren, französische Verben eher die Ortsveränderung beschreiben, zwingt zu besonderer Sorgfalt beim Übersetzen: dt. *schwimmen über* entspricht frz. *traverser en nageant*:



Der Unterschied zwischen beiden Sprachen geht so weit, daß der deutsche Satz *Er bimmelte die Straße hinauf* durch den völlig anders konstruierten französischen Satz *Il montait la rue au son des grelots* (wörtl.: ‚Er stieg beim Geläute der Glöckchen die Straße hinauf.‘) übersetzt werden muß. (Tesnière 1980: 213–4)

Die interlingualen Unterschiede betreffen aber bereits die Kodierung von Source, Goal und Wegpunkten. Bohnemeyer et al. (2007: 7) klassifizieren Sprachen nach der Möglichkeit der Realisierung von Verlassen des Ausgangspunktes, Ankunft am Endpunkt und Erreichen der Wegpunkten in einem „Macro-Event“. Typ-I-Sprachen erlauben die Realisierung aller drei Einzelereignisse in einem Macro-Event (*Floyd went from Nijmegen across the river to Elst*), Typ-II-Sprachen nur von Abfahrt und Ankunft (*Floyd went from Nijmegen to Elst, crossing the river*) und Typ-III-Sprachen erfordern für jedes Subevent eine eigene Prädikation (*Floyd left Nijmegen, crossed the river, and arrived in Elst*).

Auf weitere Anmerkungen und Literaturhinweise zum Thema ‚Bewegung im Raum‘ möchte ich hier verzichten und stattdessen auf die umfangreiche Bibliographie von Matsumoto & Slobin (2002) verweisen. Das Thema wird auch vielfach in Kap. 2.5 wieder aufgegriffen.

Routenbeschreibungen kombinieren Ausdrücke für Bewegung im Raum. Laut Taylor & Tversky (1992) sind sie immer durch die Angabe der Positionen von Bezugsobjekten mittels des intrinsischen Bezugssystems des Beobachters, der die Route beschreitet und dabei seine Position und Ausrichtung im Raum ändert, angegeben. Navigationssysteme für das Auto verfahren auf diese Weise. Die absolute Position im Raum ist über die GPS-Daten bekannt, die Ausrichtung des Fahrzeugs und damit des Fahrers über die vorausgegangene Bewegung rekonstruierbar. Daraus lassen sich dann Anweisungen konstruieren, die auf das intrinsische Bezugssystem des Fahrers referieren, z.B. *rechts abbiegen, weiter geradeaus*. Das ist von Vorteil, da das intrinsische System für den Fahrer schnell verfügbar ist. Dass natürlichsprachliche Wegbeschreibungen nur das intrinsische System nutzen, ist dagegen nicht korrekt. Vielmehr wird eine Mischung aller Bezugssysteme verwendet, siehe Beispiel (2.4), (2.88) auf S. 45 und in Levinson (2003: 216–79).

- (2.4) *So wenden wir uns südwärts. Über den freien Hang hinunter gibt es wieder schöne Aussicht. Wir kommen nach Heilinghaus. Die von Pfrentsch kommende Straße queren wir geradewegs. Das Anwesen und die Weide bleiben links. Unten angekommen führt unser Weg halblinks am Hang entlang aufwärts. Eine Buschreihe begleitet uns. Plötzlich springt unser Weg scharf nach rechts weg. Man achte auf die Markierung auf einem Stein und halte auf einen einzelnen Baum zu. Dem Teersträßchen folgen wir nach links in den Talgrund.*

2.2 Lokativa im Inhalt⇔Text-Modell

Das Inhalt⇔Text-Modell (engl. ‚Meaning Text Theory‘, russ. ‚model’ smysl ⇔ tekst‘) teilt sich eine Reihe von Problemfeldern mit der Lexikongrammatik: Prädikate und Aktanten, Stützverben, Bedeutungsdisambiguierung in Abhängigkeit von den Aktanten. Im Unterschied zur Lexikongrammatik basiert das Inhalt⇔Text-Modell auf einer dependentiellen Grammatik und bezieht die Bedeutung von sprachlichen Ausdrücken als wesentlichen Bestandteil des Modells mit ein. Einführungen in das Inhalt⇔Text-Modell existieren mehrfach, erwähnt seien Mel’čuk (1974), Kahane (2003), Mel’čuk (2006: 4–11) und Milićević (2006). Die wichtigsten Eigenschaften des Modells sind:

- es versucht eine Übersetzung (Transfer) zwischen einem Inhalt (Bedeutung) und einem (oder mehreren) Texten, die diesen Inhalt ausdrücken, zu modellieren
- Der Transfer erfolgt über mehrere Ebenen von der Ebene der Semantik, über Tiefen- und Oberflächensyntax, morphologische Repräsentation, bis zur Ausgabe des Textes in orthographischer oder phonetischer Form. Für viele Anwendungen, z.B. die Paraphrasierung (Mel’čuk 1974: 141–176, DEC: III,9–58) oder das Übersetzungssystem ĖTAP (Apresjan et al. 1989, 1992) ist jedoch keine vollständige Transformation eines Textes bis zur semantischen Ebene notwendig: der Transfer erfolgt bereits auf der Ebene der Syntax. Möglich wird dies durch eine Reihe von Normalisierungen in der Tiefensyntax, insbesondere die Anbindung der Aktanten an das Prädikat und die lexikalischen Funktionen (s.u.)
- Dependentielle Strukturen werden auf allen Ebenen (Semantik, Syntax, Morphologie) angenommen, vgl. Mel’čuk (2003)
- Zentraler Bestandteil ist das Lexikon (Erklärend-Kombinatorisches Wörterbuch, Explanatory Combinatorial Dictionary) mit den drei wichtigsten „Zonen“ eines Eintrags:
 - Bedeutungsexplikation
 - Rektionsmodell (syntaktische Realisierung der Valenzen)
 - Lexikalische Funktionen (Kap. 2.2.1)

Syntagmatische Eigenschaften eines Lexems (Valenzen und Kollokationen) nehmen also eine wichtige Position ein, was sich auch darin äußert, das ein sprachliches Zeichen als Tripel $\langle 'X'; /X/; \Sigma_X \rangle$ (signified, signifier, syntactics) definiert ist (Mel’čuk 1995: 169, Mel’čuk 2006: 384).

Beispiele für Lokativa im Erklärend-Kombinatorischen Wörterbuch sind im Anhang, Kap. D zu finden.

- Die Semantik eines Wortes wird zur Bestimmung von dessen Valenz herangezogen. In den Angaben zur syntaktischen Realisierung der Aktanten unterscheidet sich das Rektionsmodell im Erklärend-Kombinatorischen Wörterbuch kaum von üblichen Valenzwörterbüchern oder den Tabellen der Lexikongrammatik. Der Unterschied liegt vor allem darin, wie die Anzahl der Aktanten bestimmt wird. Während sich die Autoren von Valenzwörterbüchern und Lexikongrammatiken auf Tests an der syntaktischen Oberfläche (Kap. 2.3.3) stützen, steht im Inhalt⇔Text-Modell die Bedeutungsexplikation im Mittelpunkt. Alle Variablen in der Bedeutungserklärung sind auch Aktanten des jeweiligen Lexems. So hat das russ. Verb *arendovat’* ‚pachten‘ fünf²⁵ Aktanten gemäß der Definition „*die Person X pachtet die*

25 Maurice Gross behauptet dagegen, dass es keine Verben mit mehr als vier Aktanten gibt, siehe S. 16.

*Immobilie Y von der Person Z für den Geldbetrag W für die Zeitdauer T*²⁶ = gegen Bezahlung eines Geldbetrags W erwirbt die Person X von einer anderen Person Z das Recht, die Immobilie Y für eine Zeitdauer T zu nutzen“ (Apresjan 1974: 120). Würde man die Zeitdauer T nicht als Aktanten ansetzen und folglich aus der Bedeutungserklärung entfernen, würde sich die Bedeutung von *pachten* nicht von *kaufen* unterscheiden. T ist also unverzichtbarer Bestandteil der Bedeutungserklärung, was nicht mit dem obligatorischen Ausdruck von T im Satz zu verwechseln ist (Mel’čuk 2004b: 10–14).

Zentral für das Verständnis des Valenzmodells wie der lexikalischen Funktionen ist das Konzept der ‚linguistischen Situation‘ (Mel’čuk 1974: 85, Mel’čuk 2004b: 10, vgl. Pantel 1993: 33), eine in der gegebenen Sprache vorhandene Lexikalisierung eines Ausschnitts der außersprachlichen Wirklichkeit, ein (komplexes) Faktum, d.h. eine Menge von Fakten und Entitäten, die involviert sind und sprachlich durch Dependenzrelationen an das Prädikat, das die Situation bezeichnet, angeschlossen werden. Jede Situation wird durch eine lexikalische Einheit C₀, oder auch durch eine Menge von Stichwörtern (Synonyme, Konversen, syntaktische Derivate) realisiert. Z.B. kann die Situation ‚kaufen‘ durch die folgenden Lexeme sprachlich realisiert sein: *kaufen*, *erwerben* (Synonym), *verkaufen* (Konverse), *Kauf*, *Käufer*, *Ware*, *Verkäufer*, *Preis* (Derivate S₀, S₁, S₂, S₃, S₄).

2.2.1 Lokative lexikalische Funktionen

Lexikalische Funktionen²⁷ assoziieren ein Stichwort/Argument mit einer Menge von Ausdrücken/Werten. Lexikalische Funktionen verteilen sich auf zwei Typen:

1. paradigmatische LF, mit Werten, die die gleiche Situation wie das Stichwort/Argument beschreiben oder involvieren, z.B. *Dampfer* → *Schiff*, *Kapitän*
2. syntagmatische LF zur Beschreibung von Kollokationen, d.h. lexikalischen Einheiten, die im Kontext des Stichworts eine modifizierte Bedeutung haben, oder exklusiv im Kontext des Stichworts verwendet werden, z.B. *Schiff* → *an Bord*, *vom Stapel laufen*, *auf Grund laufen*

Anwendungen lexikalischer Funktionen sind die Paraphrasierung (Mel’čuk 1974: 141–176, Apresjan 1974: 316–44, DEC: III,9–58), die maschinelle Übersetzung, wo sie die Funktion einer reduzierten Metasprache übernehmen (Magn(*Raucher*) = *stark* vs. Magn(*smoker*) = *heavy*) und der Fremdspracherwerb (Apresjan et al. 2002, Boguslavsky, Barrios Rodríguez & Diachenko 2006).

Im Folgenden werden nur einige wenige lexikalische Funktionen genannt und diskutiert, um ihre Möglichkeiten bei der Beschreibung und Organisation lokativer Lexeme zu demonstrieren.

26 Die Bedeutungsexplikation beginnt typischerweise mit einer schematischen elementaren Phrase, die alle Aktanten als Variablen enthält. Für Verben entspricht dies dem elementaren Satz der Lexikogrammatik, für nominale Einheiten ist dagegen eine Phrase üblich und kein vollständiger Satz mit Stützverb oder Kopula.

27 Eingeführt von Žolkovskij & Mel’čuk 1965, siehe Mel’čuk 1974: 78–109 und Mel’čuk 1996 für eine detaillierten Beschreibung und einen Überblick über Typen lexikalischer Funktionen.

Paradigmatische lexikalische Funktionen

Conv₂₁(*behind*) = *in front of* – die Konversenrelation – hier als LF Conv₂₁ ausgedrückt – assoziiert Stichwörter, die die gleiche Bedeutung, aber eine andere Reihenfolge bei der Realisierung der Aktanten haben. Wie in Kap. 2.1.1 beschrieben, ist die Wertemenge der LF Conv₂₁ für das intrinsische *vor*¹ leer.²⁸ Neben den relativen Präpositionen *vor*, *hinter*, *links*, *rechts* sind weiter Konversen: *nördlich/südlich*, *westlich/östlich*, usw. (falls präpositional verwendet); Conv₃₁(*exportieren*) = *importieren*: *Australien exportiert Kohle nach Japan* = *Japan importiert Kohle von Australien*.

S_{loc} bezeichnet einen typischen Ort für ein Ereignis: S_{loc}(*fight* [as of two armies]) = *battlefield*, S_{loc}(*Schlacht*) = (*Schlacht*)*feld*; oder S_{loc}(*bestatten*) = *Grab*.²⁹

Die Funktion Cap liefert die Bezeichnung für den „Chef“ einer Organisation, die auch eine geopolitische Einheit sein kann, z.B. Cap(*empire*) = *emperor*, Cap(*Bundesrepublik Deutschland*) = *Bundespräsident*.³⁰

Syntagmatische lexikalische Funktionen

Die präpositionalen lexikalischen Funktionen Loc_{in} (≈ ‚being in‘), Loc_{ad} (≈ ‚moving into‘), Loc_{ab} (≈ ‚moving away from‘) sind dadurch motiviert, dass das regierte Lexem oft die Wahl der Präposition zur Angabe des Ortes bzw. Ziel- und Herkunftsortes beeinflusst oder determiniert.³¹ Diese ‚Objektdichotomie‘ bei der Wahl der Präposition ist in manchen Sprachen (so im Russischen) stark, in anderen (z.B. im Deutschen) eher schwach ausgeprägt. Entscheidend ist aber, dass für eine Menge von Lexemen die „passende“ Präposition explizit im Lexikon kodiert werden muss. Betroffen sind lokative Nomina (2.5), Ereignisse (2.6), ebenso wie Toponyme (2.7).³² Auch auf Präpositionen mit von *in/an/ab* verschiedener Bedeutung (2.8) und die Wahl des Kasusgrammems in Sprachen wie dem Ungarischen (2.9) lässt sich die Idee der LF Loc_{in} ausdehnen.

- (2.5) Loc_{in}(*Garten*) = *in*
Loc_{in}(*Hof*) = *auf, in*
Loc_{ab}(*Dachboden*) = *von*
Loc_{ab}(*Keller*) = *aus*

- (2.6) Loc_{in}(*Konzert*) = *in*
Loc_{in}(*Party*) = *auf*

- (2.7) russ. *Ja živu v Krymu / na Urale*
 in / auf
 Ich lebe auf der Krim / im Ural
Loc_{in}(*Krym*) = *v*
Loc_{in}(*Ural*) = *na*

28 In temporaler ‚Lesart‘ gilt: Conv₂₁(*vor*³) = *nach*, siehe S. 424 für die Lesarten von *hinter*.

29 Für Beispiele zur Paraphrasierung mit S_{loc} siehe Apresjan (1974: 321,334) und DEC: III,41.

30 Maurel et al. (2006) verwenden die LF Cap entgegen der Definition bei Mel’čuk, um auf die Hauptstadt eines Landes zu verweisen.

31 Später unterscheidet Mel’čuk (1996: 58) zwischen den lokativen und einer temporalen Funktion Loc_{temp}.

32 Im Deutschen sind Toponyme nur sehr selten betroffen: (1.) als (dialektale) Variante, z.B. Loc_{in}(*Schalke*) = *auf, in*, oder (2.) bei inkorporiertem Klassenbezeichner, der mit der semantischen Klasse konkurriert, z.B. Loc_{in}(*Elfenbeinküste*) = *in, an*.

(2.8) $\text{Loc}_{\text{in}}[\text{nahe}](\text{Nordsee}) = \text{an}$ (*bei)
 $\text{Loc}_{\text{in}}[\text{nahe}](\text{Hamburg}) = \text{bei}$ (*an)³³

(2.9) ungar. *Budapest-en, Szeged-en* (Superessiv)
Párizs-ban, Berlin-ben (Inessiv)

in Budapest, Szeged, Paris, Berlin

(Kracht 2002: 211)

Die Funktionen Loc_{in} , Loc_{ab} , Loc_{ad} bzw. deren Werte können als regierte Präposition im Rektionsmodell eines Lexems verwendet werden,³⁴ um so das Problem der doppelten Abhängigkeit der Präposition (vgl. S. 46) zu modellieren.

Die Werte der Stützverbfunktionen Func , Oper und Labor (Mel'čuk 1974: 92–5, Mel'čuk 1996: 59–64, Mel'čuk 2004a, vgl. auch S. 15) ermöglichen, dass Nomina die Rolle des semantischen Hauptprädikats eines Satzes einnehmen. Das Stützverb fungiert dann als syntaktisches Prädikat, trägt aber selbst nicht zur Bedeutung des Satzes bei.³⁵ Stützverben werden meist im Zusammenhang sog. ‚prädikativer Nomina‘ genannt, also Wörtern wie *Frage*, *Kontrolle*, die Handlungen oder Ereignisse bezeichnen, und zu denen meist ein paralleles Verb (V_0 : *fragen*, *kontrollieren*) existiert. Doch auch Quasiprädikate (siehe Kap. 2.6 für eine Definition) können mit Stützverben verbunden werden und dabei Aktanten anschließen, so z.B. *Straße* (vgl. Straße I.1, S. 434) über $\text{Func}_{23}(\text{Straße}) = \text{verbinden}$:

- (2.10) a. *Die neue, komplett asphaltierte Straße verbindet die namibische Hauptstadt Windhoek mit Gabarone in Botswana.*
 b. = *die neue, komplett asphaltierte Straße von der namibischen Hauptstadt Windhoek nach Gabarone in Botswana*

Für Quasiprädikate weitaus häufiger und typischer sind jedoch Verben der Real-Familie (Mel'čuk 1974: 97–8, Mel'čuk 1996: 68–9, Kahane 2003: 556), die ausdrücken, dass eine Situation oder ein Objekt die in der Bedeutungserklärung gegebene ‚innere Bestimmung‘ erfüllt, bzw. erfährt oder gemäß ihr benutzt wird. Wenn also eine *Straße* ‚ein befestigter, meist asphaltierter, begeh- und befahrbarer Landverkehrsweg‘ ist, ‚dazu bestimmt, dass eine Person X vom Ort Y über W nach Z gelangen, oder Objekte befördern kann‘, dann sind *benutzen* und *nehmen* Werte der Funktion Real_1 . In *Max nimmt die Landstraße nach Apfeldorf* nutzt Max als der erste Aktant X die *Landstraße* gemäß ihrer Bestimmung, um nach Apfeldorf zu gelangen.

Die Beschreibung von regulären, paradigmatischen Ersetzungen einerseits und die Irregularitäten in der Kombinatorik von lexikalischen Einheiten andererseits bilden den Überschneidungsbereich lokaler Grammatiken und lexikalischer Funktionen. So lässt sich der Ausdruck

33 Der Versuch, die Unterschiede in der Verwendung von *an* und *bei* allein aus Objekteigenschaften der regierten Lexeme zu erklären, muss hier scheitern. Vgl. Li (1994) und Carstensen (2000: 246), der die Distribution von *an* über ‚Kontakt‘ und *bei* über ‚Nähe‘ zum Bezugsobjekt abzuleiten versucht, dann aber über Beispiele wie **Norderstedt ist an Hamburg* u.a. feststellt, „dass jenseits rein räumlicher und/oder funktionaler Aspekte weitere bzw. andere Kriterien eine wesentliche Rolle in der Semantik lokaler Präpositionen spielen.“

34 Z.B. im TKSR: *passažir* ‚Passagier‘, *peč'2* ‚Ofen (zum Heizen eines Zimmers)‘

35 Abgesehen von grammatischen Bedeutungen wie Tempus usw. Das bedeutet aber nicht, dass das Stützverb semantisch wirklich ‚leer‘ ist: „words we call empty need not be devoid of semantic content in all possible contexts. A word may become empty in a particular context where it duplicates a part of the meaning of another word syntactically linked to it.“ (Mel'čuk & Pertsov 1987: 34) Vgl. engl. *to ask a question*.

im hohen Norden sowohl als lokale Grammatik darstellen, wie auch mittels der lexikalischen Funktionen $\text{Magn}(\text{Norden}) = \text{hoch}$ und $\text{Loc}_{\text{in}}(\text{Norden}) = \text{in}$ als Dependenzbaum repräsentieren (Abb. 2.3). Beide Repräsentationen, die Stringdarstellung in der lokalen Grammatik und die dependentielle mit lexikalischen Funktionen haben Vor- und Nachteile. Die dependentielle Darstellung im $\text{Inhalt} \leftrightarrow \text{Text}$ -Modell ist einer semantischen Repräsentation näher, lexikalische Funktionen können direkt unter dem Stichwort im Wörterbuch (*Norden*) verzeichnet werden, und weit auseinanderliegende, nicht lokale Kollokationen sind beschreibbar. Andererseits ist die Stringdarstellung intuitiver und natürlicher. Irregularitäten in der seriellen Abfolge werden ebenfalls beschrieben. Wie das „Raising“ von *hoch* in die Position eines Attributs der Präposition (*hoch im Norden*) dependentiell modelliert werden soll, ist unklar. Die beiden Ausdrücke *im hohen Norden* und *hoch im Norden* sind Paraphrasen, aber durch eine allgemeine syntaktische Regel ist die Umformung nicht zu beschreiben, vgl. *weit im Norden* (**im weiten Norden*) und *im äußersten Norden* (**äußerst im Norden*).

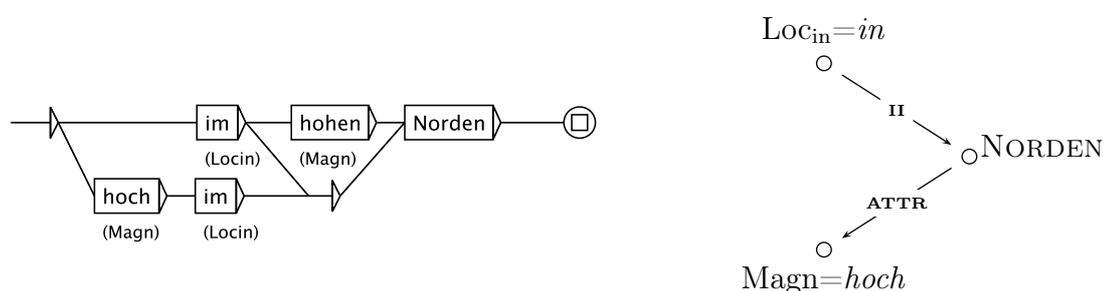


Abbildung 2.3: Repräsentation des Ausdrucks *hoch im Norden* als lokale Grammatik (links, vgl. Abb. B.3) und in der tiefensyntaktischer Darstellung mit lexikalischen Funktionen (rechts)

2.3 Lokative Prädikate: Kriterien und Realisierung

2.3.1 Die Suche nach dem Prädikat

Die Frage was das Prädikat eines Satzes ist, ist eine der wichtigsten in der Syntax, wenn nicht die wichtigste überhaupt. Das Prädikat bestimmt ganz wesentlich, in welchen Formen (Satzschemata, Satzrahmen und Transformationen) ein Satz auftreten kann. Hier zwei Beispiele, die dies deutlich machen und dabei auch zeigen, wieso die Bestimmung des Prädikats oft nicht einfach ist:

- (2.11) a. i. *Paris ist die französische Hauptstadt.*
 ii. = *Paris ist die Hauptstadt von Frankreich.*
 b. i. *Paris ist die schönste Hauptstadt Europas.*
 ii. = *Paris ist die schönste unter den Hauptstädten Europas.*
 iii. = *Unter den Hauptstädten Europas ist Paris am schönsten.*
 iv. = *Paris, die schönste europäische Hauptstadt ...*
 c. ? *Paris ist die schönste französische Hauptstadt Europas.*

In (2.11a-i) ist *Hauptstadt* das Prädikat mit *Paris* und *Frankreich* als Aktanten, in (2.11b-i) dagegen *die schönste* mit den Aktanten *Paris* und *die Hauptstädte Europas*, wie durch den

Vergleich mit (2.11b-ii) deutlich wird. (2.11c) ist ungrammatisch, weil es zwei unterschiedliche Prädikate auf syntaktisch unzulässige Weise miteinander verknüpft. Nur eine genaue Betrachtung, wo das Prädikat im Satz zu suchen ist, kann hier eine Lösung bieten. Die Suche ist nicht immer leicht, das Prädikat nicht offensichtlich. Jede Grammatik, die nicht nach dem Prädikat sucht, weil sie der oberflächensyntaktischen Struktur zu sehr verhaftet ist, muss hier versagen und „semantische Inkompabilitäten“ als Ausrede vorschieben.

Viele solcher „unkonventionellen“ Prädikate sind in Arbeiten, die in der Tradition der Lexikongrammatik stehen, zu finden. Sehr fundiert ist die Behandlung von Prädikaten und ihren Argumenten in den Arbeiten zum Inhalt⇔Text-Modell, z.B. Mel'čuk (2004b,c), der auch klar zwischen semantischen und syntaktischen Prädikaten unterscheidet.

2.3.2 Kriterien für lokative Prädikate

Die Frage nach einem Kriterienkatalog zur Bestimmung lokativer Prädikate bzw. der Lokativität eines Prädikats, lässt sich am besten über die Frage ‚Was ist ein lokativer Satz?‘ beantworten. Ein lokativer Satz ist:

- ein einfacher Satz,
- dessen Prädikat mind. einen lokativen Aktanten selektiert,
 - wobei lokative Aktanten beschränkt sind auf Orte, wo sich ein Mensch aufhalten kann.
 - [Der Fokus liegt auf Toponymen in der Aktantenposition.³⁶]
- Ein lokativer Satz muss dabei semantisch ausdrücken:
 1. die räumliche Relation zwischen zwei oder mehr Objekten (*Max steht hinter Peter*),
 - 1a. eine nicht eindeutig räumliche Relation zwischen einem Objekt und einem eindeutig räumlichen Objekt (*Max hat seinen Wohnsitz in München*),
 - 1b. (räumliche) Eigenschaft eines eindeutig räumlichen Objekts (auch einstellige Prädikate)
 2. die Bewegung eines Objektes im Raum, oder
 3. die Relation zwischen mind. zwei definitiv räumlichen Objekten (*Paris ist die Hauptstadt von Frankreich*)

Klar ist bei einer solchen Definition, die sich nahezu ausschließlich auf semantische Eigenschaften stützt, dass es keinen operationalisierbaren syntaktischen Test oder eine Testbatterie gibt, vgl. die Diskussion in Kap. 2.5.1.1.1.

Das Merkmal ein lokatives Argument regieren zu können, ist nicht ausreichend, da Lokativa (Nomina wie Eigennamen) metonymisch verwendet werden können (siehe S. 81 und Kap. 3.1.4), oder ganze Klassen wie die Kommunikationsverben (vgl. S. 48) zu lokativen Prädikaten würden. Auch leidet dieses Kriterium daran, dass die Unterscheidung von Aktanten (Komplementen) und Supplementen eine der schwierigsten in der Linguistik ist (Kap. 2.3.3).

36 Die Konzentration auf Toponyme ist keine definatorische Eigenschaft, sondern Ergebnis rein praktischer Überlegungen im Hinblick auf computerlinguistische Anwendungen wie die Informationsextraktion.

2.3.3 Abgrenzung zwischen (lokativen) Komplementen und Supplementen

Die Abgrenzung zwischen valenzgebundenen Komplementen und außerhalb des Valenzrahmens stehenden Supplementen ist das zentrale Problem jeder Abhängigkeits-/Valenzgrammatik und überhaupt jeder Grammatik, die auf Prädikat-Argument-Strukturen als wichtigen Aspekt der Beschreibung setzt.

Die Frage der Bestimmung von Valenz wurde vielfach diskutiert, z.B. in Ágel (2000: 167–213), Storrer (2003), (Mel'čuk 2004b: 38–53, 2004c: 265–9), um nur einige der neueren Arbeiten zu nennen. Zahlreiche Testverfahren wurden vorgeschlagen, aber keines oder auch keine Gruppe davon vermag immer eindeutige Antworten geben, deshalb „ist die Abgrenzung der valenzgebundenen von den valenzfreien Gliedern Hauptproblem der Valenzgrammatiker und Hauptangriffsziel ihrer Kritiker.“ Eroms (1981: 27)

Im Fokus steht hier die Valenz der Lokativa. Deren Valenzgebundenheit (Aktantenstatus) ist aus zwei Gründen schwer zu bestimmen.

1. Der erste Grund liegt in einer generellen semantischen Eigenschaft der Lokativa: „A situation normally takes place somewhere and at a given moment: space-time coordinates characterize most situations. But exactly because of that, location and time as such are not obligatory participants of linguistic situations — they are not distinctive. Location or time can be an obligatory participant only of some very specific linguistic situation: for instance, *Alain is based in Singapore* (X is based in location Y) or *2001 saw Alain in Singapore* (X is the time in which P(Y)).“ (Mel'čuk 2004b: 13)
2. Lokative Argumente eines Prädikats werden im Deutschen häufig als Adverbialkomplement in Form einer Präpositionalphrase realisiert, also in syntaktisch der gleichen Weise wie die meisten Supplemente. In der Folge scheiden viele an der syntaktischen Oberfläche operierende Tests aus.

Doch nun zu einer Reihe von Tests und der Frage ihrer Anwendbarkeit auf lokative Komplemente oder Supplemente. Vgl. dazu Gutermann (1996: 137–46), die Tests speziell für Lokativa bespricht, sowie Kap. 2.5.1.1.1 und 2.5.1.2. Dort wird auch eine zweite Frage angesprochen, die hier im Hintergrund bleibt: Gibt es einen operationalisierbaren Test, der lokative von nicht-lokativen Argumenten trennt?

- der Präsuppositionstest (Schnorbusch 1999: 93, Ulrich 2002: 36):

(2.12) * *Hein hat sein Boot verkauft, aber es gab niemanden, der es gekauft hat.*

(2.13) * *Hein hat sein Boot an Fiete verkauft, aber Fiete musste nichts bezahlen.*

Demnach hat *verkaufen* vier Aktanten (vgl. S. 26). Auf Lokativa ist der Präsuppositionstest nicht anwendbar, da ja jedes Ereignis und jede Handlung einen Ort impliziert (s.o.). Damit wäre dann auch der Ort ein Aktant von *verkaufen*:

(2.14) * *Hein hat sein Boot verkauft, und das geschah nirgendwo.*

- (Panevová 1994: 228–231, vgl. Mel'čuk 2004b: 53) schlägt einen Test zur Bestimmung der Obligatheit von semantischen Aktanten vor, der sich als hinreichendes (aber nicht notwendiges!) Kriterium pro den Aktantenstatus verwenden lässt. Getestet wird, was der Sprecher über eine Situation wissen muss. Äußert er den grammatischen Satz *Peter ist*

weggefahren, kann er auf die Frage *Wohin fährt Peter?* zwar antworten, er wisse es nicht. Auf die Frage *Von wo ist Peter weggefahren?* kann der Sprecher dagegen in jedem Fall Auskunft geben. Umgekehrt bei *ankommen*: hier muss der Sprecher den Zielpunkt der Bewegung kennen, nicht aber den Ausgangspunkt. Dieses Testverfahren liefert also neben der sicheren Bestimmung einiger Aktanten auch eine Aussage darüber, auf welchen Teil des Pfades ein Bewegungsverb den Fokus setzt.

- einige Tests versuchen Supplemente auf Sätze zurückzuführen.³⁷ Ist dies möglich, kann damit eine Konstituente als Supplement bestimmt und als Komplement (Aktant) ausgeschlossen werden.

– Temporalsatzparaphrasen (Eroms 1981: 33–5):

(2.15) a. *Er wohnte in Dresden.*

b. =* *Er wohnte, als er in Dresden war.*

(2.16) a. *Er starb in Dresden.*

b. = *Er starb, als er in Dresden war.*

(2.17) a. *Die Kinder spielten hinter dem Haus.*

b. = *Die Kinder spielten, als sie hinter dem Haus waren.*

(2.18) a. *Der Obstgarten lag hinter dem Haus.*

b. =* *Der Obstgarten lag, als er hinter dem Haus war.*

– der *und zwar*-Test (Tarvainen 1981: 29–31, Storrer 2003: 770) testet, ob sich eine Konstituente mittels *und zwar* in eine rhematische Position ausgliedern lässt:

(2.19) a. *Er starb, und zwar in Dresden.*

b. * *Er wohnte, und zwar in Dresden.*

c. *Er baut, und zwar ein Einfamilienhaus.*

d. ?? *Max fährt, und zwar nach Hamburg.*

e. *Max fährt zu seiner Oma, und zwar nach Hamburg.*

Der Test ist unbrauchbar, da er *Einfamilienhaus* in Satz (b) falsch als Supplement klassifiziert. Die Ausgliederung ist auch immer bei Mehrfachbesetzung (Quasikoordination) von Argumentstellen möglich (2.19d) und (2.19e).

- eine Reihe von Tests versuchen auf die Handlung oder ‚linguistische Situation‘ mittels eines Proverbs (*tun, machen, geschehen*) zu referieren. Diese Tests gehören zu den vielversprechendsten (vgl. Mel’čuk 2004c: 287, FN 11). Ein Nachteil ist, dass nur Verben, und keine anderen Wortarten getestet werden können.

– der *do so*-Test (Mel’čuk 2004c: 266–8):

(2.20) * *I live in Manchester and Jock does so in Salford.* (*Salford* ist Komplement)

(2.21) *I have lived in Manchester since 1995 and Jock has done so since 1999.*
(*since 1995* ist Supplement, da die Ersetzung durch ein Proverb möglich ist)

37 Eine Methode, die zwar unabhängig aber äquivalent zur Theorie von Harris ist, dass komplexe Sätze auf eine Menge einfacher Sätze abgebildet werden können. Bei Harris findet er sich als „*occur*-Test“: *Jo arrived at noon.* = *Jo arrived. It occurred at noon.*

– der *tun/geschehen*-Test:

- (2.22) a. *Er starb, und das geschah in Dresden.*
b. * *Er wohnte, und das tat er in Dresden.*

– und eine Variante des Proverb-Tests als „Fragetest“ (Storrer 2003: 769):

- (2.23) a. *Was machte er in der Stadt?*
b. *Er ging in der Stadt spazieren.*

- (2.24) a. * *Was machte er in die Stadt?*
b. *Er ging in die Stadt.*

- Formmerkmale, wie die Kasusreaktion oder konstante, „regierte“ Präpositionen werden ebenfalls als Testfälle herangezogen (Storrer 2003: 770–1). Doch ist kein Formmerkmal hinreichendes Kriterium: erstens können Komplemente wie Supplemente in der gleichen Form auftreten (zweimal Akkusativ in *Er bearbeitete das Werkstück eine Stunde*), zweitens sind gerade lokative Objekte dadurch gekennzeichnet, dass die Form (und Bedeutung) der lokativen Präposition variieren kann. Von den französischen Lexikongrammatikern wird die Variabilität der Präposition als ausreichendes Kriterium dafür angesehen, dass ein Komplement ein lokatives ist (S. 46).
- die Weglassprobe oder der Eliminierungstest (Tarvainen 1981: 24–9, Storrer 2003: 771–3) testen, ob sich eine Konstituente aus dem Satz eliminieren lässt, ohne dass er ungrammatisch wird. Falls nicht, handelt es sich in jedem Fall um ein Komplement. Nun sind Komplemente häufig optional, und damit weglassbar. Der Test sagt damit mehr über die Optionalität von Argumenten aus, als das er Komplemente zu bestimmen vermag. Im Falle der bei lokativen Prädikaten häufigen Adverbialkomplemente (PPs) scheitert der Test, denn diese sind praktisch durchweg elidierbar (Maienborn 1991).
- auch Transformationen eines Satzes mit nur einem ausgetauschten Bestandteil, sind als Kriterium im Gespräch:

- (2.25) a. *This bed has been slept in.*
b. ! *This town has been slept in.*

Auf diese Transformation wird noch zurückzukommen sein (S. 63 und S. 71), zur Bestimmung von Aktanten ist sie in diesem Fall nicht geeignet – *Bett* ist kein Aktant von *schlafen*. Hier ein Kommentar von (Seuren 2002: 181, vgl. Mel'čuk 2004c: 269–70): „Clearly, the relation between sleeping and beds is as close as can be, much closer than between sleeping and towns. [...] In that case it is treated grammatically *as if it were a lexically defined prepositional object*, thus allowing for passivization. Thus ‘mongrel’ argument terms may be called *pseudoarguments*.“

- auch durch Besonderheiten in der Wortstellung lassen sich Komplemente gegenüber Supplementen abgrenzen (Tarvainen 1981: 30). Die Kriterien gelten, aber immer nur für eine Sprache, hier das Deutsche:

– ein valenzgebundenes steht hinter einem freien Adverbial:

- (2.26) a. *Du hast das Buch am Vormittag auf den Schrank gelegt.*
b. * *Du hast das Buch auf den Schrank am Vormittag gelegt.*

– ein freies Adverbial kann im Nachfeld stehen:

- (2.27) a. *Du hast das Buch auf den Schrank gelegt am Vormittag.*
b. **Du hast das Buch am Vormittag gelegt auf den Schrank.*

– die Negation steht vor einer Adverbialergänzung (Komplement), kann aber vor und nach einer Adverbialangabe (Supplement) stehen:

- (2.28) a. *Er wohnte nicht in Berlin.*
b. **Er wohnte in Berlin nicht.*
c. *Er traf sie nicht in Berlin.*
d. *Er traf sie in Berlin nicht.*

2.3.4 Bestimmung lokativer Komplemente

Die zweite Frage dieses Abschnitts, wie lassen sich lokative von nicht-lokativen Aktanten abgrenzen, trägt zur Bestimmung lokativer Prädikate bei, da diese ja u.a. darüber definiert sind, einen lokativen Aktanten zu selektieren (Kap. 2.3.2). Folgende Tests kommen in Frage (vgl. auch Kap. 2.5.1.2 und 2.5.1.1.1):

- der Pronominalisierungstest trennt lokative Adverbialergänzungen von eigentlichen Präpositionalobjekten (vgl. Zifonun, Hoffmann & Strecker 1997: 1367–74). Dabei wird versucht, das Komplement durch ein Pronomen, Präpositionaladverb oder Adverbien wie *dort* zu substituieren:

(2.29) a. *Max steuert auf die Insel / *dorthin / darauf / auf sie zu.*

(2.30) a. *Max stammt aus Hamburg.*

b. *Max stammt von dort / *daraus / *aus ihm/ihr.*

(2.31) a. *Max wohnt in München / dort / *darin.*

(2.32) a. *Max stellt das Glas auf den Tisch / dorthin / darauf / ?auf ihn.*

Wie das Beispiel zeigt, kann der Test zwar Präpositionalobjekte mit einer stark regierten Präposition klar erkennen – eine Ersetzung durch *dort* ist nicht möglich. Der Test trennt jedoch nicht zwischen lokativem und nicht-lokativem Gebrauch (vgl. Gutermann 1996: 140). Alle Beispiele oben sind lokative Sätze, aber bezüglich der Substituierbarkeit des Ortskomplements variieren sie erheblich.

- der Fragetest, also die Abfrage eines Aktanten durch eine Fragepronomen oder -adverb, ähnelt dem Pronominalisierungstest: ist ein Komplement durch *wo*, *wohin*, *woher*, *wodurch* (*où* für die franz. Lexikogrammatik, siehe S. 46) abfragbar, handelt es sich um eine lokative Ergänzung. Auch dieser Test lässt sich nur auf Präpositionalphrasen und Adverbien mit Satzgliedstatus anwenden, sowie einige Nomina mit Direktionalergänzung (*die Fahrt wohin?*). Kombiniert mit Stützverb aber auch auf andere: *Wo ist Peter ansässig? Wohin geht die Reise?*

Der Test erlaubt es Präpositionalergänzungen in konditionaler Funktion auszugrenzen: *... gestattet den Wechsel der Magazine in allen Gerätelagen* → **Wo gestattet ...* (Gutermann 1996: 139–40). Er trennt aber nicht zwischen konkreter und abstrakter räumlicher Verwendung

(2.33) *Wohin legt er den Fall? – Ad acta / zu den Akten*

(2.34) *Wohin führt diese Spur? – In eine Sackgasse oder zur Ergreifung des Täters?*

und erlaubt auch Infinitive und satzwertige Komplemente.

(2.35) *Wohin geht Max? – Angeln.*

Der Fragetest fragt im Deutschen auch den semantischen Typ des lokativen Komplements (statisch oder direktional) und den Fokus (Ausgangs-, Ziel- oder Wegpunkt) ab. Allerdings schlägt dies auch manchmal fehl, z.B. bei Bewegungsverben, die eine statische Präposition verlangen (*Wohin fährt er?* vs. *Wo kommt er an?*).

2.3.5 Realisierung und Synonymie der Prädikate

Prädikate können durch verschiedene Wortarten realisiert sein:³⁸

- als Verb

(2.36) *Die Isar **mündet** in die Donau.*

(2.37) *Die Isar **speist** die Donau.*

(2.38) *Max **reist** durch Europa.*

- als Nomen

(2.39) *Die Isar ist ein **Nebenfluss** der Donau.*

(2.40) *Max unternimmt eine **Reise** nach Europa.*

(2.41) *Max macht eine **Europareise**.*

(2.42) *Die **Entfernung** zwischen München und Windach beträgt nur $\frac{1}{2}$ Autostunde.*

- als Adjektiv/Adverb

(2.43) *Max ist in Europa **unterwegs**.*

(2.44) *Windach liegt $\frac{1}{2}$ Autostunde von München **entfernt**.*

(2.45) *das eine halbe Autostunde von München **entfernte** Windach*

- als Adposition

(2.46) *München liegt **in** Bayern.*

(2.47) *Ingolstadt liegt **im Herzen** Bayerns.*

(2.48) *Windach liegt $\frac{1}{2}$ Autostunde **außerhalb** Münchens.*

38 Betrachtet man als Prädikation jegliche **sprachliche Kodierung** räumlicher Relationen oder von Bewegung im Raum, und bezieht verschiedene Sprachen mit ein, sind nicht nur Wörter aller Wortarten, darunter auch deiktische Pronomina (*dieser, jener*), sondern auch Grammemme, z.B. Kasusgrammeme (Lokativ, Illativ, Elativ, Inessiv, Adessiv) beteiligt (vgl. Levinson 2003: 98–110 und Kap. 2.1.2).

Wie aus den Beispielen bereits klar wird, sind Paraphrasen nicht nur über synonyme Prädikate gleicher Wortart, sondern auch über die Wortarten hinweg möglich. Deshalb werden diese vier Wortarten im Folgenden auch als gleichwertig behandelt.

Weiter kann die Prädikation natürlich auf verschiedenen Phrasenebenen realisiert sein:

- durch einen (elementaren) Satz (s.o.): dies ist die primäre Ebene der Beschreibung
- als Nominalphrase: *die Mündung der Isar in die Donau*
- (teilweise) in einer Wortform: *die Isarmündung in die Donau, der Donauzufluss, in der Kosovo-Hauptstadt Priština*

2.4 Räumliche Präpositionen

In der linguistischen Literatur werden räumliche Präpositionen häufig behandelt, häufiger noch als Bewegungs- oder Lokalisierungsverben. Eine Bibliographie gibt Breindl (o.J.), für eine Besprechung auch nur einzelner Arbeiten fehlt hier der Platz. Deshalb werden hier nur Themen angesprochen, die im Hinblick auf die syntaktische Beschreibung räumlicher Präpositionen – mit minimaler semantischer Klassifikation – relevant sind:

- Definition und Bestimmung von Adpositionen (Prä-, Post-, Zirkumpositionen, der Einfachheit halber hier meist „Präposition“ genannt) und Abgrenzung gegenüber anderen Wortarten (insbesondere Adverbien und Verbpunkteln)
- die Klassifikation (räumlicher) Präpositionen nach
 - syntaktischen Eigenschaften,
 - der Form, insbesondere die Unterscheidung einfach – komplex (Wortform – Phrase), und
 - dem semantischen Typ
- die Bildung von Präpositionalphrasen, also der Anschluss des regierten Nomens bzw. der nachgeordneten NP und die Modifikation der Präposition selbst
- die Präposition im Satz:
 - die syntaktischen Funktionen von Präpositionen oder Präpositionalphrasen
 - die Beziehung von Präpositionen zum Verb
 - die Verwendung von Präpositionen als Hauptprädikat im elementaren Satz mit einem Stützverb

Alle diese Fragen werden nachfolgend behandelt, oder auch im Rahmen der Beschreibung deutscher lokativer Präpositionen und Präpositionalphrasen (Kap. 3.5.2, 4.4 und C.1) wieder aufgegriffen.

Bestimmung und Abgrenzung

Präpositionen sind Relatoren mit mindestens zwei (semantischen) Aktanten, wovon mindestens einer syntaktisch als Dependens realisiert sein muss. Über diese Eigenschaft sind sich die Linguisten einig:

Prepositions and conjunctions are of course prototypically bivalent (rarely, trivalent)
(Mel'čuk 2004b: 54)

Präpositionen sind semantisch zweistellige Relatoren, die prototypisch Relationen zwischen Entitäten herstellen [...] Präpositionen sind syntaktisch einstellig; Breindl (2006: 937)

Präpositionen sind Relatoren; sie sind deswegen mindestens zweistellige Prädikate.
(Eroms 1981: 187)

Prepositional phrases are usually conceived of as two-place relations that establish a relation between the reference object in the PP and the content of the sentence (adverbial use) or the noun phrase (attributive use). Within the prepositional phrase, the preposition is the main carrier of the meaning of the relation.
(Grote 2001: 239)

Die Abgrenzung einer Wortart ‚Präposition‘ gegenüber anderen Wortarten ist aber im Detail schwierig und in vielen Fällen kaum möglich, bzw. erfordert Mehrfachzuordnungen (vgl. Schweisthal 1971: 15, Di Meola 2000, DUDEN2006: §896, Volk 2006: 84–5). Gegenüber folgenden Wortarten gibt es Grenzfälle bei der Abgrenzung:

- (2.49) Adverb: *abseits, außerhalb, jenseits, ...*
a. *Das Haus lag abseits der Straße.*
b. *Das Haus lag abseits.*
- (2.50) Adjektiv (Adverb): *nah, fern, nördlich, ...*
a. *Das Hotel Acropolis Select liegt nahe der Plaka von Athen.*
b. *das nahe Hotel*
c. *Das Hotel liegt nah.*
- (2.51) Substantiv: *Richtung, Ecke* (nicht lokative: *dank, ...*)
a. *Das Haus steht Ecke Tengstraße. Gegenüber liegt der "Alte Nördliche Friedhof".*
b. *Der Baustellenverkehr wird über die Ecke Heilbronner Str./Georg-Wilhelm-Str. abgewickelt.*
- (2.52) Verbpartikel: *durch, entlang, ...*; prädikativ verwendete „Präpositionen“
a. *Entlang der Straße stehen die kleinen Fischerhäuser auf Stelzen am Ufer.*
b. *Die Straße entlang stehen schnuckelige Häuser.*
c. *Wir fahren (die Straße) entlang. Wir sind (die Straße) entlanggefahren.*
d. *Die Kaffeemaschine ist an/aus.*
e. *Die Tür ist auf/zu.*
- (2.53) Konjunktion (keine lokativen Präpositionen betroffen)
a. *Während des Regens bleiben wir im Haus.*
b. *Während es regnet, bleiben wir im Haus.*
- (2.54) Verbformen: *während, entsprechend, ungeachtet* (keine lokativen Präpositionen)

Will man auf eine halbwegs scharfe Abgrenzung nicht verzichten, wie z.B. Meibauer (1995: 51–2), die nur einen prototypischen Kriterienkatalog angibt, muss zumindest der Versuch einer Abgrenzungen unternommen werden. Unproblematisch ist die Abgrenzung gegenüber Substantiven (vgl. Di Meola 2000: 93–101): es sind nur sehr wenige Einheiten betroffen (unter den Lokativa z.B. *Ecke, Richtung, Nähe*), so dass nichts gegen eine doppelte Klassifikation als Präposition und Substantiv bzw. Konjunktion spricht.

Schwieriger ist die Abgrenzung gegenüber Adverbien und Adjektiven. Prototypische Präpositionen haben mindestens einen syntaktisch abhängigen Aktanten, der auch immer realisiert sein muss. Bei Adverbien und Adjektive, wenn sie denn überhaupt einen Aktanten syntaktisch regieren, sind diese optional:

- (2.55) a. *Max wohnt in München.*
 b. *Max wohnt außerhalb Münchens.*
 c. *Max wohnt außerhalb.*
 d. **Max wohnt in.*
 e. *in und außerhalb Münchens*

Präpositionen und Adverbien unterscheiden sich nicht darin, wie sie ihren syntaktischen Aktanten anschließen. Beide lassen sich koordinieren (2.55e), wenn sie den Aktanten gemeinsam haben. Auch deshalb ist die Obligatheit des zweiten semantischen Aktanten ein schwaches Kriterium zur Abgrenzung zwischen Präposition und Adverb. Aufschlussreicher ist das Stellungskriterium, d.h. ob es möglich ist, die Präposition weg vom regierten Nomen zu schieben:

- (2.56) a. *Das Hotel liegt nur 2 Gehminuten vom Stadtzentrum entfernt.*
 b. *Vom Stadtzentrum liegt das Hotel nur 2 Gehminuten entfernt.*
 (2.57) a. *Das Hotel liegt fern von Hauptstrassen und ist doch so zentral, dass ...*
 b. ?* *Von Hauptstrassen liegt das Hotel fern und ist doch so zentral, dass ...*

Damit lässt sich *entfernt* als Adverb, *fern* als Präposition klassifizieren, vgl. auch (2.69). Aber zuletzt noch ein Gegenargument: Formen wie *rheinabwärts* inkorporieren einen Aktanten (=den Rhein abwärts), eine morphologische Eigenschaft, wie sie für Präpositionen untypisch ist, aber bei deutschen Adjektiven häufiger vorkommt: *cholesterinfrei* (=frei von Cholesterin), *abwehrbereit* (=bereit zur Abwehr). Das würde dafür sprechen, zumindest *abwärts* und andere Adpositionen – (*nil*)*aufwärts*, (*stadt*)*auswärts*, (*dorf*)*einwärts* – als Adverbien zu klassifizieren. Allerdings kann das Produkt der Inkorporation wiederum selbst als Präposition verwendet werden:

- (2.58) *Das romantische Rotweindorf Assmannshausen, wenige Kilometer rheinabwärts der Burg Ehrenfels, gehört heute zur Stadt Rüdesheim.*

Deutlich schwieriger gestaltet sich die Grenzziehung zwischen den drei Kategorien Präposition (insbesondere Zirkumposition), Direktionaladverb und Verbpartikel. Diese lassen sich durch Transformationen ineinander überführen (Beispiele z.T. aus Wunderlich 1985b und Olsen 1997):

- (2.59) a. *Sie bauen Hochhäuser um den Platz.*
 b. *Sie umbauen den Platz mit Hochhäusern.*
 (2.60) a. *Er will durch den Park fahren.*
 b. *Er will den Park durchfahren.*
 c. *Er will durchfahren.* (Präfix betont und abtrennbar)
 (2.61) a. *durch München hindurchfahren*

- b. *die Fahrt durch München hindurch*
- c. *... dann durch München hindurch, und immer Richtung Salzburg ...*
- d. *Durch die Stadt durch hatten sie die Autofahrer geleitet.*

In den Beispielen (2.61) ist es im Gesamtbild nicht entscheidbar, ob die Phrase *durch ... hindurch* eine selbständige Einheit (Konstituente) ist, oder *hindurch* als Verbpartikel kategorisiert werden muss (vgl. Olsen 1997: 20–1).

Lokative Zirkumpositionen

Lokative Zirkumpositionen sind häufig. Es existieren zwei Bildungsmodelle:

(2.62) *Prep ... Prep: vom Hauptbahnhof aus, von Norddeutschland aus*

(2.63) *Prep ... DirAdv: unter der Brücke hindurch, in das Tal der Mangfall hinunter*

Vor allem das zweite Bildungsmodell ist produktiv. Von den vielen Direktionaladverbien (*her, hin, (hin+her)(ab+an+auf+aus+durch+ein+unter), herum, hervor, hindurch, hinweg, entlang, weg, vorbei, vorüber*), die wiederum stark variiert werden können (*den Hang hinunter+herunter+runter+nunter, auf das Bett hinauf+herauf+drauf+rauf*), können einige selbst als Postpositionen verwendet werden (*den Berg hinauf/hinunter, den Weg hinein*). Di Meola (2000: 118–22) schlägt deshalb vor, diese Zirkumpositionen nicht als solche handzuhaben, sondern als „Präpositionalphrase mit nachgestelltem adverbialen Modifikator“. Fraglich ist allerdings, wie dann zu beschreiben ist, welche der möglichen 2 000 Kombinationen aus den knapp 100 „echten“ Präpositionen und den 20 Direktionaladverbien tatsächlich realisiert werden können. So lässt sich das Direktionaladverb *hindurch* nur mit drei räumlichen Präpositionen kombinieren: *durch den Tunnel hindurch, unter der Brücke hindurch, zwischen den Bäumen hindurch*, temporal auch *über Jahrzehnte hindurch*. Diese extremen Einschränkungen in der Kombinatorik müssen beschrieben werden, und dazu ist das Konzept der Zirkumposition gut geeignet.

Klassifikation (räumlicher) Präpositionen

Präpositionen lassen sich einmal nach ihrer **Form** klassifizieren, wobei die Unterscheidung einfach vs. komplex hier zentral ist. Komplexe Einheiten, also Einheiten die aus mehr als einer Wortform bestehen, werden in der Literatur vielfach als „Präpositionen“ oder „präpositionsartig“ beschrieben (Beneš 1974, G. Gross 1981, Constant 2003b). Inhaltlich sind die komplexen Einheiten gleichwertig zu „echten“ Präpositionen (2.64), mit anderen Präpositionen können sie koordiniert werden (2.65). Lexikalisierung und Idiomatisierung der komplexen Präpositionen sind fließend, insbesondere im Fall komplexer Präpositionen der Form *Pred (Det) N*, also mit einem Nomen als zentralem Bestandteil (vgl. Di Meola 2000: 102–18). Das zeigen eine Reihe von Tests, die versuchen, das Nomen außerhalb der präpositionsartigen Konstruktion zu verwenden (2.66)–(2.68). Aber auch die Semantik des Ausdrucks spielt eine Rolle, vgl. (2.66c).

- (2.64) a. *nahe der Straße*
- b. *≈ in der Nähe der Straße*

(2.65) a. *im Laufe und vor der Konferenz*

- b. *Seit der Bronzezeit nutzen Menschen das fruchtbare Land **am FuÙe und nrdlich** des Landrckens.*
- (2.66) a. *im Herzen Mnchens* (?das Herz Mnchens)
 b. *am FuÙe des Matterhorns* (?der FuÙ des Matterhorns)
 c. *am Ufer des Chiemsees* (\approx *am Chiemsee*, aber: *das Ufer des Chiemsees, das Chiemseeufer*)
 d. *in der Innenstadt von Paris* (kein komplexe Prposition!)
- (2.67) (2.66a') ? *Hier/Das ist das Herz.*
 (2.66b') ? *Hier/Das ist der FuÙ.*
 (2.66c') *Hier/Das ist das Ufer.*
 (2.66d') *Hier/Das ist die Innenstadt.*
- (2.68) a. *im Norden Mnchens, im Norden von Mnchen* (als Prposition)
 b. *in Mnchens Norden* (als Zirkumposition beschreibbar)
 c. *im Mnchner Norden* (nicht als Prposition beschreibbar)

Adpositionen sind (im Deutschen) durch die folgenden **syntaktischen Eigenschaften** charakterisiert:

- Adpositionen sind indeklinabel, wenn man von den kontrahierten Formen aus Prposition und bestimmtem Artikel absieht
- Stellung: Adpositionen werden nach ihrer Position zum Komplement unterteilt: Prpositionen stehen vor, Postpositionen nach, Zirkumpositionen vor und nach dem Komplement. Variationen zwischen Pr- und Postposition sind mglich, aber aus der Position unmittelbar vor oder nach dem regierten Nomen bzw. der NP, deren Kopf das Nomen ist, kann die Adposition nicht verschoben werden:

- (2.69) a. *Das Hotel Monaco liegt im Herzen von Mnchen.*
 b. * *Von Mnchen liegt das Hotel Monaco im Herzen.*

- Kasusreaktion: Prpositionen bestimmen den Kasus ihres Komplements
- Numerusreaktion: die Prposition *zwischen* regiert ein Nomen im Plural oder zwei oder mehr koordinierte Nomina: *zwischen den Bergen, zwischen der Zugspitze und der Alpspitze* (vgl. *zwischen* Habel 1989a, *entre* Guillet & Leclre 1992: 85)
- einige Prpositionen verlangen, dass das Nomen ohne Artikel steht: *bis Mnchen, via Bebra, Richtung Bayrischer Wald*

Details zu den syntaktischen Eigenschaften deutscher lokativer Prpositionen sind in Kap. C.1 zu finden.

Lokative Prpositionen lassen sich in semantische Typen untergliedern, je nachdem welchen Typ rumlicher Relationen oder welchen Teilaspekt einer Bewegung im Raum sie ausdrcken oder fokussieren (siehe Kap. 2.1). Im Deutschen sind die wichtigsten Klassen (vgl. Zwarts 2004), die sich an die auf S. 24 beschriebenen Modi einer Bewegung anlehnen:

- ‚statische Prpositionen‘: *im Haus, nrdlich von Mnchen* (auch ‚stativ‘, ‚lokal‘, ‚nicht indiziert‘)

- ‚direktionale Präpositionen‘ (auch ‚direktiv‘, ‚dynamisch‘, ‚transformativ‘, ‚vektorial‘), mit Ausdruck
 - des Ausgangspunktes (Source): *aus dem Haus, von/aus München* (auch ‚initial‘, ‚exogen‘)
 - des Endpunktes (Goal): *ins Haus, nach München* (auch ‚direktional‘, ‚final‘)
 - passierter Wegpunkte: *durch das Haus, durch/über/via München* (auch ‚vial‘)

Syntaktische Funktion und Realisierung

Präpositionen, oder genauer Präpositionalphrasen als Kombinationen von Präposition und deren syntaktischen Aktanten, übernehmen folgende syntaktischen Funktionen im Satz (vgl. Breindl 2006: 936):

- Präpositivkomplement (Präpositionalobjekt). Den drei Verben *warten (auf)*, *erinnern (an)*, *zusteuern (auf)* ist gemeinsam, dass die Präposition stark regiert³⁹ ist. Sie ist nicht oder sehr eingeschränkt variierbar und ist entsemantisiert, trägt also nichts (oder nicht eigenständig) zur Bedeutung des Satzes bei.
- Adverbialkomplement (adverbiale Ergänzung). Das Verb verlangt ein Komplement, das durch eine Präposition eingeleitet ist. Die Präposition ist aber paradigmatisch austauschbar und trägt zur Gesamtbedeutung des Satzes bei:

(2.70) *Max wohnt (in+bei+nördlich von) München.*

(2.71) *Das Konzert endet (um+gegen+nach) zehn Uhr.*

In dieser Funktion sind lokative oder temporale Präpositionen zu finden. Dabei kann das Verb einen bestimmten semantischen Typ (oder Modus, vgl. Kracht 2002: 54) unabhängig von der Semantik selektieren:

(2.72) a. *Max gelangt (nach+vor+in die Nähe) Freising.* (direktional final)

b. *Max kommt (in+vor+nahe+in der Nähe von) Freising an.* (statisch)

- Adverbialsupplement (freie adverbiale Angabe), als Satzadverbiale (2.73) oder adnominales Adjunkt (2.74) und (2.75)

(2.73) *Max arbeitet in München.*

(2.74) *ein See in Bayern*

(2.75) *das Haus in der Goethestraße*

Adverbialkomplemente wie -supplemente lassen sich durch Adverbien ersetzen: *Max (wohnt +arbeitet) hier, das Haus dort drüben*. Gleichfalls gilt nur für adverbial verwendete Präpositionen, dass die Präposition durch ein Adverb oder eine Maßphrase modifiziert werden kann (vgl. Breindl 2006: 942–3):⁴⁰

(2.76) a. *Das Kind spielt ganz dicht am Zaun.*

39 Zur Unterscheidung ‚stark regiert‘ vs. ‚schwach regiert‘ siehe Peškovskij (1956: 285–288) und Apresjan (1964).

40 Die Beschränkung ergibt sich aus der Tatsache, dass stark regierte Präpositionen entsemantisiert sind, und in der Folge auch nicht semantisch modifiziert werden können.

- b. *Das Kind stand ganz dicht am Zaun.*
- c. *Das Kind kam bis dicht an den Zaun.*
- d. **Ich dachte dicht an nichts.*

(2.77) *zehn Kilometer nördlich von München*

Die Kombinationsmöglichkeiten von Präpositionen mit Modifikatoren unterliegen starken Beschränkungen, eine Bestimmung über einige wenige semantische Merkmale von Präpositionen (z.B. Rauh 1996, 1997) ist nicht möglich. Das gilt auch für die These von Wunderlich & Herweg (1991), Bierwisch (1988), nur Richtungspräpositionen (*vor, rechts, unter, ...*) seien durch Maßangaben modifizierbar, topologische Präpositionen (*in, an, auf, bei*) dagegen nicht. Gegenbeispiele sind schnell gefunden:

(2.78) *7 km im Landesinneren auf der Höhe von Bjerge liegt der viertgrößte See Dänemarks.*

(2.79) *Das Phosphat wurde 300 km im Landesinneren von Mauretanien abgebaut und zu einem Anlegehafen an der Atlantikküste gebracht, wo es verschifft werden sollte.*

Klar ist aber auch, dass diese Modifikatoren nicht auf Ebene der Präposition *in* zu beschreiben, da nur die Kombination aus *in* und *Landesinnere* durch Maßangaben modifiziert werden kann (*?100 km in Deutschland*).

Typischerweise bestehen Präpositionalphrasen aus Präposition und Nomen/Nominalphrase (siehe Kap. 3.5.2). Präpositionen können aber auch mit Adverbien (*ab da, von oben, seit heute*) oder einer Präpositionalphrase selbst (*bis vor München, bis hinter die Grenze, für nach dem Essen, seit vor der Wende*) verbunden werden. Beide Fälle unterliegen starken kombinatorischen Beschränkungen und werden deshalb nicht als Präpositionalphrasen, sondern als komplexe Adverbien, komplexe Präpositionen oder als lexikalisierte Modifikatoren beschrieben.

Räumliche Präpositionen als Prädikat im elementaren Satz

Soll eine räumliche Präpositionen als Hauptprädikat im elementaren Satz auftreten, benötigt sie dazu ein Stützverb. Satzschema ist $N_0 V_{sup} Loc N_1$, siehe S. 47–48 und Kap. 2.5.1.4. Die deskriptive Herausforderung besteht darin, dass sowohl die Präposition *Loc* als das semantische, als auch das Verb V_{sup} als (oberflächen)syntaktisches Prädikat durch die beiden beteiligten Aktanten – lokalisiertes Objekt (Figure) und Bezugspunkt (Ground) – determiniert sein können. *Loc* ist Prädikat des Satzes und beschreibt, in welcher räumlichen Relation sich ein Objekt N_0 in Bezug auf einen Ort N_1 befindet. *Loc* kann semantisch variiert werden (2.80), kann aber als Wert der lexikalischen Funktion $Loc_{in}(N_1)$ determiniert sein (2.81):

(2.80) *Das Hotel liegt (in+nördlich+in der Nähe) der Stadt.*

(2.81) *Das Hotel liegt (auf+*in) der Insel.*

Die Auswahl des Verbs ist dagegen weitgehend durch N_0 bestimmt und kann nur bei Lebewesen und manchen beweglichen Gegenständen variieren – und ist dann (2.85) auch kein wirkliches Stützverb mehr. Dass Objekteigenschaften die Wahl von V_{sup} bestimmen oder beeinflussen können (Wunderlich 1985a: 41–2, Habel 1991: 605), und dass ein Zusammenhang zwischen V_{sup} und der Kombination mit Dimensionsadjektiven (Carstensen & Simmons

1991, vgl. (2.86)) besteht, trifft sicher zu, bedarf aber einer intensiveren Untersuchung und Beschreibung (vgl. auch Kap. 2.5.3.2).

- (2.82) *Freising liegt nördlich von München.*
- (2.83) *Der Eiffelturm steht in Paris.*
- (2.84) *Das Hotel liegt+steht in der Nähe des Strandes.*
- (2.85) *Max (ist+liegt+steht+sitzt) am Strand.*
- (2.86) *Der Eiffelturm ist 320 Meter hoch.*

Lokative Präpositionen im Zusammenspiel mit Bewegungsverben

Im Zusammenspiel mit Bewegungsverben übernehmen lokative Präpositionen die wichtige Funktion, Teile der komplexen Semantik einer Bewegung zu kodieren. Deutlich wird das, wenn man verschiedene Sprachen einander gegenüberstellt, wie Englisch und Bahasa in (2.87). In Bahasa erfüllen Sequenzen aus Verben die gleiche Funktion wie im Englischen die Kombination aus Verb und Präposition, was Becker & Arms (1969) zur Aussage „prepositions share many basic features with verbs and may be represented at a fairly abstract level of grammar as predicates“ verleitet.

- (2.87) a. *He ran to the river.*
- b. *Dia berlari sampai sungai.*
 He ran arrived river

Die Idee, Präpositionen als Prädikate oder Teile des Prädikats zu betrachten, wurde bereits 1824 von John Fearn formuliert (vgl. Eroms 1981: 185–7, Brekle 1976):

If we say ‘He comes to London’—‘He resides in England’, the Verbes ‘comes’ and ‘resides’ are Major Verbs, as being the Signs of the Grammatically Principal Actions asserted in these two Sentences; and the So-called Prepositions, ‘to’ and ‘in’ are Minor Verbs, as being the Signs of the dependent and defining Actions of finishing London and innng England
(Fearn 1824-1827: 274, nach Eroms 1981: 185–7)

Fearn begründet seine Überlegungen mit einer Reihe von Paraphrasen, z.B. *He goes to London = He goes ending London* oder *Figs come from Turkey to England = Figs come, – Turkey beginning the Coming, – The coming ending England*, (333,321), die allerdings eher der Glosse aus Beispiel (2.87) als einem englischen Satz ähneln.

Nimmt man Präpositionen als Teil einer komplexeren Prädikation an, stellt sich umgekehrt die Frage nach der Selbständigkeit der Präposition. In Bezug auf die deutschen direktionale Präpositionen und die Präpositionen mit doppeltem Kasus, wird vielfach behauptet, dass diese nur mit Bewegungsverben vorkommen und die direktionale Lesart erst im Zusammenwirken mit dem Verb zustande komme“ (Eroms 1981: 200, siehe auch Abraham 2003).

Fakt ist, dass auch Bewegungsnomina (*Fahrt*) und Quasiprädikate (*Straße, Fähre, Zug*) Ak-tanten über direktionale Präpositionen anschließen können. Zum anderen können direktionale Präpositionalphrasen, wenn sie Teilstrecken benennen, temporal verwendet werden: *Max a dormi de Lyon à Gap* (Guillet & Leclère 1992: 79). Und schließlich ist es möglich, direktionale Präpositionen in Wegbeschreibungen ganz ohne Verb zu verwenden:

- (2.88) *Vom Rathaus Sigmaringen Richtung Bahnübergang, direkt vor dem Bahnübergang links, auf dem befestigten Weg über die Donaubrücke und dann links leicht bergab auf den Donauradweg, dort rechts weiter bis zum Kiosk Bootshaus mit Teich und Spielplatz (km 1), dort rechts am Spielplatz vorbei, Straße geradeaus **überqueren** und geradeaus zw. Autohaus Zimmermann und Autohaus Keller **fahren**, nach ca. 200 Meter rechts unter schmaler Unterführung hindurch immer geradeaus weiter, leicht bergan, vorbei an der Brauerei Zoller-Hof bis Straße eine Rechtskurve beginnt, dort links unter einer Straße hindurch (km 1,5), dann auf meist gut befestigtem Weg, überwiegend bergan abwechslungsreich bis ca. km 4. Dort, im Bereich des unbefestigten Parkplatzes bei einer Gaststätte links weiter mehrere hundert Meter bergan auf befestigtem Weg, dann rechts unter der Straße hindurch und gleich wieder rechts auf meist befestigtem Weg, überwiegend bergab in Richtung Jungnau. An einem weißen Kreuz (km 6,5) links weiter, teilweise steil bergan bis km 7,5, im Bereich einer Wegquerung. Dort rechts bergab Richtung Ortsmitte Jungnau (Vorsicht rechts vor links-Regelung), im Ort (am Briefkasten) rechts, der Beschilderung Radweg R. Veringenstadt **folgen**, Hauptstraße geradeaus **überqueren**, weiter geradeaus über Bahnübergang, rechts und dann links weiter, auf befestigter Straße, überwiegend bergan Richtung Hochberg durch das sog. Geißental. (Vorsicht Kfz-Verkehr). Wenn auf der Albhochfläche die Häuser von Hochberg zu sehen sind, bei einem Kreuz nach rechts auf überwiegend befestigtem Weg, mal kurz bergan, aber meist bergab, immer auf dem erkennbaren Hauptweg (geradeaus) **bleiben**. . . .*

In dieser keineswegs kurzen Wegbeschreibung sind nur fünf Verben zu beobachten, der Ausdruck der Route erfolgt meist durch Adverbien und Präpositionen. Auch wenn man jeden der Sätze über eine Elision eines finiten (Bewegungs)Verbs herleiten kann: direktionale Präpositionen implizieren bereits eine Bewegung, insofern ist die Überlegung der Lexikongrammatiker (vgl. (2.93b) und (2.94d) auf S. 47) auch semantisch minimale Bewegungsverbene als Stützverben zu betrachten, nicht ganz abwegig. Der Ausdruck einer Bewegung ist natürlich nicht semantisch leer, aber Verben wie *sich begeben* oder *gelangen* verdoppeln (siehe Fußnote 35 auf S. 29) nur Bedeutungskomponenten, die auch durch die direktionale Präposition ausgedrückt werden.

2.5 Lokative Verben

2.5.1 Im Rahmen der Lexikongrammatik des LADL

Im Rahmen der Lexikongrammatik des LADL wurden viele lokative Verben klassifiziert (Gross 1975, Boons, Guillet & Leclère 1978, Guillet & Leclère 1992), wobei viel Mühe auf die genaue, operationalisierbare Bestimmung und Klassifikation verwandt wurde (ibid. und Boons 1985). Der Ansatz der Lexikongrammatiker des LADL wurde von Volker Caroli (Caroli 1984a,b, Caroli & Figge 1989) auf deutsche Verben übertragen (siehe Kap. 2.5.1.2).

2.5.1.1 Französische Verben

2.5.1.1.1 Kriterien und Verfahren zur Klassifikation

Als jeweils ausreichende, aber nicht notwendige Bedingungen für ein Ortskomplement und die Lokativität eines Verbs werden von den französischen Lexikongrammatikern folgende Kriterien vorgeschlagen:

1. die Abfragbarkeit des lokativen Arguments durch das Pronomen *où* ‚wo‘ (Boons 1985: 196–200, Boons, Guillet & Leclère 1976: 194–7, Guillet & Leclère 1992: 11–4, Gross 1995a: 176–7), bzw. analog analog dazu die Pronominalisierung oder Ersetzung durch *y* ‚dort‘ (Gross 1995a: 176):

- (2.89) a. *Le pot est caché (sur + sous + dans) le lit*
 b. *Le pot y est caché*
 c. *Où est caché le pot ?*

Allerdings können durch *où* nicht nur lokative Argumente, sondern auch andere Komplemente (hier satzwertige ‚Kompletive‘) abgefragt werden (Boons, Guillet & Leclère 1978: 203–4, Guillet & Leclère 1992: 13).

- (2.90) a. *Où vas-tu ? – Chercher du pain.*
 b. *D’où reviens-tu ? – De chercher du pain.*

Deshalb muss der Fragetest auf Konkreta beschränkt bleiben: „Nous avons en conséquence limité notre étude aux compléments *Loc N* où *N* est un substantif concret“ (Boons, Guillet & Leclère 1978: 203–4).

2. die Variabilität der lokativen Präposition,⁴¹ siehe Beispiel (2.89) (Boons 1985: 202, Boons, Guillet & Leclère 1976: 203–6, Gross 1995a: 176, 189–91, Gutermann 1996: 141). Dabei ist die Wahl der Präposition nicht allein durch das Verb determiniert, sondern auch von weiteren Parametern abhängig:

Rappelons que la préposition *Loc* varie avec le verbe:

- Max charge l’échelle sur le camion*
 ?* *Max charge l’échelle contre le camion*
Max pose l’échelle sur le camion
Max pose l’échelle contre le camion

et, pour un verbe donné, *Loc* varie avec les distributions de N_1 et N_2 :

- Max pose les oranges dans la boîte*
Max pose l’échelle contre le mur
 ? *Max pose le paquet sous le lit*
Max a envoyé Luc (dans le Midi, au sud de la France, en Méditerranée)

Notre position sur l’analyse des prépositions de lieu peut se résumer de la façon suivante:

- chaque *Loc* possède des éléments de sens particuliers (sa sélection) ;
- chaque verbe en relation avec *Loc N_i* induit des interprétations spéciales des *Loc* ;
- dans le cas de *dans*, *sur*, *sous*, *contre* au moins, la géométrie particulière de chaque N_i apporte une contribution spéciale au sens des phrases.

(Guillet & Leclère 1992: 39–40)

41 Die Präposition kann auch ausfallen: *Luc habite (E + à) Paris* (Gross 1995a: 190, 198), was entsprechend als Merkmal *Prép = E* in den Tabellen (z.B. 35L, 35ST) kodiert wird.

Dass der lokative Aktant (*Nieu*, *Nloc*) als Dependent der Präposition diese determinieren kann, wurde bereits anlässlich der lexikalischen Funktionen *Loc_{in}*, *Loc_{ab}*, *Loc_{ad}* (S. 29) angesprochen. Die Beispiele für den Einfluss des lokalisierten Objekts (*Figure*) sind jedoch zweifelhaft: (1.) bei der Übersetzung ins Deutsche müssen verschiedene Verben gewählt werden: *Max legt die Leiter auf den Wagen* vs. *Max lehnt die Leiter an den Wagen / an die Wand*. (2.) semantisch hat *poser* in der Bedeutung ‚lehnen‘ einen zusätzlichen Aktanten: die Leiter hat ja Kontakt mit dem Boden und Wand/Wagen und wird auch von beiden in ihrer Position gehalten.

3. es muss ein elementarer Satz vorliegen (Boons 1985:202–5), wodurch Komplemente von Supplementen unterschieden werden. Guillet & Leclère (1992:14–7) verwenden auch die Begriffe ‚compléments locatifs de verbe‘ und ‚compléments locatifs de phrase‘. In

(2.91) *Max est tombé dans une mare à la campagne*

ist *dans une mare* Komplement des Verbs, und *à la campagne* ‚Komplement der Phrase‘ oder ‚complément scénique‘. In Operator-Argument-Schreibweise würde (2.91) entsprechend durch zwei Prädikationen repräsentiert und ist damit kein elementarer Satz.

(2.91') avoir lieu(tomber(Max, dans une mare), à la campagne)

4. die Explikation der räumlichen Relation durch einen einfacheren Satz (Guillet 1984, Boons 1985:206–9, Guillet & Leclère 1992:22–7,61–78, vgl. Caroli & Figge 1989:244–6):

- (2.92) a. *Max a enfoncé le clou dans le mur*
 b. *Le clou est dans le mur*

Der vereinfachte Satz enthält also nur noch die NPs von *Figure* und *Ground*, die räumliche Präposition und ein stützverbartiges Verb. Für Bewegungs- und Transportverben ist es auch möglich, das Verb durch ein weniger spezifisches Bewegungs- oder Transportverb („verbe-support dynamique“) zu ersetzen ((2.93b) und (2.94d)), oder den Zustand vor und nach der Verbalhandlung zu betrachten ((2.94a)–(2.94c)).

- (2.93) a. *Max marche vers la tour*
 b. *Max va vers la tour*

- (2.94) a. *Max fährt zur Schule.*
 b. *Max ist nicht in der Schule.* (Zustand vor dem Prozess)
 c. *Max ist in der Schule.* (Zustand nach dem Prozess)
 d. *Max gelangt zur Schule.* (Caroli & Figge 1989:244–6)

Durch die Explikation der räumlichen Relation werden selbst Verben wie *manger* ‚essen‘ zu lokativen Verben,⁴² obwohl z.B. die Abfrage mit *où* (2.95b) scheitert:

42 Verben wie *manger* stehen deshalb in Tabelle 38L0 (Guillet & Leclère 1992:192–202). Auch Rogets Thesaurus (1852), einer der älteren Versuche einer hierarchischen Untergliederung des Wortschatzes, ordnet *eat/food* ein unter: ‚space in general‘ > ‚motion‘ > ‚motion with reference to direction‘ > ‚food‘. Zugegeben *manger* teilt sich viele Eigenschaften mit lokativen Verben, was auch in Paraphrasen wie *Max verspeist ein Törtchen* → *Max schiebt sich ein Törtchen in den Mund* deutlich wird. Aber der semantische Kern von *manger* ist sicher nicht, dass ein Objekt in den Mund/Magen „wandert“ – zu wichtig sind die weiteren Bedeutungsaspekte.

- (2.95) a. *Max a mangé sa soupe dans une assiette bleue*
 b. *Où Max a-t-il mangé sa soupe ? – *Dans une assiette bleue*
 c. *Sa soupe est dans une assiette bleue* (Boons 1985: 198, 206, 208)

Im Zusammenhang mit dem Explikationstest steht die Unterscheidung von vier Typen von Komplementen, und zwar je nachdem, ob die Explikation mit *être* zeitlich vor, während und nach der Handlung möglich ist (Boons 1985: 226–232, Guillet & Leclère 1992: 22–33, 61–78):

- Ausgangspunkt/Source (,complément locatif initial‘): Explikation mit *avant* ‚vor‘: *Max quitte la chambre* ‚Max verlässt das Zimmer‘
- Zielpunkt/Goal (,destination‘, ,complément locatif final‘): Explikation mit *après* ‚nach‘: *Max entre dans la chambre* ‚Max betritt das Zimmer‘
- Wegpunkt(e) (,trajet‘, ,médián‘): Explikation mit *pendant* ‚während‘: *Max passe par la chambre* ‚Max durchquert das Zimmer‘
- statische Ortsangaben (,statique‘): eine Paraphrase/Explikation mit *avant* und *après* ist nicht möglich (Boons, Guillet & Leclère 1978: 235)

Nun stellt sich die Frage, ob es Zusammenhänge zwischen den vier Typen und dem Argumentstatus (Komplement oder Supplement) eines lokativen Bestandteils in einem Satz besteht. Boons (1985: 232) wagt hier eine angeblich empirisch nicht widerlegte Hypothese:⁴³

1. Tout complément locatif initial ou final est nucléaire.
2. Tout complément locatif apparaissant dans un emploi non statique relève d’une relation locative qui est soit initiale, soit médiane, soit finale.
3. Tous les compléments circonstanciels d’emplois non statiques sont médians.

Damit sind Ausgangs- und Zielpunkt immer Komplemente, während für statische Ortsangaben und (perlative) Wegpunkte der Argumentstatus getestet werden muss. Nun kritisiert Gutermann (1996: 145) Boons’ Hypothese zurecht, da auch Kommunikationsverben (*Er brüllte* ‚Feuer‘ in den Saal) oder Wahrnehmungsverben (*in den Spiegel blicken*) „direktionale Bestimmungen“ annehmen können und damit in der Logik Boons’ zu lokativen Verben würden.⁴⁴

43 „L’expérience semble montrer que rien ne contredit la validité empirique du système formé par les règles“ (Boons 1985: 233)

44 Die lokativen „Komplemente“ der Kommunikationsverben resultieren aus dem Raising einer Ortsangabe, die einen der Kommunikationspartner modifiziert in die Position einer Direktionalergänzung (1), ev. verbunden mit einer Tilgung des Kommunikationspartners (3). Hier einige Beispiele:

- (1) *Emanuel Geibel ... telegraphierte an seinen Freund Rittershaus in Barmen: ...*
- (2) *Kortner telegraphiert mir nach Wien: ...*
- (3) *Auf dem Berggipfel ist ein Wetterwart stationiert, der seine Messungen täglich ins Tal telegraphiert.*
- (4) *So telegraphierte der US-Diplomat Herschel Johnson am 31. August 1943 aus Stockholm, Brandt habe "enge Kontakte zur sowjetischen Gesandtschaft".*
- (2') *Wohin telegraphiert Kortner?*

Die Abfrage mit *wohin/woher* ist möglich (2'), ebenso wie im Französischen:

- (5) a. *Il a télégraphié à Washington.*
 b. *Où a-t-il télégraphié ? – À Washington.*

Tab.	Schema	Beispiel	Kommentar	Qu
2	$N_0 V (Loc N) V_{inf} W$	<i>Paul court (à l'épicerie) cherche du sel</i>	Bewegungsverben, Verben der Ortsveränderung mit satzwertigem Komplement	[1]
3	$N_0 V N_1 (Loc N) V_{inf} W$	<i>Paul envoie Luc (au village) faire le marché</i>	Verben mit satzwertigem Komplement, Ortsveränderung von N_1	[1]
34L0	$N_1 V de N_0$	<i>Le toit dégouline d'eau (src)</i> <i>La planche se couvre de poussière (goal)</i> <i>Cette ville grouille d'espions (stat)</i>	Resultat der Transformation „The garden is swarming with bees“ aus $N_0 V Loc N_1$ (vgl. 35 ST)	[2]
35ST	$N_0 V Loc N_1$	<i>Luc habite dans cette ville</i>	N_1 = statische Ortsangabe	[2]
35L	$N_0 V Loc N_1$	<i>Max sort de la chambre (src)</i> <i>Max passe par la chambre (trans)</i> <i>Max entre dans la chambre (goal)</i>	N_1 = direktionale Ortsangabe	[2]
36SL	$N_0 V N_1 (Loc + avec + et) N_2$	<i>Max a collé le carton sur la feuille</i>	Symmetrie der Objekte N_1 und N_2	[3]
37E	$N_0 V N_1 de N_2src$	<i>Max débarrasse la table des livres</i>		[3]
37M	$N_0 V N_1 de N_2dest$	<i>Max couvre la table de livres</i>		[3]
38L	$N_0 V N_1 Loc N_2src Loc N_3dest$	<i>Luc transvase le vin de la cuve dans le tonneau</i>	N_2 =: source, N_3 =: destination	[3]
38LD	$N_0 V N_1 Loc N_2dest$	<i>Max pose le verre sur la table</i>		[3]
38LS	$N_0 V N_1 de N_2src$	<i>Max enlève le verre de la table</i>		[3]
38LH	$N_0 V N_1hum Loc N_2$	<i>On a rapatrié Luc de Goa (src)</i> <i>On a hospitalisé Max à Cochin (goal)</i>		[3]
38LR	$N_0 V N_1 Loc N_2$	<i>Max ajuste la mire sur la tour</i> <i>Max a scellé le piton dans le mur</i>	Résiduels: N_2 nicht als Source oder Goal interpretierbar	[3]
38L0	$N_0loc V N_1$	<i>Le métro engouffre les voyageurs (src)</i> <i>Le volcan crache de la lave (src/Partitiv)</i> <i>L'éponge absorbe l'eau (goal)</i>		[3]
38L1	$N_0 V N_1loc$	<i>Le train quitte la gare (src)</i> <i>Max traverse la chambre (trans)</i> <i>La fumée envahit la pièce (goal)</i> <i>Max visite la ville (process)</i>		[3]

Tabelle 2.2: Übersicht über die Klassen von Verben mit lokativen Komplementen in der französischen Lexikongrammatik des LADL (vgl. Leclère 2002: 41–6). Quellen: [1] = Gross (1975); [2] = Boons, Guillet & Leclère (1976); [3] = Guillet & Leclère (1992).

2.5.1.1.2 Klassen französischer lokativer Verben

Gross (1975: 165–9, 237–44) behandelt nur Verben mit satzwertigem Komplement, darunter finden sich aber zwei Gruppen, die auch lokative Komplemente zulassen: 130 Bewegungsverben bzw. Verben der Ortsveränderung (Tabelle 2) und 44 Verben, die eine Bewegung oder Ortsveränderung kausieren (Tabelle 3). Beiden gemeinsam ist, dass sie eine Infinitivkonstruktion regieren.

- (2.96) *Il court voir Marie.*
,Er läuft, um Marie zu sehen.‘
- (2.97) *Pierre envoie Paul voir Marie.*
,Pierre schickt Paul, damit er Maria sieht.‘

Als Kriterium zur Abgrenzung gegenüber anderen Verben mit Infinitivkonstruktionen bzw. den Modalverben (*pouvoir*, *vouloir* etc.), die auch einen Infinitiv fordern können, dient Gross der Fragetest mit *où* (s.o.):

- (2.98) *Où court-il ?*
Voir Marie.
- (2.99) *Où Paul envoie-t-il Pierre ?*
Voir Marie.

Transitive Bewegungsverben wie *traverser* ‚durchqueren‘ behandelt Gross als Untergruppe der Tabelle 2. Für sie fällt der Fragetest negativ aus:

- (2.100) *Paul franchit la rivière retrouver Marie.*
,Paul überquert den Fluss, um Marie wiederzufinden.‘
- (2.100′) * *Où Paul franchit-t-il la rivière ?*

Präpositional kodierte Ortsargumente zusätzlich zur Infinitivergänzung sind für die meisten Verben möglich. Ihre Realisierbarkeit, ebenso wie die Form der Realisierung (durch eine bestimmte Präposition, *ici* oder *là*) sind als Merkmale in den Tabellen kodiert.

Boons, Guillet & Leclère (1976) beschreiben intransitive⁴⁵ Verben, darunter auch eine Reihe lokativer Verben.

Transitive lokative Verben sind dagegen Gegenstand von Guillet & Leclère (1992). Die Klassifikation basiert dabei auf der Überlegung, dass das lokalisierte Objekt (Figure), der Ort der Lokalisierung (Ground) und ein mögliches drittes kausatives oder agentives Argument in den syntaktischen Position Subjekt, direktes Objekt oder Präpositionalobjekt realisiert werden können. Das ergibt sechs lokative Konstruktionen – nicht alle sind transitiv (Tab. 2.3). Jede Klasse lässt sich wiederum in drei Subklassen unterteilen, in Abhängigkeit von der Zuordnung des Ground durch den Explikationstest zu einem der Typen Ausgangs-, Weg-, oder Zielpunkt. Ergebnis sind 18 Subklassen, von denen aber nicht alle als elementarer Satz belegt sind (Tab. 2.4).

Zu den Klassen aus Tab. 2.3 und 2.4 kommt noch Tabelle 38L für vierwertige Verben, in denen sowohl Ausgangs- wie Zielpunkt als Komplemente akzeptiert sind. Wegpunkte werden

45 Für die Unterscheidung transitiv vs. intransitiv, siehe Boons, Guillet & Leclère (1978: 59–168).

		Ground					
		Subjekt		dir. Obj.		Präp.obj.	
		src	goal	src	goal	src	goal
	Subjekt	—		38L1		35L	
Figure	dir. Obj.	38L0		—		38LS	38LD
	Präp.obj.	34L0		37E	37M	—	

Tabelle 2.3: Einteilung der transitiven lokativen Verben nach der syntaktischen Realisierung von lokalisiertem Objekt (Figure) und Ort der Lokalisierung (Ground) als Subjekt, direktes Objekt und Präpositionalobjekt, und dem semantischen Typ (Ausgangs- oder Zielpunkt) des Ground als dritter Dimension (nach Guillet & Leclère 1992: 28, Leclère & Brisbois-Lenehardt 2004: 395)

<i>Narg V Nlieu</i>	av	<i>Max quitte la chambre</i>
	pe	<i>Max traverse la chambre</i>
	ap	<i>La fumée envahit la pièce</i>
<i>Narg V Prép Nlieu</i>	av	<i>Max sort de la chambre</i>
	pe	<i>Max passe par la chambre</i>
	ap	<i>Max entre dans la chambre</i>
<i>Nlieu V Narg</i>	av	<i>La cheminée crache de la fumée</i>
	pe	<i>La gouttière canalise l'eau</i>
	ap	<i>L'éponge absorbe l'eau</i>
<i>N₀ V Narg Loc Nlieu</i>	av	<i>Max enlève la verre de la table</i>
	pe	<i>Max traîne la caisse sur le sol</i>
	ap	<i>Max pose le verre sur la table</i>
<i>Nlieu V de Narg</i>	av	<i>Le vase déborde d'eau</i>
	pe	<i>Le jardin grouille d'abeilles</i>
	ap	<i>(La table se couvre de livres)</i>
<i>N₀ V Nlieu de Narg</i>	av	<i>Max débarasse la table des livres</i>
	pe	<i>(Max parcourt la page du doigt)</i>
	ap	<i>Max couvre la table de livres</i>

Tabelle 2.4: Mögliche Satzchemata lokativer Konstruktionen, mit kausativem oder agentivem N_0 , zu lokalisierendem Objekt *Narg* und Ort *Nlieu*, jeweils unterteilt durch die Explikation der lokativen Relation vor (avant = Ausgangspunkt), während (pendant = Wegpunkt) und nach (après = Zielpunkt) der Verbalhandlung (aus Guillet & Leclère 1992: 31, vgl. Tab. 2.3)

als Aktanten akzeptiert, erhalten aber keinen eigenen Index, sondern werden als zweiter Zielpunkt indiziert:⁴⁶

Par N₃ :

Cette colonne note la possibilité d'un complément de rang 3, distinct de la source et de la destination, et qui désigne un lieu de passage :

Max projette la valise du salon dans le jardin par la fenêtre

(Guillet & Leclère 1992: 226)

Eine Bewegung impliziert einen Pfad mit Ausgangs- und Zielpunkt. Syntaktisch müssen aber nicht immer beide realisiert oder realisierbar sein. Mit dieser Fragestellung beschäftigen sich Guillet & Leclère (1992: 80–1; vgl. Boons, Guillet & Leclère 1978: 222–4, Gross 1995a: 180,191) und kommen zum Schluss, dass es vom Verb abhängig ist, welche Kombinationen erlaubt sind: Ausgangspunkt oder Zielpunkt allein, nur beide in Kombination, oder eine Kombination dieser drei Varianten (Tab. 2.5). Von den sieben möglichen Kombinationen sind zwei nicht belegt.

	O/D	O	D		
<i>descendre</i>	+	+	+	A	‚hinuntergehen‘
<i>poser</i>	+	-	+	B	‚stellen, legen‘
?	+	+	-	C	
<i>promener</i>	+	-	-	D	hier: ‚(den Scheinwerfer) führen‘
?	-	+	+	E	
<i>enterrer</i>	-	-	+	F	‚eingraben, vergraben‘
<i>retirer</i>	-	+	-	G	‚zurückziehen‘

Tabelle 2.5: Realisierungsmöglichkeiten von Ausgangspunkt (Origine) und (Destination) oder beiden gemeinsam (O/D) bei französischen Verben (Guillet & Leclère 1992: 81)

Fazit

Die Klassen der französischen Lexikongrammatik – nicht nur die der lokativen Verben – sind aus semantischer Perspektive sehr heterogen: „les classes de prédicates sont souvent hétérogènes du point de vue du sens“ (Le Pesant & Mathieu-Colas 1998: 9). Gerade weil die Kriterien empirisch festgelegt werden und sich dabei auf überwiegend syntaktische Eigenschaften stützen, werden Verben als lokativ klassifiziert, deren Semantik wenig oder keine lokativen/räumlichen Bestandteile enthält: ‚buttern, mit Butter bestreichen‘ (37M5: *Max beurre le pain de beurre salé*), das bereits besprochene *manger*, aber auch ‚übersetzen‘ (38L: *Max transpose ce texte du latin en grec*) oder ‚abstrahieren, lösen‘ (*Max abstrait son esprit des problèmes actuels*), die sich zwar vielleicht über lokative Metaphern herleiten lassen, aber eindeutige Nicht-Lokativa sind.

46 Vgl. S. 16 zur These der Lexikongrammatiker, kein Verb habe mehr als vier Aktanten.

2.5.1.2 Deutsche Verben

Deutsche Verben wurden bislang nur wenig im Rahmen der Lexikongrammatik behandelt. Treig (1977) behandelt in Anlehnung an Gross (1975) Verben mit satzwertigen Komplementen. Caroli (1984a,b) und Caroli & Figge (1989) beschäftigen sich mit der Klassifikation deutscher Verben mit Lokalgängung. Leider war es nicht möglich, die erarbeiteten Tabellen von den Autoren zu erhalten, nur ein Teil ist in Caroli (1984b) publiziert, eine weitere Tabellen, die Transportverben enthält und sich teilweise mit Caroli (1984b) überschneidet, findet sich ohne ausdrücklichen Zusammenhang zur Thematik in Caroli et al. (1990).

Caroli (1984a) gründet seine Klassifikation deutscher Bewegungsverben auf die Anzahl der Komplemente und die Art ihrer Realisierung. Die Realisierung beschreibt er anhand der zwei Parameter: (1.) mit/ohne Präposition, (2.) Kasus, in dem das Komplement steht. Auch wenn er ausschließlich Bewegungsverben klassifiziert, reichen die beiden Parameter nicht aus, um Ausgangs- und Endpunkt der Bewegung eindeutig zu bezeichnen. Sowohl *aus dem Haus*, als auch *zum Arzt* wird auf *Prep N_{dat}* reduziert (1984a:25). *Herein*, *heraus*, *entlang* und andere Richtungsadverbien behandelt Caroli als solche und nicht, wie vielerorts üblich als abtrennbares Verbalpräfix. Damit ergeben sich folgende Klassen:

$N_0 V$	<i>Max schwimmt.</i>
$N_0 V N_{akk}$	<i>Max schwimmt hundert Meter.</i>
$N_0 V N_{akk} Adv$	<i>Max schwimmt hundert Meter weit.</i>
$N_0 V P N_{akk}$	<i>Max fährt in den Hof.</i>
$N_0 V P N_{dat}$	<i>Max geht zum Arzt. (Endpunkt)</i> <i>Max geht aus dem Haus. (Ausgangspunkt)</i> <i>Max schwimmt im See. (Ort der Bewegung)</i>
$N_0 V P N_{akk} Adv$	<i>Max fährt in die Garage hinein. (Endpunkt)</i> <i>Max fährt über die Brücke hinüber. (Wegpunkt)</i> <i>Max fährt um das Haus herum. (Wegpunkt)</i>
$N_0 V P N_{dat} Adv$	<i>Max fährt aus der Garage heraus.</i>
$N_0 V N_{akk} Adv$	<i>Max fährt die Straße entlang/hinunter. (Wegpunkt)</i> <i>Max fährt den Berg hinauf. (Endpunkt)</i>
$N_0 V P N_{dat} P N_{akk}$	<i>Das Glas fällt vom Tisch auf den Boden.</i>

Anzumerken ist, dass eine Schematisierung der Form ' $P N_{(Kasus)}$ ' dem zu beschreibenden Sachverhalt nicht gerecht wird. Zum einen ist die Kasusforderung auch eine Eigenschaft der Präposition und damit nur im Fall der Präpositionen mit doppeltem Kasus relevant für die Beschreibung des Verbs. Zum anderen ist die semantische Typisierung der Ortsargumente in bewegtes Objekt (N_{arg}), Ausgangs- und Endpunkt der Bewegung bzw. statisches Ortsargument notwendig und sinnvoll. Für die (später erfolgte) Klassifikation der entsprechenden französischen Verben ist sie maßgebliches Kriterium.

Caroli (1984b) beschränkt sich auf die Klassifikation und tabellarische Beschreibung transitiver Verben mit Lokalgängung, also Verben in Sätzen des Schemas $N_0 V N_{1arg} Loc N_{2lieu}$:

- (2.101) a. *Max zieht eine Zigarette aus der Schachtel.*
 b. *Max legt das Buch auf den Tisch.*
 c. *Max gießt Wasser über den Tisch.*
 d. *Max schreibt eine Nachricht auf den Zettel.*

e. *Max mariniert den Braten in einer Rotweinbeize.* (Caroli 1984b: 234)

Beschreibungskriterien der Verben sind die Form und semantische Rolle der Lokalgängung, die Tilgung von Aktanten, Passiv- und Reflexivtransformationen, sowie mögliche Derivationen des Verbs. Für die Form der Lokalgängung unterscheidet Caroli neun Typen, davon sieben als semantische Rollen (Caroli 1984b: 255–7):

LI Komplement bezeichnet Ausgangspunkt der Bewegung (,lieu initial‘)

LF Komplement bezeichnet Endpunkt der Bewegung (,lieu final‘)

LS Komplement betont die räumlichen Eigenschaften des lokalisierten Objekts (*Narg*) und der Lokalgängung:

(2.102) *Max gießt Wasser über den Tisch.*

LOE Komplement kann sowohl im Dativ (statisch) als auch im Akkusativ („Endpunkt“) realisiert sein (,lieu de l’objet effectué‘ = „effizientes Objekt“ bei Caroli & Figge 1989: 253):

(2.103) a. *Max baut ein Haus auf das Grundstück.*
 b. *Max baut ein Haus auf dem Grundstück.*

LM Komplement kann auch als Werkzeug oder Mittel interpretiert und mit der Präposition *mit* realisiert werden (,lieu-moyen‘):

(2.104) a. *Max schöpft Wasser in einem Eimer.*
 b. *Max schöpft Wasser mit einem Eimer.*

LPC Komplement bezeichnet Körperteil (,lieu partier du corps‘); die Transformation $N_0 V N_{1_{akk}} Loc N_{2_{akk}} \rightarrow N_0 V Loc N_{2_{akk}} N_{1_{gen}}$ ist möglich:

(2.105) a. *Der Hund beißt den Briefträger in das Bein.*
 b. *Der Hund beißt in das Bein des Briefträgers.*

Ist das Komplement ein Körperteil, sind zwei weitere Transformationen als Beschreibungskriterium relevant:

$N_{1_{akk}} = N_{1_{dat}}$ Das direkte Objekt wird vom Akkusativ in den Dativ (,Pertinenzdativ‘) gesetzt:

(2.106) a. *Der Hund beißt den Briefträger in das Bein.*
 b. *Der Hund beißt dem Briefträger in das Bein.*

$Loc N_{2_{dat}} = N_{2_{akk}}$ Das direkte Objekt wird in den Dativ gesetzt, das Ortsargument wird zum Akkusativobjekt:

(2.107) a. *Max streichelt die Katze am Kopf.*
 b. *Max streichelt der Katze den Kopf.*

LST Komplement bezeichnet statischen Ort (,lieu statique‘)

Ein weiterer Punkt in Carolis Beschreibung ist die Tilgung von Aktanten, die Caroli durch das Schema des verbleibenden Satzes markiert:

- (2.108) a. *Max bohrt ein Loch in das Brett.*
 b. *Max bohrt ein Loch.* ($N_0 V N_{1_{akk}}$)
 c. *Max bohrt in das Brett.* ($N_0 V Loc N_2$)
 d. *Max bohrt.* ($N_0 V$)
- (2.109) a. *Max klebt ein Plakat an die Wand.*
 b. ? *Max klebt ein Plakat.*
 c. * *Max klebt an die Wand.*
 d. ? *Max klebt.* (Caroli 1984b: 239)

Die Tilgung des Ortsarguments kann an die Präfigierung des Verbs gebunden sein, wobei die getilgte Präposition das Präfix darstellt:

- (2.110) a. *Max klebt den Zettel* $\left\{ \begin{array}{l} \textit{an die Wand} \\ \textit{auf die Schachtel} \end{array} \right\}$
 b. *Max klebt den Zettel* $\left\{ \begin{array}{l} (\textit{an die Schachtel}) \textit{an} \\ (\textit{auf die Schachtel}) \textit{auf} \end{array} \right\}$ (nach Caroli 1984b: 240)

Diese Art der Tilgung ist nicht für jedes Verb und auch nicht immer für jede Präposition, jedes Präfix möglich. Caroli markiert sie deshalb unter ' $N_0 V N_{1_{akk}} Préf$ ' und nennt außerdem, welche Präfixe (*ab*, *an*, *auf*, *aus*, *hin* oder andere) möglich sind.

Zusammen mit der Präfigierung des Verbs kann das direktionale Argument in ein statisches transformiert werden ($Loc N_{2_{akk}} \rightarrow Loc N_{2_{dat}}$):

- (2.111) a. *Max hängt den Mantel an den Haken.*
 b. *Max hängt den Mantel am Haken auf.* (nach Caroli 1984b: 258)

Weitere Beschreibungskriterien sind die Passiv- und Reflexivbildung: das Passiv mit *werden* ('PW') und *sein* ('PS'), das „intentionale Reflexiv“ ('RI': *Max legt (sich + für sich) 20 Flaschen Rotwein in den Keller.*), das Reflexiv der Koreferenz ('RC': *Das Kind legt sich in das Bett.*) und eine Reflexivtransformation, die zur Tilgung des Subjekts führt ('RP'):

- (2.112) a. *Max legt das Papier um die Schachtel.*
 b. *Das Papier legt sich um die Schachtel.*

Vier „Derivationen“⁴⁷, drei „Intransitivderivationen“, sowie die Präfigierung des Verbs mit *be-* beschließen den von Caroli aufgestellten Kriterienkatalog:

IT-LD Das direkte Objekt rückt in die Subjektposition, die direktionale Ortsangabe (,lieu directionnel') bleibt:

- (2.113) a. *Max tropft Wasser auf den Tisch.*
 b. *Wasser tropft auf den Tisch.*

IT-ST Die direktionale Ortsangabe wird zur statischen (,lieu statique'):

47 Den Unterschied zwischen Transformation und Derivation definiert Caroli (1984b) nicht.

- (2.114) a. *Max hängt den Mantel an den Haken.*
 b. *Der Mantel hängt am Haken.*

IT-LM Bei Argumenten, die ein Mittel bezeichnen (,lieu-moyen‘):

- (2.115) a. *Max mariniert den Braten in einer Rotweinbeize.*
 b. *Der Braten mariniert in einer Rotweinbeize.*

Préf *be* Präfigierung mit *be* (vgl. für das Französische Tsuruga 2005)

- (2.116) a. *Max pflanzt Blumen auf das Beet.*
 b. *Max bepflanzt das Beet mit Blumen.*

Im Anhang an Carolis Aufsatz findet sich eine Tabelle mit 88 Verben (255–267), die nach den beschriebenen 29 Kriterien klassifiziert sind. Tab. 2.7 zeigt einen Auszug daraus.

Caroli & Figge (1989) bemühen sich um eine „Vergleichende Klassifikation deutscher und französischer Verben mit Lokalgängung“ auf Basis der Lexikogrammatik des LADL. Insbesondere bei der Frage nach der Unterscheidung von Komplementen und Supplementen (vgl. Kap. 2.3.3) beziehen sie auch Arbeiten aus der Valenzgrammatik mit ein. Der Aufsatz basiert auf den Vorarbeiten Carolis (1984a, 1984b), bringt aber einige neue Transformationen. Um Wiederholungen zu vermeiden, werden hier nur neue Aspekte wiedergegeben.

Als Methode verwenden Caroli & Figge (1989:244–6) die Explikation der lokalen Relation analog zu den Arbeiten der französischen Lexikogrammatiker (vgl. S. 47). Basierend auf der Einteilung von Guillet & Leclère (siehe Tab. 2.3) gelangen Caroli & Figge (1989:245–251) zu einer klaren Strukturierung in einige Hauptklassen:

1. Transitive Verben, wobei ,transitiv‘ in der traditionellen Bedeutung (,mit direktem bzw. Akkusativobjekt‘) verwendet wird. Je nachdem, wie lokalisierter Gegenstand und Ort realisiert werden, unterscheiden Caroli & Figge vier Subklassen:
 - a) Ort=PP, lok. Ggstd.=Obj. (*Max stellt das Glas auf den Tisch*)
 - b) Ort=Obj., lok. Ggstd.=PP (*Max bepflanzt das Beet mit Blumen*)
 - c) Ort=Obj., lok. Ggstd.=Subj. (*Max passiert das Tor*)
 - d) Ort=Subj., lok. Ggstd.=Obj.: nur Verben der Nahrungsaufnahme oder -ausscheidung (*Max verschlingt das Essen, Max spuckt das Kaugummi auf den Boden*)
2. Intransitive Verben verteilen sich auf zwei Subklassen:
 - a) Ort=PP, lok. Ggstd.=Subj.: Diese teilen „sich in eine Klasse von Bewegungsverben und in eine Klasse von statischen Verben auf“ (250):
 - i. *Max geht ins Kino.*
 - ii. *Max weilt im Hotel.*
 - b) Ort=Subj., lok. Ggstd.=PP (*Der Käse wimmelt von Würmern*): Sätze dieser Klasse können als Transformationen des Typs *The garden is swarming with bees* aufgefasst werden (von *Die Würmer wimmeln im Käse*)

Die Lokalgängungen unterteilen Caroli & Figge (1989:251–8) in direktionale und nicht-direktionale. Diese sind im Deutschen anders als im Französischen „eindeutig voneinander unterschieden“, was daran liegt, „dass die jeweiligen Präpositionen entweder voneinander

	Typen Ortsargumente								Tilgung von Aktanten							Passiv, Refl.					Intr.der.								
	LI (Ausgangspunkt)	LF (Endpunkt)	LS	LOE	LM (Mittel)	LPC (Körperteil)	$N_{1_{akk}} = N_{1_{dat}}$	$Loc N_{2_{dat}} = N_{2_{akk}}$	LST (statisch)	$N_0 V$	$N_0 V Loc N_2$	$N_0 V N_{1_{akk}}$	$N_0 V N_{1_{akk}} Préf$	<i>ab</i>	<i>an</i>	<i>auf</i>	<i>aus</i>	<i>hin</i>	andere Präfixe	$Loc N_{2_{akk}} \rightarrow Loc N_{2_{dat}}$	PW (<i>werden</i>)	PS (<i>sein</i>)	RI (<i>sich/für sich</i>)	RC (<i>sich</i>)	RP (Subjekttilgung)	IT-LD	IT-ST	IT-LM	<i>be-</i>
<i>bauen</i>	-	-	-	+	-	-	-	-	+	+	+	-	-	-	-	-	-	-	andere Präfixe	$Loc N_{2_{akk}} \rightarrow Loc N_{2_{dat}}$	-	+	+	-	-	-	-	-	+
<i>beißen</i>	-	-	-	-	-	+	+	-	+	+	+	-	-	-	-	-	-	-	andere Präfixe	$Loc N_{2_{akk}} \rightarrow Loc N_{2_{dat}}$	+	-	-	+	-	-	-	-	-
<i>bohren</i>	-	-	-	+	-	-	-	-	+	+	+	+	-	-	-	-	+	-	andere Präfixe	$Loc N_{2_{akk}} \rightarrow Loc N_{2_{dat}}$	+	+	+	-	-	-	-	-	-
<i>brennen</i>	+	+	-	-	+	-	-	-	-	-	+	+	-	-	-	+	+	-	andere Präfixe	$Loc N_{2_{akk}} \rightarrow Loc N_{2_{dat}}$	+	+	+	+	-	-	-	-	-
<i>gießen</i>	+	+	+	+	-	-	-	-	+	-	+	+	+	+	+	+	+	+	andere Präfixe	$Loc N_{2_{akk}} \rightarrow Loc N_{2_{dat}}$	+	-	+	-	+	-	-	-	+
<i>hängen</i>	-	+	+	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+	andere Präfixe	$Loc N_{2_{akk}} \rightarrow Loc N_{2_{dat}}$	+	+	+	-	+	-	-	-	+
<i>kleben</i>	-	+	+	-	-	-	-	-	-	-	-	+	-	+	+	+	-	-	andere Präfixe	$Loc N_{2_{akk}} \rightarrow Loc N_{2_{dat}}$	+	+	+	-	-	-	-	-	-
<i>legen</i>	+	+	+	+	-	-	-	-	-	-	-	+	+	+	+	+	+	+	andere Präfixe	$Loc N_{2_{akk}} \rightarrow Loc N_{2_{dat}}$	+	+	+	+	+	-	-	-	+
<i>marinieren</i>	-	-	-	-	+	-	-	-	-	-	+	-	-	-	-	-	-	-	andere Präfixe	$Loc N_{2_{akk}} \rightarrow Loc N_{2_{dat}}$	+	-	-	-	-	-	-	+	-
<i>pflanzen</i>	-	-	-	+	-	-	-	-	+	+	+	+	-	-	-	-	-	-	andere Präfixe	$Loc N_{2_{akk}} \rightarrow Loc N_{2_{dat}}$	+	+	+	-	-	-	-	-	+
<i>rasieren</i>	+	-	-	-	-	+	+	+	-	-	+	+	+	+	-	-	+	-	andere Präfixe	$Loc N_{2_{akk}} \rightarrow Loc N_{2_{dat}}$	+	+	+	+	-	-	-	-	-
<i>reisen</i>	+	+	-	+	-	-	-	-	-	-	-	+	+	+	-	+	+	-	andere Präfixe	$Loc N_{2_{akk}} \rightarrow Loc N_{2_{dat}}$	+	+	+	-	-	-	-	-	-
<i>schöpfen</i>	+	+	-	-	+	-	-	-	-	-	+	+	+	+	-	+	+	-	andere Präfixe	$Loc N_{2_{akk}} \rightarrow Loc N_{2_{dat}}$	+	-	+	-	-	-	-	-	-
<i>schreiben</i>	+	-	-	+	-	-	-	-	+	+	+	+	+	+	+	+	-	-	andere Präfixe	$Loc N_{2_{akk}} \rightarrow Loc N_{2_{dat}}$	+	+	+	-	-	-	-	-	+
<i>streicheln</i>	-	-	-	-	-	+	+	+	-	-	+	-	-	-	-	-	-	-	andere Präfixe	$Loc N_{2_{akk}} \rightarrow Loc N_{2_{dat}}$	+	-	-	+	-	-	-	-	-
<i>tropfen</i>	+	+	+	-	-	-	-	-	-	-	-	+	-	-	+	-	-	-	andere Präfixe	$Loc N_{2_{akk}} \rightarrow Loc N_{2_{dat}}$	+	+	+	-	-	-	-	-	+
<i>ziehen</i>	+	+	+	-	-	-	-	-	+	-	+	+	+	+	-	-	-	-	andere Präfixe	$Loc N_{2_{akk}} \rightarrow Loc N_{2_{dat}}$	+	+	+	+	-	-	-	-	-

Tabelle 2.7: Transitive deutsche Verben mit Direktionalergänzung. Auszug aus Caroli (1984b: 260–7)

verschieden sind oder, wo dies nicht der Fall ist, wenigstens unterschiedliche Rektionen haben: Dativ in nicht-direktionalen Ergänzungen, Akkusativ in direktionalen Ergänzungen des Zielortes.“ (252) Zu den bei Caroli (1984b) beschriebenen Besonderheiten bei den Lokaler-gänzungen (LOE, LM, LS, LPC/Pertinenzdativ) kommen noch Infinitivergänzungen (siehe auch Kap. 4.5.3) und komplexe Lokaler-gänzungen hinzu. Letztere unterteilen Caroli & Figge (1989: 254–6) in „doppelte Ergänzung des Ausgangs- und Zielortes“ und Ergänzungen mit Angabe des Erstreckungsbereichs (*vom einen Ende des Beets bis zum anderen*). Die Verben mit „doppelter Ergänzung“ gliedern sich in fünf Gruppen:

1. Ausgangs- und/oder Zielort (*holen*),
2. komplex oder nur Zielort (*fahren*),
3. Ausgangs- oder Zielort, nicht komplex (*Die Birne (in die + aus der) Fassung drehen*),
4. nur Zielort (*kleben*),
5. nur Ausgangsort (*vom Tisch nehmen*).

An Transformationen (zuzätzlich zu Caroli 1984b) nennen Caroli & Figge (1989: 261–2) nur die „symmetrische Konstruktion“:

- (2.117) a. *Max klebt den Deckel an die Schachtel.*
 b. *Max klebt den Decken und die Schachtel aneinander.*
 c. *Max verklebt Decken und Schachtel (miteinander).*

Dem Derivationspräfix *be-* stellen die Autoren *ent-* gegenüber, durch das der Ausgangsort in die Objektposition gerückt wird (262).

Fazit

Leider sind nur kleine Tabelle deutscher Verben zugänglich, die mit transitiven Verben auch nur einen Typ aus den Verben mit Lokaler-gänzung enthalten. Abgesehen von den Transport-
 verben in Caroli et al. (1990), handelt es sich um eine semantisch sehr inhomogene Gruppe von Verben, die sich freilich an der syntaktischen Oberfläche gleich/ähnlich verhält, z.B. (*ins Haus*) *bringen* vs. (*ins Papier*) *wickeln*. Beschrieben wurden zahlreiche Transformationen der betreffenden Verben, nur angeschnitten dagegen die Frage komplexer Ortsargumente (Pfade). Diese kritischen Punkten, die Behandlung von Wegpunkten (Trajektorie) und die semanti-sche Inhomogenität der lokativen Verben, teilen sich Caroli und die Lexikogrammatiker des LADL.

2.5.1.3 Garrigues (1993, 1995)

Mylène Garrigues beschreibt mit lokalen Grammatiken französische Sätze des Inhalts ‚Mensch *X* begibt sich zu einem Ort *Y*‘: „un humain effectue un déplacement à destination explicite ($N_{dest} = N_{pays}, N_{cont}$ ou *Nile*)“ (Garrigues 1993: 282). Sie beschränkt sich dabei auf ca. 60 Bewegungs-
 verben und bei den Ortsargumenten auf die Namen der Kontinente, 160 Staaten und etwa 70 Inseln. Das Ergebnis sind 15 manuell erstellte, hierarchisch angeordnete Graphen, die Sätze dreier Typen beschreiben:

- (2.118) a. *Nhum V Prép Nloc: Jo va en Italie.*

- b. *Nhum V Nloc: Jo regagne l'Italie.*
- c. *Nloc V Nhum: L'Italie attire les touristes.*

Obwohl damit nur einen winziger Ausschnitt des Französischen beschrieben wird, ist die Anzahl der beschriebenen Pfade mit über 30 000 hoch, wobei noch kein Subjekt (*Nhum*) eingefügt ist.

Garrigues' Beschreibung von Sätzen ausgehend von einer semantischen und nicht syntaktischen Ähnlichkeit, ist selten in der Linguistik.

In der folgenden Diskussion möchte ich mich von den drei Schemata auf (2.118a) beschränken, also auf Konstruktionen, deren lokales Element (Kontinent, Staat, Insel) als Präpositionalphrase vorliegt. Garrigues betrachtet Kombinationen aus Verb, Präposition und der ihr untergeordneten NP als „unteilbare Phraseneinheiten“ („entité phrastique insécable“, 1993: 286). Die „Wahl“ der Präposition hängt dabei sowohl vom Verb als auch der lokativen NP ab. Garrigues verdeutlicht das, indem sie sowohl den Kern (Nomen) der NP als auch das Verb variiert und die so erzeugten Sätze nach ihrer Grammatizität (s.u.) bewertet. Anschließend beschreibt Garrigues in 15 hierarchisch angeordneten Graphen alle korrekten (grammatischen) subjektlosen Teilsätze (*V Prép Nloc* für die ca. 60 Verben und 300 geographische Namen). Betrachtet man den Graphen der obersten Ebenen (Abb. 2.4), fällt auf, dass alle Pfade in der Mitte des Graphen durch nur zwei Teilpfade führen – ein Pfad, wo die Präposition ausschließlich durch das Verb determiniert ist, und einer wo das Verb *en* oder *à* vorgibt, und das lokative Argument (der Zielpunkt) bestimmt, welche der beiden zu wählen ist.

Die Wahl zwischen den Präpositionen *en* und *à* ist also durch das regierte Objekt bestimmt. Allerdings bedeutet das nicht, dass es keine Regularitäten gibt: die Präposition ist herleitbar aus der Kombination von Genus, vokalischem oder konsonantischem Anlaut des Toponyms und dessen Gebrauch mit/ohne Artikel. Eine regelhafte Beschreibung wäre bis auf wenige wirkliche Ausnahmen möglich,⁴⁸ was Garrigues (1993: 294) selbst zugibt.

Die Arbeit Garrigues' ist motiviert durch viele kontrastierende Gegenüberstellung von grammatischen und nicht grammatischen Ausdrücken. So können z.B. nicht alle Inselnamen zusammen mit dem Klassenbezeichner (*l'île de*) verwendet werden (1993: 295). Stellt man den Beispielen Garrigues' jedoch Korpusbelege gegenüber, legen diese oft eine gegenteilige Bewertung der Grammatizität nahe, so im Fall von *l'île de Crète* vs. **l'île de Corse* oder *les îles de Philippines* / **les îles de Comores*. Dass außerhalb des argumentativen Teils der Arbeit in den Grammatiken selbst einige fragwürdige Pfade erscheinen (*?*arriver sur Israël*) ist zu verzeihen und als Übergeneralisierung nachvollziehbar.

Garrigues' Arbeit kontrastiert mit den Methoden der Lexikogrammatik und stellt in der Konsequenzen und ausschließlichen Verwendung lokaler Grammatiken ein Novum dar. Mit ähnlicher Methodik und ähnlichen Ausdrücken befassen sich Constant (2003b, siehe Kap. 2.5.1.4) und Chaves (2005a,b). Letzter nutzt zur Extraktion von Definitionen für und Relationen zwischen geographischen Entitäten im Portugiesischen ebenfalls lokale Grammatiken in einer Weise, die Garrigues ähnelt, aber in der Tiefe deutlich zurückbleibt.

48 So bei Glättli (1966), Boons (1985:201), Grimaud (1988), Gross (1995a:173–4)

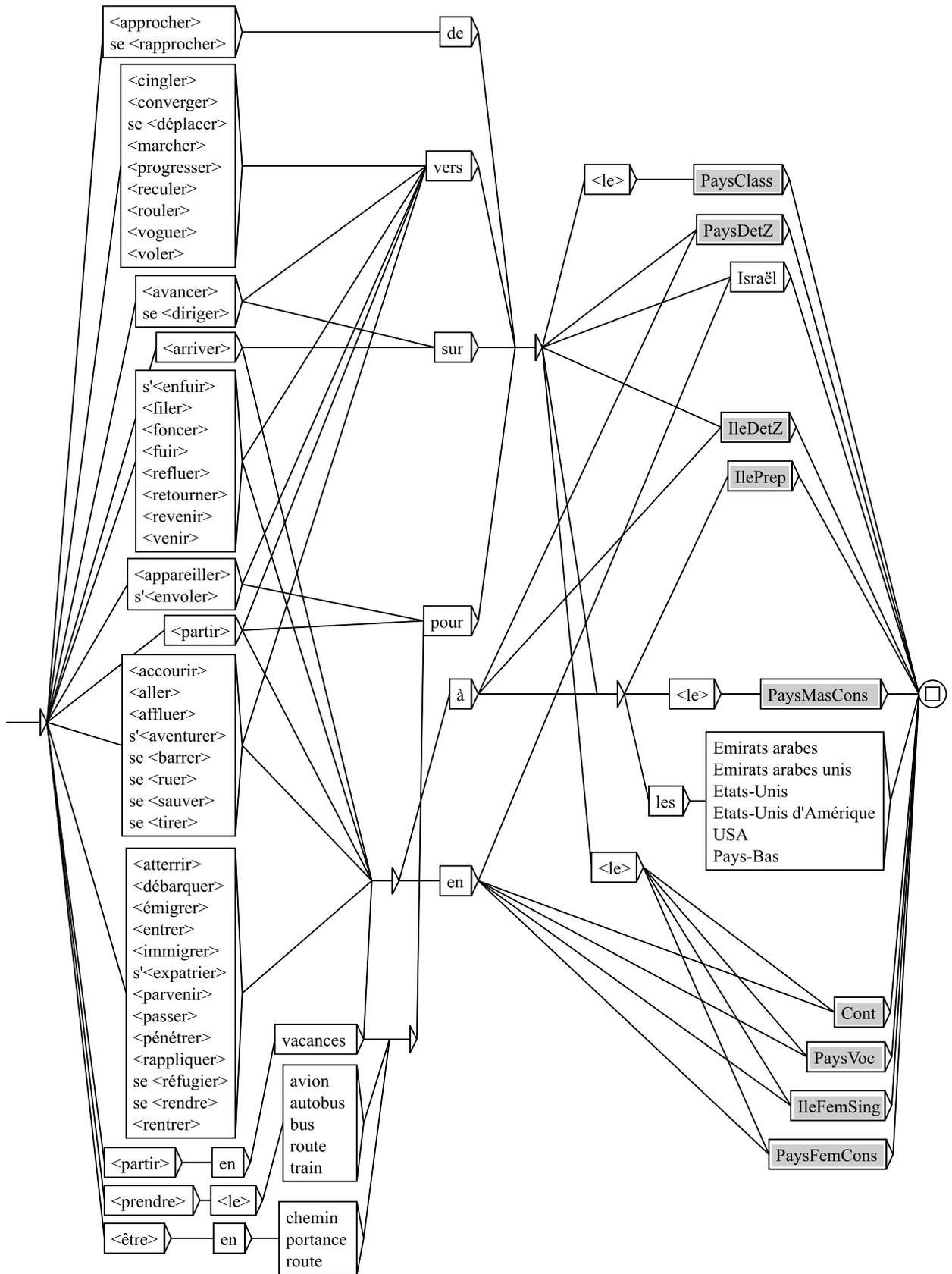


Abbildung 2.4: Französische Sätze des Typs *Nhum V Prép Nloc* (Garrigues 1993: 288)

2.5.1.4 Constant (2003b)

Constant (2002, 2003b) beschreibt Ortsadverbiale der Form *Prep Det N Modif*, die er in elementaren Sätzen des Typs $N_0 V_{sup} Prep Det N Modif$ ⁴⁹ betrachtet, wobei V_{sup} durch *être*, *se passer* oder *se dérouler* realisiert werden kann. Constant beschreibt zum einen komplexe Präpositionen (*dans la zone ouest de*, *à la frontière de*) durch lokale Grammatiken, der Fokus der Beschreibung liegt aber auf der Kombinatorik von 650 Toponymen und deren Adjektivderivationen, 70 Klassenbezeichnern, Präposition und Artikel. Die Idiosynkrasien bei der Kombination der Bestandteile zu einem Adverbial modelliert Constant in 50 Tabellen der Lexikongrammatik:

- der Klassenbezeichner als Namensbestandteil: *république (fédérale/populaire), département*, usw.
- Adjektivableitungen: *France* → *française* (im Genus des Klassenbezeichners)
- Verwendung mit/ohne Artikel und Form des Artikels: *le Chili* vs. *la Chine*
- Realisierungsmöglichkeiten und Varianten:

<i>le Nc Adj1 de Det Npr</i>	<i>la république démocratique de le Congo</i>
<i>le Nc Adj1 de Npr</i>	<i>la république arabe d'Égypte</i>
<i>le Nc de Det Npr</i>	<i>la république fédérale de l'Allemagne</i>
<i>le Nc de Npr</i>	<i>la république de Hongrie</i> , aber: * <i>la république de France</i>
<i>le Nc Npr-e</i>	<i>la république française</i>
<i>Det Npr</i>	<i>la Hongrie</i>

Hier werden Lexikongrammatiken also zur lexikographischen Beschreibung von variablen Formen von Toponymen eingesetzt. Um die Erstellung der Schablonengraphen für alle 50 Tabellen zu vereinfachen, werden diese über eine Metatabelle aus einem Metaschablonengraphen generiert (Constant 2003b: 149–52). Die Methode, die syntaktischen Eigenschaften der Toponyme wie Genus oder Artikelverwendung in einer Lexikontabelle und nicht als Merkmal im Lexikon zu kodieren, hat jedoch einen Nachteil: die Koordination von Toponymen in Nominalphrasen (*les états de Californie et du Texas*) ist nicht beschreibbar, da die Anzahl der kombinatorischen Möglichkeiten und damit die der instantiierten Graphen explodiert (hier 650^2 Möglichkeiten allein bei der Kombination von zwei Toponymen). Constant (2003b: 152) schlägt dazu vor, die Informationen aus den Tabellen in ein Wörterbuch wie PROLEX zu konvertieren, und dann damit die notwendigen Grammatiken zu konstruieren.

Zur Beschreibung vollständiger Präpositionalphrasen, die als Adverbiale verwendet werden können, fehlen noch die Idiosynkrasien bei der Auswahl der Präposition, und zwar:

- mit welchen der Präpositionen *à* oder *dans*, *sur* und *en* werden die 70 Klassenbezeichner verwendet? Dies wird in einer eigenen Tabelle beschrieben.
- kann das Toponym allein in Kombination mit Präpositionen und Artikel in den Konstruktionen *en Npr*, *dans Det Npr*, *à Det Npr* und *à Npr* verwendet werden? Hierfür werden die 50 Tabellen der Toponyme um diese vier Spalten erweitert.

49 Eigentlich müsste es $N_0 V_{sup} Prep N_1$ (so bei Gross 1995a: 181–4) heißen, da die Präposition das Prädikat eines solchen Satzes ist. Constant (2003b: 115) lässt diese Frage beiseite: „Pour certain linguistes [...] la préposition locative est un prédicat. Cette discussion n'est pas pertinente à notre propos.“

Die auf die verschiedenen Tabellen verteilten Informationen werden dann mittels einem in Constant (2003a) und (2003b: 169–77) beschriebenen Algorithmus in einer Tabelle zusammengeführt und über Metatabelle und Metaschablonengraph (s.o.) in eine lokale Grammatik überführt.

2.5.2 Lokative Verben außerhalb der Lexikogrammatik

Linguistische Beschreibungen lokativer Verben sind so zahlreich, dass es hier unmöglich wäre, sie darzustellen, vgl. z.B. die Bibliographie zu Bewegung(sverben) von Matsumoto & Slobin (2002). Deshalb kann hier nur eine kleine Auswahl an Beschreibungen außerhalb der Lexikogrammatik behandelt werden. Der Schwerpunkt liegt dabei auf Darstellungen zu deutschen lokativen Verben.

2.5.2.1 Alternationen und Klassifikation

Ausgangspunkt für die Arbeit von Beth Levin (1993) ist die These, die Bedeutung eines Verbs bestimme weitgehend dessen syntaktisches Verhalten (die Argumentstellen und deren Interpretation). Daraus folgt, dass das Verhalten eines Verbs als Ausgangspunkt für eine Untersuchung der Verbsemantik dienen kann. Als Mittel zur syntaktischen Beurteilung dienen Levin Alternationen (Transformationen). Verben, die dem gleichen Satz an Alternationen unterliegen, bilden eine Klasse. Mit etwa 80 solcher Alternationen unterteilt sie die gut 3000 untersuchten Verben in 49 Klassen mit zahlreichen Untergruppen, insgesamt etwa 190 Gruppen.

Die intensive Beschreibung und Listung von Transformation (oder ‚Alternationen‘) bilden auch den Schnittpunkt zwischen Levin und der Lexikon-Grammatik. In der Methodik sind die Unterschiede dagegen groß:

Levin's study implicitly meets three of the criteria set up in [(Gross 1994: 212–4)] for the construction of lexicon grammars: namely, that the linguistic unit of meaning is the elementary sentence, not the word: that verbs cannot be considered without their subjects and possible complements; and that internal syntactic properties of other constituents may depend on the main verb. [...] Gross' requirement that „verbs have to be described individually and not in terms of intensional classes“ (ibid.) is explicitly rejected by Levin.

[...] the research methodology she advocates contradicts nevertheless the lexicon-grammatical principle of describing verbs and other lexical entries individually and exhaustively by experimental procedures. [...] Without having first enumerated all the facts she assumes in the generative vein that „the behaviour of a verb, particularly with respect to the expression and the interpretation of its arguments, is to large extent determined by its meaning.“ (1993: 1). Despite her awareness that „native speakers can make extremely subtle judgments concerning the occurrence of verbs with a range of possible combinations of arguments and adjuncts in various syntactic expressions“, (2) she proposes that „the ideal lexical entry for a word should minimize the information provided for that word“ (11) because a verb's syntactic behaviour can predicted from the interaction of its meaning components and general principles of grammar. (Boas 1998: 12)

Levin beschreibt folgende Alternationen, an denen lokative Argumente beteiligt sind:

- Lokative Alternation

- (2.119) a. *The farmer loaded apples into the cart.*
 b. *The farmer loaded the cart with apples.*
- (2.120) a. * *Gina filled lemonade into the pitcher.*
 b. *Gina filled the pitcher with lemonade.*

- Wegfall der Präposition, Locative Preposition Drop Alternation

- (2.121) a. *Martha climbed up the mountain.*
 b. *Martha climbed the mountain.*

- Search-Alternations

- (2.122) a. *Ida hunted the woods for deer.*
 b. *Ida hunted for deer in the woods.*
 c. *Ida hunted deer in the woods.*

- Alternationen mit „obliquem“ Subjekt

- Locatum Subject Alternation

- (2.123) a. *I filled the pail with water.*
 b. *Water filled the pail.*

- Location Subject Alternation

- (2.124) a. *We sleep five people in each room.*
 b. *Each room sleeps five people.*

- Passiv-Transformation (Prepositional Passiv bei intransitiven Verben)

- (2.125) a. *George Washington slept in this bed.*
 b. *This bed was slept in by George Washington.*

- There-Insertion

- (2.126) a. *A ship appeared on the horizon.*
 b. *There appeared a ship on the horizon.*

- Lokativ-Inversion

- (2.127) a. *A cat jumped on/onto the table.*
 b. *Onto the table jumped a cat.*
 c. *On the table jumped a cat.*

- Direktionale Phrasen alternierend mit ungerichteten Bewegungsverben

- (2.128) a. *The truck rumbled.*
 b. *The truck rumbled into the driveway.*

Mit dem Vergleich von Alternationen im Englischen und Deutschen beschäftigen sich Olsen (1994, 1995), Frense & Bennet (1996), mit ausschließlich deutschem lokativem Material Wunderlich (1985b), vgl. auch die bereits besprochenen Arbeiten von Caroli (Kap. 2.5.1.2). Die Beschreibung von Alternationen und Transformationen ist ohne Zweifel eine Notwendigkeit für Anwendungen wie die Informationsextraktion, was Levins z.T. seltsam anmutende Beispiele verdecken. Deshalb hier noch zwei realistischere Beispiele, die Levins „Passiv“-Transformation (2.125) ähneln, aber auch eine klare Verbindung zum Phänomen der regulären Polysemie (Kap. 3.1.4) aufweisen:

- (2.129) a. *Das Parlament diskutiert den Gesetzentwurf.*
 b. *Vom Parlament wird der Gesetzentwurf diskutiert.*
 c. *Im Parlament wird der Gesetzentwurf diskutiert.*
- (2.130) a. *Deutschland wählt (einen neuen Bundestag).*
 b. *In Deutschland wird (ein neuer Bundestag) gewählt.*

Die Idee Levins, dass syntaktische Verhalten und die Semantik eines Verbs in enger Korrelation zueinander stünden, entwickelt Schulte im Walde (2003, 2004, 2006) weiter. Mit Hilfe eines statistischen Parsers extrahiert sie die syntaktischen Rahmen deutscher Verben aus einem Korpus und versucht anschließend ein großes Sample an Verben zu clustern. Dazu verwendet sie drei Typen syntaktischer Rahmen: (D1) 38 syntaktische Subkategorisierungsrahmen, z.B. $na = N_{nom} V N_{akk}$, (D2) D1 mit instantiierten Präpositionen und (D3) D2 mit Selektionsklasse des Objekts (15 Objektklassen auf Basis von GERMANET). Trotz der interessanten und aufwändigen Methodik ist eine linguistisch brauchbare Klassifikation nicht das Ergebnis der Arbeit Schulte im Waldes. Ambiguitäten verursachen zu viel Rauschen, gerade auch, weil Phraseologismen nicht berücksichtigt werden und dem enthaltenen verbalen Kernen zugeschlagen werden, wie die Beispiele (2003: 27–103) nahelegen, z.B. Stützverbkonstruktionen *unter Kontrolle bringen*, *Preise klettern*, Idiome *eine Lawine ins Rollen bringen* oder Unzulänglichkeiten des Parsers [*kein Wort*]_{moving_entity} *gehev* *verloren*_{path}.

2.5.2.2 Valenzorientierte Beschreibung

Gerling & Orthen (1979) erarbeiteten eine dependenzorientierte Klassifikation deutscher Zustands- und Bewegungsverben. Zentrales Kriterium ist die Verbvalenz (für Bewegungsverben: SOURCE, GOAL, PATH), die durch weitere semantische Merkmale (Schnelligkeit, Medium der Fortbewegung, usw.) ergänzt wird. Die zweite Hälfte des Buches enthält ein Lexikon mit über 300 Verben, und jeweils Angaben zu Valenz, Selektionsbeschränkungen, semantischen Merkmalen und Modifikatoren, Beispielsätze und Bedeutungserklärung. Das Lexikon stellt nicht nur eine gute Materialsammlung dar, die Valenzorientierung ermöglicht auch eine semantisch sinnvolle Klassifikation inklusive der Lesartenunterscheidung, z.B. für *gehen* (s.u.) oder *wohnen* (S. 68).

¹gehen – ‚sich (zu Fuß) irgendwohin fortbewegen‘: *Paul geht durch den Wald zur Kirche. Tante Klara geht an die Luft.*

²gehen – ‚eine Einrichtung regelmäßig besuchen‘: *Beatrix geht zur Schule.*

³gehen – ‚in eine bestimmte Richtung fahren‘: *Der Zug geht nach Berlin. Die M.S. Hanseatic geht nach New York.*

⁴gehen – ‚sich (zu Fuß) entfernen, verlassen‘: *Karl geht aus dem Haus. Martin geht aus dem Unterricht. Er geht.* (Gerling & Orthen 1979: 177–8)

Das Verblexikon von Schröder (1993) enthält rund 250 deutsche Bewegungsverben, und listet dazu noch gut 1700 präfigierte Verben. Der Beschreibungsansatz orientiert sich stark an dem von Helbig & Schenkel (1975). Die Bedeutungserklärungen in Kombination mit der Valenzbeschreibung und den Beispielen wirken aber vielfach etwas wirr (vgl. z.B. die Lesarten von *gehen* 49–50). Bezüglich der Stelligkeit der Verben entscheidet sich (Schröder 1993:12) dafür, nur die Stellen Agens, Pfad, und optional Transportmittel oder Medium anzunehmen. D.h. der Pfad mit den Bestandteilen LOC SOURCE, LOC GOAL und LOC PATH wird als eine Einheit gesehen, die nur einen Valenzslot füllt. Daneben gibt es für Schröder Verben, die nicht alle dieser Bestandteile realisieren, z.B. *fallen* (LOC GOAL und/oder LOC SOURCE), *bummeln* (LOC PATH).

2.5.2.3 Kognitive, konzeptuelle und formale Semantik

Hier sollen kurz einige Arbeiten vorgestellt werden, die kognitiv oder konzeptuell orientiert sind, und/oder eine formale semantische Beschreibung lokativer Verben versuchen.

Rauh (1981) und Radden (1989) vergleichen kontrastiv deutsche und englische Bewegungs-
verben (*kommen, gehen* vs. *come, go*) mit dem Ergebnis, dass die deutschen Verben auf einen Punkt der Bewegung (Ausgangs- oder Zielpunkt des Pfades) fokussieren, während im Englischen die deiktische Situation wiedergegeben wird. Dabei drückt *go* aus, dass der Sprecher zum Zeitpunkt der Äußerung nicht am Zielpunkt ist, *come* dagegen, dass der Zielpunkt auch Ort des Sprecher oder Hörers (oder ‘home base’) ist. Wichtig ist die Feststellung Raddens (1989: 232), dass Bewegungs-
verben „typischerweise inhärent eine Orientierung bzw. einen Fokus auf eines der Strukturelemente des Pfadschemas“ haben.

Im Rahmen des Projektes LILOG (Herzog & Rollinger eds. 1991) entstanden eine Reihe von Arbeiten, die sich mit Raumausdrücken des Deutschen, ihrer Syntax, Semantik und dahinterliegenden räumlichen Konzepten (Pfadkonzepte, Depiktion) beschäftigen. Claudia Maienborn (1990a,b, 1991) beschäftigte sich dabei mit lokativen Verben. Diese teilt sie in die zwei Klassen Positions- und Bewegungs-
verben ein. In der formalsemantischen Repräsentation ähneln sich beide strukturell sehr stark:

(2.131) $\lambda p \lambda x[\text{LOC}(x, p) \ \& \ \text{MOD}_{\text{POSITION}}(x)]$

(2.132) $\lambda w \lambda x[\text{MOVE}(x, w) \ \& \ \text{MOD}_{\text{MOTION}}(x)]$

Die Variable x steht für das lokalisierte Objekt, p für den Ort, w für den Pfad. Zentrale Annahme ist also, dass Bewegungs-
verben nur ein lokatives Argument haben, den Pfad.

Bei der Beschreibung deutscher Verben im Rahmen des CISLEX-Projektes greift Schnorbusch (1999: 101–2) diese Lösung bei der Beschreibung von Bewegungs-
verben am Beispiel *segeln* auf: „Eine andere Lösung bestünde darin, Bewegung als eine zweistellige Relation zwischen dem sich bewegenden Gegenstand und einem Weg aufzufassen und Startpunkt, Pfad und Zielpunkt nicht als Relata der Bewegungsrelation zu begreifen [. . .] Im Einklang mit jüngeren Vorschlägen aus der formalen Semantik [Maienborn (1990b)] habe ich mich für diese zweite Lösung entschieden und habe nur ein Argument für den Weg vorgesehen.“

Ehrich (1996) beschäftigt sich mit der Subklassifikation von Transport-
verben, deren formal-semantischer Beschreibung und deren Beschreibung von Alternationen.

Tschander (1999) untersucht die Auxiliarselektion von Bewegungsverben in Anlehnung an Arbeiten von Jackendoff (1990). Sie unterscheidet drei Klassen von Verben:

- (2.133) a. *Willy *hat/ist gewatschelt.*
 b. *Willy *hat/ist entlang des Baches gewatschelt.* (PATH)
- (2.134) a. *Debbie hat/?ist getanzt.*
 b. *Debbie *hat/ist durch den Saal getanzt.* (MANNER)
- (2.135) a. *Debbie hat/ist gehumpelt.*
 b. *Debbie *hat/ist nach Hause gehumpelt.* (MANNER-PATH)

Die Auxiliarselektion sieht Tschander (1999: 29) als Fokussierungseffekt. Bei den MANNER-PATH-Verben (2.135), „wird mittels der *haben*-Selektion die Bewegungsart und durch die *sein*-Selektion der Bewegungsweg fokussiert. Damit ist es nicht mehr notwendig, dass für Verben wie *humpeln*, *joggen* oder *schwimmen* zwei Lexikoneinträge postuliert werden müssen.“

Heintzeler (1992) vergleicht Kombinationen aus lokativem Verb und transformativer Präpositionen im Deutschen mit transformative Verben (*entrer*, *sortir*) im Französischen und konzentriert sich dabei auf die aspektuelle und temporale Komponente. Sie stellt fest, dass transformative lokative Präpositionen die Zeitkonstitution des Verbs oder der Verbalphrase verändern, und zwar von kursiv zu transformativ,⁵⁰

- (2.136) a. *Elsa schwimmt stundenlang*
 b. * *Elsa schwimmt in zwei Stunden*
- (2.137) a. * *Elsa schwimmt stundenlang nach Meersburg*
 b. *Elsa schwimmt in zwei Stunden nach Meersburg*

und kommt zum Schluss (122), „dass deutsche transformative lokale Präpositionen Verben in gewisser Weise französischen transformativen Verben wie *entrer* oder *sortir* entsprechen.“⁵¹

Eschenbach et al. (2000) stellen eine Klassifikation von Bewegungsverben (Tab. 2.9) auf, die sich auf zwei Unterscheidung stützt: (1.) das Verhältnis zwischen Agens und Figure, was gleichbedeutend ist mit der Valenz des Verbs, und (2.) der Fokus des Verbs auf den gesamten Pfad oder den Zielpunkt („verbs of change of position focus on the change into the final position, and neglect the rest of the path“).

2.5.2.4 Fictive motion: statische Lesarten von Bewegungsverben

Bewegungsverben verfügen vielfach auch über eine statische Lesart, in der sie mit unbeweglichem Subjekt (Gross 1995a: 192) eine „fiktive Bewegung“ (Talmy 1996) beschreiben.

- (2.138) a. *That mountain range lies between Canada and Mexico.*
 b. *That mountain range goes from Canada to Mexico.*

50 Siehe Fabricius-Hansen (1975: 19) zu den Begriffen ‚kursiv‘ und ‚transformativ‘, die Vendlers 1968 ‚activity‘ bzw. ‚accomplishment‘ entsprechen.

51 Vgl. das Beispiel Tesnières, S. 25.

	path	change of position
agens = figure	verbs of locomotion <i>schlendern, wackeln, rennen, gehen, rollen, kommen</i>	verbs of change of position <i>treten, eintreten, betreten, ankommen, verlassen</i>
agens ≠ figure	verbs of transport <i>werfen, tragen, bringen</i>	verbs of caused change of position <i>stellen, setzen, legen, hängen</i>

Tabelle 2.9: Klassen von Bewegungsverben aus Eschenbach et al. (2000)

c. *That mountain range goes from Mexico to Canada.* (Talmy 1996: 215)

(2.139) *La route entre dans la forêt*
Le canal va à la mer

(2.140) *La falaise plonge dans la mer*
,Die Felswand fällt ins Meer ab‘ (Gross 1995a: 179)

Nicht alle Bewegungsverben vermögen so eine „fiktive Bewegung“ auszudrücken, z.B. **fly* in (2.138). Die Verben können auch nicht immer 1:1 übersetzt werden – im Deutschen sind *sich erstrecken, reichen* Entsprechungen zu *go* aus (2.138). Deshalb müssen diese Verben – abseits aller Überlegungen zur Herleitung dieser Lesarten über Metaphern und kognitive Funktionen – werden als Einheiten beschrieben, die Eigenschaften von oder Relationen zwischen Orten (geographischen Entitäten) ausdrücken.

Das Konzept der „fictive motion“ oder „abstrakter Pfade“ wird von Pribbenow (1991: 609–10) ausgedehnt auf Objekte wie Wege und Flüsse: „But it is worth mentioning that objects of a special elongated form such as streets and rivers [...] or a collection of objects forming a line [...] can also induce an abstract path“. Das Beispiel, das die Autorin nennt (*Die neue Eisenbahnlinie führt von Hamburg nach München.*), ist aber denkbar ungeeignet. Eisenbahnlinien sind Entitäten, die gebaut werden, um Orte zu verbinden (hier *Hamburg* und *München*), und nicht, um sich beliebig durchs Gelände zu schlängeln. Die beiden Orte müssen also als Aktanten von *Eisenbahnlinie* betrachtet werden.

2.5.3 Zwei Fallbeispiele: *wohnen* und *liegen*

Zum Abschluss des Abschnitts über lokative Verben sollen zwei Beispiele besprochen werden: *wohnen* und *liegen*. Beide werden vielfach in der Literatur erwähnt, weil sich in ihnen eine Reihe von Problemen manifestieren: Valenz und Selektionsklassen, Adverbialergänzung vs. stark regierte Präposition, sowie die Problematik der ‘appropriate words’ (Napp).

2.5.3.1 *wohnen*

In meinem Bett hab ich bestimmt schon mehr als 10⁵⁸ Tonnen Weltall hingeträumt, denn ich wohn schon lange hier. Wohnen ist eine sonderbare Tätigkeit. Man wohnt und wohnt und merkt es nicht. Wohnen ist juristisch das, was biologisch atmen ist, obgleich man seinen Atem doch manchmal zur Kenntnis nimmt, wenn man sich verschluckt oder nach der Bahn rennt. Wohnen müsste ein Geräusch machen, knacken oder leise singen, damit es als Aktion

bemerkbar würde. Man kann auch nicht ausschließlich wohnen, man raucht, schläft oder steht dabei. Ich würde gern mal versuchen, eine Minute lang nur zu wohnen, das wäre eine wunderbare Meditation. Wohn, wohn, wohn. Man müsste intensiver wohnen. Schließlich ist das ein Grundrecht, das der gütige Staat mir gewährt. Grundrecht, ick genieße dir. Der Genuß von Grundrechten führt leider rasch zu Ermüdungserscheinungen, eine Erkenntnis, die die Väter und Mütter der Demokratie wohl gründlich verdrösse.

(Max Goldt „Zehn hoch achtundfünfzig“ aus „Die Radiotrinkerin“)

Das Verb *wohnen* dient in der Literatur als Paradebeispiel für ein Verb, das (1.) ein obligatorisches lokatives Komplement erfordert und (2.) dessen Valenz schwierig zu beschreiben ist. Vor allem die Valenzgrammatik hat sich mit *wohnen* beschäftigt, deshalb ein Blick in Valenzwörterbücher und Valenz-Literatur:

Helbig & Schenkel (1975) ordnen *wohnen* zwei obligatorische Aktanten zu, deren zweiter als prädikatives Adjektiv oder als präpositionale Ortsangabe, eingeleitet durch eine Reihe lokaler Präpositionen, realisiert werden kann:

- I. wohnen₂ [= zwei obligatorische Aktanten]
- II. wohnen → Sn, pS/Adj
 - Sn → Hum (*Der Lehrer wohnt in der Schule.*)
 - p = bei, neben, in ... (lokale Präpositionen)
- III. pSd → Loc (*Er wohnt bei seinem Freund, in der Schule, neben dem Gasthaus.*)
 - Adj → Mod (*Mein Vater wohnt gut.*)

Die Beschreibung in Gerling & Orthen (1979: 299) erfolgt ebenfalls als zweiwertiges Verb:

¹wohnen – ‚Heim, Wohnung, Wohnsitz, ständigen Aufenthaltsort haben‘

zweiwertig: (+LEB), (LOC)

²wohnen – ‚innewohnen‘: *In ihm wohnt unbändiger Ehrgeiz*

Die Adjektivergänzung grenzen Gerling & Orthen dagegen aus: *Wir wohnen zur Miete, Meyers wohnen gut* „stellen Varianten dar, die als Funktionsverb, bzw. Adjektivverb (Heringer) nicht in den Darstellungsbereich gehören.“ Der erste Aktant werde typischerweise durch einen Menschenbezeichner realisiert, die Autoren betonen aber die Erweiterung per Analogie auf andere Lebewesen: „In der semantischen Beschränkung der Mitspieler ist für a meist (MENSCH) typisch. (TIER) gilt nur für einige Tierarten, die ebenfalls einen ‚ständigen, überdachten Aufenthaltsort‘ haben.“ (*Bären wohnen in Höhlen. *Schmetterlinge wohnen auf Blüten*).

Schumacher et al. (2004: 850–1) unterscheiden zwei Lesarten: 1. ‚irgendwo vorübergehend eine Unterkunft haben; untergebracht sein‘ und 2. ‚irgendwo irgendwie seine Wohnung haben und dort ständig leben‘. Der erste Aktant wird in beiden Lesarten durch eine Person realisiert. Als zweiter Aktant wird für 1. ein Adverb des Ortes angesetzt, repräsentiert durch „meist Hotel o.Ä.“ mit der Präposition *in* oder „indirekt“ durch eine Person mit *bei*. Für 2. besteht die Wahl zwischen einem Ortsadverb und einem Adverb der Art und Weise (*möbliert, zur Untermiete, ganz allein*). Das Ortsadverb kann dabei durch einen „Ortspunkt“ mit beliebiger Präposition (*an, auf, in der Nähe*) oder durch eine Person und *bei* realisiert werden. Die Adverbialergänzung kann für 2. entfallen wenn das Verb „im Kontrast verwendet wird: *Der Werbespruch von IKEA lautete: ‚Wohnst Du noch oder lebst Du schon?‘*“.

Der Verweis von Schumacher et al. auf den Ikea-Spruch trägt nicht zur Klärung der Verhältnisse bei, da *wohnen* den Eliminierungstest besteht, ebenso wie den *tun/geschehen*-Test, die Temporalsatzparaphrase und den *und zwar*-Test (vgl. Kap. 2.3.3).

Eroms (1981: 213–5) beschreibt zunächst die Valenz von *wohnen* als zweiwertig, die „Lokal- oder Modalphrase“ sei dabei („bis auf Kontrast- oder Nektionsfälle“) „obligatorisch“. Die Ortsangabe ist nicht durch (2.141) paraphrasierbar:

(2.141) * *Karl befindet sich wohnend auf dem Lande.*

(2.142) a. *Karl wohnt ruhig auf dem Lande.*

b. = *Karls Wohnort ist auf dem Lande. Karls Wohnort (Wohnung) ist ruhig.*

Mit der Analyse der Paraphrase in (2.142) trifft er den Kern des Problems:

Der Grund, warum *wohnen* stets zweiwertig verwendet werden muss, lässt sich nun angeben: Sätze, die ein Verb des Wortfeldes *wohnen* enthalten, sind noch um eine Stufe kompakter als reine Situativa. Diese setzen Individuenterme und Ortsausdrücke in (direkte) Beziehung, das heißt, sie lokalisieren Individuen und bezeichnen darüber hinaus die Art der Befindlichkeit der genannten Individuen. Sätze mit Verben des Wortfeldes *wohnen* sind kompakte Ausdrücke, die über Orte aussagen, an denen Individuen dauernd situiert sind (=,wohnen‘). Mit ihnen verbundene Modalangaben sind kompakte, das heißt verkappte, Prädikationen über den Wohnort z. (Eroms 1981: 214)

Auch wenn die syntaktische Struktur ein zweiwertiges Prädikat vermuten lässt, ist die semantische Struktur, wie sie in den Paraphrasen aufscheint, komplexer. Wenn man davon absieht, dass unbekannt ist, ob Karl ein Haus (oder eine Wohnung, einen Wohnwagen, o.ä.) bewohnt, lässt sich Eroms’ Beispiel weiter umformulieren:

(2.142a’) *Das Haus, in dem Karl wohnt, liegt ruhig auf dem Lande.*

Erklären lässt sich das Phänomen auch über Harris’ ‚appropriate words‘ (Napp). *Wohnung* oder *Haus* sind so naheliegende Argumente von *wohnen* – und dabei so wenig informativ, dass sie eliminiert werden können. Offen bleibt aber, wie die Valenz von *wohnen* nun linguistisch zu beschreiben ist, oder genauer, ob und wie das besondere syntaktische Verhalten lexikographisch erfasst werden soll. Im Detail geht es um die Frage, ob sich die verschiedenen syntaktischen Konstellationen auch in verschiedenen Bedeutungen oder Lesarten widerspiegeln, d.h. ob *wohnen* polysem ist oder nicht. Dazu äußert sich Le Pesant (2000: §II.5.2), der die Situation ansonsten ähnlich analysiert:

[...] ces prédicats (*habiter, s’installer, s’établir, se retirer, se fixer, être un habitant de, habité, inhabité, désert, abandonné ...*) sélectionnent aussi bien les <logements>, que les <terrains naturels> (e.g. *vallée, colline*), les <territoires> (e.g. *région, pays, ville, massif montagneux*), voire les <voies urbaines> :

J’habite dans cette maison depuis un an

J’habite dans cette vallée depuis un an

J’habite dans cette région depuis un an

J’habite dans cette rue depuis un an

[...] Le fait remarquable est que les compléments d’objet locatif peuvent être cumulés :

J'habite dans une région rurale, dans un appartement, au fond d'une vallée, dans une rue tranquille

[...] on observe une *association ordinaire cible/site* :

Mon appartement est situé dans une rue tranquille, dans une vallée

[...] Faut-il en déduire que les prédicats d'*habitation* sont polysémiques ? Y aurait-il quatre emplois différents, suivant que ces prédicats sélectionnent les <logements>, les <voies urbaines>, etc. ? C'est contraire à l'intuition linguistique. Ces mots ne semblent pas changer de sens suivant qu'elles sélectionnent telle ou telle classe d'objets.

Le Pesants Argument, es widersprache der linguistischen Intuition, *wohnen* als polysem anzunehmen, ist ohne jegliche empirische Grundlage – mit „linguistischer Intuition“ lässt sich alles begründen. Versucht man eine detaillierte lexikographische Beschreibung des Verbs *wohnen* und verwandter Lexeme, kommt man nicht umhin, mehrere Bedeutungen für *wohnen* anzusetzen. Die Bedeutungsexplikationen der Lesarten unterscheiden sich, haben aber auch Bestandteile gemeinsam. Schließt man Synonyme, Derivate und Transformationen mit in die Beschreibung ein, wird außerdem offensichtlich, dass die Lesarten ganz unterschiedliche Möglichkeiten zur Paraphrasierung bieten:

- (2.143) a. *Max wohnt in einer Dachgeschosswohnung.*
b. *Max bewohnt eine Dachgeschosswohnung.*
c. *Max ist **B**ewohner einer Dachgeschosswohnung.*
- (2.144) a. *Max wohnt in München.*
b. * *Max bewohnt München.*
c. *Max' Wohnsitz ist München.*
d. *Max ist in München wohnhaft/ansässig.*
e. *≈ Max ist in München polizeilich gemeldet.*
f. *Max hat seinen ständigen Aufenthalt in München.*
g. *Max ist ein **Ein**wohner Münchens.*
h. *Max ist Münchner. (!)*
- (2.145) a. *Max wohnt nördlich von München.*
b. *Max' Wohnort liegt nördlich von München.*
c. *Das Haus, das Max bewohnt, steht nördlich von München.*

Die Bedeutungen von *wohnen* in (2.143) und (2.144) müssen voneinander unterschieden werden, vgl. die entsprechenden Wörterbucheinträge in Kap. D.1. Für (2.145) lässt sich diskutieren, ob hier eine eigene Lesart vorliegt, oder nur eine Variante mit *Napp* von *wohnen* in (2.143). Wie auch immer: *wohnen* hat mehrere Lesarten, die einander nahe sind und viele Bedeutungsaspekte teilen, aber (1.) über keinen gemeinsamen Satz von Paraphrasen verfügen und (2.) unterschiedlichen Transformationen unterliegen. Transformationen sind ein Aspekt, der insbesondere die Passivtransformation des englischen Verbs *inhabit* betrifft – dt. *wohnen/bewohnen* werden ja eher als zwei Lexeme aufgefasst.

Consider the verb *to inhabit*, which has no passive in

- (3) *Max inhabits Manhattan.*

* *Manhattan is inhabited by Max.*

But when the subject is plural or ‘collective’, we have

- (4) *Rich politicians inhabit Manhattan.*
Manhattan is inhabited by rich politicians.

[...] Such pairs show that, for a given verb, existence of passive forms depends on use, metaphorical or not, of the verb—a use linked to particular pairs of NP’s. (Gross 1979: 863–4)

Für *inhabit* besteht das deskriptive Problem in der Verwendung der Passivformen, für (*to*) *live* ist es das sog. „preposition stranding“ (vgl. (2.125)):

The problem of whether only governed prepositions can appear in this pseudopassive construction [...], remains, in our view, unsolved. [...] T. Langendoen (1970: 38–39 [=Essentials of English grammar]) points out that the sentence

- (7) *This house was lived in by George Washington.*

is grammatically correct, while

- (7') * *Virginia was lived in by George Washington.*

and, likewise,

- (7'') * *This house was remained in by George Washington.*

was not. One possible explanation of the above discrepancy is that in *live in a house* the verb LIVE is the value of the lexical function Real₁ of *house*, and the preposition *in* in that phrase is not only governed but also semantically empty. As for the phrases *live in Virginia* and *remain in a house*, in which the verbs are not values of any lexical function, the preposition *in* that occurs in them, while also being governed by the verb, is not semantically empty, cf. *live near* ⟨*far away from*⟩ *Virginia*, *remain behind* ⟨*in front of*⟩ *a house*, etc. Thus one might easily conclude that for a preposition to be admissible in the construction described [...], it is not enough that it be governed; it is also necessary that it be semantically empty.

(Mel'čuk & Pertsov 1987: 344)

Damit fügen Mel'čuk & Pertsov noch einen Baustein zur Überlegung Eroms' hinzu. Wenn Sätze wie *Karl wohnt in München* „Prädikationen über den Wohnort“ sind, kann die Präposition, die darin eine räumliche Relation ausdrückt, nicht semantisch leer sein, was sich auch darin äußert, dass sie variiert werden kann. Ist *wohnen* aber lexikalische Funktion (Real₁(*Haus*)=*wohnen in, bewohnen*), ist die Präposition *in* stark regiert und nicht variiierbar.

Um damit die Diskussion abzuschließen:⁵² die Beschreibung der Valenzen von *wohnen* und ihrer syntaktischen Realisierung ist und bleibt eine Herausforderung für den Linguisten. Zwei Prinzipien sollten Berücksichtigung finden:

1. für verschiedene Konstellationen aus dem Prädikat und verschiedenen Typen von Argumenten muss genau bestimmt werden, welche Bedeutung das Prädikat hat. Lässt sich also z.B. für die vier folgenden Sätze die Bedeutung von *wohnen* auf die gleiche Weise einem Nicht-Muttersprachler erklären?⁵³

52 Die Diskussion ist damit sicher nicht zu Ende, insbesondere Kollokationen und phraseologischer Ausdrücke (*unter der Brücke wohnen*) blieben unbehandelt.

53 Für den Muttersprachler: Bezeichnen die vier Sätze die gleiche ‚linguistische Situation‘, von der Variation der Aktanten abgesehen?

- (2.146) a. *Boris Becker wohnt in einer schicken Villa.*
 b. *Boris Becker wohnt in München, nicht in Monaco.*
 c. *Boris Becker wohnt im Sheraton.*
 d. *Die Etrusker wohnten südlich des Apennin.*

Wohnen muss dann möglicherweise als polysem beschrieben werden. Aber Bedeutungsexplikationen sind immer ein strittiger Punkt, und manche Linguisten werden den gemeinsamen Kern der verschiedenen Lesarten als die eine und „wesentliche“ Bedeutung ansetzen.⁵⁴

2. die nächste Aufgabe besteht darin, für die Konstellationen oder die bereits voneinander abgegrenzten Lesarten, verschiedene Transformationen und Paraphrasen zu testen.

Bei jeglicher Art von sprachlicher Informationsverarbeitung laufen die beiden Punkte Bedeutungsexplikation und syntaktische Beschreibung zusammen: verschiedene Bedeutungen bedingen ja verschiedene Informationen, Verb- und Objektklassen dienen vor allem zur Bedeutungsdisambiguierung, und die Beschreibung von Paraphrasen fasst die gleiche Information, die in verschiedenen Formen realisiert wird, zusammen.

2.5.3.2 *liegen*

Zur sprachlichen Situation von ‚*liegen*‘ gehört, dass ein Objekt *X* in horizontaler Lage sich oben auf einer waagrechten Oberfläche *Y* befindet und mit dieser Kontakt hat. *Y* ist ein Aktant von *liegen* (Mel’čuk 2004b: 11) und wird über die Präposition *auf* angeschlossen. In allen anderen Fällen handelt es sich entweder um Supplemente bzw. Tilgungen von Komplementen in ursprünglich komplexeren Sätzen,

- (2.147) a. *Die Zeitung liegt in der Küche.*
 b. = *Die Zeitung liegt in der Küche auf dem Tisch.*

oder um Verwendungen von *liegen* als Stützverb:⁵⁵

- (2.148) a. *München liegt in Bayern.*
 b. **München liegt.*

Maienborn (1991: 627) versucht die Problematik anhand der Objekteigenschaften des lokalisierten Objekts (Figure) zu erklären: Wenn Objekte wie *Zimmer* oder *Berge* keine alternativen Positionen (*liegen*, *stehen* etc.) erlauben, kann der Focus nicht auf die Art der Position gelegt werden. Damit ist **Das Zimmer liegt* nicht möglich. Aber wie sind dann nachfolgende Dubletten (2.149) zu erklären, die die gleiche Situation beschreiben. Hier bleibt eigentlich nur die Beschreibung von *liegen/stehen* als Stützverb, offen bleibt die Frage, was die Wahl des Stützverbs steuert (siehe Kap. 2.4 auf S. 43).

54 Linguisten sind bei Nomina weitaus großzügiger mehrere polyseme Bedeutungen anzusetzen, vgl. Kap. 3.1.4 und klassische Beispiele wie *Kirsche* als ‚Frucht‘, ‚Baum‘, ‚Holz‘ und ‚Farbe‘. Der Grund dafür ist, dass man versucht, Objekte – die meist als Nomina realisiert sind – über die Hyponymierelation zu Taxonomien anzuordnen. Das macht Polysemien offensichtlich. Bei Verben wird die Hyponymierelation dagegen kaum verwendet, was auch daran liegt, dass sie für Verben kaum zur Klassifikation taugt (vgl. Fellbaum ed. 1998).

55 Siehe Kap. 4.4, das Satzchemata (elementare Sätze) mit Präposition und Stützverb behandelt.

- (2.149) a. *Die Burg Boetzelaer steht in Appeldorn, einem Stadtteil von Kalkar im Kreis Kleve.*
 b. *Die Jugendherberge Burg Schwaneck liegt in dem wunderschönen Ort Pullach direkt oberhalb der Isar.*

2.6 Lokative (quasi-)prädikative Nomina

Nomina, die unter die Definition eines lokativen Prädikats (Kap. 2.3.2) fallen, verteilen sich auf zwei Gruppen. (1.) Nomina, die traditionell als prädikativ bezeichnet werden, insbesondere deverbale (*die Reise nach Afrika*) und (2.) „quasiprädikative“ (Mel’čuk 2004b: 8), die Aktanten an sich binden, aber Objekte oder Entitäten bezeichnen (*die Fähre nach Husum*).

Bezüglich der syntaktischen Realisierungsmöglichkeiten der semantischen Aktanten unterscheiden sich die quasiprädikativen Nomina von den prädikativen. Letztere bezeichnen ja unmittelbar eine (linguistische) Situation, z.B. *Peters_X Aufenthalt_{SIT} in Moskau_Y*. Bei den Quasiprädikaten, kann es sein, dass der erste Aktant X, nicht als syntaktischer Dependent des Nomens realisiert werden kann, d.h. syntaktisch geblockt ist (Mel’čuk 2004b: 55). Z.B. bezeichnet *Präsident* eine Relation zwischen einer Person X und einem Land/Staat Y. Aber nur Y wird vom Nomen unmittelbar regiert (*der Präsident der Russischen Föderation*). Die Person X kann nur per Apposition (*der russische Präsident Putin*) oder Stützverbkonstruktion (*Putin bekleidet das Amt des Präsidenten in Russland*) angeschlossen werden. Ähnliche Zwitterposition – *Präsident* bezeichnet einerseits die Relation, referiert andererseits auf die Person X – sind auch bei „echten“ prädikativen Nomina zu beobachten. So bezeichnet *Ausrüstung* zum einen die Situation selbst, zum anderen den zweiten Aktanten der Situation: *die Ausrüstung der Armee mit Atomwaffen – Atomwaffen gehören zur Ausrüstung der Armee* (vgl. Apresjan 1973: 21 und Kap. 3.1.4). Oftmals entstehen so aber verwirrende Situationen, wo es schwer ist die Stelligkeit eines Nomens zu bestimmen. Hier ein Fallbeispiel:

Die Nomina *Mündung* und *Zufluss* können als Ableitungen zweier synonymen Verben (*zufließen, münden*) die gleiche Situation bezeichnen:

- (2.150) a. *Die Tauber fließt dem Main zu.* (selten, veraltet)
 b. *Der bedeutendste Zufluss des Mains auf dem Gebiet Baden-Württembergs ist die Tauber.*
 (2.151) a. *In Wertheim mündet die Tauber in den Main.*
 b. *Wertheim liegt an der Mündung der Tauber in den Main*

Allerdings bezeichnet *Mündung* nicht ‚die Tatsache, dass ein Wasserlauf X in ein Gewässer Y fließt‘, und auch nicht ‚den Wasserlauf X, der in ein Gewässer Y fließt‘, sondern ‚die Stelle, wo ein Wasserlauf X in ein Gewässer Y fließt‘. Die Stelle des Einmündens ist ein Ort, Orte sind Entitäten, damit stellt sich die Frage, ob die Stelle selbst ein Aktant von *Mündung* ist. Sätze, in denen die Zuflussstelle realisiert ist, lassen sich finden, sind aber selten:

- (2.152) *Die Boca ist die Mündung des Riachuelo in den Rio de la Plata*
 (2.153) *Der Meeresarm von Burry ist die Mündung des Loughor-Flusses.*
 (2.154) *Die Camargue ist das Mündungsgebiet der Rhône ins Mittelmeer.*

Letztlich ist es eine Frage der Ontologie (der Weltsicht), ob Entitäten der Klasse MÜNDUNG existieren. In Tab. C.6 wurde die Frage positiv beantwortet und *Mündung* als trivalent beschrieben.

Syntaktisch realisiert werden können die Aktanten eines Nomens als (vgl. Sommerfeldt & Schreiber 1983: 19–24):

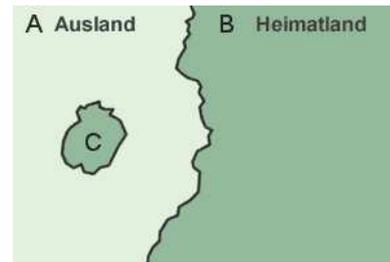
- Substantiv im Genitiv: *Peters Reise*
- Präposition und Substantiv: *die Reise nach Afrika*
- Beziehungsadjektiv: *der amerikanische Präsident*⁵⁶
- Possessivpronomen: *seine Reise*
- Fugenform: *Peters Afrikareise*; auch mit Bindestrich: *der US-Präsident*

Die lokativen (quasi)prädikativen Nomina verteilen sich im Wesentlichen auf die folgenden Gruppen:

- deverbale Nomina (*Fahrt, Abflug, Aufenthalt*) und einige sich syntaktisch sehr ähnlich verhaltende prädikative Nomina (*Urlaub, die Sommerfrische in Hohenschwangau*)
- durch Quasiprädikative ausgedrückte Relationen zwischen Menschen (oder Organisationen) und Orten, wobei entweder Mensch oder Ort Referent des Nomens sein können:
 - Mensch als Referent: *Einwohner, Präsident* (eines Landes)
 - Ort als Referent: *Wohnsitz, Firmensitz*
- Relationen zwischen Orten, meist zwei (2.155)–(2.156), manchmal auch drei Orten (2.157):
 - (2.155) *Moroni ist die Hauptstadt der Komoren.*
 - (2.156) *Thüringen ist Bundesland von Deutschland.*
 - (2.157) *Bremerhaven ist eine Exklave des Bundeslandes Bremen in Niedersachsen.* (vgl. Abb. 2.5)
- (öffentliche) Verkehrsmittel: *der Zug nach Berlin*. Als Paradebeispiel für Quasiprädikate (Mel'čuk 2004b: 8), blieben die vielfachen Möglichkeiten, Nomina argumentartig innerhalb einer NP anzuschließen, nicht unbeachtet. Vgl. die Beispiele bei G. Gross (1998b: 223, 1994: 23–6, 1996: 154–155): *le train de 9h, le train de+pour Marseille*, die Synonymie mit Wörtern wie *Verbindung, Reisemöglichkeit*, und die Kombination mit Verben, die auch mit anderen Nomina als Stützverben auftreten, z.B. *den Zug nehmen, der Zug geht um 9 Uhr*, siehe Fußnote 63 auf S. 83.
- Verkehrswegen den Status eines Prädikats zuzuschreiben erscheint exotisch. In der Linguistik werden sie entweder als abstrakte Pfade (siehe Kap. 2.5.2.4) beschrieben, oder wie bei Mathieu-Colas (1998) als Objekte ohne Aktanten betrachtet. Verben wie *aller* ‚gehen‘ oder *conduire* ‚führen‘ werden dann als ‚prédicats appropriés‘ (siehe S. 80) beschrieben. Dass Verkehrswege auch im Französischen Argumente direkt an sich binden können (*première pause à l'entrée de l'autoroute à Nancy*), übersieht Mathieu-Colas. Sinnvoller erscheint eine Beschreibung von Verkehrswegen als Quasiprädikate, wie sie für einige wenige in Kap. D.1.3 versucht wurde.

56 Vgl. das Beispiel von Sommerfeldt & Schreiber *der Rat der Schwester – der schwesterliche Rat* und Mel'čuk (2006: 186) für Adjektive in der Rolle des tiefsyntaktischen Aktanten. Adjektivisch kodierte Aktanten spielen bei lokativen (quasi)prädikativen Nomina eine wichtige Rolle, vgl. die Paraphrasierungsbeispiele in Kap. 3.5.1.5.

Abbildung 2.5: *Enklave* und *Exklave* bezeichnen eine ternäre Relation: *C* ist *Exklave* von *B* in *A*, vgl. (2.157) (Quelle: WIKIPEDIA.DE)



2.7 Adjektive

Adjektive haben, wenn sie nicht paradigmatisch für Nomina stehen (z.B. in *die französische Regierung*) immer mindestens eine semantische Valenz, die aber syntaktisch nicht als Valenz realisiert wird, sondern durch das regierende Substantiv, also ein externes Argument ausgefüllt wird (Zifonun 2006: 374). Im Deutschen verfügen Adjektive über voneinander verschiedene Formen für attributive und prädikative Stellung. In letzterer können sie Prädikat eines elementaren Satzes sein, siehe Kap. 4.7 und Tab. C.7 für die Beschreibung solcher elementarer Sätze.

Klammert man Dimensionsadjektive aus (vgl. Bierwisch & Lang eds. 1987), wurden räumliche Adjektive im Deutschen nicht gesondert beschrieben. Im Rahmen der franz. Lexikogrammatik existiert mit Laporte (2005) ein teils missglückter Beschreibungsversuch (s.u.). In allgemeinen Arbeiten zur Adjektivvalenz (z.B. Lee 1994) werden lokative Adjektive zwar behandelt, doch scheint diese Gruppe, vor allem wenn man nur die mind. zweiwertigen zählt, sehr klein. Sommerfeldt & Schreiber (1977: 153–4) beschreiben z.B. das Adjektiv *entfernt* mit zwei Bedeutungen *entfernt*¹ = ‚fern gelegen, entlegen‘ und *entfernt*² = ‚sich in einer bestimmten Entfernung von einem Ort befindend‘.⁵⁷ Für die zweite Lesart geben die Autoren drei Realisierungsmöglichkeiten an:

- (2.158) attr *der 5 km von hier entfernte Hafen*
 präd *Der Hafen ist 5 km von hier entfernt.*
 adv *Der Hafen liegt 5 km von hier entfernt.*

Laporte (2005) beschreibt eine Gruppe adjektivischer Prädikate, die durch Konstruktionen mit den folgenden Eigenschaften charakterisiert sind:

1. N_1 bezeichnet einen Teil von N_0 :

- (2.159) *la façade de la maison*
 N_1 N_0
 ,die Fassade des Hauses‘

2. dem liegt der elementare Satz zugrunde:

⁵⁷ Interessant in Bezug auf die Arbeit der Autoren ist ein Vergleich mit dem später erschienen Wörterbuch der Substantive (Sommerfeldt & Schreiber 1983). Hier werden für *Entfernung* = ‚kürzeste Verbindung zwischen zwei Punkten‘ nur zwei Aktanten angesetzt, nämlich die beiden Punkte, nicht aber die tatsächliche Entfernung. Das unterscheidet die oberflächlichen Arbeiten der deutschen Valenzgrammatiker von den gründlichen Analysen im Rahmen des Inhalt \Leftrightarrow Text-Modells. Dort wäre $A_0(\textit{Entfernung}) = \textit{entfernt}$, die Anzahl der Aktanten muss gleich sein, da beide Lexeme die gleiche linguistische Situation bezeichnen.

(2.160) *Ceci est la façade de la maison.*

3. diese Konstruktion lässt sich um ein lokatives Adjektiv ergänzen, das als Prädikat angenommen wird, mit dem Schema *être N₁ Adj de N₀*:

(2.161) *Ceci est la façade sud de la maison.* ‚Das ist die südliche Fassade des Hauses‘

4. *N₁* ist *N_{app}* (siehe Kap. 3.1.3) des Adjektivs und kann deshalb in vielen Fällen entfallen:

(2.162) a. *Ceci est la région australe du Chili.* ‚der südliche Bereich Chiles‘
b. = *Ceci est le Chili austral.* ‚das südliche Chile‘

Auf den ersten Blick handelt es sich in (2.162) um appropriate nouns (*N_{app}*). Allerdings lässt sich der Effekt auch erklären, wenn man zwei separate Bedeutungen (neben weiteren) von *südlich* ansetzt, auf die sich die beiden Sätze in (2.162) verteilen lassen: (1.) *das östliche X von Y = X* bezeichnet eine Teilregion von Y, X liegt im Osten von Y; und (2.) *das östliche X =* im Osten gelegener Teil oder Teilregion von X.

Laporte bezieht noch von Meeresnamen abgeleitete Adjektiven mit ein, die in an der syntaktischen Oberfläche in identischen Konstruktionen vorkommen:

(2.163) *la côte atlantique de la France*

Da natürlich auch die „Nominalisierung“ der Adjektive wie in (2.163) konsequent als Prädikate behandelt werden müssen, würden Schwarzes Meer (*pontique* → *Pont-Euxin*), Atlantik u.a. zu zweiwertigen Prädikaten. Als „une difficulté“ beobachtet Laporte (2005: 8–9), dass sich die Nominalisierungen nur schwer als Prädikat in einem Satz verwenden lassen. Einzig Konstruktionen mit *avoir/posséder* sind möglich:

(2.163') a. * *Ceci est l'Atlantique de ce département*
b. *Ce département possède un littoral atlantique.*
c. *Ce département possède un littoral sur l'Atlantique.*
N₀ avoir un N₁ Loc Adj-n

Diese Schwierigkeit umgehen lässt sich nur, indem man *côte sud* und *côte atlantique* verschieden behandelt. *Sud* ist ein adjektivisches Prädikat, *atlantique* füllt dagegen einen Aktantenslot von Prädikaten wie *littoral* oder *côte*: *N₀ ist die Küste von N₁* oder *N₀ ist die N₁adj Küste* (vgl. Kap. 3.5.1.5 und 4.6 für weitere Beispiele).

2.8 Adverbien

Adverbien definiert Becker (1994: 33) als: „Adverbien gehören zu derselben syntaktischen Kategorie wie Präpositionen [...], binden aber kein internes Argument an sich. Das Relatum bleibt implizit und muss aus dem Kontext geschlossen werden.“ Dass eine solche Definition zu Schwierigkeiten bei der Abgrenzung von Adverbien gegenüber Präpositionen führt, darauf wurde bereits in Kap. 2.4 eingegangen. Die Definition ist aber auch insofern falsch, da es tatsächlich Wortformen gibt, die weder als Präposition noch als Adjektiv – Adverbien sind indeklinabel und können nicht als vorangestelltes Attribut eines Nomens realisiert werden – klassifiziert werden können, aber Aktanten auch syntaktisch an sich binden:

(2.164) *Max ist unterwegs nach Italien.*

(2.165) *Max ist auf dem Weg nach München.*

Nimmt man komplexe Adverbien wie in (2.165) hinzu, ist die Gruppe solcher mindestens zweiwertiger Adverbien nicht ganz so klein. Beschrieben werden als Prädikat verwendete lokative Adverbien in Kap. 4.7, für Adverbien als Füller einer Argumentstelle siehe Kap. 3.5.3.

3 Lokative Argumente

Drei Fragen werden in diesem Kapitel behandelt:

(1.) Was sind Lokativa? Diese Frage ist zentral für die vorliegende Arbeit, da lokative Sätze ja dadurch definiert sind, dass ihr Prädikat mindestens einen lokativen Aktanten verlangt (siehe Kap. 4). Dieser ist im Normalfall durch ein Lokativum, d.h. ein Element einer Objekt- oder Selektionsklasse LOKATIVA, realisiert.

Lokativa werden die Lexeme genannt, die „Orte, die als Aufenthaltsorte für Menschen geeignet sind“ bezeichnen. Die Gründe für diese pragmatische und anthropozentrische Definition sind in Kap. 3.1.1 zu finden.

(2.) Wie ist die Gruppe Lokativa weiter untergliedert? Weshalb eine hierarchische Organisation sinnvoll, ja notwendig ist, wird in Kap. 3.1.2 besprochen. Wie und nach welchen Kriterien die Untergliederung erfolgt, dazu sei auf die Kap. 3.1.2, 3.4.1.1 und 3.4.2.1 verwiesen.

(3.) Wie werden lokative Argumente lexikalisch und syntaktisch realisiert, d.h. durch welche Wortarten und Konstruktionen? Dieser Problematik ist der weitaus größte Teil des Kapitels gewidmet.

(3a.) Lokative Argumente können als Nomina, Eigennamen, Adjektive, Adverbien oder als Teil eines Nominalkompositums realisiert sein. Lokativa im engeren Sinn sind natürlich die lokativen Nomina (franz. ‚noms locatifs‘ – *Nloc* in der Notation der Lexikongrammatiker), z.B. *Zimmer*, *Haus*, *Garten*, *Stadt*, *Fluss*.

Die Lokativa lassen sich hierarchisch organisieren, durch Hyponymie- (*Gewässer* → *Wasserlauf* → *Fluss*) oder Meronymierelationen (*Zimmer* ist Teil von *Haus*). Viele der untergeordneten Klassen von *Nloc* haben die Eigenschaft, dass ihre Instanzen durch Eigennamen (‚Ortsnamen‘ oder ‚Toponyme‘) benannt sind: *Frankreich*, *Paris*, *Seine*. Die Toponyme bilden den Schwerpunkt dieses Kapitels (Kap. 3.3 und 3.4): zum einen übertreffen sie die anderen Lokativa zahlenmäßig weit, zum anderen sind sie sehr informativ und damit relevant für Anwendungen im IR. Argumentstatus haben auch Adjektive (**französische** *Regierung*), Fugenformen (**Italien***reise*) oder Adverbien (*Max wohnt hier*), siehe Kap. 3.3.2.11, 3.3.2.12 bzw. 3.5.3. In Paraphrasen (*französische Regierung* = *Regierung Frankreichs*, *Italienreise* = *Reise nach Italien* = *nach Italien reisen*, siehe Kap. 2.3.5 und 3.5.1.5) wird dies deutlich.

(3b.) Abschnitt 3.5 widmet sich den syntaktischen Konstruktionen (Phrasen), in die lokative Argumente eingebettet sind, vornehmlich Nominal- und Präpositionalphrasen, die wiederum kombiniert werden können. Aber auch sehr spezifische Phrasentypen wie z.B. Adressen oder Koordinatenangaben müssen behandelt werden.

3.1 Einführung

3.1.1 Was sind Lokativa? – Versuch einer Definition

Was sind Lokativa? Eine anthropozentrische Definition, wie sie Ackermann (1999: 8) vorschlägt („Lokativa sind alle Orte, an denen sich Menschen aufhalten können.“),⁵⁸ erscheint hier angemessen. Diese Definition schließt zwar viele vorwiegend über ihre räumlichen Eigenschaften definierten Lexeme aus (z.B. BEHÄLTER), wodurch auch eine Reihe von primär lokativen Prädikaten (*X legt Y in Z*) ausgeklammert werden. Aber da nahezu jedes Konkretum auch lokativ verwendet werden kann, wird so eine klare Grenzlinie gezogen. Diese Definition schließt die informativ wichtigste Klasse der Lokativa, die geographischen Namen, vollständig ein.⁵⁹

In der Lexikongrammatik bilden die Lokativa (*Nloc*) eine der großen Objektklassen⁶⁰ und werden vielfach in der syntaktischen Beschreibung und zur Verbklassifikation herangezogen (vgl. Tab. 2.2 auf S. 49). Umgekehrt kann auch die Bestimmung, was ein Lokativum ist, durch eine Reihe von Prädikaten bzw. elementaren Sätzen erfolgen, gemäß der anthropozentrischen Definition (s.o.): *Der Mensch X hält sich (in + am Ort) Y auf* oder *X reist nach Y*, usw.

Diese Prädikate oder Operatoren, die alle Elemente einer Klasse als Argumente nehmen und damit charakteristisch für diese Klasse sind, werden spezifische Prädikate („*prédicats appropriés*“) genannt.⁶¹ Für die Klasse der Lokativa sind laut Le Pesant (2000: §I.4.2) die wichtigsten spezifischen Prädikate, die diese von anderen dimensionalen Objekten (geometrischen Formen und Konkreta) abgrenzen:

- die Relation Bewohner – Umgebung, u.a. *X ist Bewohner von Y*
- die Relation Objekt – Aufbewahrungsort: *X lagert/parkt in Y*

58 Vgl. Langer (1996: 104–5):

Ich sehe als LOKATIVA einerseits Objekte an, die in erster Linie die Funktion haben, als Ort für den Aufenthalt von Menschen oder anderen Objekten zu dienen [...] Andererseits sind Lokativa Bezeichner von Teilen von Objekten (*Gipfel, Ende*) oder Räumen (*Decke, Boden*), die primär auf deren räumliche Struktur Bezug nehmen, und die unterschiedliche Abstraktionsgrade aufweisen. [...] Eine weitere, allerdings relativ kleine Gruppe von Lexemen, die ebenfalls bereits bei der Grobklassifizierung den Lokativa zugeordnet wurden, sind die RICHTUNGSANGABEN [...] Hinzu kamen einige Lexeme wie *Ferne, Nähe*, die sich relativ idiosynkratisch verhalten, die aber eindeutig primär auf eine räumliche Situation abheben. Die Klasse der LOKATIVA ist eine Gruppe von Lexemen, die aufgrund der mangelnden Kohärenz als Gesamtklasse nur ein Übergangsstadium zur Feinklassifizierung der zugeordneten Lemmata darstellen kann. Sie ist weder eine wirklich relevante Selektionsklasse noch eine taxonomische Einheit.

59 Vgl. die ebenfalls sehr pragmatische Definition eines topographischen Objekts auf S. 93. Selbstverständlich muss die Definition „Orte, an denen sich Menschen aufhalten können“ großzügig ausgelegt werden, bei Objekten wie Geysiren, Vulkanen, manchen Bauwerken usw., im Sinne: „in deren unmittelbarer Nähe“.

60 Gross (G. 1994: 20) nennt als Objektklassen der obersten Ebene: *humain, animal, végétal, inanimé concret, inanimé abstrait, locatif, temps, événement*.

61 Gross & Clas (1997) unterscheiden zwischen ‚generischen‘ Prädikaten, die eine der Objektklassen der obersten Ebene charakterisieren, und ‚spezifischen‘ zur weiteren Untergliederung:

- les *prédicats généraux* qui caractérisent tous les termes qui ont le même trait syntactico-sémantique, par exemple les *concrets* ;
- les *prédicats appropriés* qui caractérisent les divers éléments d’une classe d’objets, par exemple les *vêtements*.

- Wettererscheinungen, Wetterereignisse und Naturkatastrophen: diese selektieren GEWÄSSER und GEOGRAPHISCHE ZONEN

Ein weiterer Punkt, den die Klassifikation über spezifische Prädikate leistet, ist, verschiedene Lesarten eines Lexems auseinanderzuhalten. Dies wird an einem Beispiel von Guillet & Leclère (1992: 8) deutlich:

(3.1) *Les Gaulois ont envahi Rome* ‚Die Gallier sind in Rom eingedrungen.‘

(3.2) *Les Gaulois ont vaincu Rome* ‚Die Gallier haben Rom besiegt.‘

Rome bezeichnet in (3.1) einen Ort, in (3.2) bezeichnet es eine Gruppe von Menschen (die Römer). Die Entscheidung, ob ein *N* lokativ ist, ist also von dem Verb abhängig, dessen Komplement es ist. Dies führt natürlich zu einem Problem bei der Definition lokativer Prädikate, siehe Kap. 2.3.2. Weitere Kriterien sind zu einer endgültigen Definition notwendig. Außerdem stellen solche Fälle von (regulärer) Polysemie natürlich ein Problem für jegliche Art von Informationsverarbeitung dar (siehe Kap. 3.1.4).

3.1.1.1 Beziehung zu hyperonymen und Abgrenzung gegenüber kohyponymen Klassen

Le Pesant versucht mit Hilfe spezifischer Prädikate die Beziehungen der Lokativa zu hyperonymen Klassen zu klären und die Lokativa gegenüber kohyponymen Klassen abzugrenzen. Lokativa definiert Le Pesant (2000: §I.3.1–2) als diejenigen DIMENSIONALEN OBJEKTE (darunter geometrische Formen, Konkreta, und eben die Lokativa), deren Meronyme (Teilräume: *intérieur*, *surface*) Hyponyme des Wortes *lieu* (‚Ort, Stelle‘) sind. Diese zweifache Einbindung in die Taxonomie (siehe Abb. 3.1.1.1) soll erklären, wieso *maison* in (3.3), aber nicht in (3.4) durch *lieu*, das wiederum Hyperonym von *intérieur* ist, substituiert werden kann.

- (3.3) a. *J’habite dans cette maison*
 b. = *J’habite à l’intérieur de cette maison*
 c. = *C’est dans ce lieu que j’habite*

- (3.4) a. *Je me suis fait construire cette maison*
 b. =* *Je me suis fait construire l’intérieur de cette maison*
 c. =* *C’est ce lieu que je me suis fait construire*

Le Pesant will damit eine operationalisierbare Abgrenzung der Lokativa von anderen dimensionalischen Objekten (darunter alle Konkreta) erreichen. Was Le Pesant dabei übersieht: (1.) die Nomina der ‚localisation interne‘ wie *surface* oder *intérieur* sind selbst Prädikate, die Lokativa (im Grunde aber alle dimensionalischen Objekte)⁶² als Argumente nehmen: *das Innere der Kirche, die Oberfläche der Kugel* etc. Erst das Ergebnis dieser Prädikation (im Dt. dann *Kircheninneres, Kircheninnenraum*) ist ein Meronym des Lokativums. (2.) Die von

62 *Innere* ist also spezifisches (eigentlich generisches, vgl. Fußnote 61) Prädikat vieler zwei- und dreidimensionalen Objekte. Le Pesant (2000: §I.2.1) definiert selbst die den Lokativen übergeordneten DIMENSIONALEN OBJEKTE über die spezifischen Prädikate Dimension (*longueur, largeur, usw.*), Gestalt und Form.

Le Pesant (2000: §I.3.1) aufgeführten Beispiele (*habiter dans une maison*, *emprunter un couloir*, *cultiver un champ*, *assécher un lac*) repräsentieren fast durchweg lexikalische Funktionen der betroffenen lokativen Nomina: $\text{Real}_1(\text{Haus})=\text{wohnen in}$, $\text{CausFunc}(\text{Haus})=\text{bauen}$, $\text{Real}_1(\text{Feld})=\text{bestellen}$, $\text{LiquFunc}(\text{See})=\text{trockenlegen}$. Zwar können die Werte lexikalischer Funktionen auch vererbt werden, allerdings geschieht dies nicht regelmäßig und ist nur schwer vorhersagbar (vgl. Mel'čuk & Wanner 1996).

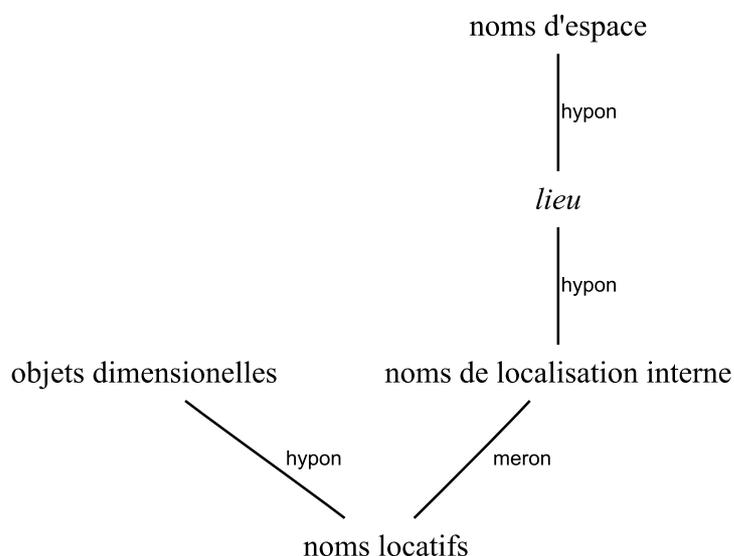


Abbildung 3.1: Übergeordnete Klassen der Lokativa bei Le Pesant (2000: §I.3.1–2)

Damit sind die Lokativa Hyponyme der DIMENSIONALEN OBJEKTE. Dass sie als solche Akтанten bestimmter räumlicher Prädikate sein können, die Teilräume und -regionen auswählen, ist selbstverständlich, reicht aber als Definition einer Klasse der Lokativa, wie Le Pesant sie versucht, nicht aus. Für eine Abgrenzung gegenüber den kohyponymen Klassen, vor allem den Konkreta, fehlt damit ein klares Kriterium. Die von Le Pesant (2000) genannten spezifischen Prädikate (s.o., *wohnen*, *lagern* usw.) sind dazu nicht geeignet – mit ihrer Hilfe ist nur eine feinere Unterteilung der Lokative möglich. Die Abgrenzung von den meisten Konkreta wird aber durch die Inkompabilität der Lokativa mit Prädikaten gewährleistet, die die Masse eines Objekts oder dessen Bewegung ausdrücken (Le Pesant & Trégouët 1997-2000).

3.1.2 Untergliederung der Lokativa

Ebenso wie die Bestimmung der Klassenzugehörigkeit, muss eine hierarchische Untergliederung der Lokativa über die Kombinatorik mit spezifischen Prädikaten erfolgen. Sie dient dem gleichen Zweck: elementare Sätze lassen sich besser beschreiben, da Prädikate in ihren Akтантenslots Elemente bestimmter Objektklassen fordern. Wird eine Objektklasse verlangt, sind alle ihre Elemente und auch Elemente hyponymer Klassen möglich. Die Kombinatorik mit bestimmten (spezifischen) Prädikaten wird als Eigenschaft an untergeordnete Klassen vererbt (3.5)–(3.6).

- (3.5) a. *Max wohnt in einer Siedlung.*
 b. *Max wohnt in (einer Stadt + einem Dorf + einem Weiler).*
 c. *Max wohnt in München+Hausen.*

- (3.6) a. *Max wohnt an einem Gewässer.*
 b. *Max wohnt an einem Meer+See+Fluss.*
 c. *Max wohnt am Mittelmeer+Chiemsee+Rhein.*
- (3.7) a. *Max wohnt auf einem Bauernhof.*
 b. *Max wohnt in+auf einem Schloss.*
 c. *Max wohnt in einem Kloster.*

Die Vererbung der spezifischen Prädikate an untergeordnete Klassen ist transitiv, wie Le Pesant & Mathieu-Colas (1998: 13) an der Kette *PRENDRE(moyen de transport) → PRENDRE(avion) → PRENDRE(airbus)* deutlich machen.⁶³ Diese Vererbung „nach unten“ an Hyponyme gilt allerdings nicht notwendig für Werte lexikalischer Funktionen (vgl. Kap. 3.1.1.1). Hier können Eigenschaften auch durch die jeweiligen konkreten Lexeme überschrieben werden, z.B. ist der Wert der Funktion *Loc_{in}* in (3.7) individuell verschieden für die drei Lexeme *Bauernhof*, *Schloss* und *Kloster*. Er lässt sich auch nicht herleiten aus dem Wert für die hyperonyme Klasse *Siedlung*: *Loc_{in}(Siedlung)=in*.

Die Untergliederung der Lokativa von Le Pesant bleibt vage, die beiden zugänglichen Quellen (Le Pesant & Trégouët 1997-2000, Le Pesant 2000) widersprechen sich häufig. Die wichtigste Untergliederung nimmt Le Pesant (2000: §VI) durch die Einteilung in Lokativa, die durch den Einfluss des Menschen entstanden sind („lieux artificiels ou noms locatifs fonctionnels“), und natürliche Orte („lieux naturels“):⁶⁴

1. noms de lieux artificiels
 - a) les BÂTIMENTS (Gebäude)
 - b) les PIÈCES (Zimmer)
 - c) les LOCAUX (Räumlichkeiten)
 - d) les TERRAINS AMÉNAGÉS (kultiviertes Land)
 - e) les AIRES (Spiel-, Sport-, Park-, Rast-, Lagerplätze)
 - f) les PROPRIÉTÉS (Anwesen)
2. noms de lieux naturels
 - a) les ÉTENDUES DE TERRE (Landflächen)
 - b) les ÉTENDUES D’EAU (Wasserflächen)
 - c) les TERRITOIRES (Regionen, Territorialeinheiten)
 - d) les TERRAINS NATURELS (Naturräume)
 - e) les ZONES GÉOGRAPHIQUES (geographische Zonen)
 - f) le MONDE (die Welt, Erde)

(Le Pesant 2000: §VI)

Die weitere Untergliederung ist (1.) geprägt durch starke Überlappungen der einzelnen Klassen. Nun führt die Einteilung (eines Teilbereichs) des Wortschatzes in „Selektionsklassen“ spezifischer Prädikate nicht notwendig zu einer Baumstruktur, d.h. Objekte können zwei Klassen angehören, und Klassen sich damit überlappen.⁶⁵ (2.) Die Wahl und Einteilung der

63 Mel'čuk (1996: 69) (vgl. Kahane 2003: 536) würde *prendre* als lexikalische Funktion (siehe Kap. 2.2.1) betrachten: *Real₂(bus) = [to] ride [on ART~]*, und analog dazu *to take a bus*. G. Gross (1994: 22–3, 28) zählt die Verwendung von *prendre* als Kollokation der Real-Familie nicht zu den lexikalisierten Verwendungen als Kollokation, Stützverb o.ä., und behandelt diese (*prendre l'autobus+une médicament+...*) jeweils als separate Einträge in der Lexikongrammatik.

64 Vgl. das Zitat von Bauer (S. 131), der sich kritisch zur Unterteilung Kultur- vs. Naturnamen äußert.

65 Vergleiche die Untergliederung der Verkehrsmittel bei G. Gross (1994: 22–6) und dazu Langer (1996: 58–65).

Klassen ist nicht unmittelbar nachvollziehbar – was sich auch darin äußert, dass sich teils nur schwer adäquate Übersetzungen für die Klassen finden lassen. So versucht Le Pesant (2000: §VI.7.1) eine Klasse *noms d'aires* (Spiel-, Sport-, Park-, Rast-, Lagerplätze) zu etablieren, die mittels der spezifischen Prädikate *asphalter* und *bitumer* von der Klasse der *terrains aménagés* (Kulturland) abzugrenzen sein soll (3.8). Als alleinige klassenbildende Eigenschaft wird die angebliche semantische Inkompabilität mit einem Prädikat *pflastern/asphaltieren* angesetzt, denn es ist ja kein Kriterium eines Park- oder Spielplatzes gepflastert zu sein. Allerdings hindert nichts daran, auch Gärten zu *asphaltieren* (so in McEwan's Roman „Zementgarten“), insofern ist die Bewertung von (3.8b) fraglich.

- (3.8) a. *Ce (parking, terrain de jeu) est pavé.* ‚Dieser Park-/Spielplatz ist gepflastert.‘
 b. * *Ce (champ, jardin) est pavé.* ‚Dieser Acker/Garten ist gepflastert.‘
- (3.9) a. *Ce district est surpeuplé.* ‚Dieser Distrikt ist überbevölkert.‘
 b. * *Cette colline est surpeuplée.* ‚Dieser Hügel ist überbevölkert.‘

Ähnlich problematisch gestaltet sich die Grenzziehung zwischen den Klassen *territoires* und *terrains naturels*, die Le Pesant (2000: §VI.9.1) über Prädikate der Gruppe *peuplement* und *domiciliation* (3.9) zu definieren versucht. Satz (3.9b) ist nicht (wirklich) ungrammatisch – so liest man in einem Text über den Montmartre: *Mais la colline est surpeuplée!* Richtig ist: für die Territorialeinheiten stellt die Kombinierbarkeit mit Prädikaten wie *bewohnen*, *bevölkern* eine notwendige Eigenschaft dar, ist ihre Bedeutung doch wesentlich durch die beiden Bestandteile ‚räumliche Einheit‘ und ‚Gruppe von Menschen‘ bestimmt. Dies bedeutet aber nicht, dass diese Prädikate nicht mit anderen, hier natürlichen räumlichen Regionen kompatibel sind.

Kritisch erscheint auch, dass Le Pesant tendenziell versucht, Formen und nicht Bedeutungen (Lesarten) zu klassifizieren. Deutlich wird das bei der Behandlung des Wortes *propriété* (Le Pesant 2000: §VI.8.2):

- (3.10) a. *un château avec parc* ... *cette propriété* ...
 b. *un supermarché avec ses parkings* * ... *cette propriété* ...
 c. *un supermarché avec ses parkings* ... *cette propriété privée* ...

Bei der Übersetzung der Beispiele wird klar, dass zwischen den Bedeutungen ‚Anwesen‘ [=größeres Grundstück mit (Wohn-)Haus] und ‚Eigentum, Besitz‘ unterschieden werden muss. Eine Klassifikation von Wörtern, deren Bedeutung(en) noch nicht unterschieden sind, ist zum Scheitern verurteilt. Stattdessen müssen Bedeutungseinheiten klassifiziert werden, die sich natürlich durch eine Menge von Wörtern oder Lexemen repräsentieren lassen, vgl. ‚Synsets‘ in WordNet (Fellbaum ed. 1998).

Le Pesant (2000) präsentiert umfangreiches Material – eine große Anzahl lokativer Nomina mit vielen spezifischen Prädikaten, Hypo-, Hypero- und Meronymen. Eine schematische oder graphische Präsentation der Ergebnisse fehlt jedoch. Dadurch bleibt seine Arbeit schwer überschaubar. Der Methodenapparat, so wie in Le Pesant (2000) angewendet, erscheint aber nicht ausreichend für eine operationale und nachvollziehbare Untergliederung der Lokativa. Doch bleibt keine wirkliche Alternative zum Versuch, eine Klassifikation über Prädikate und ihre Argumente zu erarbeiten. Bei Le Pesant bleibt stellenweise der Eindruck, dass mit Tests, die an der syntaktischen Oberfläche operieren, Aussagen über die Semantik der Lokativa

ohne das notwendige Fingerspitzengefühl erzwungen werden. Schließlich wird am Beispiel von verschiedenen Taxonomien geographischer Entitäten (Kap. 3.4.2.1), einer Teilmenge der Lokativa, deutlich, dass keine a-priori-Klassifikation existiert, wie Harris es ganz allgemein formuliert hat:

[...] everybody who has used Roget's *Thesaurus* knows, there is no usable classification and structure of meanings per se, such that we could assign the words of a given language to an a priori organization of meanings. Meanings over the whole scope of language cannot be arranged independently of the stock of words and their sentential relations. They can be set up independently only for kinship relations, for numbers, and for some other strictly organized parts of the perceived world. (Harris 1988: 60)

3.1.3 Appropriate nouns

Als ‚appropriate‘ bezeichnet Harris (1964, 1972: 134–7, 1976: 113–4, siehe auch S. 15 in Kap. 1.2) Bestandteile eines Satzes, die mit höchster Wahrscheinlichkeit in einer gg. Position auftreten, so dass ihre Tilgung immer ohne Informationsverlust möglich ist.

Übertragen auf Prädikat-Argument-Strukturen wird als ‚appropriate noun‘, franz. ‚nom/substantif approprié‘, ‚complément approprié‘, dt. ‚angemessenes Komplement‘ (Geyken 2001: 22–3) ein möglicherweise getilgtes, aber rekonstruierbares Komplement eines Prädikats bezeichnet.

- (3.11) a. *Max ist blond.*
b. = *Die **Haare** von Max sind blond.*
c. = *Max hat blonde Haare.*

Motiviert ist die Tilgung des *Napp* durch Metonymie, vor allem Meronymie.⁶⁶ In Bezug auf die Problematik der (spezifischen) Prädikate, Objektklassen und deren taxonomischer Anordnung führt das Phänomen *Napp* zu einer Vererbung „nach oben“ über Meronymie – im Gegensatz zur Vererbung „nach unten“ an hyponyme Klassen. D.h. ein spezifisches Prädikat einer Objektklasse kann auch mit einer übergeordneten holonymen Klasse kombiniert werden. So „erbt“ *Wohnung* die Kombinierbarkeit mit Prädikaten wie *streichen* von dem Meronym *Wand*: *eine Wohnung streichen = die Wände einer Wohnung streichen*.

‚Appropriate nouns‘ wurden mehrfach im Rahmen der Lexikogrammatik behandelt,⁶⁷ mit dem Ziel, Transformationen wie in (3.11a)–(3.11c) zu modellieren. Parallelen ergeben sich zur regulären Polysemie – beide sind meist durch Meronymie motiviert. Gesteuert werden ‚appropriate nouns‘ durch das sie regierende Prädikat, sie gehören deshalb auch nicht in den Lexikoneintrag des Nomen/Komplements: ‚die Haare von Max‘ ist sicher keine Lesart des Lexems *Max*. Dies ist anders im Fall der regulären Polysemie.

66 *Napp* lässt sich aber auch selbst als Prädikat betrachten: um (3.11a) transformationell aus (3.11c) herzuleiten, werden dann entsprechend *Napp* (*Haare*) und *V_{sup}* (*haben*) getilgt. Vgl. auch die Diskussion in Kap. 2.5.3.

67 Vgl. Guillet & Leclère (1981), Laporte (1995a,b, 2004b), um nur wenige zu nennen, und S. 75, wo Laporte (2005) behandelt wird.

3.1.4 Reguläre Polysemie

Regulär (Apresjan 1973, Le Pesant & Mathieu-Colas 1998: 22) oder systematisch (Blutner 1995) wird die polyseme Verwendung eines Wortes dann bezeichnet, wenn parallel zu diesem Wort weitere Wörter existieren, die in analoger Weise polysem verwendet werden.⁶⁸ So können die meisten Bezeichnungen für Organisationen und Institutionen, auch das Gebäude bezeichnen, in denen die Institution ihren Sitz hat. Ausnahmen sind hier eher selten:

- (3.12) a. *Die Schule / Sparkasse / Anwaltskanzlei / der Bundestag / der Hertie brennt.*
b. *? Die FDP brennt. / Die Grünen brennen.*

Der Prozess, der diesen Transfer motiviert, ist vor allem die Metonymie, weniger die Metapher, die eher symptomatisch für irreguläre Bedeutungsübertragungen ist. Auch semantische Analogie, die Verkürzung von Phrasen und Wortbildungsprozesse können beteiligt sein (vgl. Apresjan 1973: 16–7). Als Phänomen steht die reguläre Polysemie der Wortbildung nahe: beide sind produktiv.

In der Praxis wirft die reguläre Polysemie in zwei Punkten Fragen und Probleme auf. (1.) „Für den Psycholinguisten und Computerlinguisten steht vor allem das Problem der Verarbeitung im Mittelpunkt: Durch welche Prozesse und Steuerungsmechanismen gelingt es, im jeweiligen Kontext die angemessene Lesart herauszufinden?“ (Blutner 1995: 34) (2.) Die reguläre Polysemie bzw. die Art ihrer Behandlung hat erheblichen Einfluss auf die Organisation des Lexikons. Sollen die Interpretationsvarianten als Lexikoneinträge oder einzelne Lesarten angenommen werden? Oder sollen im Hinblick auf die Ökonomie des Lexikons Regeln eingeführt werden, die die Interpretation der Varianten systematisch beschreiben? Und schließlich: Ist es im Fall von zwei Bedeutungen, die klar in deutlich verschiedenen Domänen liegen (z.B. ‚Nationalität‘ → ‚Schiff‘, s.u. (3.32)), überhaupt gerechtfertigt von Polysemie zu sprechen? Müssen stattdessen nicht zwei homonyme Lexikoneinträge angenommen werden?

Reguläre Polysemie bedeutet für eine taxonomische Einteilung des Wortschatzes, dass entweder multiple Hyperonyme⁶⁹ sehr zahlreich werden, oder – trennt man die polysemen Be-

68 Eine formale Definition ist bei Apresjan (1973: 16) zu finden:

Polysemy of the word A with the meanings a_i and a_j is called regular if, in the given language, there exists at least one other word B with the meanings b_i and b_j , which are semantically distinguished from each other in exactly the same way as a_i and a_j and if a_i and b_i , a_j and b_j are nonsynonymous.

69 Umgekehrt bedeutet eine Zuordnung zu mehreren Hyperonymen nicht automatisch, dass reguläre Polysemie vorliegt. Le Pesant & Trégouët (1997-2000) werfen unter der Überschrift „Les intersections de classes d’objets (classes dites parfois « polysémiques »)“ zwei Dinge in einen Topf:

Certains noms sont à l’intersection de deux classes d’objets (par exemple le mot *bateau-phare* est à l’intersection des noms de moyens de *transport maritime* et de *signaux*). D’autre part, certaines classes d’objets sont des intersections de classes. C’est ainsi que les différentes classes de noms d’*entreprises*, d’*établissements*, de *cabinets* et de *collectivités territoriales* sont à l’intersection de deux catégories sémantiquement très différentes :

- les noms humains, plus précisément les noms d’institutions (organismes, services, associations, communautés, organes)
- les noms locatifs, plus précisément les noms de terrains, de bâtiments, de locaux (= pièces ou bâtiments), de propriétés (=ensemble de bâtiments et terrains) et de territoires.

(Le Pesant & Trégouët 1997-2000)

Firmen (*entreprises*) und Territorialeinheiten (*collectivités territoriales*) sind Paradebeispiele regulärer Polysemie, allerdings ist ein Feuerschiff (*bateau-phare*) nicht polysem, sondern hier liegt eine Klassifikation

deutungen – die Zahl der Knoten sich signifikant erhöht. Eine konsequente Behandlung ist aber notwendig, vgl. die Problematik in WORDNET (S. 135).

Reguläre Polysemie und Lokativa

Reguläre Polysemie ist im Zusammenhang mit Lokativa in zwei Kontexten zu beobachten: Einerseits werden Bezeichnungen von Organisationen, bestimmten Berufsbezeichnungen und Ereignissen etc. lokativ verwendet, andererseits sind lokative Nomina und Toponyme selbst häufig Träger regulärer Polysemie. In die erste Gruppe fallen z.B.

- Organisation → Gebäude (Sitz oder Filiale der Organisation), vgl. (3.12)

(3.13) *Die Veranstaltung in der Sparkasse findet in Kooperation mit ... statt.*

- Staatsform → Staat mit dieser Staatsform (Apresjan 1973: 24)

(3.14) *In der Republik leben fast eine Million Menschen.*

- Person/Beruf → Ort/Laden, an der die Person arbeitet:

(3.15) *Max war beim Metzger* (= Max war in der Metzgerei)

(3.16) *Max hat beim Bäcker ein Brot gekauft* (= in der Bäckerei)

(3.17) *Bei mir um die Ecke hat ein neuer Bäcker aufgemacht.*

Zahlreiche französische Beispiele und eine ausführliche Diskussion des Problems in Bezug auf Lokativa gibt Le Pesant (2000: §V) anhand von zwei Gruppen: Organisation → Ort (Verwaltung, Schule, etc.) und der Schnittmenge Vegetation \cap *Nloc* (z.B. Wald).

Besonders wichtig, da relevant in Bezug auf Lokativa im Information Retrieval, sind die Fälle regulärer Polysemie, an der Toponyme oder deren Derivationen beteiligt sind und dabei keine lokative Bedeutung aufweisen, vgl. die Beispiele in Apresjan (1973), Markert & Nissim (2002) und Leveling & Hartrumpf (2006):

- Die Namen von Siedlungen und geopolitischen Einheiten (Verwaltungseinheiten, Gebietskörperschaften) können systematisch auch für

– die darin lebenden Menschen

(3.18) *Deutschland hat gewählt.*

– oder deren Regierung oder Verwaltung verwendet werden,

(3.19) *der Vertrag zwischen Frankreich und Deutschland*

– wobei die Unterscheidung zwischen beiden oft offen bleibt:

(3.20) a. *ein Brief an München*

b. = *ein Brief an die Münchner*

c. = *ein Brief an den Münchner Bürgermeister/Stadtrat*

- Die Hauptstadt steht für die dort ansässige Regierung eines Landes.

nach verschiedenen Kriterien (Form vs. Funktion, vgl. Langer 2003) vor.

(3.21) *Paris rejected the “logic of ultimatums.”*

(3.22) *Washington dementierte nur halbherzig.*

- Mit der Bezeichnung für die Bewohner einer Stadt oder Region wird auf eine dort beheimatete Organisation (Fußballmannschaft, Firma usw.) referiert.

(3.23) *Mit 12 800 Mitarbeitern setzten die Walldorfer im vergangenen Jahr rund sechs Milliarden Mark um.*

(3.24) *Die Bayern überzeugten in den ersten 45 Minuten gegen die schwachen Mailänder.*

- Ein Toponym steht für eine dort ansässige Organisation oder dort lebende Menschen.

(3.25) *Dann dampfte der vollbesetzte Zug ab, ohne daß dies nach Firwood-Camp telegraphiert worden war.*

- Ein Toponym steht für ein Ereignis oder eine Situation, die dort stattgefunden hat. Neben okkasionellen und kontextabhängigen Verwendungen, ist hier eine starke Lexikalisierung zu beobachten:

(3.26) *ein zweites **Vietnam**, **Pearl Harbor**, ...*

(3.27) ***Watergate** changed our politics.* (Markert & Nissim 2007b)

(3.28) *Neben **Wimbledon** findet 2009 mit dem Saisonfinale der acht besten Profis ein weiteres hochkarätiges Tennisereignis in London statt.*

- Ein Toponym bezeichnet ein dort gefertigtes Produkt.

(3.29) *I bought a real Meissen.* (Markert & Nissim 2007b)

(3.30) *a smooth Bordeaux that was gutsy enough to cope with our food* (Markert & Nissim 2007a)

Die Herkunftsbezeichnungen von Weinen sind im Deutschen jedoch kein Fall regulärer Polysemie, da sie das Ergebnis eines morphologischen Prozesses sind:

- entweder wird die Bezeichnung für einen männlichen Bewohner (*Burgunder*, *Elsässer*, *Bergsträßer*) verwendet,
- oder, falls das Toponym (*Bordeaux*, *Beaujolais*) in der Form unverändert verwendet wird, ändert sich dessen Genus, also eine syntaktische Eigenschaft. Die Bezeichnung für den Wein ist damit das Ergebnis einer ‚Konversion‘ (Mel’čuk 2006: 304–15):

(3.31) a. ***Das** Bordeaux ist eines der größten und ältesten Wein-Anbaugebiete der Erde.* (neutr.)

b. ***Dieser** Bordeaux hat ein schönes Bukett und ist ausgewogen und harmonisch.* (mask.)

Kein synchron erkennbarer räumlicher Bezug besteht in Wendungen wie *French fries*, *Berliner (Pfannkuchen)*, *Wiener (Würstchen)*. Hier ist es falsch, von regulärer Polysemie zu sprechen.

- Die Bezeichnung für einen männlichen Bewohner eines Landes steht für ein unter der Flagge des Landes fahrendes Schiff, oder für ein Kraftfahrzeug, das in diesem Land registriert ist.

(3.32) ... er wollte jetzt nach England mit einem holländischen Schiffe, das im Hafen seinen erlittenen Seeschaden ausgebessert hatte; von da wollte er nach Indien gehen. Er hatte aber noch nicht das Land verlassen, denn **der Holländer** hatte des Sturmes halber Zuflucht im Außenhafen gesucht ... (Alexander Kielland, Schiffer Worsé)

(3.33) Auf dem Rückweg hatte ich dann einen Holländer vor mir, der fuhr mit 60 über die Landstraße.

- Daneben sind weitere metonymische Verwendungen okkasionell möglich. Markert & Nissim (2002) erwähnen:

(3.34) *The thing about the record is the influences of the music. The bottom end is very New York/New Jersey and the top is very melodic.* ‚Stil, Duktus der in New York und New Jersey gespielten Melodien‘

Die Grenzen zu kontextabhängigen Ellipsen sind dabei offen, vgl. Kap. 5.1.5.1.3.2.

Im Information Retrieval wurde das Problem der regulären Polysemie lange ignoriert (vgl. die Standards zur Annotierung, Kap. 5.1.5.1.3). Erst spät wird es thematisiert (z.B. Sekine, Sudo & Nobata 2002, Schilder, Versley & Habel 2004, Leveling, Hartrumpf & Veiel 2005), und als gesonderte Aufgabe entdeckt, z.B. Markert & Nissim (2002, 2007a,b), Poibeau (2006), Leveling & Hartrumpf (2006), Leveling & Veiel (2006).

Laut Leveling & Hartrumpf (2006) werden im deutschen Korpus des CoNLL-2003-Shared-Task (Testset A) 25% aller Toponyme metonymisch verwendet. Genau genommen setzen sich die 25% aus 17% rein metonymischer Verwendung und 8% gemischter („mixed“) Verwendung zusammen. Als letztere annotieren Leveling & Hartrumpf (2006) Toponyme, die gleichzeitig literal und metonym verwendet werden, so in:

(3.35) *Ein großer Teil der Nahrungsmittel stammt aus der Ukraine, die entschieden hat, den Weizenexport weiter zu erhöhen.*

In Fällen wie diesem, wo *Ukraine* gleichzeitig auf die administrative Region und die Regierung dieser referiert, wird klar, dass für spezifische Aufgaben wie die Entitätenerkennung (Kap. 5.1) eine Aufteilung in die beiden Interpretationsvarianten/Lesarten eher nachrangige Priorität hat. Eine Grenzziehung zwischen literaler und metonymischer Lesart fällt hier schwer. Denn bei administrativen territorialen Einheiten („Gebietskörperschaften“) gehören Land, Menschen und Regierung bzw. Verwaltung untrennbar zusammen.⁷⁰ Einige Prädikate lassen hier auch keine Unterscheidung zu: der *Präsident* eines Landes „herrscht“ gleichzeitig über die Landfläche und die Bewohner (Staatsbürger), und ist oft auch Chef der Regierung. Andere Prädikate selegieren dagegen klar eine der metonymischen Lesarten einer Territorialeinheit X: *Fläche* die ‚Region‘, *verhandeln* die ‚Regierung von X‘, *wählen* die ‚Bewohner/Bürger von X‘.

70 Wegen der Untrennbarkeit der Aspekte, wäre es auch absurd, eine semantische „Unterspezifikation“ anzunehmen, analog zu dem bekannten Phänomen in der Syntax: *Er findet und hilft Frauen*_{Dat/Akk} (Crysmann 2005).

Demgegenüber stehen eindeutige Fälle regulärer Polysemie (*Moskau gab bekannt ...*)⁷¹ und weiter die Fälle, in denen die Polysemie irregulär zu werden beginnt, d.h. schwer vorhersagbar ist und sich mit morphologischen Prozessen zur Wortbildung vermischt (s.o. Weine: *Bordeaux* vs. *Burgunder*). Eine wichtige Teilmenge aus diesem Übergangsbereich ist die Verwendung von Toponymen und Derivationen, um auf Sportvereine, Fußballmannschaften usw. zu verweisen, die dort beheimatet sind oder das Toponym in ihrem Namen führen. Zum einen ist die Verwendung von Toponymen als Referenz auf die Mannschaften sehr häufig, zum anderen gibt es eine Reihe von Prädikaten, die zwei Mannschaften als Argumente nehmen, wobei es beliebig ist, wie jedes der beiden Argumente realisiert wird, ob als Name des Vereins oder der Mannschaft, oder repräsentiert durch Toponym, Wohnernamen, generischen Bestandteil des Vereinsnamens:

- (3.36) *1997 war er bereits beim Finale im Europapokal der Pokalsieger zwischen dem FC Paris St. Germain und dem FC Barcelona im Einsatz.* (zweimal Vereinsname)
- (3.37) *Am Samstag piff er die Partie zwischen Nürnberg und Bayer Leverkusen.* (Toponym, Vereinsname)
- (3.38) *Verhandlungen zwischen Nürnberg und Leverkusen über den Wechsel des Trainers Klaus Augenthaler ...* (zweimal Toponym)
- (3.39) *... nach dem 0:1 der Hamburger gegen den FC Bayern ...* (Wohnernamen, Vereinsname)

Natürlich sollten bei der Erkennung oder Annotierung von Entitäten alle diese Fälle gleich gehandhabt werden, vgl. die Diskussion von Annotierungsrichtlinien für „Named Entities“ (Kap. 5.1.5.1.3). Allerdings ist die metonymische Verwendung stark konventionalisiert und weist zahlreiche Irregularitäten auf. Zum einen kann nur das Toponym, das im Vereinsnamen enthalten ist oder dort in prominenter Position steht, den Verein repräsentieren, aber nicht der Ort, wo der Verein zuhause ist: *Hamburg* kann nicht für den *FC St. Pauli*, *Bayern* nur für den *FC Bayern München*, nicht aber für den ebenfalls in der bayerischen Hauptstadt ansässigen *TSV 1860 München* stehen.⁷² Dies beschreibt die Situation besser und schließt auch die Fälle ein, wo andere Namensbestandteile auf den Sportverein referieren: *Juventus* für *Juventus Turin*, usw. Zusätzliche Evidenz liefert eine Paraphrase: *(der) SC Freiburg*, *(der) Freiburger SC*, *(der) SC aus Freiburg*; dagegen **der bayrische FC*, aber *(der) Münchner FC Bayern*. Dass die Paraphrase möglich ist, ist freilich ein Indiz, dass der Ortsbezug von *Freiburg* und *München*, weniger von *Bayern* in den Beispielen noch besteht, vgl. Kap. 5.1.5.1.2.

Insgesamt sind die Möglichkeiten, wie die Vereine bezeichnet werden können, stark konventionalisiert. Obwohl z.B. viele Vereine ein *FC* im Namen führen und damit *Clubs* sind, kann die Bezeichnung *der Club* nur auf den *1. FC Nürnberg* referieren. *Schalke gewinnt gegen den*

71 Diese sind vor allem dadurch charakterisiert, dass der „Name“ der einen Entität stellvertretend für eine andere Entität steht, die selbst benannt ist, aber mit einem anderen „Namen“. Diese Fälle werden von LDC (2008a) „nickname metonymy“ genannt.

72 Anders verhält es sich für die Wohnernamen (*gegen die Hamburger* = *gegen den FC St. Pauli*) und Phrasen wie *die Gäste aus Nürnberg*:

(1) *Vor dem Spiel gegen den FC St. Pauli wurde Andreas Boshack gestern abend unruhig. ... Dennoch war Boshack vor der Partie gegen die Hamburger zuversichtlich ...*

Club ist in seiner Bedeutung nicht ambig. Deshalb erscheint es sinnvoller, die Konventionen der Subsprache⁷³ der Sport- und Fußballnachrichten eigenständig zu beschreiben. Das Lexikon listet dann entsprechend die variablen Bezeichnungen für die Vereine auf, z.B.:

<i>FC Bayern München</i>	:	<i>Bayern München, Bayern, FC Bayern, die Münchner, der Club von der Säbener Straße, der FC, *?München</i>
<i>TSV München von 1860</i>	:	<i>TSV München 1860, 1860 München, die Löwen, die 60er, 60, TSV, die Münchner, *?München</i>
<i>SpVgg Unterhaching</i>	:	<i>Unterhaching, Haching, die Unterhachinger, die Hachinger, die SpVgg</i>
<i>1. FC Nürnberg</i>	:	<i>der Club, FC Nürnberg, die Nürnberger, 1. FCN</i>

Zuletzt ist die Referenz auf ein Objekt nicht das einzige Kriterium, das herangezogen werden sollte. Wird ein Toponym metonymisch, also in einer anderen Bedeutung, verwendet, kann dessen Syntax modifiziert sein (Verwendung der Präposition in (3.40), Modifikation durch Adjektive in (3.41)) oder auch erhalten bleiben, so *Washington* in (3.42), obwohl es auf die amerikanische Regierung referiert. Deshalb ist eine Auflösung der Metonymie auf Ebene der textuellen Oberfläche, z.B. bei der Annotierung der ‚Named Entities‘ (siehe Kap. 5.1.5.1.1), nicht immer wünschenswert.

(3.40) *Kahn spielt bei Bayern (*in Bayern)*

(3.41) *der neue FC Bayern (≠das neue Bayern)*

(3.42) *Clement teilte mit, **in Washington** herrsche eine „unverändert große Aufmerksamkeit für Deutschland“ ... (Clement nach Gesprächen mit US-Vizepräsident Dick Cheney)*

3.1.5 Zusammenfassung: Beschreibung und Klassifikation der Lokativa

Prädikate selektieren Aktanten bestimmter Objektklassen. Sind die Selektionsrestriktionen der Prädikate und die Zugehörigkeit der potentiellen Aktanten zu den Selektionsklassen beschrieben, sollte dadurch das Potential einer automatischen syntaktischen Analyse erheblich verbessert werden. Doch werden die Beziehungen zwischen Prädikat und Aktant durch eine Reihe sehr häufiger Effekte überlagert, bei denen Selektionseigenschaften Objekte anderer Klassen übertragen („vererbt“) werden:

- „nach unten“ an hyponyme Nomina:
X mündet in Y mit *Y* als GEWÄSSER > SEE, FLUSS, MEER > DONAU, ..., ATLANTIK
- „nach oben“ über Metonymie/Meronymie (appropriate nouns, Napp):
Der Maler streicht die Wohnung = die Wände der Wohnung
- an metonyme Bedeutungen/Lesarten (reguläre Polysemie): *beim Metzger = in der Metzgerei* (nicht Person, sondern Geschäft)
Paris rejected the “logic of ultimatums.” (Hauptstadt → Regierung)

73 Vgl. Kittredge & Lehrberger eds. (1981), Kittredge (1982) für eine Definition des Konzepts Subsprache.

Diese Effekte machen die Organisation des Wortschatzes zu einem in seiner Komplexität immer wieder unterschätzten Unterfangen. In der Praxis betroffen sind insbesondere die Konstruktion von Thesauri und Ontologien (Kap. 3.4) und die Erkennung und Klassifikation von Entitäten in Dokumenten (Kap. 5.1.5.1).

3.1.6 Ein Lexikon lokativer Nomina und der Klassenbezeichner von Toponymen

Um die Aktanten lokativer Prädikate im Text automatisch erkennen zu können, ist ein Lexikon lokativer Nomina (noms locatifs – *Nloc*) notwendig. Gemäß Definition (siehe Kap. 3.1.1) enthält es Nomina, die „Orte, an denen sich Menschen aufhalten können“ bezeichnen. Die aus praktischen Gründen wichtigste und größte Gruppe der lokativen Nomina bilden die Klassenbezeichner von Toponymen bzw. geographischen Entitäten, die einerseits zusammen mit einem Toponym eine Nominalphrase (*die Hauptstadt Paris*) bilden, aber auch alleine als Kopf einer NP stehen können (*die Landeshauptstadt*). Neben den Klassenbezeichnern (s.u.) sind im Lexikon vor allem Bezeichnungen für Gebäudeteile und räumliche Teilregionen (*Stockwerk, Südende, Inneres*), überwiegend funktionale Bezeichnungen (*Arztpraxis*), Werte der lexikalischen Funktion S_{loc} (*Tatort*), und weitere Begriffe (z.B. *Immobilie*), auf die die Definition für Klassenbezeichner nicht anwendbar ist, aufgenommen worden.

Klassenbezeichner werden hier Nomina genannt, die ein Toponym einer bestimmten Klasse zuordnen, oder eine Eigenschaft eines Toponyms beschreiben, und zwar in elementaren Sätzen des Schemas $N_0 V_{sup} N (W)$ mit $N_0 = \text{Toponym}$:

(3.43) *München ist eine Stadt.*

(3.44) *Der Rhein ist ein Fluss.*

(3.45) *Der Yalu ist ein Grenzfluss zwischen der Volksrepublik China und Nordkorea.*

Damit sind Klassenbezeichner als substantivische Prädikate mit einem Toponym als erstem Aktanten und möglichen weiteren (so in (3.45)) definiert (vgl. Kap. 2.6). Entscheidend ist aber – und deshalb werden sie auch an dieser Stelle behandelt –, dass Klassenbezeichner vor allem zusammen mit Toponymen in Nominalphrasen auftreten, die als Nominalisierungen elementarer Sätze betrachtet werden können:

(3.43') *die Stadt München*

(3.44') *der Fluss Rhein*

(3.45') *der chinesisch-koreanische Grenzfluss Yalu*

Klassenbezeichner übernehmen dabei eine disambiguierende Funktion, z.B. *der Fluss Gera* vs. *die Stadt Gera*.

Nominalphrasen mit Klassenbezeichner und Toponym sind in einer Reihe von Varianten möglich (siehe Kap. 3.5.1.3). Meist stehen Klassenbezeichner vorangestellt in Apposition zum Toponym. Als Kriterium zur Definition der Klassenbezeichner ist diese syntaktische Eigenschaft aber nicht geeignet, da sie auch Kombinationen wie *Hallenbad Kaufering* (vs. *Stadt Kaufering*) oder *Stadtsparkasse München* (vs. *Stadt München*) miteinschließen würde.⁷⁴

74 Eine Abgrenzung wäre über eine Unifikation der semantischen Merkmale von Klassenbezeichner und Toponym möglich: *Stadt* und *Kaufering* haben das Merkmal '+Siedlung'/'+'Stadt' gemeinsam, *Hallenbad* und *Kaufering* dagegen nicht.

Das Ziel beim Aufbau des Lexikons der Klassenbezeichner war, möglichst viele Kontexte von Toponymen zu erfassen, die Anwendung der oben genannten Definition erfolgte im Zweifelsfall großzügig. Insgesamt wurden über 6000 Klassenbezeichner zusammengetragen, neben einfachen (*Stadt, Fluss*) vor allem Nominalkomposita (*Gebirgsbach, Landeshauptstadt, Alpentransversale*), aber auch MWLs (*kreisfreie Stadt, gemeindefreies Gebiet, Gewässer dritter Ordnung*), einschließlich aus anderen Sprachen entlehnter (Fach-)Termini (*Rift Valley* ‚Grabenbruch‘, *Powiat* ‚Verwaltungseinheit in Polen‘). Die Aufnahme ins Lexikon erfolgte mit allen notwendigen Eigenschaften, d.h. Flexionsklasse, Genus und der Angabe der Klasse analog zur Klasseneinteilung des Eigennamenlexikons (siehe Kap. 3.4). Quellen für das Lexikon waren das Wörterbuch von Langer (1996), WBIMMOBILIEN, WIKIPEDIA.DE sowie die laufende Arbeit mit lokalen Grammatiken und Texten. Beispiele für das Lexikon sind im Anhang (Kap. A.2) zu finden.

3.2 Geographische (Eigen)Namen

Geographische (Eigen)Namen oder ‚Toponyme‘ sind Namen für geographische (topographische) Objekte oder Entitäten:

each toponym designates a particular topographic feature (Kadmon 2001: 36)

Ein topographisches Objekt ist ein Gegenstand, der für die Darstellung auf allgemein-geographischen und damit auch topographischen Karten relevant ist. Damit sind (i. d. R.) feststehende, mit der Erde verbundene, auf ihr eindeutig festgelegte, erkennbare oder bestimmbare Objekte gemeint. (Definition ‚topographisches Objekt‘ in WIKIPEDIA.DE)

Unter diese Definition⁷⁵ fallen Kontinente, Länder und Regionen, Städte, Dörfer, einzelne Gebäude und Bauwerke, Verkehrswege, Gewässer jeder Art, Berge, Wälder, Parks usw. und schließlich Orte mit religiöser oder kultischer Bedeutung.⁷⁶ Die Eigennamen sind Links, Identifikatoren, Labels für diese Objekte. Und viele, wenn auch nicht alle dieser Objekte tragen Namen. Innerhalb einer Sprachgemeinschaft werden die Namen konsistent verwendet, um auf das Objekt zu referieren. Die Identifikation der Objekte, etwa durch die Angabe ihrer

75 Vgl. die Definition von UNGEGN (2006: 9):

The United Nations Group of Experts on Geographical Names defines a geographical name as a name applied to a feature on Earth (Glossary, 216). In general, a geographical name is the proper name (a specific word, combination of words, or expression) used consistently in language to refer to a particular place, feature or area having a recognizable identity on the surface of the Earth. Named features include:

1. Populated places (for example, cities, towns, villages)
2. Civil divisions (for example, States, cantons, districts, boroughs)
3. Natural features (for example, streams, mountains, capes, lakes, seas)
4. Constructed features (for example, dams, airports, highways)
5. Unbounded places or areas that have specific local (often religious) meaning (for example, grazing lands, fishing areas, sacred areas)

A geographical name may also be referred to as a topographical name or toponym (a term that in a wider context can also include extraterrestrial names, such as names applied to features on the Moon or on other planets).

76 Vgl. „geokodierte“, d.h. mit geographischen Koordinaten versehene Objekte, wie sie z.B. in der WIKIPEDIA.DE zu finden sind. Neben den „klassischen“ Toponymen gehören auch Schulen, Museen, Kirchen, Theater, Denkmäler, Gebäude und Gebäudeteile (!), Friedhöfe, Straßen, Plätze, Parks, usw. dazu.

geographischen Koordinaten, stellt keinen wirklichen Ersatz dar.⁷⁷ Auch Zahlen, Buchstaben bzw. Kombinationen der beiden sind offensichtlich nicht geeignet die Funktion von Eigennamen zu übernehmen. Es gibt vereinzelte Beispiele zur Benennung von Bergen (*K2*), Straßen (*42nd Street*), oder Stadtbezirken,⁷⁸ die meisten topographischen Objekte aber tragen einen natürlichsprachigen Namen.

Eigennamen können ein einzelnes Wort, mehrere Wörter oder auch einen größeren sprachlichen Ausdruck umfassen. Aber nicht alle sprachlichen Ausdrücke, die auf ein topographisches Objekt verweisen und es identifizieren, sind als „Eigennamen“ zu bezeichnen. Die Ausdrücke *Paris* und *die Hauptstadt Frankreichs* bezeichnen beide dasselbe topographische Objekt. Dennoch möchte man intuitiv nur den ersten einen Eigennamen nennen. Das Argument, der zweite Ausdruck verfüge über keine konstante Referenz im Fall eines Hauptstadtwechsels (Bonn, Berlin), zählt nicht in Anbetracht der gerade in der jüngeren Geschichte zahlreichen Namensübertragungen und Umbenennungen von Städten. Entscheidend ist, dass sich Ausdrücke der Form *Hauptstadt von X* syntaktisch systematisch beschreiben lassen. Betrachtet man den elementaren Satz *Paris ist die Hauptstadt Frankreichs* wird klar, dass es sich bei *Hauptstadt* um ein Prädikat handelt, das eine bestimmte Relation ausdrückt.⁷⁹

Der hier angesetzte Eigennamenbegriff ist also linguistisch motiviert. Eine Definition, die sich lediglich auf ‚Monoreferentialität‘ (s.u.) stützt, und damit Ausdrücke wie *die Hauptstadt Frankreichs* oder gar *ich* als Eigennamen bezeichnet (vgl. Wimmer 1995: 373–5) ist unbrauchbar, weil sie das zu beschreibende Objekt nicht eingrenzt.

Ohne auf die von Philosophen und Linguisten seit der antike geführte Diskussion über den Status von Eigennamen einzugehen (siehe Wimmer 1995, Summerell 1995, Lötscher 1995, Koß 1995, Šrámek 1995, Vater 2005: 102–5, Kamianets 2000: 41–2), möchte ich an dieser Stelle die wichtigsten Merkmale von (geographischen) Eigennamen anführen.

- Singularität/Monoreferentialität: jedes Objekt trägt einen individuellen Namen bzw. ein Eigenname referiert auf nur ein Objekt, nicht auf eine Klasse von Objekten. Ambiguitäten/Homonyme (*Frankfurt* für *Frankfurt am Main* / *an der Oder*) oder Synonyme („Allonyme“, z.B. *K2*, *Godwin Austen*, *Dapsang*) sind damit nicht ausgeschlossen. Sie sind ja auch nicht auf Eigennamen beschränkt, auch Appellativa (z.B. *Schloss*) sind betroffen. (zur Disambiguierung geographischer Namen siehe Kap. 5.2)
- Der Inhalt (Bedeutung) des Namens ist ohne Einfluss auf die Beziehung zwischen Eigennamen und benanntem Objekt. *Mailand* ist kein Land, *Innsbruck* keine Brücke, „*Mecklenburg* keine Burg, *Düsseldorf* kein Dorf“ (Sugarewa 1974: 206), *Starnberg* kein Berg!⁸⁰

77 Vgl. die Definition für ‚gazetteer‘ auf S. 103.

78 In den Niederlanden wurden unter Napoleon *Wijk A, B, C* usw. als Bezeichnung für Stadtbezirke eingeführt. Später wieder abgeschafft, sind sie heute ausschließlich mit negativer Konnotation erhalten (Ormeling 2003b: 30).

79 Vgl. Maurel (2004: 777): „Il est clair que *le président Chirac* n’est pas un nom propre, mais une description de la fonction politique exercée par l’homme qui a pour nom propre *Chirac*. Cependant, cette distinction n’est plus vraie pour les formes longues des noms de pays où la description est figée par des considérations politiques (Piton et Maurel 1997). Le dernier exemple en date est celui du *Zaire* qui est devenu, suite à un coup d’état, la *République démocratique du Congo*. Peut-on dire que le référent (le pays africain désigné) n’est plus le même qu’avant? Cela semble difficile à admettre. Et, dans une perspective de recherche d’information, il faudrait sans doute associer les deux termes.“

80 Allerdings gibt es in Österreich einen Berg namens *Starnberg*.

- Übersetzbarkeit: Eine Konsequenz des letzten Punktes ist, dass Eigennamen üblicherweise nicht übersetzt werden: *Churchill* ≠ *Kirchhügel*. Allerdings werden manche Toponyme übersetzt – vollständig (*Great Bear Lake* → *Großer Bärensee*, *Schwarzwald* → *Black Forest*), oder nur in einzelnen Namensbestandteilen (*Ziwa Victoria/Lake Victoria* → *Victoriasee*). Durch die Übersetzung ergeben sich ‚Exonyme‘, ein Phänomen, das quer durch alle Sprachen zu finden ist (siehe Kap. 3.2.2).
- Gleiches gilt für die Paraphrase: Während *das kalte Meer* und *das Meer ist kalt* den gleichen Inhalt ausdrücken, ist das für *Schwarzes Meer* und *das Meer ist schwarz* nicht der Fall. Eigennamen, die aus mehreren Wort- oder Morphembestandteilen bestehen, sind also ein Spezialfall von (morphologischen) Idiomen oder Quasi-Idiomen (vgl. Mel’čuk 1995).
- Nur eine relativ beschränkte Menge von Objekten werden mit Eigennamen benannt: Personen, Orte (topographische Objekte), Organisationen/Firmen und deren Produkte, Bücher, Filme usw., manche Tiere, seltener auch Ereignisse – alles Objekte, die dem Menschen und der Gesellschaft wichtig sind, eine individuelle Behandlung verlangen oder im Fall der Ortsnamen durch ihre Dauerhaftigkeit als Fixpunkte zur Orientierung dienen, (vgl. Roth 2002: 10, Kamianets 2000: 41). Abzugrenzen sind hier Eigennamen (‚named entities‘) von unbenannten Entitäten wie Datums-, Zeit-, Prozent- und Währungsangaben. Beide werden im Information Retrieval oft gemeinsam behandelt, siehe Kap. 5.1.
- Eigennamen seien eine offene, nicht abgeschlossene Klasse, wird häufig als als eine wichtige Eigenschaft genannt: „[proper nouns] form an open class, making the incompleteness of gazetteers an obvious problem.“ (Cucchiarelli, Luzi & Velardi 1999: 174) Und diese Eigenschaft wird als Argument herangezogen, keine Namenslexika im Information Retrieval zu verwenden, weil deren Erstellung und Aktualisierung im Arbeitsaufwand nicht zu leisten wären (vgl. Mikheev, Moens & Grover 1999, Roth 2002: 23). Da dieses Argument nie empirisch untermauert wird,⁸¹ liegt der Verdacht nahe, dass es sich um einen Topos handelt, der unreflektiert übernommen wird.⁸²

Eigennamen haben sicher einen wesentlichen Anteil an bislang unbekanntem Wörtern eines Korpus. Sie sind spezifisch, auf andere Klassen, z.B. (Fach-)Termini, trifft dies jedoch ebenfalls zu. Vermutlich gibt es in keiner lebenden Sprache nicht eine einzige Wortklasse, die definitiv geschlossen ist.

Eigennamen bilden eine große Klasse, möglicherweise größer als die der Nicht-Eigennamen. Eine zuverlässige Schätzung ist jedoch kaum möglich, da fairerweise in beiden Gruppen Mehrworteinheiten mitzuzählen sind. Knapp die Hälfte der Namen in NGA:GNS bestehen aus mehr als einem Wort. Für den allgemeinen Wortschatz lässt sich die Zahl der Mehrworteinheiten nur schwer abschätzen, sie dürfte jedoch die Zahl der Einworteinheiten

81 Lediglich Cucchiarelli, Luzi & Velardi (1999: 173) versuchen anhand eines italienischen Korpus den Anteil der Eigennamen an den unbekanntem Wörtern eines Textes abzuschätzen:

After parsing about 600,000 words, the number of newly encountered words stabilizes at around 500 new words every 100,000. Of these, about 85% have at least one capitalized letter, which seems to suggest the presence of a proper noun. Clearly, since at this level the system does not perform any analysis of complex nominals, the actual number of PN items could be lower [...]

Die Anteil von 85% scheint überhöht. Aus dem annotierten Brownkorpus lässt sich entnehmen, dass 4,3% der Wörter Eigennamen (mit Tag NP oder -TL) sind, von den unbekanntem Wörtern eines Dokuments sind knapp 24% Eigennamen. Ansonsten sind die Ergebnisse vergleichbar: ab 600 000 gelesenen Wörtern stabilisiert sich der Anteil unbekannter Wörter, allerdings nicht auf 500 je 100 000, sondern auf 3 200.

82 Vgl. Martin (1986) und Pullum (1991: 159–71) für ein Beispiel eines Topos in der Linguistik.

deutlich übersteigen.⁸³

Die Behauptung, Eigennamen bilden eine offene Klasse, impliziert, dass diese stärkeren Veränderungen unterworfen seien, als der allgemeine Wortschatz. Dem steht entgegen, dass (1.) Eigennamen meist nicht von Änderungen (Reformen) der Orthographie betroffen sind. (2.) Gerade Toponyme sich als robust gegenüber dem Sprachwandel erweisen und sich deshalb in der historischen Linguistik großer Beliebtheit erfreuen. Für andere Klassen von Eigennamen, z.B. Produktnamen mag das weniger gelten, doch ist die Ursache dafür eher in Marketingstrategien als in a priori angenommenen linguistischen Eigenschaften zu suchen.

- Orthographie: Namen unterliegen in der Regel weniger den orthographischen Konventionen einer Sprache als Appellativa.⁸⁴ Das wird z.B. bei der Getrennt-, Zusammen- bzw. Bindestrichschreibung deutlich, wo kein einheitliches Prinzip zu erkennen ist: *Neubrandenburg*, *Neu Lübbenau*, *Neu-Bamberg* (Nerius 1995: 418). Die Orthographie eines Names ist also ebenfalls per Konvention an den einen Namen gebunden. Allerdings gibt es auch für Toponyme eine Tendenz zur Normierung in amtlichen Namensregistern. Laut Bauer (1995) erfasst diese in Deutschland nicht mehr nur Siedlungsnamen, sondern auch Gewässernamen.
- Insbesondere geographische Namen sind sehr beständig und verändern sich weniger als andere Teile des Lexikons einer Sprache: „toponyms generally (although not always) outline their creators“ (Tichelaar 2003a: 51).
- Besondere syntaktische Eigenschaften können Eigennamen auszeichnen. Allerdings sind diese kaum geeignet, um eine Klasse der Eigennamen von anderen Wortklassen abzugrenzen. Die Definition Bloomfields (1933: 205) von Eigennamen als Wörtern, die nur im Singular und ohne Artikel vorkommen, ist unzureichend und würde $\frac{1}{5}$ der deutschen Toponyme nicht als Eigennamen klassifizieren (siehe Tab. 3.4).
- Eigennamen gehören zu den wenigen echten linguistischen Universalien (Kalverkämper ed. 1978: 24): „Every human language has proper names“ (Hockett 1963). Trotzdem sind Namen Teil des Systems einer Sprache. Ihre morphologischen und syntaktischen Eigenschaften unterscheiden sich deshalb von Sprache zu Sprache.
- Eigennamen, Toponyme insbesondere, werden außer in der rein referentiellen Funktion auch als Produkt- und Markennamen (*Champagner*, *Cognac*, *Bordeaux*, *Brie*, *Edamer*) verwendet. Sie können mit Konnotationen verbunden sein, die (*Afghanistan war das Vietnam der Sowjetunion*). Und einer metonymischen Verwendung sind kaum Grenzen gesetzt (siehe Kap. 3.1.4).
- Politisch sind Toponyme Zeichen der Souveränität einer geopolitischen Einheit (Ormeling 2003b: 34). Als Teil von Postadressen und Fahrplänen sind sie unabdingbar für den Aufrechterhalt der Infrastruktur eines Landes. Meist im Zusammenhang mit politischen Konflikten stehen „Namenskriege“, z.B. in Südtirol, Zypern oder Palästina, über die Benennung von Siedlungen und anderen topographischen Objekten (vgl. Kadmon 2001: 79–85, 224–5). Diskussionen vor der UNGEGN, der für Toponyme zuständigen Behörde der UNO, ob das zwischen Arabischer Halbinsel und Iran gelegene Gewässer international nun *Persischer*

83 Mel'čuk (1995: 169) schätzt das Verhältnis zwischen „Phrasemen“ und einfachen Wörtern im Lexikon auf 10 zu 1.

84 Eine Ausnahme bildet z.B. die Rechtschreibreform des Niederländischen, die im belgischen Flandern, nicht aber in den Niederlanden, auch auf Ortsnamen angewandt wurde. (Ormeling 2003a: 13)

oder *Arabischer Golf* genannt werden sollte, sind dagegen harmlos, da keine Bewohner von einem Namenswechsel ernsthaft betroffen sind.

- ‚Proprialisierung‘ lässt sich aber auch generell als Klasseneigenschaft von Wörtern (neben Wortart, Pronominalität u.a.) betrachten. Verschiedene Meinungen existieren bezüglich der Merkmalsausprägung: Mel’čuk (1993-2000: 3,181–3) nimmt ein binäres Kategorie ‚type de nomination‘ mit den Merkmalen Eigename (‚nom propre‘) und Gattungsname (‚nom commun‘) an, Friburger (2002: 8) und Nübling (2004) sprechen sich dagegen für ein Kontinuum zwischen den Polen Eigennamen – Gattungsnamen aus.

3.2.1 Strukturelle Eigenschaften von Toponymen

Die strukturellen Eigenschaften geographischer Namen lassen sich in zwei Gruppen gliedern: (1.) die interne Struktur, d.h. die Art der Elemente (Morpheme und Wörter), aus denen der Name zusammensetzt ist, und (2.) die syntaktischen Eigenschaften eines Namens, die sich erst im Kontext (Phrase, Satz) äußern. Diese Unterscheidung ist teilweise identisch mit McDonalds Begriffspaar ‚internal‘ vs. ‚external evidence‘ (1996: 32, zitiert auf S. 179, Kap. 5.1.2.2). Allerdings werden als ‚external evidence‘ an dieser Stelle nur unmittelbar durch syntaktische Eigenschaften des Toponyms determinierte Kontexte (Form des Artikels, Genus des attributiven Adjektivs, etc.) berücksichtigt. Als weitere Fälle von ‚external evidence‘ werden in Kap. 3.5 die Einbettung von Toponymen in Nominal und Präpositionalphrasen (auch zusammen mit Klassenbezeichnern) beschrieben, und in Kap. 4 die aktantische Verwendung in Prädikat-Argument-Strukturen, die den maximalen Kontext (unter den sinnvoll beschreibbaren) darstellen.

3.2.1.1 Interne Struktur

Nur wenige Toponyme erscheinen so atomar wie *Köln* oder der *Main*. Es überwiegen Formen wie *Oberhausen*, *Mühlheim*, *Altmarkkreis*, die eine Segmentierung in Einzelbestandteile ohne weiteres zulassen. Noch deutlicher wird dies bei Toponymen wie *Wildbad Kreuth*, wo die Segmente auch graphisch voneinander abgesetzt sind, oder gar separat flektieren (*Bayrischer Wald*).

Die „Bedeutung“ des Toponyms ist im Gegensatz zur Form nicht-kompositionell. Deshalb müssen alle Toponyme, unabhängig davon, ob sie von ihrer Form einfach oder komplex sind, ins Lexikon aufgenommen werden. Allerdings lässt sich die interne Struktur gut dazu verwenden, unbekannte, d.h. nicht im Lexikon verzeichnete Toponyme in Texten zu erkennen (siehe Kap. 5.1.5.3.1.1). Für eine ausführliche Beschreibung der internen Struktur von Toponymen sei auf Kaufmann (1977: 6–28), Back (1995), Fleischer & Barz (1992: 130–5, 192–6), Fuchshuber-Weiß (1995), Kolde (1995) und Laur (1995) verwiesen. Die wichtigsten Phänomene sind:

- Insbesondere Siedlungsnamen sind durch charakteristische Bildungssuffixe (*-acker*, *-aich*, *-beuren*, *-hafen*, *-hausen*, *-kirchen*, *-münde*, *-stetten*, *-weiler*) gekennzeichnet.⁸⁵

85 Eine umfangreiche Liste mit Suffixen (und Präfixen) ist in der WIKIPEDIA.DE unter dem Stichwort „Ortsname“ zu finden.

- Eine Ausnahme bilden sog. Satznamen (Hellfritzsich 1995) wie der Berg *Schauinsland*, der Turm *Spring ins Gut*, *Trausnicht* für verschiedene Mühlen und Schlösser
- Häufig ist im Namen ein generisches Element enthalten, das gleichzeitig auf die Klasse des Toponyms verweist. Dieses kann im Deutschen sowohl als separates Wort (*Starnberger See*), als auch als Bestandteil eines Ein-Wort-Toponyms (*Chiemsee*) realisiert werden. Das generische Element ist nicht zu verwechseln mit einem Klassenbezeichner. Letzter kann nahezu beliebig durch ein Synonym ersetzt werden: *in den Ländern/Bundesländern Bayern und Thüringen*. Für das generische Element ist das kaum möglich (?*Starnberger Süßwassersee*, **Chiemsee Süßwassersee*), es stellt vielmehr einen fixen Bestandteil des Eigennamens dar, der sich auf dem Weg der Entsemantisierung befindet (*Steinhuder Meer*, ein Süßwassersee), oder bereits entsemantisiert ist (,false generics‘): *Großenmeer* (eine Siedlung/Stadt), *Mount Vernon* und *Lake Placid* (zwei Städte), *Rio de Janeiro* (span. *Rio* = ,Fluss‘) und auch zwischen Sprachen nicht übereinstimmen muss (*Aralsee* vs. *la mer d’Aral*, vgl. Grass 2006: 660,663). Ein Prozess der Lexikalisierung also, der sich auch im Artikelgebrauch (*in die Wiener Neustadt* vs. *nach Wiener Neustadt*, vgl. 3.3.2.4), und einer Abwandlung des Flexionsschemas widerspiegeln kann.

Geographische Eigennamen können sogar nur aus appellativen Elementen bestehen (*Rotes Meer*), ihre Kombination und der Kontext machen daraus einen Eigennamen, was durch Konventionen wie die Großschreibung unterstützt werden kann bzw. sich darin äußert.⁸⁶

Die generischen Elemente sind notwendiger Bestandteil eines Eigennamens, im Lexikon-eintrag des Eigennamens müssen sie mit enthalten sein. Lemmata wie *Starnberger* (=See in Südbayern), engl. *Baltic* (=Ostsee) sind weniger adäquat wie die vollständigen Formen *Starnberger See* und *Baltic Sea*. Trotzdem wurde und wird dies nicht von allen Linguisten bei der Annotierung von Eigennamen in Texten und Korpora so gehandhabt, siehe Kap. 5.1.5.1.1.

- Die Koppelung von Ortsnamen führt u.a. in folgenden Fällen zu einem neuen Namen:
 - Stadtteile: *München-Schwabing*, *Berlin-Pankow*
 - Vereinigung von Orten, Ländern etc.: *Garmisch-Partenkirchen*, *Schleswig-Holstein*
 - Ort und Himmelsrichtung: *Dresden-Nord*, *Norddresden* (Fleischer & Barz 1992: 133)
 - Gewässername und Konnotation: *Isarathen*, *Elbflorenz*, *Spreathen* (Fleischer & Barz 1992: 131–2)
 - determinative Bestimmung durch Angabe der Lage mittels Präposition und weiterem Ort (meist ein Gewässer): *Stein am Rhein*, *Mannersdorf am Leithagebirge*, *Fladnitz an der Teichalpe*, *San Paolo di Roma*. Dass es sich um einen konventionalisierten Namensbestandteil handelt, wird aus Negativbildungen (?*München an der Isar*, ?*Berlin an der Spree*) deutlich. Z.T. ist die Bedeutung der Bestimmung nur noch (sprach-)historisch erfassbar, z.B. in *Berg am Laim* (*Laim*=,Lehm‘). Die Ersetzung der determinativen Bestimmung durch einen Relativsatz führt ebenfalls zu keiner gültigen Paraphrase: **Frankfurt, das am Main liegt, ...*

86 Im Deutschen werden generische Elemente (außer Präpositionen und Artikel) großgeschrieben (vgl. Kap. A.1.1): *Hohes Venn*, *Westfriesische Inseln*, *Saint Vincent und die Grenadinen*. Andere Sprachen verfahren z.T. anders, z.B. russ. *Černoe more* ,Schwarzes meer‘, engl. *The Hague*.

3.2.1.2 Syntaktische Eigenschaften

Toponyme sind eine Subklasse der Eigennamen oder der Substantive, und können wie diese den Kern/Kopf einer Nominalphrase bilden. Als solcher determinieren sie die Form von attributiven Adjektiven und Determinativen hinsichtlich der Kategorie Genus. Da das Genus eines Toponyms nicht an der Form ersichtlich ist, muss es im Lexikon spezifiziert sein (siehe Kap. 3.3.2.2).

Anders als bei anderen Eigennamentypen oder Substantiven ist die Verwendung des Artikels mit Toponymen lexikalisiert, d.h. ob ein Toponym mit oder ohne Artikel verwendet wird ist nicht durch allgemeine syntaktische Regeln oder semantische Eigenschaften des Toponyms bestimmbar, und auch keine Frage der Stilistik: *in Frankreich* vs. *in der Türkei* (siehe Kap. 3.3.2.4).

Wie andere Eigennamen auch können Toponyme in Apposition mit einem Klassenbezeichner, d.h. einem Nomen, das das Toponym einer Klasse zuordnet, verwendet werden (siehe Kap. 3.5.1.3). Für manche Toponyme ist die Verwendung mit Klassenbezeichner in den meisten Kontexten obligatorisch (Kap. 3.3.2.5).

Zwei Gruppen von Toponymen verhalten sich problematisch in Bezug auf ihre syntaktischen Eigenschaften, nur mit Einschränkungen können sie als Unterklasse der Nomina beschrieben werden: (1.) Toponyme, die den Artikel innerhalb der Phrase, durch die sie repräsentiert sind, „inkorporiert“ haben, und (2.) Toponyme in Form einer Präpositionalphrase.⁸⁷ Beide Gruppen beinhalten Berg-, Flur- und Straßennamen: *Die Ditzl*, *Die Hex*, *Der Krumme Acker*, *Die kleine Beune* bzw. *Ober der Mauer*, *Auf der Steil*, *Im Hintern Eis*, *Auf dem Sand*, *Auf den Wäldern*, *Unter den Linden*, *Am Markt*, *An der Alten F99*, *An der Landwehr*. Toponyme der ersten Gruppe deklinieren inklusive dem Artikel, die der zweiten Gruppe sind indeklinabel. Für beide Gruppen ist eine Modifikation mit einem Adjektivattribut nur dann möglich, wenn das Toponym in Apposition zu einem Klassenbezeichner steht. Für die zweite Gruppe bestehen darüberhinaus Unklarheiten, was den Anschluss an eine Präposition angeht. Drei Strategien sind belegt:

1. Verwendung mit Präposition, das Toponym steht ohne (3.46a) oder mit (b.) Artikel

- (3.46) a. ... und ganze 3 km entfernt von "Unter den Linden" und der Humboldt-Universität
b. Ausgesetzte Gratwanderungen mit leichter, aber ungesicherter Kletterei über die Ober der Mauer (2.520m) und die Lampermahdspitze ...

2. Verwendung ohne Präposition in statisch-lokativen Kontexten (der Kontext würde eine lokative Präposition verlangen, die jedoch bereits im Toponym „enthalten“ ist)⁸⁸

- (3.47) a. wenn man Unter den Linden wohnt
b. Er residiert Unter den Linden 41

⁸⁷ Parallelen bestehen zu anderen Eigennamenklassen, insbesondere zu Buch- und Filmtiteln, z.B. in

(1) *Mario Adorf und seine berühmten Rollen, sei es beim Schattenmann, beim großen Bellheim, bei der verlorenen Ehre der Katharina Blum oder in der Blechtrommel.*

⁸⁸ Allerdings können Straßen oder Adressen in lokativen Kontexten oft generell ohne Präposition verwendet werden, siehe Kap. 3.5.5.

3. Verwendung in Apposition zu Klassenbezeichner

- (3.48) a. *Ich freute mich zu dieser Zeit auf die Bella Vista, den Gipfel „Im Hintern Eis“ und die nahe Gletscherwelt.*
b. *Die Immobilie befindet sich im Wohngebiet „An der Landwehr“*

3.2.2 Exonyme

Die Namen für ein und dieselbe geographische Einheit können von Sprache zu Sprache verschieden lauten.⁸⁹ Als ‚Endonym‘ wird dabei der Name in der vorort gesprochenen Sprache bezeichnet. ‚Exonyme‘ sind alle davon abweichenden traditionellen Bezeichnungen in anderen Sprachen. So sind z.B. franz. *Aix-la-Chapelle*, span. *Aguas Gran*, poln. *Akwizgran* oder tschech. *Cáchy* Exonyme für das deutsche Endonym *Aachen*; span. *Países Bajos* bezeichnet die *Niederlande*, engl. *Leghorn* das italienische *Livorno*.⁹⁰ Exonyme sind quer durch alle Sprachen zu finden.⁹¹

Problematisch wird obige Definition eines ‚Exonyms‘ für Regionen, die mehrsprachig sind oder in denen sprachliche Minderheiten beheimatet sind. Die Definition der UNGEGN erfolgt deshalb ausgehend vom durch die lokalen Behörden standardisierten Toponym: „In contrast with a nationally standardized geographical name, or “endonym”, an “exonym” is a conventional or traditional name in a language other than the one in which the name was standardized“ (Raper 1995: 284). Eine Definition, die das Problem nicht löst, sondern den nationalen Behörden und ihrer Sprachenpolitik überlässt. Sind Toponyme nicht amtlich standardisiert, wie in den Niederlanden (Ormeling 2003a: 14), greift sie nicht. Sie ist weiter sinnlos für geographische Objekte, die sich auf mehrere Sprachgebiete erstrecken, z.B. die Alpen.

Das Ziel der UNGEGN, den Gebrauch von Exonymen insbesondere auf Landkarten und Atlanten einzuschränken, scheint erfolgreich: „Weltweit ist ein Trend erkennbar, in Weltatlanten in verstärktem Maße die endonymischen Namenformen zu verwenden“ (StAGN:Exon-Empf) In Texten werden jedoch vermutlich immer Exonyme zu finden sein – zu stark sind Tradition und Konvention, zu groß mitunter Probleme mit Aussprache und Orthographie. Der Satz *Max fährt nach Moskva (Moskau)* klingt gewollt seltsam und wird sich in absehbarer Zeit nicht durchsetzen. Die jüngste Zwischenbilanz der UNGEGN fällt deshalb weniger optimistisch aus: „Time has, however, shown that initial ambitious attempts to rapidly decrease the number of exonyms were over-optimistic and not possible to realise in the intended way. The

89 Back (1995: 1349) spricht von „interlingualer Allonymie“.

90 Viele weitere Beispiele sind zu finden bei: StAGN (2002), Manias (1994); in der englischen WIKIPEDIA (<http://en.wikipedia.org/>) unter den Stichworten ‚Names of European cities in different languages‘, ‚List of European regions with alternative names‘ und ‚List of European rivers with alternative names‘, in der deutschsprachigen WIKIPEDIA.DE unter ‚Liste der Listen deutscher Bezeichnungen ausländischer Orte‘, ‚Exonyme für deutsche Toponyme‘; WORLD-GAZETTEER, ein multilingualer Gazetteer (engl., franz., dt., span.); ARCHINFORM eine Architekturdatenbank, nennt für Städte die Namen in vielen verschiedenen Sprachen. Auch der Gazetteer NGA:GNS enthält neben dem Endonym, (orthographische) Varianten und Exonyme des Englischen und anderer europäischer Sprachen. Insgesamt gut 1 Mio. der Einträge weist Varianten auf.

91 Vgl. La Robertie (2006) mit Beispielen aus dem Chinesischen, z.B. 勃朗山 (*Bolang Shan* = „Bolang-Gebirge“) ‚Mont Blanc‘.

reason would appear to be that many exonyms have become common words in a language and can be seen as part of the language's cultural heritage.“ (UNGEGN 2004)

Exonyme entstehen durch Adaption eines Toponyms an Grammatik, Orthographie und Aussprache der Zielsprache. Eher selten sind willentliche Adaptionen oder Umbenennungen, z.B. im Zuge der Annexion eines Gebietes durch eine anderssprachige Nation. Die Adaption kann in einem oder mehreren der folgenden Punkte erfolgen:

1. Anpassung der Aussprache: *Paris* [pari] → dt. [pa'ʁi:s], engl. [p^hæ.ʁɪs]
2. Anpassung der Orthographie
 - a) Transliteration oder Transkription für Namen aus fremden Alphabeten, z.B. russ. *Благовещенск* → dt. *Blagoweschtschensk* oder *Blagovešensk*⁹²
 - b) Verlust von Diakritika: *Orléans* → *Orleans*. Diakritika können in der Sprache des Endonyms bedeutungsunterscheidend sein, z.B. handelt es sich bei *Brezany* vs. *Břežany* um verschiedene Orte Tschechiens (Woodman 2003). Der Fall von *Rothenbürg* veranlasste einst das SZ-Magazin zu einem Bericht über den Ort in Oberfranken, in dem immer wieder ausländische Touristen stranden, deren eigentliches Ziel *Rothenburg ob der Tauber* war.
3. Adaption an das grammatische System der Zielsprache
 - a) Phonologie der Zielsprache: *Bozen* → *Bolzano*
 - b) Flexion, womit praktisch alle Namen Exonyme wären. So weist *Berlin* im Russischen (*Берлин*) sechs Formen/Kasus auf, umgekehrt realisiert *Ural* im Deutschen statt sechs nur vier Kasus.
 - c) Wechsel des Genus, vgl. franz. *le Rhône* (mask.) → dt. *die Rhone* (fem.), oder russ. *Odessa* (fem.) → dt. *Odessa* (neutr.)
 - d) Verlust/Hinzufügen eines Artikels/Determinativs: arab. *القاهرة* *al-Qāhira* („die Siegreiche“) → dt. *Kairo*
 - e) Ergänzung von generischen Elementen
4. auch Kombinationen aus den obigen Punkten sind möglich, insbesondere ein Mix aus Transskription (inkl. Einfluss von historischem Lautwandel) und Ergänzung von generischen Namensteilen: *Schweiz* → *Switzerland*.
5. (Teil-)Übersetzung mit folgenden Fällen (vgl. Kadmon 2001: 151–2):
 - a) Übersetzung eines einfachen Namens (Simplex): flämisch *Bergen* → wallonisch *Mons*
 - b) für komplexe Namen, bestehend aus mehreren spezifischen oder spezifischem und generischem Element, sind wiederum drei Fälle zu unterscheiden:
 - i. nur das generische Element wird übersetzt: *Lago di Garda* → *Lake Garda*
 - ii. nur das oder eines der spezifischen Elemente wird übersetzt: russ. *Malyj Enisej* → dt. *Kleiner Jenissej*, *Appenzell-Außerrhoden* → ital. *Appenzello Esterno*
 - iii. sowohl generisches, als auch spezifisches Element werden übersetzt: holl./flämisch *Nieuwerkerke* → franz. *Neuve-Eglise*, *Schwarzwald* → *Black Forest*, *Forêt Noire*

92 Wird ein offizielles, standardisiertes Transliterationssystem benutzt, gilt der transliterierte Name für die UNGEGN nicht als Exonym (Ormeling 2003c: 119). Ziel der UNGEGN ist deshalb, für jedes Paar Schriftsystem-Zielsprache ein einziges offizielles Transliterationssystem zu etablieren (Raper 1995: 286–7).

6. Substitution durch anderen Namen: Kadmon (2001:145) nennt als Beispiel den *Mount Everest*, den die britischen Kolonialbehörden so ohne Rücksicht auf seinen bestehenden tibetischen und nepalesischen Namen benannten.

Die Punkte 1 und 2 betrachtet die UNGEGN nicht als Exonyme. Wenn nur generische Bestandteile übersetzt werden, spricht sie von ‚partiellen Exonymen‘ (Woodman 2003).

Exonyme sind ein unvermeidliches Phänomen bei der Wiedergabe von geographischen Namen in verschiedenen Sprachen. Auch wenn politischer Wille mit der Zeit einen Teil der Exonyme obsolet macht – das englische *Leghorn* für *Livorno*, oder *Langensee* für *Lago Maggiore* findet man nur in „historischen“ Kontexten –, werden alltäglich benutzte Exonyme nicht zu verdrängen sein. Schwache Formen der Exonymbildung (Transliteration, Diakritika, Artikel, Flexion, s.o.) sind weiter notwendig, da die Namen an die Grammatik der Zielsprache angepasst werden müssen.

Eine Wertung durch den Linguisten ist nicht angebracht. Im vorgestellten Toponymlexikon werden Exonyme deshalb schlicht als Namensvarianten geführt. Die im Deutschen üblichere Bezeichnung wird als primärer Name, andere Formen unabhängig davon, ob sie Exo- oder Endonyme sind als Varianten geführt, z.B. *Brüssel* vs. *Bruxelles*. Entscheidend ist allerdings, dass die Variante auch in deutschen Texten anzutreffen ist. Eine wirklich konsistente Behandlung von Exonymen ist ohnehin nur im Rahmen eines mehrsprachigen Lexikons oder einer multilingualen Ontologie (vgl. Tran, Grass & Maurel 2004, Maurel et al. 2004, Grass 2006) möglich.

3.3 Ein Lexikon geographischer Eigennamen

Ohne Lexikon keine Grammatik – ein Lexikon geographischer Eigennamen ist zur Beschreibung lokativer Sätze unbedingt notwendig. Geographische Namen stellen die größte Gruppe der Argumente in solchen Sätzen und tragen wesentlich zum Informationsgehalt eines Satzes bei. Deshalb sind für das Information Retrieval umfangreiche Toponymlexika unverzichtbar. „Resource-light“-Methoden, die auf eine lexikalische Analyse vollständig verzichten oder sich lediglich auf einen Stemmer stützen (z.B. Petras & Gey 2005), erzielen speziell bei der Erkennung von Toponymen im Text deutlich schlechtere Werte.⁹³

Es sind zwei Disziplinen, Linguisten und Geographen, die Toponymlexika erstellen. Linguisten haben gewaltige⁹⁴ Namenwörterbücher erarbeitet, allerdings fast durchweg mit etymologischer Ausrichtung und historischen Fragestellungen (Siedlungsgeschichte, sprachliche Straten) im Hintergrund. Toponymlexika, die morphologische und syntaktische Eigenschaften ihrer Lemmata und deren Derivationen erfassen, sind rar. Vorbildliche Beispiele aus dem Ausland sind z.B. das elektronische PROLINTEX, oder papiere Levašov (2000). Für das Deutsche behandeln papiere Wörterbücher (z.B. Liedtke 2002, Meynen ed. 1966) nur Teilaspekte (Genus und Varianten). Bei den elektronischen morphologischen Lexika, z.B. dem CISLEX (Maier-Meyer 1995), machen geographische Namen nur einen Teil des lexikalischen Bestandes aus. Relativ gut im Hinblick auf die genannten Anforderungen (s.u.) ist das Wörterbuch von CANOO.NET, das für einige hundert bis wenige tausend Toponyme und deren Derivationen alle Flexionsformen listet. Die meisten Initiativen, geographische Namen zu

93 Vgl. auch die Diskussion zu Mikheev, Moens & Grover (1999) auf S. 178.

94 Z.B. die 17 großformatigen Bände Vasmer (1961-1969) und RGN.

sammeln, gehen jedoch von Geographen aus. Erwartungsgemäß legen diese den Schwerpunkt auf die geographische „Vermessung“ eines Ortes und nutzen den Namen lediglich als Schlüssel zum Zugriff auf die geographischen Daten, die als ‚Gazetteer‘ organisiert sind. Allerdings ist das Interesse der Geographen an linguistischen Eigenschaften von Toponymen z.T. erstaunlich hoch (vgl. GENTILÉSDUQUÉBÉC, Kadmon (2001), UNGEGN).

Einen Gazetteer⁹⁵ definieren Hill, Goodchild & Janée (2004) folgendermaßen: „A basic gazetteer entry is formally defined as a triple $\langle N,t,g \rangle$, where N denotes a feature name, t denotes a feature type, and g denotes coordinates for the location and extent of the feature. The content of g might be a simple coordinate pair (no information on extent), or a bounding box (generalized representation of extent), or a representation of a complex polygonal or polylinear shape (detailed representation of extent). [...] This three basic elements of a gazetteer support discovery by names, coordinates, and types (e.g., searching for schools in or near Goleta, CA).“ Als Anforderungen an einen Gazetteer formuliert Leidner (2004a):

- Zugänglichkeit: offen zugängliche Gazetteers sind zu bevorzugen
- Skopus: die erfasste Region eines Gazetteers kann eine einzelne Kommune (cadastral scope) sein, aber auch die ganze Welt umfassen (global scope). Für linguistische Anwendungen und im Information Retrieval sollte der Skopus global sein.
- Vollständigkeit: Obwohl sicher kein Gazetteer existiert, der alle existenten Toponyme erfasst, variieren verschiedene Gazetteers hierin.
- Korrektheit: Gazetteers enthalten immer unkorrekte oder veraltete Einträge. Das betrifft z.B. auch den Gazetteer NGA:GNS und zwar trotz 20 000 Änderungen/Ergänzungen pro Jahr.
- Auflösung (‚granularity‘): in Abhängigkeit von verschiedenen Aufgaben oder Zielregionen muss der Gazetteer in der räumlichen Auflösung unterschiedlich genau sein. Hobbs (2006): „A city can be viewed as a point, a region, or a volume.“ Welche der Darstellungen für die gg. Aufgabe die geeignete ist, sollte dem Benutzer überlassen bleiben. Für bestimmte Objektklassen, z.B. Wasserläufe ist eine Repräsentation als Punkt außerdem nicht geeignet (vgl. Hauff, Trieschnigg & Rode 2006, Ferrés Domènech 2007: 43).

In der Konsequenz ist bei einer groben Auflösung auch die Vollständigkeit eine andere, da kleine Objekte nicht enthalten sind.

- Balance: Korrektheit und Vollständigkeit sollten gleichmäßig für alle erfassten Regionen sein.
- Richness: Die Detailgenauigkeit der Einträge, ggf. mit weiteren Merkmalen (z.B. der Bevölkerungszahl) kann unterschiedlich sein.

Gazetteers unterscheiden sich vor allem in den beiden Punkten Skopus und Vollständigkeit und weiter darin, wie die drei elementaren Angaben zu Name, Klasse und räumlicher Repräsentation im Detail aussehen: (1.) Werden Namensvarianten angegeben? Wenn ja, welche? (2.) Welche Klassen gibt es, und wie sind sie in einer Taxonomie organisiert (vgl. Kap. 3.4.2.1)? (3.) Wie sieht die Repräsentation durch Angabe von Koordinaten aus?

95 Der Begriff ‚gazetteer‘ wurde so erstmals von Laurence Echard benutzt, einem englischen Historiker, der 1693 den Titel „*The Gazetteer’s: or, Newsman’s Interpreter; Being a Geographical Index*“ herausgab (Kadmon 2001: 19).

Die Anforderungen an ein maschinenlesbares Toponymlexikon, das in Grammatiken zum Information Retrieval verwendet wird, decken sich z.T. mit den von Leidner formulierten Anforderungen an einen Gazetteer. Das Toponym und die Klasse, der es oder genauer die bezeichnete geographische Entität angehört, sind essentielle Bestandteile eines Eintrags, die räumliche Repräsentation hingegen nicht. Neu hinzu kommt dafür die linguistische Dimension, d.h. die sprachlichen Eigenschaften des Toponyms – dessen „Syntax“⁹⁶, die die Kombinationsmöglichkeiten mit anderen Wörtern beschreibt. Im Detail bedeutet das:

- eine geographische Repräsentation ist nicht notwendig
- der Skopus des Lexikons ist global, aber die Balance (s.o.) ist stark durch die Objektsprache bedingt, und an die Verwendungshäufigkeit der Toponyme in Texten/Korpora anzupassen. In einem deutschsprachigen Toponymlexikon sollte eine im deutschen Sprachraum gelegene Kleinstadt enthalten sein, nicht unbedingt aber eine vergleichbar kleine Stadt Asiens.
- die Relevanz eines Eintrags ist durch dessen Häufigkeit in Texten definiert und nicht durch dessen geographische Eigenschaften (Fläche, Einwohnerzahl etc.)
- für jedes enthaltene Toponym sollten kodiert sein (siehe Kap. 3.3.2):
 - die semantische Klasse, der das Toponym angehört
 - alle Formen, d.h. Flexionsformen und Varianten
 - die syntaktischen Eigenschaften des Toponyms bzw. der Flexionsform
 - von dem Toponym abgeleitete Derivationen, z.B. Adjektive, Bewohnernamen. Für die meisten Siedlungen, geopolitischen Einheiten und manche Landschaftsformen wie Inseln und Täler sind Adjektive und Bewohnerbezeichnungen lexikalisiert!
- der Name in korrekter Orthographie der Objektsprache inklusive der Diakritika.⁹⁷
- alle in der Objektsprache gebräuchlichen Namensvarianten

Als Format für das Lexikon wurde das DELA-Format gewählt. Das Lexikon liegt in zwei Formen vor, als Grundformenlexikon und Vollformenlexikon. Beispiele sind im Anhang, Kap. A.1.2, zu finden. Außerdem wurde das Lexikon versuchsweise in eine Ontologie konvertiert (siehe Kap. 3.4). Gegenwärtig enthält das Grundformenlexikon 50 000 Toponyme und 13 000 Derivationen von 48 000 geographischen Entitäten. Das Vollformenlexikon enthält über 300 000 Flexionsformen und Varianten, sowie 1 700 Fugenformen.

3.3.1 Quellen für ein Toponymlexikon

Dieses Kapitel gibt einerseits einen Überblick über vorhandene Gazetteers und Lexika und beschreibt andererseits die Quellen, mit denen das Toponymlexikon erarbeitet wurde. Der lexikalische Bestand des Lexikons wurde auf zweierlei Weise aufgebaut und erweitert.

(1.) Die Toponyme aus einigen zuverlässigen, aber tendenziell kleinen Lexika oder Listen wurden komplett übernommen, und zwar (a) dann wenn die Listen hochrelevantes Material, d.h. wichtige und allgemein bekannte Orte, enthielten, oder wenn (b) damit sichergestellt

96 Vgl. die Definition eines sprachlichen Zeichens als Tripel (Bezeichnetes, Bezeichnendes und Syntax) durch Mel'čuk, siehe S. 26.

97 Diakritika werden von vielen englischsprachigen Gazetteers nicht oder nur unvollständig berücksichtigt.

werden konnte, eine möglichst breite Anzahl an Phänomenen (linguistische Eigenschaften oder verschiedenen Klassen von Toponymen) erfasst werden konnten.

(2.) Toponyme aus großen, breit angelegten, aber fehlerbehafteten Quellen wurden als Kandidaten für einen Lexikoneintrag betrachtet und zunächst ihre Vorkommen in geeigneten Kontexten aus einem 6 Milliarden Wortformen umfassenden Textkorpus (siehe Tab. E.2 auf S. 441) extrahiert. Damit erfolgte auch eine Klassifikation der syntaktischen Eigenschaften des Toponyms. Steht das Toponym z.B. im Kontext *in der X* wurde der Zähler für die Merkmale ‘Femininum’ und ‘mit Artikel’ erhöht, im Kontext *in X* für ‘Neutrum’ und ‘ohne Artikel’. Der höchste Wert wird dann als Vorschlag ins Lexikon übernommen. Meist kann die vorgeschlagene Klassifikation direkt übernommen werden (in Tab. 3.2 für die Toponyme *Türkei*, *Afghanistan* und *Donau*). Immer wieder kommt es jedoch zu falschen Vorschlägen, was eine Nachkorrektur unbedingt erforderlich macht. So werden die Werte für die Stadt *Halle* durch das homonyme Nomen „gestört“. Beginnend mit den häufigeren wurden so alle

Toponym	F _{tot}	F _{ctx}	+Det	+DetZ	M	F	N	m	Vorschlag
Türkei	179105	201926	66310	975	12	53446	985	12913	+Det:F
Afghanistan	40931	58523	367	19132	171	4	19344	6	+DetZ:N
Main	269360	665306	221019	740	110207	533	110107	941	+Det:M
Donau	56668	81769	25560	1695	54	18773	1739	6693	+Det:F
Halle	281572	239772	50950	27743	230	49459	27978	4719	+Det:F

Tabelle 3.2: Syntaktische Vorklassifikation der Toponyme durch ihr Vorkommen in Kontexten. Im einzelnen bezeichnen: F_{tot} = Frequenz im Korpus, F_{ctx} = Frequenz in allen Kontexten, +Det=verwendet mit Artikel, +DetZ=ohne Artikel, M=mask., F=fem., N=neutr., m=Plurale tantum. Aufgrund ambiguer Kontexte ist F_{ctx} > F_{tot}.

Toponyme mit einer Frequenz > 100 manuell nachkorrigiert. Neben der Korrektur der syntaktischen Merkmale, waren dabei noch die Informationen zur Flexion zu ergänzen. Weiter musste das Toponym einer Klasse zugeordnet werden, die meist von der Quelle übernommen werden konnte. Und schließlich waren „falschen Freunde“ zu kennzeichnen, also Toponyme wie *Halle*, die zu Nicht-Toponymen homonym sind (vgl. S. 108). Die automatische Klassifikation erwies sich hier als hilfreich: liefert sie falsche Werte, ist das ein starkes Indiz für einen „falschen Freund“.

(3.) Um sicherzustellen, dass im kleinen Rahmen auch weniger zentrale Klassen von Toponymen (Mikrotoponyme und verschiedene Klassen von Bauwerken) im Lexikon enthalten sind, wurden entsprechende Listen aus WIKIPEDIA.DE übernommen und gemäß den Anforderungen kodiert.

Im folgenden sollen nun die vorhandenen Quellen (Gazetteers, Toponymlexika und -listen) kurz vorgestellt werden. Wenn eine Quelle für den Aufbau des Lexikons verwendet wurde, ist sie mit ‘[+]’ gekennzeichnet.

StAGN (1981) offizieller Gazetteer der Bundesrepublik Deutschland: listet Toponyme mit Angabe von Bezirk, administrativer Zuordnung, Koordinaten. Für „physisch-geographische Objekte“ (Berge, Gebirge, Landschaften, Gewässer) sind auch Genus und Numerus (Pluralia tantum) angegeben.

- STAGN (2004) [+] Liste mit Staatennamen, enthält Kurz- und Langformen, sowie abgeleitete Adjektive und die Namen der Bewohner.
- Meynen ed. (1966) Toponyme weltweit mit deutscher Bezeichnung, phonetischer Transkription, Orthographie, Erklärung, Koordinaten. In einem kurzen Kapitel sind die Konventionen beschrieben, die sich bei der Schreibung geographischer Namen herausgebildet haben, siehe Kap. A.1.1.
- Zikmund ed. (2000) liefert für Orte im Baltikum und der Gemeinschaft Unabhängiger Staaten (GUS) sowohl deren deutsches Exonym (*Moskau*), als auch das Endonym in Originalorthographie und Transliteration, und weiter Betonung, administrative Zuordnung, Erklärung, Koordinaten, z.B.:
- Ismajil** -ji- Ukr.: Geb. Odessa (*ukr.*: Izmaïl/Ізмаїл -ï-, *russ.*: Izmail/Измаил -il; Stadt am li. Ufer des Kilija-Nebenarms der Donau; 94 700 Einwohner; Hptst. des gnam. Kreises)
ö 28/n 45
- Spieß (1994), StAGN (2002) [+] Liste mit häufig verwendeten Exonymen im Deutschen.
- Liedtke (2002) [+] listet 700 Namen von Regionen und Landschaften inklusive Varianten und Genus.
- Müller (2003) listet alle deutschen Gemeinden mit administrativer Zuordnung.
- EURCOM:LAU [+] administrative Untergliederung von EU-Ländern in tabellarischer Form.
- GEODATENZENTRUM Die Datenbank des Geodatenzentrum des Bundesamtes für Kartographie und Geodäsie beinhaltet alle Namen von Siedlungen und verschiedenen Landschaftsformen, die auf den amtlichen Karten im Maßstab 1:250.000 verzeichnet sind – ca. 46.000 Einträge. Seit Ende 2006 sind die Daten on-line zugänglich, als vollständiger Datensatz aber weiterhin nur gegen Bezahlung.
- CANOO.NET listet für einige hundert bis wenige tausend Toponyme und deren Derivationen alle Flexionsformen.
- WIKIPEDIA.DE [+] bietet zahlreiche Listen mit einzelnen Typen topographischer Objekte (Städte, Flüsse, Seen, Gletscher, etc.).
- HISTORTSBUCH [+] Namen von Gemeinden des Deutschen Reiches und Österreichs mit administrativer Zuordnung aus dem Zeitraum 1871–1945.
- ÖSTERREICHISCHEBERGNAMEN [+] eine ziemlich vollständige Liste der Oronyme Österreichs.
- Manias (1994) [+] Hervorragende Liste mit über 2000 Toponymen vornehmlich aus Deutschland, Österreich und der Schweiz mit ihren Entsprechungen (Exonymen) in zahlreichen europäischen Sprachen.
- aot.ru [+] enthält Toponyme in einer groben Vorklassifikation (Stadt, Staat, Gewässer, sonstige) mit Angabe von Genus und Flexion. Es enthält auch Listen mit Adjektivableitungen und Bewohnern.
- IMSLEX enthält nach eigenen Angaben (Fitschen 2004: 105) ca. 5000 Toponyme mit Flexionsformen.
- CISLEX (Maier-Meyer 1995) enthält zwar Toponyme, allerdings sind diese nicht durchgehend als solche gekennzeichnet (z.T. auch als Nomina), eine Feinklassifizierung (Siedlung, Gewässer) fehlt. Die Klassifikation in Flexionsklassen ist teils fehlerhaft. Alle Einträge sind kleingeschrieben, was es schwierig macht, die korrekte Schreibung (*USA* vs. *Ufa*)

wiederherzustellen. Obwohl das Lexikon am CIS vorlag, war es als Vorlage deshalb nicht geeignet. Ein Lexikon aufmerksam Korrektur zu lesen, ist ähnlich aufwändig, wie es neu zu erarbeiten.

USGS:GNIS [+] frei verfügbarer Gazetteer, erfasst 2,3 Millionen Objekte auf dem Gebiet der USA.

NGA:GNS [+] das internationale Pendant zu USGS:GNIS, wie dieser frei verfügbar, mit bald sechs Millionen Einträgen einer der größten Gazetteers. Enthält in der Regel die Namen als Endonyme, aber auch als englisches Exonym (*Bad Toelz*, *Isar Valley*, *Black Forest*). Z.T. sind fälschlich als Toponyme aufgenommene geographische Bezeichnungen eingestreut: *Baggersee*, *Bahnbrücken*, *Wasserleitung*, *Wasserfall*, *Übersee-Hafen*, *Bauernhof* – alle mit Klassifikation und Koordinaten. Die Geokodierung erfolgt immer durch ein Koordinatenpaar, auch z.B. für Regionen und Flüsse.

WORLD-GAZETTEER [+] Gazetteer mit 170 000 Einträgen. Enthält auch Namensvarianten (andere Sprachen, andere Schreibung), admin. Zuordnung, Koordinaten und Einwohnerzahl.

Law (1999) : administrative Untergliederung aller Staaten der Erde, in Deutschland bis zur zweiten Ebene (Regierungsbezirke). Enthält neben den englischen Bezeichnungen auch Endonyme und weitere Exonyme.

GEONAMES.ORG enthält 6 Millionen Einträge, davon 2,2 Millionen Siedlungen und speist sich aus u.a. folgenden Quellen: NGA:GNS, USGS:GNIS, WORLD-GAZETTEER, WIKIPEDIA, GEOBASE.CA (vollständige Liste unter <http://www.geonames.org/about.html>). Die Daten können als Wiki editiert werden, ein Export der Daten ist möglich.

UN/LOCODE enthält 40 000 geographische Namen (weltweit) mit „genormtem“ Kürzel, Länderkürzel und Koordinaten. Alle Namen sind auf den Zeichensatz ISO-8859-1 reduziert, was zum Verlust von Diakritika führt. Auch andere Zeichen fehlen mitunter (*St Johann in Tirol*).

Alexandria Digital Library Gazetteer (ADL, Hill 2000, Hill, Goodchild & Janée 2004) enthält derzeit (Version vom 15. 3. 2002) 5,9 Millionen geographische Namen. Die Namenliste ist ohne Diakritika, aber frei zugänglich. Der Gazetteer mit umfangreicher Zusatzinformation, dagegen nicht. Die Klassen sind als Thesaurus organisiert (siehe S. 135).

GETTY Thesaurus of Geographic Names (TGN) enthält 1,3 Millionen geographische Namen, die über die administrative Untergliederung thesaurusartig organisiert sind. Dazu kommen (siehe Abb. 3.2 auf S. 128): Namensvarianten, Exo- und Endonyme, Koordinaten und Quellenangaben. Jeder Name ist einer oder mehreren von 1 700 Klassen zugeordnet. Der Thesaurus ist zwar zugänglich, eine Nutzung der Daten ohne Lizenzierung aber untersagt.

3.3.2 Kodierte Eigenschaften

Als Eigenschaften eines Toponyms wurden im Lexikon erfasst: semantische Klasse, Genus, Verwendung mit/ohne Artikel, Flexion, (ortho)graphische Variationen und Derivationen (für Details siehe die folgenden Kapitel, Beispiele im Anhang Kap. A.1.2).⁹⁸ Alle diese Eigenschaften sind notwendig für die exakte Beschreibung, wie sich Toponyme im Kontext verhalten.

98 Vgl. auch für das Französische Grass (2006: 668–9).

Sie ermöglichen die Disambiguierung zwischen Toponymen und gleichlautenden Nomina, Eigennamen oder anderen Wortarten bzw. zwischen homonymen Toponymen:

- (3.49) a. *in Essen* (+DetZ)
 b. *im Essen* (kein Toponym)
- (3.50) a. *bei Singen* (+DetZ)
 b. *beim Singen* (kein Toponym)
- (3.51) a. *in Gera* ‚in der Stadt Gera‘ (+DetZ)
 b. *in der Gera* ‚im Fluss Gera‘ (fem., +Det)

Die Liste der Toponyme, denen ein homonymes Nicht-Toponym (Nomen, andere Eigennamenklasse oder Phrase) gegenübersteht, ist lang und enthält auch einige Mehrwort-Toponyme:

Aalen, Altavista, Atlas, Berg, Bergen, Bogen, Bordeaux, Buch, Buchen, Corpus Christi, Darwin, Deutz, Erlangen, Essen, Feld am See, Franklin, Gilbert, Haar, Haindling, Halle, Haus im Wald, Jersey, Langen, Lauf, Laufen, Leiden, Leine, Lippe, Mauern, Mobile, Monroe, Mulde, Mutter, Nixon, Pereira, Pierre, Pilatus, Po, Pocahontas, Posen, Regen, Ried, Rohr, Rumford, Schönberg, Seeleuten, Seine, Singen, Spielfeld, Stegen, Stein, Steinbrück, Stockhausen, Stolberg, Superior, Übersee, Victoria, Warburg, Washington, Weiden, Weil, Weiler, Werden, Zelle, Zimmern, Zug

Die homonymen Pendants dieser Toponyme sind einerseits Nomina, andererseits andere Klassen von Eigennamen, insbesondere Personen und Firmennamen. Im Lexikon sind solche „falschen Freunde“ mit dem Merkmal ‘+!’ markiert. Homonyme zu Verben, Adjektiven, Präpositionen wurden nur teilweise erfasst, da sie in der Regel durch die Regeln bzgl. der Äquivalenz von Groß- und Kleinbuchstaben⁹⁹ eliminiert werden können. Doch wenn die Homonyme am Satzanfang großgeschrieben werden, und in einigen anderen Fällen reichen diese Regeln nicht aus – so ist z.B. der Fluss *Ob* „homonym“ zu *OB* (= *Oberbürgermeister*). Eindrucksvoll ist auch die Liste aus Pouliquen et al. (2006), Tab. 5.1 auf S. 178 und die Experimente von Nadeau (2007: 78–82), die zeigen, dass etwa die Hälfte der englischen Städtenamen ambigüe. Die häufigste Ambiguität – 13% aller Städtenamen sind betroffen – ist die zu Nachnamen von Personen. Zum Teil treten Ambiguitäten außerdem erst für Flexionsformen auf, z.B. für die Stadt *Au* (*in der Hallertau*) mit einer Genitivform, die zur Präposition *aus* homonym ist (z.B. in *Aus neuer Bürgermeister*).

Nicht immer lassen sich Toponyme in den erwähnten Merkmalen eindeutig klassifizieren. Varianten mit unterschiedlichen syntaktischen Eigenschaften werden ins Lexikon als getrennte Einträge aufgenommen (3.52). Erst in der Ontologie (Kap. 3.4) werden beide Lemmata als Namensvarianten einer einzigen geographischen Entität zugeordnet.

- (3.52) a. *Wedding*, .EN+Topon+Oikon+Det:MNS13 → *im Wedding* (ein Berliner Bezirk)
 b. *Wedding*, .EN+Topon+Oikon+DetZ:NNS2 → *in Wedding*

99 Eine gängige Praxis in der Sprachverarbeitung und im IR ist, Großbuchstaben zwar als Äquivalente von den jeweiligen Kleinbuchstaben zu betrachten, aber nicht umgekehrt. So verfährt UNITEX, aber auch z.B. die Suchmaschine yandex.ru. Eine Query *Weil* (bzw. ein Lexikoneintrag) matcht also nicht *weil* im Dokument. Bei einer Suchmaschine ist dieses Feature nützlich, um gezielt nach Eigennamen zu suchen. Bei der Verarbeitung deutscher Texte kann leicht zwischen Nomina und anderen Wortarten unterschieden werden.

Variationen in der Flexion werden direkt im Flexikonskode beschrieben (siehe Kap. 3.3.2.6).

Mehrwort-Toponyme – knapp $\frac{1}{3}$ der Toponyme im Lexikon – werden als Strings repräsentiert. Diese einfache Art der Repräsentation hat den Vorzug, dass alle Toponyme im Lexikon gleich repräsentiert sind. In nur ganz wenigen Fällen ist sie problematisch, nämlich dann, wenn der Name bereits Artikel oder Präposition enthält, d.h. als „vollständige“ NP oder PP vorliegt (Beispiele siehe S. 99 und Fußnote 102 auf S. 111).

3.3.2.1 Lemma und Grundform

Die Konventionen zur Wahl der Grundform eines Lemmas folgen Maier-Meyer (1995: 44):

- Nom. Sing. für Nomina (Bewohner) und Eigennamen
- Nom. Pl. bei ‚Pluralia tantum‘: *Niederlande, Vogesen, Kanaren, Dodekanes, Trobriand-Inseln, etc.*
- *e*-Form bei Nomina und Eigennamen mit adjektivischer Flexion: (*der*) *Deutsche*; (*der*) *Indische Ozean*, (*die*) *Weißer Elster*, (*das*) *Hohe Venn*; *Vereinigte Staaten von Amerika*, *Äußere Hebriden*, *Vereinigte Arabische Emirate*. D.h. im Singular steht die schwache Form, im Plural die starke Form als Grundform.¹⁰⁰
- die unflektierte Positivform bei Adjektiven:
 - die prädikative Form bei flektierenden Adjektiven: *deutsch, südafrikanisch*
 - die einzig existierende Form bei indeklinablen Adjektive: *Bukarester, Idar-Obersteiner*

Das Lemma bzw. dessen ‚Zitierform‘ ist meist mit der Grundform identisch. In einigen Fällen weicht die Zitierform jedoch von der Grundform ab, und zwar dann, wenn die Grundform, die ja als Ausgangsbasis zur Generierung des Paradigmas herangezogen wird, nicht der allgemein üblichen oder offiziellen Zitierform des Toponyms entspricht. Dies ist z.B. bei Lemmata mit adjektivischer Flexion und Lemmata mit Variationen in der orthographischen Repräsentation der Fall. So wird der *Große Arber* üblicherweise als *Großer Arber* gelistet, also nicht in der CISLEX-konformen Grundform. Der Eintrag enthält dementsprechend sowohl die Grundform, die zur Bildung des Paradigmas verwendet wird, als auch die Zitierform des Lemmas:

- (3.53) a. *Große Arber, Großer Arber*.EN+Topon+Oron+Det:M{ADJ+NS13}
b. *Appenzell=Innerrhoden, Appenzell Innerrhoden*.EN+Topon+Choron+admin+...
c. *Russland*, .EN+Topon+Choron+admin+Cl(Staat)+DetZ:N{NS13}
d. *Rußland, Russland*.EN+Topon+Choron+admin+Cl(Staat)+DetZ:N{NS13}

Beispiel (3.53b) gibt die offizielle Schreibweise in der Zitierform, in der Grundform wird von der DELA-Konvention Gebrauch gemacht, ‚=‘ als Symbol für Leerzeichen oder Bindestrich

100 Damit weicht das CISLEX von der in Wörterbüchern und Enzyklopädien üblichen Konvention ab, die starke (artikellose) Form als Grundform zu verwenden: *Persische Golf* vs. *Persischer Golf*, *Hohe Venn* vs. *Hohes Venn*. Um dies zu umgehen, unterscheidet das hier vorgestellte Lexikon zwischen ‚Lemma‘ und ‚Grundform‘.

zu setzen. (3.53d) stellt eine orthographische Variante von (3.53c) dar. Beide werden unter dem Lemma *Russland* im Lexikon geführt.¹⁰¹

Diese Konvention, mehrere Grundformen einer Zitierform zuzuordnen, bleibt aber beschränkt auf Flexionsformen (*Große Arber* ist eine Flexionsform von *Großer Arber*), orthographische Varianten und z.T. Abkürzungen (vgl. Kap. 3.3.2.7). Zur Behandlung von „echten“ Allonymen ist diese Konvention nicht geeignet, da z.B. *Russland* und *Russische Föderation* voneinander verschiedene syntaktische Eigenschaften (wie die Verwendung mit und ohne Artikel) aufweisen. Diese Eigenschaften werden als globale Eigenschaften eines Lemmas betrachtet, einzelne Formen können sich darin nicht unterscheiden. Zur Allonymbehandlung ist deshalb ein anderer Mechanismus vorgesehen (siehe Kap. 3.4).

3.3.2.2 Genus

Alle deutschen Toponyme determinieren, wenn sie als Kopf einer Nominalphrase stehen, das Genus von dependenten attributiven Adjektiven und Determinativen (*der Rhein* vs. *die Donau*). Auch anaphorische Pronomina kongruieren in der Kategorie Genus mit dem Toponym: *der Rhein ... Er fließt durch vier Länder*. Anders als bei den attributiven Wortarten ist das Genus eines Toponyms nicht an dessen (morphologischer) Form zu erkennen. Es muss deshalb im Lexikon spezifiziert sein. Dass manche Toponyme im Genus variieren, stellt dazu keinen Widerspruch dar. Variationen im Genus sind verhältnismäßig selten, meist handelt es sich um Toponyme nicht-deutscher Herkunft (*der/die Peloponnes*), oder die Genusvariation geht einher mit einer Variation im Artikelgebrauch (*Wedding*, siehe Kap. 3.3.2.4). Das Genus eines Toponyms lässt sich sicher bestimmen, indem man es in geeignete Kontexte einsetzt (z.B. *Peter badet im ...* vs. *in der ...*) und die Grammatikalität der resultierenden Phrase überprüft.

Für zwei Gruppen von Toponymen lässt sich das Genus so jedoch nicht bestimmen: (1.) ‚Pluralia tantum‘ (*die Alpen*, siehe Kap. 3.3.2.3) und (2.) Toponyme, die ausschließlich mit Klassenbezeichner, also immer in enger Apposition verwendet werden (*Landkreis München*, siehe Kap. 3.3.2.5). Das Genus von Toponymen dieser beiden Gruppen wird im Lexikon als ‚unspezifiziert‘ (‘U’) gekennzeichnet.

	M	F	N	U
Oikonyme	32 (0%)	50 (0%)	32999 (100%)	0
Oronyme	5766 (68%)	1990 (24%)	705 (8%)	4 (0%)
Choronyme	714 (11%)	1143 (18%)	3974 (61%)	644 (10%)
Hydronyme	1154 (56%)	823 (40%)	90 (4%)	1 (0%)
(Seen)	561 (27%)	22 (1%)	10 (0%)	0
(Fließgew.)	473 (23%)	731 (35%)	28 (1%)	0
(Meere)	57 (3%)	66 (3%)	52 (3%)	1 (0%)
Bauwerke	61 (18%)	103 (30%)	181 (52%)	1 (0%)
Hodonyme	129 (66%)	62 (32%)	3 (2%)	0
<i>insges.</i>	7856 (16%)	4171 (8%)	37952 (75%)	650 (1%)

Tabelle 3.3: Verteilung des Genus auf semantische Typen von Toponymen (M=mask., F=fem., N=neutr., U=unspezifiziert/Pluralia tantum)

101 Zur Erweiterung des DELA-Formats vgl. Silberztein (2005) und NOOJ-MANUAL mit dem Beispieleintrag *csar, tsar .N+FLX=Pen*. Silberztein verwendet den Begriff „Super-Lemma“.

3.3.2.3 Pluralia tantum

Einige Toponyme sind Pluralia tantum (*die Alpen, die Niederlande, die Dardanellen*). Diese bilden nur Pluralformen und sind ohne Genus. Allerdings ist für einige der Pluralia tantum die Referenz auf ein einzelnes Element des „kollektiven“ Toponyms möglich (3.54b). Für diesen Fall ist auch das Genus des Plurale tantum im Lexikon verzeichnet.

- (3.54) a. * *eine der Niederlande*
b. *Curaçao ist eine der niederländischen Antillen.*

3.3.2.4 Artikelgebrauch bei Toponymen

Eigennamen werden gewöhnlich ohne Artikel verwendet, darin unterscheidet sich das Deutsche nicht von anderen europäischen Sprachen, wie dem Englischen, Französischen oder Italienischen. Allerdings gilt diese Regel nicht generell für Toponyme. Für diese ist der Artikelgebrauch „lexikalisiert“ (vgl. Fleischer 1964: 377, 1967: 148, Gallmann 1997, Sturm 2005: 80) und wird deshalb im Lexikon mit den Merkmalen +DetZ und +Det kodiert. Der Artikel in (3.55b) ist obligatorisch und kann nur in Texten im Telegrammstil, z.B. in Überschriften (3.56), in Aufzählungen (3.57), oder bei Angabe der Grundform, z.B. als Stichwort in einem Lexikon (*Bayerischer Wald*, vgl. Kap. 3.3.2.1) weggelassen werden. Mit bestimmten Verben wie *heißen* (3.58) muss er ausgelassen werden. Er ist aber auch nicht Bestandteil des Namens,¹⁰² sondern lediglich Bestandteil der Nominalphrase, dessen Kopf das Toponym ist: Adjektivmodifikatoren stehen zwischen Artikel und Eigennamen, er ist durch andere Determinative austauschbar und kontrahiert mit Präpositionen.¹⁰³

- (3.55) a. *in Frankreich* (+DetZ)
b. *in der Türkei* (+Det)

(3.56) *Vier Tote bei Terroranschlag in Türkei*

(3.57) *Maestros Yachting – Segeln in Griechenland, Türkei, Italien und Karibik*

(3.58) *Das neue Boom-Land heißt Türkei*

Verschiedene Klassen von Toponymen unterscheiden sich hinsichtlich des Artikelgebrauchs (Tab. 3.4). Nur wenige Namen von Siedlungen (Oikonyme) werden mit Artikel verwendet (*die Wiener Neustadt*, eine eigenständige Stadt, kein Stadtteil von Wien), meist handelt es sich um Bezeichnungen von Stadtteilen (*die Schwarze Pumpe, die Blumenau, der Friedrichshain*). Diese variieren mitunter im Gebrauch des Artikels (*im Wedding* vs. *in Wedding*), allerdings lässt sich dies auch damit erklären, dass außerhalb Berlins die syntaktischen Eigenschaften des *Wedding* nicht bekannt sind. Bei den Bezeichnungen für Länder und Regionen (Choronyme) ist das Verhältnis obligatorischer Artikel – kein Artikel in etwa ausgewogen. Für die

102 Der Name kann den Artikel bereits selbst enthalten, z.B. im Fall des österreichischen Berges *Die Ditzl* (weitere Beispiele auf S. 99). Im Lexikon wurde hier das Merkmal '+DetZ' gesetzt, da die Konsequenzen für die Syntax des Namens ähnlich sind wie für „artikellose“ Toponyme. '+DetZ' ist also nicht gleichbedeutend mit Verwendung ohne Artikel.

103 Vgl. Leys (1967) in der Argumentation gegen Sørensen (1958: 151,166), der *the* in *the thames* als „a mere syllable“ bezeichnet.

Namen von Gewässern (Hydronyme), Bergen und Gebirgen (Oronyme) sowie Verkehrswegen (Hodonyme) wurde bislang ausschließlich die Verwendung mit Artikel beobachtet – eine empirische Beobachtung, die Anlass zu einer Regel geben könnte.¹⁰⁴ Ebenso gilt das für die Beobachtung, dass mit Ausnahme einiger pluralischer, da koordinierter Toponyme (*auf Saint Vincent und den Grenadinen*) alle Toponyme mit dem Merkmal +DetZ Neutra sind.

	DetZ	Det	inkorpDet
Oikonyme	32914 (99%)	167 (1%)	0
Oronyme	0	8456 (100%)	8 (0%)
Choronyme	3275 (54%)	2801 (46%)	2 (0%)
Hydronyme	0	2068 (100%)	0
Bauwerke	57 (17%)	288 (83%)	0
Hodonyme	0	193 (99%)	1 (1%)
<i>insges.</i>	36246 (72%)	13973 (28%)	11 (0%)

Tabelle 3.4: Artikelgebrauch bei Toponymen, Verteilung auf semantische Typen (+Det=mit Artikel, +DetZ=ohne Artikel, +inkorpDet mit „inkorporiertem“ Artikel, siehe S. 99 und Fußnote 102 auf der vorherigen Seite)

Eine Ausnahme besteht, wenn der Eigenname genauer spezifiziert wird und dabei eventuell nicht auf eine einzelne Entität, sondern auf einen Teilaspekt dieser (*in der laizistischen Türkei, im ehemaligen Jugoslawien, im vereinten Deutschland, im spanischen Baskenland, im französischen Baskenland*) oder auch eine Menge von gleichnamigen Entitäten referiert: *München besitzt auch eine Mozartstraße*. Die Grenze zu einem neuen komplexeren Eigennamen (*französisches Baskenland*) ist dabei nicht immer klar zu ziehen. Entscheidend ist, dass hier auch bei Toponymen, die sonst ohne Artikel stehen, der bestimmte Artikel (oder ein anderes Determinativ) gesetzt wird:

- (3.59) a. *In Deutschland* begegnete man sowohl dem skandinavischen *Etatismus* als auch dem ...
 b. *Wie hat sich die Lebensqualität im vereinten Deutschland* entwickelt? ...

Eine Ausnahme bilden hier die Quantoren *ganz* und *halb* – bei artikellosen Namen steht die unflektierte Form, sonst die flektierte mit bestimmtem Artikel: *in ganz Deutschland* vs. *in der ganzen Türkei* (vgl. S. 142). In Titeln, Aufzählungen etc. (s.o.) steht auch mit Modifikator kein Artikel: *unvergessliches Frankreich*. Mit Präposition wird der Artikel, wenn möglich, kontrahiert:

- (3.60) a. *im Schwarzwald*
 b. ? *in dem Schwarzwald*

¹⁰⁴ Die Annahme einer allgemeingültigen Regel sollte mit äußerster Vorsicht erfolgen. So liegen Helbig & Buscha (1991: 231) mit der Behauptung Siedlungsnamen („Ortsnamen“) würden nie, Berg- und Gewässernamen dagegen immer mit Artikel verwendet nicht, zumindest nicht vollständig richtig. Dass sich keine eindeutigen Regeln angeben lassen, zeigt sich auch darin, dass die Dudengrammatik (DUDEN-GRAMMATIK 1973: §376, 1998: §357) andere „Faustregeln“ nennt. Die Namen von Landschaften stehen demnach immer mit, Länder mit Ausnahmen ohne, Städte immer ohne, Straßen und Plätze wiederum mit Artikel. Ein andere Ausgabe der DUDEN-GRAMMATIK (1984: 220–1, §357) gibt dagegen eine brauchbare Beschreibung: Kontinente, Länder, Gebiete und Städte stehen mit „nicht wenigen Ausnahmen“ und Variationen ohne, Berge, Gebirge, Flüsse, Seen, Meere und Sterne dagegen mit bestimmtem Artikel.

Als Besonderheit haben Toponyme (und andere Eigennamen) im Genitiv verschiedene Formen, je nachdem ob sie mit oder ohne Artikel (Determinativ) benutzt werden (vgl. Gallmann 1997: 77, Sturm 2005: 83, Mel'čuk 2006: 129):

- (3.61) a. *Türkeis Außenminister Gül bleibt hart bei Zypern-Status.*
 b. *An den Verhandlungen nahmen die Außenminister der Türkei, Abdullah Gül, und Zyperns, George Lillikas, teil.*

3.3.2.5 Toponyme mit obligatorischem Klassenbezeichner

Verwaltungseinheiten (Landkreise, Grafschaften etc.) sind häufig nach einer bedeutenden Stadt bzw. nach der Stadt, in der die Verwaltung ihren Sitz hat, benannt. In diesem Fall kann nur in Kombination mit einem Klassenbezeichner oder einem eine Einheit dieses Typs regierenden Prädikat auf die Region/Verwaltungseinheit referiert werden. Der Name allein ist nicht ausreichend.

- (3.62) a. *in München* immer ‚in der Stadt München‘
 b. *im Landkreis München* (+oblCl)
 c. *der Münchner Landrat / der Landrat von München* (München = ‚Landkreis München‘)

Der Klassenbezeichner kann variieren (*Landkreis, Kreis*) oder kann bei koordinierten Toponymen ausgeklammert werden (*in den Landkreisen Erding, Freising und Ebersberg*). Um mit einer Grammatik auch diese beiden Fälle beschreiben zu können (vgl. Constant 2003b: 152–3), ist es nicht ausreichend im Lexikon einen Eintrag, der Klassenbezeichner und Toponym kombiniert, aufzunehmen (**Landkreis München, .EN. . .**). Solche Kombinationen finden sich im Lexikon nur, wenn sie offizielle Bezeichnung eines Toponyms sind (*Freistaat Bayern, Haschemitisches Königreich Jordanien, Volksrepublik China, Republik Irland*).

3.3.2.6 Flexion

When an Irishman returns from Liverpool to the home country, Éire, he goes hÉirinn, i.e. to Ireland. His Muscovite friend living в Москве, in Moscow, comes из (from) Москвы to visit him — and then goes back в Москву. Similarly, a Magyar who left his capital city comes from Budapestől, from Budapest, but later returns Budapestre, to Budapest. Hungarian even has special forms for names when comparing places: Budapesttel is larger than Szegeddel; and, when speaking of distances between locations, one finds Budapesttől not far from Wien, Vienna, which knows no such complications. (Kadmon 2001: 59)

$\frac{2}{3}$ der Toponyme im Lexikon verfügen über nur zwei Formen: von der Grundform unterscheidet sich lediglich der Genitiv durch ein angefügtes *-s* (*Berlin* → *Berlins*). Für das verbleibende Drittel der Toponyme ist die Flexion komplexer als die anderer Eigennamenklassen.¹⁰⁵ Zurückzuführen ist dies (1.) auf generische Elemente, die teils wie gewöhnliche

105 So werden praktisch alle Vor- und Familiennamen nur im Genitiv durch Anhängen von *-s* in der Oberflächenform modifiziert, von veraltet klingenden Ausnahmen (*Fritzens, Voß* → *Vossens*, ebenso *Grazens*). Auch im Vergleich zu anderen z.B. slavischen Sprachen (vgl. Mańczak 1995) ist die Flexion deutscher Toponyme relativ überschaubar.

Nomina flektieren: *Friedrichshain*, *Blumenau* (vgl. Kap. 3.2.1.1). Der generische Bestandteil vererbt häufig auch Genus und Numerus an das Toponym (siehe Kap. 3.3.2.2). (2.) sind Toponyme häufig aus mehreren flektierenden Elementen zusammengesetzt, in der Kombination sind so bis zu acht distinkte Formen möglich: *Persischer Golf*, *der Persische Golf*, *des Persischen Golfes/Golfs*, *Persischem Golf/e*, *dem Persischen Golf/e*.

Die Flexionsformen sind nach Kasus spezifiziert, im Fall von Toponymen mit adjektivisch flektierenden Bestandteilen auch nach der Kategorie Deklination. In Kombination mit dem bestimmten, unbestimmten oder ohne Artikel werden verschiedene Formen verwendet:

- (3.63) a. *im Persischen Golf*
 b. *zwischen Persischem Golf und Rotem Meer*

Die Flexion der Toponyme wird mittels der Flexionsklassen des CISLEX beschrieben. Einige Erweiterungen ermöglichen die effiziente Beschreibung (1.) von Flexionsvarianten und (2.) der Flexion komplexer Einträge (siehe Tab. 3.5). Letztere werden zunächst an den Leerzeichen in Einzelwörter gesplittet - im Deutschen kann an Positionen, wo ein Bindestrich steht (oder stehen kann), keine Flexionsendung auftreten.¹⁰⁶ Anschließend wird jeder Bestandteil nach dem ihm zugewiesenen Flexionskode flektiert. Alle Formen der Einzelwörter werden dann, sofern die grammatischen Merkmale unifizieren, mit den Formen der anderen Bestandteile wieder mit Leerzeichen aneinandergesetzt.¹⁰⁷ Zu den Grundformen *Frankfurt am Main* wird so die abweichende Flexionsform *Frankfurts am Main* erzeugt, *Ostfriesische Inseln* die Formen *Ostfriesischer Inseln* (Genitiv) und *Ostfriesischen Inseln* (Dat., Akk.).

106 Im Gegensatz zu Silberztein (1990), der die Trennung mit dem Zeichen ‘/’ explizit macht, oder Savary (2005), die die Mehrwortigkeit anhand eines vordefinierten Alphabets segmentiert. Als ein Segment gilt entweder eine Sequenz aus Buchstaben des Alphabets oder jedes einzelne Nicht-Buchstaben-Zeichen. *Amerikanisch-Ozeanien* würde demnach in drei Bestandteile “*Amerikanisch*”, “*-*” und “*Ozeanien*” aufgeteilt.

Lediglich ein Toponym ist mir bekannt, wo die Zerlegung in Einzelbestandteile anhand der Leerzeichen nicht möglich ist: *Washington, D.C.* mit der Genitivform *Washingtons, D.C.*. Hier musste zur korrekten Flexion eine besondere Konvention eingeführt werden: das Unicodezeichen U+200C (zero width non-joiner) markiert Grenzen zwischen Einheiten, die nicht bereits durch eine Leerzeichen markiert sind. Er steht also zwischen *Washington* und dem folgenden Komma. Bei der Flexion der komplexen Einheit wird er wieder entfernt.

107 Das Verfahren entspricht also im Wesentlichen dem von Silberztein (1990):

- (1) `cousin(N32)/germain(A32),NA :ms/++`
`abaissée(N21)/de//aile,NDN :fs/-+`

Flexionskodes stehen in Klammern, `:ms` bzw. `:fs` bezeichnen Genus und Numerus der Grundform. `+` und `-` geben an, ob der Lexikoneintrag in Genus bzw. Numerus flektiert. Bei der Expansion werden die Formen unifiziert. Dann ergeben sich die Formen:

- (2) `cousin germain.NA:ms`
`cousine germaine.NA:fs`
`cousins germains.NA:mp`
`cousines germaines.NA:fp`
`abaissée d'aile.NDN:fs`
`abaissées d'aile.NDN:fp`

Demgegenüber beschreibt Agata Savary (Savary 2005, 2000, Chrobot 1998) die Flexion anhand von Graphen. Zusammen mit ihrer Methode der Tokenisierung (siehe Fußnote 106) ist das mit Sicherheit der universellere Ansatz – geeignet für viele Sprachen. Der Nachteil besteht jedoch in der hohen Zahl der Flexionsgraphen, die gezeichnet werden müssten. Für das hier vorgestellte Lexikon ist mit etwa 400 zu rechnen.

Symbol	Verwendung	Beispiel
Operatoren		
	trennt Varianten	<i>Marina del Rey</i> {1:NS2 NS2} → <i>Marinas del Rey</i> oder <i>Marina del Reys</i>
+	Trenner für Teile eines Mehrwortlexems	<i>Hohes Venn</i> {ADJ+NS13} → (<i>das</i>) <i>Hohe Venn</i> , (<i>des</i>) <i>Hohen Venns</i> , ...
1:	(auch 2: etc.): nur erster (zweiter etc.) Bestandteil flektiert	<i>Frankfurt am Main</i> {1:NS2} → <i>Frankfurt am Main</i> , <i>Frankfurts am Main</i> ; <i>Sankt Veit an der Glan</i> {2:NS2} → <i>Sankt Veit an der Glan</i> , <i>Sankt Veits an der Glan</i>
Flexionsklassen		
0	indeclinabler Bestandteil/Teilwort	<i>Bosnien und Herzegowina</i> {NS2+0+NS2}
ADJ00	indeclinables Adjektiv	<i>Starnberger See</i> {ADJ00+NS13}
ADJ	alle flektierenden Adjektive ¹⁰⁸	<i>Große Syrte</i> {ADJ+NS0} → * <i>Größere Syrte</i>
DET4	flektierender bestimmter Artikel als Namensbestandteil	<i>Sankt Vincent und die Grenadinen</i> {0+NS2+0+DET4+NPO} → <i>Sankt Vincents und der Grenadinen</i> , <i>Sankt Vincent und den Grenadinen</i>
NS22	Genitivendung <i>-ens</i> (veraltet)	<i>über Grazens Dächer und Turmesspitzen</i>
NS113	Genitiv Sg.: -∅, -s, -es, sonst endungslos (<i>Westnordwestgrat des Aletschhorns/ Aletschhorns/ Aletschhorn</i>) ¹⁰⁹	
NP120	<i>Powiat</i> , Plural <i>-y: Powiaty</i> (administrative Untergliederung der zweiten Ebene in Polen)	

Tabelle 3.5: Erweiterungen der CISLEX-Flexionsklassen zur vollständigen Beschreibung der Flexion lokativer Nomina und deutscher Toponyme

Der Plural von Toponymen, auch wenn er gelegentlich verwendet wird (3.64), wurde im Lexikon nicht kodiert.

- (3.64) a. *die beiden Amerika* (Nord- und Südamerika)
 b. *die beiden Deutschland* (BRD und DDR)
 c. *das Königreich beider Sizilien*
 d. *Universitätsstandorte wie Göttingen, Marburg, Gießen, Erfurt, die beiden Frankfurts oder Greifswald*

Ebenfalls im Flexionsmodul, bzw. einem unmittelbar nachgeschalteten Modul, werden zwei Phänomene behandelt, die nicht der Flexion zuzuordnen sind, sondern allgemein die Variation von Formen betreffen: (1.) der Verlust von Diakritika. So ist *Orleans* in deutschen Texten etwa viermal häufiger zu finden als *Orléans*. (2.) die Abkürzung des Toponyms oder seiner Bestandteile (*Frankfurt am Main* → *Frankfurt a. M.* oder *Frankfurt/Main*) Die Beschreibung erfolgt mit einer Reihe von Ersetzungsregeln, die kaskadenartig angewendet werden: *Sankt Anton am Arlberg* → *St. Anton am Arlberg* → *St. Anton (Arlberg)*, *St. Anton a. A.*

Die resultierenden Formen werden als „Flexionsformen“ dem Lemma (hier *Orléans* und *Sankt Anton am Arlberg*) zugeordnet (vgl. auch die Beispiele in Kap. A.1.2.3):

(3.65) *Orleans,Orléans*.EN+Topon+Oikon+DetZ:aeN:deN:geN:neN

(3.66) *St. Anton a.A.,Sankt Anton am Arlberg*.EN+Topon+Oikon+DetZ:aeN:...

Allerdings lassen sich nicht alle abgekürzten Formen durch Regeln beschreiben, und müssen deshalb, wie im folgenden Kapitel erläutert, separat beschrieben werden.

3.3.2.7 Abgekürzte Formen

Abkürzungen¹¹⁰ und Kurzformen von Toponymen sind in einer Vielzahl von Typen zu finden, wobei Typen hier als Muster (Pattern) verstanden werden und nicht als Klassen wie Initialismen/Alphabetismen (*USA, DDR*), Akronyme (*MeckPomm*), „eigentliche“ Abkürzungen (*Mecklbg.-Vorp.*) und Mischformen (*D'dorf*). Eine solche linguistisch motivierte Klassifikation ist erst notwendig, wenn der orthographischen Repräsentation eine phonetische zugeordnet werden muss. Aus Perspektive der Orthographie ist allen Abkürzungen die triviale Charakteristik gemeinsam, dass Buchstaben getilgt werden. Allerdings werden gegebenenfalls Interpunktionszeichen eingefügt, oder einzelne Buchstaben ändern sich bzgl. Groß-/Kleinschreibung.

108 Da für adjektivische Derivationen und adjektivisch flektierende Bestandteile von Toponymen weder Komparativformen existieren, muss keine Unterscheidung der adjektivischen Flexionsklassen nach Stammveränderungen getroffen werden. Eine adjektivische Flexionsklasse ADJ genügt zur Beschreibung, neben den indeklinablen Adjektiven ADJ00.

109 Die endungslosen Genitivformen sind in Korpora etwa ebenso häufig wie die Formen auf *-s* oder *-es*, z.B. *auf dem Südgrat des Aletschhorn, die Erstbesteigung des Aletschhorn, die Überschreitung des Aletschhorns, Nordostgrat und weiter zum Gipfel des Aletschhornes*. Dies rechtfertigt eine neue Flexionsklasse.

110 Vgl. Belleil & Maurel (1997: 275) für die Situation bei französischen Toponymen.

Die gesammelten Abkürzungsmuster unterscheiden sich darin, wie regulär sie sind, d.h. auf welche Menge von Toponymen sie anwendbar sind. Am einen Ende der Skala stehen „lexikalisierte“ Pattern, die nur ein einzelnes Toponym betreffen (*Düsseldorf* → *D'dorf*, *Kurfürstendamm* → *Ku'damm*, *Württemberg* → *Wttbg*), am anderen Ende dagegen Pattern, die auf eine Vielzahl von Toponymen zutreffen, z.B. die Abkürzung von Bestandteilen wie *am Main* (→ *a.M.*) usw., wo sich Regeln wie die nachfolgende formulieren lassen: ‘Wenn auf die Wortform *am* ein großgeschriebenes Wort folgt, kürze *am* zu *a.* und das folgende Wort auf seinen ersten Buchstaben.’

Auch wenn die Regularität der Muster eine graduelle Eigenschaft ist, muss aus operationellen Gründen eine Trennlinie gezogen werden zwischen regulären und „lexikalisierten“ Abkürzungsmustern. Letztere werden explizit im Grundformenlexikon aufgeführt, erstere in einem der Flexion nachgeschalteten Modul (vgl. Kap. 3.3.2.6) behandelt. Ziel muss sein, einen Kompromiss zu finden zwischen einem kleinen Lexikon mit nur einem Minimum an Lemmata und einer möglichst kleinen Menge an Ersetzungsregeln. Dies garantiert die Wartbarkeit von Lexikon und Regeln. Darüberhinaus sind eine Reihe von Faktoren zu berücksichtigen, womit sich folgende Bedingungen dafür ergeben, dass Abkürzungen durch reguläre Pattern synthetisiert werden:

1. das Pattern muss für mehr als ein Lexem im Lexikon zutreffen, und für diese Lexeme auch hinreichend gut in einem Textkorpus belegt sein. Unikale Pattern wie *Düsseldorf* → *D'dorf* sind damit ausgeschlossen, auch wenn diese okkasionell für andere Toponyme belegt sind (z.B. *Zehlendorf* → *Z'dorf*)
2. die syntaktischen Eigenschaften der Form (Σ_X) ändern sich durch die Abkürzung nicht, d.h. alle syntaktischen Eigenschaften (Kasus, Genus, Artikelverwendung) bleiben konstant
3. das Pattern muss für alle zutreffenden Lemmata zu einem vernünftigen, nicht ambiguen Ergebnis führen. D.h. die Ergebnisse einer regulären Ersetzung müssen im Hinblick auf mögliche Anwendungen des Lexikons kontrollierbar bleiben. So kann zwar das Suffix *-heim* zu *-h.* oder *-stadt* zu *-st.* abgekürzt werden: *Rüsselsheim* → *Rüsselsh.*, *Veitshöchheim* → *Veitshöhh.*, *Eisenhüttenstadt* → *Eisenhüttenst.* Diese Regel ist jedoch nicht auf alle Ortsnamen sicher und ohne Nebeneffekte anwendbar: *Stroheim* → *Stroh.*, *Herbstadt* → *Herbst.*
4. es sind keine Derivationen von der abgekürzten Form möglich. Die Ableitung von Adjektiven und Wohnernamen, so wie in Kap. 3.3.2.11 beschrieben, erfolgt ausgehend von Lemma bzw. Grundform und nicht von Flexionsformen. Folglich müssen Lemmata wie *DDR*, *U'haching*, *Meck-Pomm* im Lexikon vorhanden sein, um davon die Derivation *DDRler*, *U'Hachinger*, *Meck-Pommmler* bilden zu können.¹¹¹
5. das Pattern sollte mit allgemeinen orthographischen und typographischen Konventionen in Einklang stehen. Dadurch werden z.B. folgende Formen ausgeschlossen: *Frankfurt a./M.*, *Frankfurt(Main)* (ohne Leerzeichen vor der öffnenden Klammer).

Kap. A.1.3 im Anhang zeigt eine Übersicht über alle regulären Pattern und erläutert deren Anwendung.

¹¹¹ Im Prinzip gilt gleiches für Abkürzungen von Derivationen wie in *Brandenbg. Konzerte*, *Regensbg. Motorboot- u. Wassersportverein* (für *Brandenburgische*, *Regensburger*). Diese lassen sich als „Nullderivationen“, d.h. als Derivationen durch Suffigierung mit einem leeren Suffix, behandeln.

Per Konvention wird „lexikalisierten“ Abkürzungen als Zitierform (siehe Kap. 3.3.2.1) die nicht abgekürzte Form zugeordnet:

(3.67) Hann.=Münden,Hannoversch Münden.EN+Topon+Oikon+abbr+DetZ:N{NS13}

(3.68) Bln.,Berlin.EN+Topon+Choron+admin+Cl(Bundesland)+abbr+DetZ:N{NS0}

Damit wird erreicht, dass im resultierenden Vollformenlexikon und darauf zugreifenden Anwendungen kein Unterschied zwischen „lexikalisierten“ und „regulären“ Abkürzungen zu spüren ist. Beide verfügen auch über die gleichen syntaktischen Eigenschaften. Aus linguistischen (wie sieht die phonetische Form solcher Ortsnamen aus?) wie praktischen (Assymetrie in der Organisation des Lexikons) Gründen wird im Lexikon als Grundform immer die unabgekürzte Form geführt. Auch dann wenn abgekürzte Formen häufiger sind (vgl. die Beispiele in Kap. A.1.3), oder als offizielle, amtliche Form eines Ortsnamens definiert sind (*Hann. Münden, Heiligenstadt i.OFr.*).

3.3.2.8 Kürzel, Ländercodes

Eine weitere Möglichkeit um in Texten auf Orte (Städte, Länder und Regionen) zu referenzieren, ist die Verwendung verschiedener Kürzel und Ländercodes: ISO 3166, Internationales Olympisches Komitee [IOC], United Nations Code for Trade and Transport Locations [UN/LOCODE], Amtlicher Gemeindeschlüssel [AGS], Amtliche Gemeindekennzahl [GKZ], etc. Die gleiche Funktion übernehmen auch Telefonvorwahlen, Kfz-Kennzeichen und Top-Level-Domains. Allerdings ist die Verwendung der Kürzel und Codes in sprachlichen Äußerungen eher eingeschränkt – einzelne Textsorten wie Kleinanzeigen in Zeitungen ausgenommen. Von einer Aufnahme ins Lexikon wurde deshalb abgesehen.

3.3.2.9 Nicknames, Epitheta, Un-Orte

Nicknames von Orten, d.h. scherzhafte, metaphorische, nur teilweise transparente Bezeichnungen verschiedener Orte (vgl. Kadmon 2001: 49–50), wurden, da sie lexikalisiert und monoreferentiell sind, auch ins Lexikon der Ortsnamen aufgenommen:

Mainhattan, Bankfurt, (The) Big Apple, Spreethen, Isarathen, Elbflorenz

Verschiedene Standards zur Annotierung von Named Entities (siehe Kap. 5.1.5.1) betrachten diese ebenfalls als „vollwertige“ Ortsnamen. Allerdings ist der Übergang zu epitheton-artigen Klassenbezeichnern fließend. Diese können attributiv mit Ortsnamen verwendet werden oder auch – wie im Grunde alle Klassenbezeichner – stellvertretend für den Ortsnamen stehen. Epitheta wurden ins Lexikon der Klassenbezeichner aufgenommen. Ausschlaggebendes Kriterium zur Abgrenzung ist, ob das fragliche Wort in der Konstruktion in <DET> <Epitheton/Nickname> <Toponym> stehen kann oder nicht. Falls ja, handelt es sich um einen Klassenbezeichner. Das Kriterium der Monoreferentialität kann nur manchmal einen Hinweis geben. So hat *Lutherstadt* zwei Referenten (*Wittenberg* und *Eisleben*).

Als „Un-Orte“ wird hier schließlich eine kleine Gruppe von Ortsnamen bezeichnet, die sich zwar nach interner morphologischer Struktur und syntaktischem Verhalten nicht von anderen Oikonymen unterscheiden, aber die Bedeutung ‚kleiner, abgelegener Ort‘ haben.

Kleinkleckersdorf, Pusemuckel, Hintertupfingen, Kleinsiehstenich, Hinterpumuckel; mit den Varianten *Hintertupfing, Pusemuckeldorf, Pusemuckelsdorf, Kleinkleckerdorf*

Meist stehen sie in der Konstruktion *in X*, mit der Bedeutung ‚an einem abgelegenen Ort‘. Eine Ausnahme bilden *Buxtehude* und einige andere Toponyme, die in dieser oder ähnlichen Konstruktion (*in Buxtehude, in der Wallachei*) stehen können, aber daneben auch auf reale Orte referieren. Siehe auch Kap. 3.5.3 (Ortsadverbiale) und Kap. 5.1.5.3.1.3 (Antigrammatiken).

3.3.2.10 Koordinationsellipsen: *Ober-, Mittel- und Unterfranken*

Koordinationsellipsen gehören mit zu schwierigsten Problemen der automatischen Sprachverarbeitung, da sie als syntaktisches Phänomen nicht nur ober-, sondern auch unterhalb der Wortebene, also auf Morphemebene ansetzen.¹¹² Voraussetzung für eine vollständige Behandlung ist deshalb der Zugriff auf einzelne Morpheme und Wortbestandteile.

- (3.69)
- a. *Slowakische und Tschechische Republik*
 - b. *zwischen Atlantischem und Pazifischem Ozean*
 - c. *zwischen Ammer- und Starnberger See*
 - d. *Mitter- und Untersending*
 - e. *Adelz- und Odelzhausen*
 - f. *Hecken- und Schlehengäu*
 - g. *Kirch-, Forst- und Osterseeon*
 - h. *Kreuz- und Ödenpullach*
 - i. *Glas-, Berg-, Mitter- und Niederlern*
 - j. *Pitz- und Kaunertaler Alpen*
 - k. *Haid- und Kirchstockacher* (Bewohner)

Von einer Auflösung der Ellipsen in einem Vorverarbeitungsschritt wurde abgesehen, da dadurch der Text verändert wird.¹¹³ Stattdessen wurden einige koordinierte Toponyme der Form

112 Prinzipiell ließen sich auch die elliptischen Formen als Wortformen annehmen (vgl. Mel’čuk 1997-2001: I,189–195): *Chiem-* wäre dann eine Wortform (ein Lex) des Lexems *Chiemsee*, allerdings mit sehr beschränkter Distribution. Eine solche Lösung erscheint nicht praktikabel, da (1.) die erlaubten Kontexte solcher Formen nur sehr schwer zu beschreiben sind, und (2.) Formen wie *Ober-, Nord-*, die ja Präfixe vieler Toponyme sind, zu einer Explosion der Ambiguitäten (Homonyme) im Lexikon führen würde.

113 Zwei mögliche Vorgehensweisen seien hier kurz skizziert:

- die Zerlegung erfolgt anhand von Pattern und Listen der beteiligten Prä- und Suffixe: *Ober-, Unter-, Mitter-, Zentral-, Mittel-, Groß-, Klein-, Alt-, Neu-, Nord-, Ost-, Nordwest-, ...; -see, -straße, -gässchen, ...; -hausen, -gäu, ...*
Mit diesem Verfahren lassen sich die Beispiele (3.69a)–(3.69f) behandeln. Für (3.69g)–(3.69i) sind sehr umfangreiche Prä- und Suffixlisten notwendig, die für Fälle wie (3.69j) auch Mehrwort-Suffixe einschließen müssen.
- finde ein Suffix des zweiten vollständigen Bestandteils, so dass der erste Bestandteil konkatentiert mit dem Suffix ein gültiges Toponym ist. Diese Lösung erfordert den Zugriff auf ein möglichst vollständiges Toponymlexikon. Um in Fällen mit ambigen Zerlegungsmöglichkeiten wie (3.69h) keine falschen Lösungen (*Kreuzach, Kreuzlach*) zu generieren, muss diese Strategie noch mit weiteren Heuristiken kombiniert werden, z.B. der ‘geometric minimality’ (S. 216), die die Lösung bevorzugt, bei der beide Orte in größtmöglicher räumlicher Nähe liegen. Damit wäre *Kreuz-* in (3.69h) eindeutig als *Kreuzpullach* aufzulösen.

Ober-, Mittel- und Unterfranken, Hecken- und Schlehengäu oder *Ost- und Westberlin* direkt ins Lexikon aufgenommen und durch das Merkmal '+CTopon' gekennzeichnet. Die Lösung ist insofern praktikabel, da nur eine kleinere Zahl von Toponymen auf diese Weise tatsächlich koordiniert wird. Voraussetzung ist, dass die koordinierten Einheiten auch einander nahestehen und als größere Einheit gesehen werden können wie in (3.70). Nicht aufgenommen wurden dagegen Kombinationen, die außer einem gemeinsamen Prä- oder Suffix (semantisch) wenig verbindet (3.71).

- (3.70) a. *Ober-, Mittel- und Unterfranken*
 b. *Ober- und Unterammergau*
- (3.71) *am Chiem- und Ammersee*

3.3.2.11 Derivationen: Adjektive und Bezeichnungen für Bewohner

Das Lexikon enthält weiter die Derivationen von Toponymen, d.h. die Beziehungsadjektive und die Bezeichnungen für Bewohner (sog. ‚Ethnika‘, franz. ‚noms ethniques‘ oder ‚gentilés‘). Andere deonymische Ableitungen (*Afrikanist, Amerikanismus, Finnlandisierung, schwäbeln*, vgl. Schweickard 1995) wurden wegen ihres unsystematischen Charakters und der meist unvorhersagbaren zusätzlichen Bedeutungskomponente nicht ins Lexikon aufgenommen. Als Einzelfälle fanden jedoch folgende unsystematischen Adjektivableitungen Aufnahme: *links- und rechtsrheinisch, innerdeutsch, atlantisch, ostelbisch, trans- und cisbaikalisch, transatlantisch, transsibirisch, subsaharisch*. In Phrasen übernehmen sie die gleichen Funktionen wie Beziehungsadjekte. Sofern sie in einem Korpus belegt waren, wurden auch adverbiale Bildungen mit dem Suffix *-seits* aufgenommen (*deutscherseits, sowjetischerseits, schweizerischerseits/schweizerseits*).

Gebrauch der Derivationen

Die Adjektivableitungen weisen im Wesentlichen fünf Verwendungen auf (vgl. Sugarewa 1974, Fleischer & Barz 1992: 239, Schilder, Versley & Habel 2004):

1. Beziehungsadjektive (*die französische Regierung, die deutsche Nordseeküste*) ersetzen syntaktisch das ursprüngliche Toponym und sind damit wichtig für die Paraphrase: *die Regierung Frankreichs = die französische Regierung*, siehe Kap. 3.5.1.5
2. Nationalität, Geburts- oder Wohnort: *der Münchner Autor, seine englischen Verwandten*
3. Ziel- oder Herkunftsort: *indischer Seeweg* (= ‚Seeweg nach Indien‘), *Glatzer Neiße* (= ‚die von Glatz kommende Neiße‘)
4. als Teil eines Firmennamens, den (ursprünglichen) Firmensitz oder -mittelpunkt bezeichnend: *Dresdner Bank, Hallesche Krankenversicherung*
5. an gg. Ort gesprochene Sprache/Dialekt: *bayrisch, berlinerisch*. Diese Funktion übernehmen nur flektierende Adjektive.

Meist übernimmt ein Adjektiv mehrere Funktionen (*die französische Küste, der französische Schriftsteller, er spricht französisch*, etc.), wobei oft nicht klar zwischen den Funktionen unterschieden werden kann: Lebt ein *französischer Schriftsteller* in Frankreich oder schreibt

er französisch? Mitunter liegt jedoch für die beiden letzten Funktionen ein eigenes Adjektiv vor. Während *Berliner* nur für die drei ersten Funktionen verwendet wird (*Berliner Bürgermeister*, *Berliner Autobahn*), ist *berlinerisch* nur auf die Sprache bezogen, ebenso *berlinisch*, das sich aber auch in Namen von Firmen und Institutionen findet (*Berlinische Galerie/Lebensversicherung/Monatsschrift*). Ins Lexikon wurden alle Adjektive aufgenommen. Da es ausschließlich zur Analyse und zur Korpusarbeit verwendet wird, stellt das kein Problem dar. Für die Synthese von Sprache müsste selbstverständlich zwischen den Funktionen unterschieden werden, falls diese durch verschiedenen Adjektive repräsentiert werden.

Die Ethnika sind immer mit *Bewohner von X* paraphrasierbar: *Münchner* = *Bewohner Münchens*, *Franzose* = *Bewohner Frankreichs*. Es kann allerdings zu Überschneidungen mit den Bezeichnungen von Ethnien oder Trägern einer Sprache kommen (z.B. *Bengale*).¹¹⁴ Eher selten, aber im Einzelfall ärgerlich für den Lexikographen sind die Fälle, in denen die Bezeichnung für ein Lebensmittel (*Champagner*, *Bolognese*) oder eine Hunderasse (*Weimeraner*, *Dalmatiner*) etc. eine mögliche regelmäßige Derivation überlagern oder „blockieren“. Als Folge ist dann keine Derivation möglich – ein Bewohner der Champagne muss *Bewohner der Champagne* genannt werden –, oder ein alternatives Bildungsmodell (*Weimarerer*) springt ein. Für viele Toponyme sind Derivationen schließlich aus morphologischen Gründen nicht möglich oder ungebräuchlich. Beispiele sind *Montana*, *Berkley*, *Colorado*, *Massachusetts*, *Nancy*, *Ohio*, der *Taunus*. Bildungen wie *Massachusetter* sind zwar denkbar, waren aber in den benutzten Korpora nicht oder nur okkasionell belegt. Andere wie *Taunese*, *Tauneraner* sind inakzeptabel, es bleibt nur die Umschreibung durch Nominalkomposita (*Taunusbewohner*), ev. mit Bindestrich (*Taunus-Bewohner*), oder als Phrase (*Bewohner des Taunus*, *Bewohner/ Einwohner von Massachusetts*). Die bislang bearbeiteten Toponyme (s.u.) vermitteln den Eindruck, dass für alle deutschen Dörfer und Städte auch Derivationen (Adjektiv und Bewohner) existieren, aber bei weitem nicht für alle ausländischen Städte.¹¹⁵ Für Ableitungen von Regionen und anderen Toponymgruppen lassen sich aber kaum Vorhersagen über die Existenz der Deonomastika treffen.

Abgrenzung Adjektiv vs. Bewohnername

Z.T. unterscheiden sich die Formen von Adjektiven und Bewohnernamen (*französisch* vs. *Franzose/Französin*). Aber auch für die Bildungen auf *-er* (*Berliner*) ist entgegen anderen Behauptungen (Fuhrhop 2001, 2003) eine Abgrenzung zwischen Adjektiv und Bewohnerbezeichnung möglich. Auf den ersten Blick mag in einigen Phrasen eine Unterscheidung schwerfallen:

(3.72) *der Berliner Bürgermeister*

(3.73) *der Berliner Klaus Wowereit*

(3.74) *der Potsdamer Tobias* (Fuhrhop 2003)

Aber während das Adjektiv indeklinabel ist, verfügt die Bewohnerbezeichnung über eine Form im Genitiv Singular und Dativ Plural, die sich vom Rest des Paradigmas unterscheidet. Setzt

114 Einen Sonderfall nimmt *Russländer* als Bezeichnung für die Bewohner Russlands in Abgrenzung zu *Russe*, dem Mitglied der Ethnie ein. Als Lehnübersetzung von *rossijskij* vs. *russkij* ist es bislang nur schwach im Deutschen etabliert.

115 Vgl. die Situation im Französischen, dargestellt bei Grass (2000: 659): *Hannover* → *Hannoveraner*, aber *Hanovre* → *habitant de Hanovre*. Auch Mel'čuk (2006: 392) berichtet, dass nicht für alle russischen Städtenamen Bewohnernamen existieren.

man also die Phrasen in den Gen. Sg. oder den Dativ Plural ist sofort erkennbar, ob es sich um Adjektiv oder Wohnernamen handelt:

(3.72') *des Berliner_ Bürgermeisters_*

(3.73') *des Berliners_ Klaus Wowereit*

(3.72'') *... erläuterte Pflüger, denn er sehe sich in der Tradition mit zwei großen Berliner_ Bürgermeistern_, mit Richard von Weizsäcker und Eberhard Diepgen, ...*

(3.73'') *den Berlinern_ Klaus Wowereit und Eberhard Diepgen*

(3.74') *zusammen mit den Potsdamern_ Tobias und Peter*

(3.74'') *zusammen mit den Potsdamer Tobiassen*

Allein in Fuhrhops Beispiel (3.74) ist es möglich, sofern man eine Form *Tobiassen* (Dativ Plural) akzeptiert, *Potsdamer* als Adjektiv (3.74'') wie auch Wohnernamen (3.74') zu interpretieren. Die Semantik bleibt die gleiche, beide verweisen ja auf die Stadt Potsdam (,aus Potsdam; in Potsdam ansässig'). Lediglich in der pragmatischen Dimension ergeben sich Unterschiede hinsichtlich der Intension, ob dem Hörer weitere *Tobiasse* aus Potsdam bekannt sind.

Die eben beschriebenen Phänomene spiegeln die Abhängigkeitsverhältnisse in der Nominalphrase: das Adjektiv wird vom substantivischen Kopf der Phrase regiert, der Wohnernamen ist selbst Kopf der Phrase und kann einen Personennamen in Apposition regieren. Als Kopf der Phrase ist er „morphological contact point“ (Mel'čuk 2003: 200–201) und trägt im Genitiv Singular und Dativ Plural die Kasusmarkierung.

Bildung der Derivationen

Die Bildung der Derivationen zu deutschen Toponymen ist hochkomplex.¹¹⁶ Bildungsregeln mit breiter oder gar unbegrenzter Gültigkeit lassen sich nicht formulieren. Selbst für die größte und wichtigste Gruppe der Toponyme, die deutschen Städte nennen Fleischer & Barz (1992: 155) nur eine gute Faustregel (aber auch nicht mehr):¹¹⁷ „Bei deutschen Städtenamen auf *-en*, sofern mehr als zweisilbig, entfällt in der Regel das *-en* (abgesehen von denen auf *-hafen* und *-kirchen*): *Donaueschinger*, *Burghauser*, *Frankenhäuser*, *Solinger*; aber *Euskirchener*, *Friedrichshafener*. Bei der Mehrheit der zweisilbigen bleibt das *-en* erhalten: *Dresdener*, *Gießener*, aber *Bremer*.“ Ausnahmen zu dieser Regel existieren (*Leverkusen* → *Leverkusener*), und der Umlaut in *Frankenhausen* → *Frankenhäuser* oder *Osterode* → *Osteröder* bleibt unerwähnt.

Die Behauptung von Fuhrhop (1998), die Ableitungen von Ländern und Regionen ließen sich regulär beschreiben, ist nicht nachvollziehbar. Zwar lassen sich auch hier Regularitäten bei der Verteilung der Suffixe, die sie als „Allomorphe“¹¹⁸ bezeichnet, feststellen:

116 Die Situation in anderen europäischen Sprachen unterscheidet sich nicht vom Deutschen, vgl. für das Englische den Artikel „Demonyim“ und die Liste der „adjectival forms of place names“ in EN.WIKIPEDIA.ORG, für das Französische Eggert, Maurel & Belleil (1998), Maurel, Piton & Eggert (2000), Eggert (2005), für das Russische Mel'čuk (2006: 392,450) und Arapov (1972)

117 Ebenso DUDEN2006: §1109,§470,§1153. Für ernstzunehmende Versuche einen Regelapparat zu entwickeln, siehe Eggert (2005) oder Arapov (1972).

118 Vgl. kritisch zu den Begriffen ‚Allomorph‘ und ‚Suppletion‘ bei französischen und russischen Wohnernamenbezeichnungen: Mel'čuk (2006: 392–4,442,449–50)

[. . .] Faktoren sind die Länderendung, der Auslaut der Segmentform und die syllabische Struktur der Segmentform. [. . .]

[. . .] wenn eine Endung von der Länderbezeichnung abgespalten wird:

- land → Schwa
- istan → Schwa
- ei → Schwa
- en → Schwa, wenn letzte Segmentformsilbe betont
→ -er in den übrigen Fällen
- a → -es-e, wenn die Segmentform einsilbig ist und auf Nasal oder Liquid auslautet
→ -er, wenn der Akzent auf der letzten Segmentformsilbe liegt
→ -an-er in den übrigen Fällen
- ien → -ier/-er im Wandel

Wird nichts abgespalten, entstehen Bewohnerbezeichnungen durch Anhängen von -er, bei -land und -mark mit Umlaut. (Fuhrhop 1998: 178–9)

Offen bleibt, wie die „Segmentform“ oder „Derivationsstammform“ regulär bestimmt werden kann, es muss bekannt sein, ob und was abgespalten wird, z.B. eben nicht *England* *→ *Engle* in Analogie zu *Irland* → *Ire*, oder *Pakistan* *→ *Pake* zu *Afghanistan* → *Afghane*. Eine Reihe von Problemfällen erwähnt Fuhrhop (1998: 180) selbst und spricht dabei von „Suppletion“¹¹⁸: *Engländer* (**Engle*) ↔ *englisch* (?*engländisch*), *Pole* (**Polne*) ↔ *polnisch* (**polisch*). Fraglich ist auch, wieso dann bei der von Fuhrhop (1998: 183–4, 239–41) durchgeführten Befragung so viele Sprecher des Deutschen von den angeblichen Regeln abgewichen sind und die Toponymderivationen anders gebildet haben. Gerade die Tatsache, dass so enorme Unsicherheiten bestehen, spricht dafür, die Bildung von Adjektivableitungen und Wohnernamen als weitgehend lexikalisiert zu betrachten (so Russ 1991: 22). Auch wenn gewisse Regularitäten existieren und sich einige Generalisierungen finden lassen, gibt es keine definitiven Regeln, die ausreichend wären, die Distribution der Suffixe allein aufgrund phonologischer und morphologischer Kriterien zu beschreiben (vgl. Mel’čuk 2006: 450).

Dafür sprechen auch (1.) Dubletten aus konkurrierenden Bildungsmodellen z.B. *Jenaer* vs. *Jenenser*, auch mit mehr als zwei konkurrierenden Modellen: *Rügener* – *Rüganer* – *Rugianer*. Und (2.) homonyme Siedlungsnamen, die auf verschiedene Weise Derivationen bilden. So heißen die Bewohner von *Mühlhausen*, wenn es sich um die Stadt in Thüringen handelt *Mühlhäuser*, aber *Mühlhauser* im Falle des gleichnamigen Ortsteils von Unna.¹¹⁹

119 Vgl. die systematische Aufstellung von Homonymen bei Toponymen und Wohnernamen für das Französische bei Eggert (2005: 122):

(3.75)	<i>Jena</i>	$\left\{ \begin{array}{l} \textit{Jenaer} \\ \textit{Jenenser} \end{array} \right.$	<i>Nepal</i>	$\left\{ \begin{array}{l} \textit{Nepalese} \\ \textit{Nepali} \end{array} \right.$
	<i>Taiwan</i>	$\left\{ \begin{array}{l} \textit{Taiwaner} \\ \textit{Taiwanese} \end{array} \right.$	<i>Zypern</i>	$\left\{ \begin{array}{l} \textit{Zypriote} \\ \textit{Zyprer} \end{array} \right.$
	<i>Fehmarn</i>	$\left\{ \begin{array}{l} \textit{Fehmarn} \\ \textit{Fehmaraner} \\ \textit{Fehmeraner} \end{array} \right.$	<i>Eggenfelden</i>	$\left\{ \begin{array}{l} \textit{Eggenfeldener} \\ \textit{Eggenfelder} \\ \textit{Eggenfeldner} \end{array} \right.$
	<i>Madrid</i>	$\left\{ \begin{array}{l} \textit{Madrider} \\ \textit{Madrilene} \\ \textit{Madrileño} \end{array} \right.$	<i>Rügen</i>	$\left\{ \begin{array}{l} \textit{Rügener} \\ \textit{Rüganer} \\ \textit{Rugianer} \end{array} \right.$
	<i>Mühlhausen</i>	$\left\{ \begin{array}{l} \text{(Thüringen)} \textit{ Mühlhäuser} \\ \text{(Unna)} \textit{ Mühlhauser} \end{array} \right.$		

Kodierung der Derivationen im Lexikon

Kodiert werden die Derivationen anhand von Suffixoperationen direkt im Eintrag des Toponyms, von dem die Ableitung erfolgt. Die Beschreibung als „Derivationsparadigma“ ist einfacher zu bearbeiten, übersichtlicher und kompakter. Das Format des Basis-Lexikons bleibt damit weiter zeilenbasiert. Erst bei der Expansion zum Vollformenlexikon werden auch die Derivationen gebildet. Im Vollformenlexikon sind die Derivationen dann als eigenständige Lemmata enthalten. Beispiele sind zu finden in Kap. A.1.2.

Für das Deutsche¹²⁰ existiert bislang kein auch nur einigermaßen vollständiges Wörterbuch, das zu jedem Ort Bewohnernamen und Adjektivableitungen enthält. Um eine hinlänglich große Zahl von Derivationen zu erhalten, musste also ein anderer Weg gefunden werden. Anhand von Material in Artikeln zum Thema (Fuhrhop 1998:233–8, Sugarewa 1974, Russ

I.	un toponyme	⇒ un gentilé (rapport 1:1)
	<i>Paris</i>	<i>Parisien</i>
II.	un toponyme	⇒ deux (ou plus) gentils (= doublets)
	<i>Dinan</i>	<i>Dinannais</i> <i>Dinandois</i>
III.	deux toponyms paronyms	⇒ deux gentils différents
	<i>Orvault</i> (Loire-Atlantique)	<i>Orvaltais</i>
	<i>Orveau</i> (Esonne)	<i>Orvallois</i>
IV.	deux toponyms paronyms	⇒ deux gentils homonymiques, une forme
	<i>Bagneux</i>	<i>Bagnolais</i>
	<i>Bagnols</i>	
	ou plus directement	
	<i>Germaine</i>	
	<i>Germigny-sur-Loire</i>	<i>Germinois</i>
	<i>Saint-Germain-des-Prés</i>	
V.	deux toponyms homonyms	⇒ deux gentils homonymiques (différents ou homonymes)
	<i>Sarry</i> (Saône-et-Loire)	<i>Sarrien</i>
	<i>Sarry</i> (Marne)	<i>Sarrysien</i>
	<i>Lunas</i> (Dordogne)	<i>Lunassien</i>
	<i>Lunas</i> (Dordogne)	<i>Lunassien</i>

120 Anders z.B. für das Französische mit Prolintex, Prolex, Maurel et al. (2006), Belleil (1997), Belleil & Maurel (1997), Maurel et al. (1997), Maurel, Piton & Eggert (2000) und für das Russische mit Babkin & Levašov eds. (1975), Levašov (2000), Gorodeckaja & Levašov (2003).

1991: 22–3), Listen wie StAGN 2004 und Beispielen in Arbeiten zur deutschen Grammatik und Wortbildung (Fleischer & Barz 1992: 154–5, 238–40, Wellmann 1975: 398–402/§2:6.2.19–22, Simmler 1998: 533–4/§246.21, 564/§256.7, Berger & Drosdowski 1985: 216–8) wurden zu Beginn etwa 3 500 Derivationen gesammelt. Diese dienten als Input für einen iterativen Bootstrapping-Prozess, bei dem aus den bereits beschriebenen Derivationen für weitere Toponyme mögliche Derivationen in einem statistischen Verfahren geraten werden. Nach einem manuellen Korrekturschritt fließen diese wieder in den Prozess ein. Parameter waren (1.) die Kompabilität eines „Derivationsparadigma“ (Menge aller Formen bzw. Suffixoperationen) mit dem Toponym (morphologische Beschaffenheit, Fähigkeit zur Umlautbildung etc.). (2.) die Länge des übereinstimmenden Suffixes mit anderen bereits verifizierten Toponymen des gleichen „Derivationsparadigmas“. Damit wird insbesondere ausgeschlossen, dass seltene und spezifische Modelle auf alle Toponyme angewandt werden. und (3.) wie gut das Paradigma in einem großen Korpus belegt ist. So wird z.B. für *Zentralafrika* das Bildungsmodell „hänge -er an“ ausgeschlossen, weil die resultierenden Formen (**Zentralafrikaer*, **Zentralafrikaers*, **Zentralafrikaern*, **Zentralafrikaerin*, **Zentralafrikaerinnen*) nicht oder mit sehr geringer Frequenz im Korpus belegt sind. Das korrekte „Derivationsparadigma“ (*Zentralafrikaner*, *Zentralafrikanerin*, *zentralafrikanisch*, ...) ist dagegen gut belegt und erhält einen hohen Score. Ein hoher (oder der höchste) Score ist als Empfehlung zu werten, nicht als definitive Entscheidung. Oft verursachen häufige Wörter, die nicht mit dem Toponym in Verbindung stehen, falsche Empfehlungen. Bildet man zu dem Ort *Büchen* Adjektiv und Wohnernamen analog zu *Bremen* → *Bremer*, erhält man *Bücher*, ohne Zweifel ein häufiges Wort, aber nicht die korrekte Einwohnerbezeichnung, die *Büchener* lautet. Ebenso verhält es sich mit *England* *→ *Enge*, analog zu *Irland* → *Ire*. Eine manuelle Korrektur der automatisch gewonnenen Vorschläge ist wegen solcher Interferenzen notwendig.

Die Anzahl der im Lexikon verzeichneten Derivationen konnte so nach 12 Durchläufen des Bootstrapping-Prozesses und etwa 40 Stunden Korrekturarbeit auf über 13 000 (für 3 600 Ortsnamen) erhöht werden. Die Anzahl der Derivationsuffixe bzw. -operationen liegt derzeit bei fast 500, eine vollständige Liste findet sich in Kap. A.1.4. Das weitaus am häufigsten vorkommende Suffix ist -er, ev. kombiniert mit einer Trunkierung des Ortsnamens. Mehr als die Hälfte der Suffixe sind jedoch Unikate für deutsche (*Baden* → *Badenser*, *Fehmarn* → *Fehmeraner*) wie für ausländische Toponyme (*Teneriffa* → *Timerfeño*, *Buenos Aires* → *Porteño*).¹²¹ Die Anzahl der „Derivationsparadigmen“, also der Kombinationen aus Suffixoperationen zur Bildung der drei Derivationstypen, liegt bei 400.

3.3.2.12 Fugenformen

Fugenformen von Toponymen sind notwendig zur Beschreibung von Komposita, vorwiegend Nominalkomposita (*Rheintal*, *Donaudampfschiffahrt*), aber auch einiger Adjektive (*europa-weit*, *bayernweit*, *pekinghörig*). Für die meisten Ein-Wort-Toponyme ist die Fugenform mit der Grundform identisch: *Bayerwaldrundfahrt*, *Mittelmeerinsel*, *Nordseeküste*. Das gilt nicht für einige Mehrwortlexeme, z.B. lautet sie für *Schwarzes Meer* *schwarzmeer* in *Schwarzmeerflotte*, vgl. *Roter/Weißer Main* → *Rotmaintal*, *Weißmainquelle*, *Weißmainradweg*. Teilweise ist es möglich, die Fugenformen eines Mehrwortlexems von einer Variante des Toponyms

¹²¹ Zugegeben, der Begriff ‚Derivationsuffix‘ ist für Derivationen wie *Porteño* oder auch *Bocas del Toro* → *Bocatoreño*, *São Tomé und Príncipe* → *Santomeer* wenig angemessen. Aus operativer Sicht ist es jedoch sinnvoll alle Derivationen im Lexikon einheitlich zu behandeln.

abzuleiten. Für *Bayrischer Wald*, *Vorderes Asien* und andere existieren parallel die konkatenierten Formen *Bayerwald*, *Vorderasien*. Wegen der deutlich einfacheren Beschreibung wird die Fugenform wenn möglich von diesen abgeleitet. Meist existiert nur eine Fugenform, selten zwei oder mehr, die dann in jeweils anderen Kontexten stehen: *Brasilienreise*, *Brasilkaffee*. Die Bildung der Fugenform wird gemeinsam mit der Derivation (Kap. 3.3.2.11) behandelt.

- (3.76) a. *Alte Welt*, .EN+Topon+Choron+Det:FADJ+NS0/<<6welt.FF>>
 b. *Rote Main*, *Roter Main*.EN+Topon+Hydron+Cl(Fluss)+Det:MADJ+NS13/<<6main.FF>>

Für Ein-Wort-Hydronyme wird implizit immer eine Fugenform angenommen, die mit der Grundform identisch ist. Sie sind auch die Toponymgruppe, die am systematischsten in Komposita vertreten sind: *Donaudelta*, *Spreeufer*, *Rhôneetal*, *Zürichseeufer*, *Innbrücke*, *in Ostseenähe*. Es folgen die Bergnamen z.B. mit *Watzmannbesteigung*, *Watzmannüberquerung*, *Watzmannostwand*, *Watzmannblick*, *Zugspitzgipfel*, *Zugspitzmassiv*, *Zugspitztour*, *Zugspitzkamm*. Regionen werden weniger systematisch in Komposita verwendet (*Krimtataren*, *Krimkrieg*, *Krimsekt*, *Sibirtiger*, *in Alpennähe*), Siedlungsnamen kaum und sehr spezifisch (*Münchenpremiere*, *Münchenezuschlag*, *Berlinförderung*). Auch werden diese überwiegend mit Bindestrich geschrieben (*Helsinki-Prozess*), vgl. Fleischer & Barz (1992: 132).

Adjektivderivationen wurden entgegen Fuhrhop (2003) mit den Beispielen *Schweizerkäse* und *Altstetterplatz* nicht als Fugenformen bzw. Kompositionserstglieder angenommen. Im ersten Fall zeigt ein Blick in ein Textkorpus, dass die Form *Schweizer Käse* die deutlich häufigere ist. Für den zweiten Fall, das Hodonym *Altstetterplatz*, überwiegt zwar die zusammengesetzte Form, aber hier handelt es sich um eine konventionalisierte lexikalisierte Form, und auch weniger um ein Adjektiv als um einen Bestandteil eines neuen Toponyms.

Ambiguitäten bei der Kompositasegmentierung lassen sich hier wie auch allgemein nicht vermeiden. So stehen *Regenwald* und *Regenwurm* nicht in Bezug zum Fluss *Regen*. Die Beschreibung muss also sehr spezifisch erfolgen, indem z.B. für Flüsse nur bestimmte Zweitglieder wie *-tal*, *-brücke*, *-ufer* usw. zugelassen werden und auch die resultierende Konstruktion ist zu berücksichtigen (vgl. *in Donaunähe*, Kap. 3.5.3).

Die Kompositasegmentierung ist dabei nicht Selbstzweck. Vielmehr können eine Reihe von (quasi-)prädikativen Nomina (siehe Kap. 2.6 und 4.6) einen ihrer Aktanten als Kompositabestandteil anschließen. *Die Mittelmeerinsel Kreta* lässt sich mit *Kreta ist eine Insel im Mittelmeer* paraphrasieren, *Innbrücke* mit *Brücke über den Inn* oder *Brücke, die über den Inn führt*, die *Berlin-Reise* eine *Reise nach Berlin* usw. (vgl. Kap. 3.5.1.5).¹²² Konsequenterweise wird in der Lexikongrammatik (Kap. 4.6) für solche Nomina der Anschluss über die Fugenform als Realisierungsmöglichkeit eines Aktanten behandelt.

Z.T. entstehen durch die Komposition wiederum geographische Eigennamen, meist mit anderer Extension und auch anderer Klassenzugehörigkeit (*Inntal*, *Mittelmeerraum*, *Zürichbiet*).

122 Das schließt nicht ein, dass Kompositum und Phrase immer austauschbar sind, z.B. bei der Referenzierung/Pronominalisierung auf Kompositaerstglieder (vgl. Mel'čuk 1997-2001: I,197):

- (1) a. *die Brücke über die Donau, die mit über 2 800 km der längste Fluss Europas ist*
 b. * *die Donaubrücke, die mit über 2 800 km der längste Fluss Europas ist*

Insbesondere bei Gewässernamen sind aus Namen und Klassenbezeichner gebildete Komposita üblich: *Njassa-Fluss*, *Njassafluss*, *Shire-Fluss*, *Shire-River* (vgl. Ortner et al. 1991: 189–192).

Die Bildung neuer, komplexerer Toponyme ist allerdings nicht nur auf den morphologischen Prozess der Komposition beschränkt. Parallel können Phrasenbildungen bestehen, z.B. *Mittleres Rheintal* (= *Mittelrheintal*) Ausgehend von der Kompositumsform sind wiederum weitere Phrasenbildungen möglich (*Oberes Mittelrheintal*).¹²³ Das letzte Beispiel zeigt auch ein weiteres produktives Mittel, wie durch Kompositabildung neue Toponyme entstehen können, nämlich die Präfigierung mit „Modifikatoren“ wie *Ober-*, *Unter-*, *Zentral-*, *Mittel-*; *Nord-*, *Ost-*, *Nordwest-*.¹²⁴ *Südengland*, *Mittelschweden*. Die Grenze zur Lexikalisierung mit Toponymen wie *Oberbayern* oder *Mittelfranken*) ist auch hier fließend.

3.4 Erweiterung des Lexikons zur Ontologie

Eine Ontologie¹²⁵ enthält in einer mehr oder minder formalen Repräsentation Informationen zu geographischen Entitäten, insbesondere deren Namen, gruppiert diese in Klassen (SIEDLUNG, REGION, GEWÄSSER, usw.) und beschreibt Relationen zwischen geographischen Entitäten (z.B. *liegt in*, *ist Teil von*, *fließt in*). Der letzte Punkt unterscheidet eine Ontologie von einem Lexikon oder Thesaurus.

Die zentrale Überlegung (vgl. Tran, Grass & Maurel 2004, Maurel et al. 2006) beim Aufbau der Ontologie ist die Trennung in zwei Ebenen, die der geographischer Entitäten und die der Namen (Toponyme). Die geographischen Entitäten repräsentieren in der Ontologie reale Objekte (vgl. die Definition auf S. 93). Eine geographische Entität kann mehrere Namen („Allonyme“) tragen. Diese sind als Relation zwischen der Entität und einem Toponym (Lemma) realisiert. Zusammen mit möglichen Relationen zwischen Lemmata (z.B. Derivation) und den Relationen zwischen den geographischen Entitäten ergibt sich ein Netz mit geographischen Entitäten und deren Namen als Knoten und verschiedenen getypten Relationen als Kanten.

Zur „Ontologisierung“ des text- und zeilenbasierten Lexikons, d.h. der Konversion in eine Ontologie, wird aus jedem Eintrag eine sprachliche und eine geographische Entität. Dazu sind eine Reihe von Annotationen vorgesehen, um Allonymie und die Verlinkung geographischer Entitäten handhaben zu können, siehe Kap. A.1.2.1 und die Erläuterungen zum Lexikon auf S. 301. Auf eine räumliche Repräsentation der geographischen Entitäten wurde verzichtet (dies ist Aufgabe eines Geographen).

3.4.1 Die sprachliche Ebene

Alle Knoten auf der sprachlichen Ebenen der Ontologie sind Lexeme in der üblichen Definition: eine Menge (Paradigma) von Wortformen, die sich nur in flektivischen Bedeutungsbestandteilen unterscheiden und durch eine Grundform (siehe Kap. 3.3.2.1) und die Angabe

123 Vgl. Kap. 5.1.5.3.1.1 zur Erkennung unbekannter, nicht im Lexikon enthaltener Toponyme und S. 98 zur Koppelung von Ortsnamen.

124 Siehe auch Fußnote 113 für weitere Beispiele und S. 75 für vergleichbare Konstruktionen im Französischen.

125 Vgl. Fonseca (2001), Fonseca et al. (2002), Fonseca, Davis & Câmara (2003) und Fu, Abdelmoty & Jones (2003: 6) für eine Diskussion des Konzepts einer geographischen Ontologie. Für praktische Anwendungen einer Ontologie im Information Retrieval, siehe Kap. 5.4.3 und Kap. 5.4.1.2.

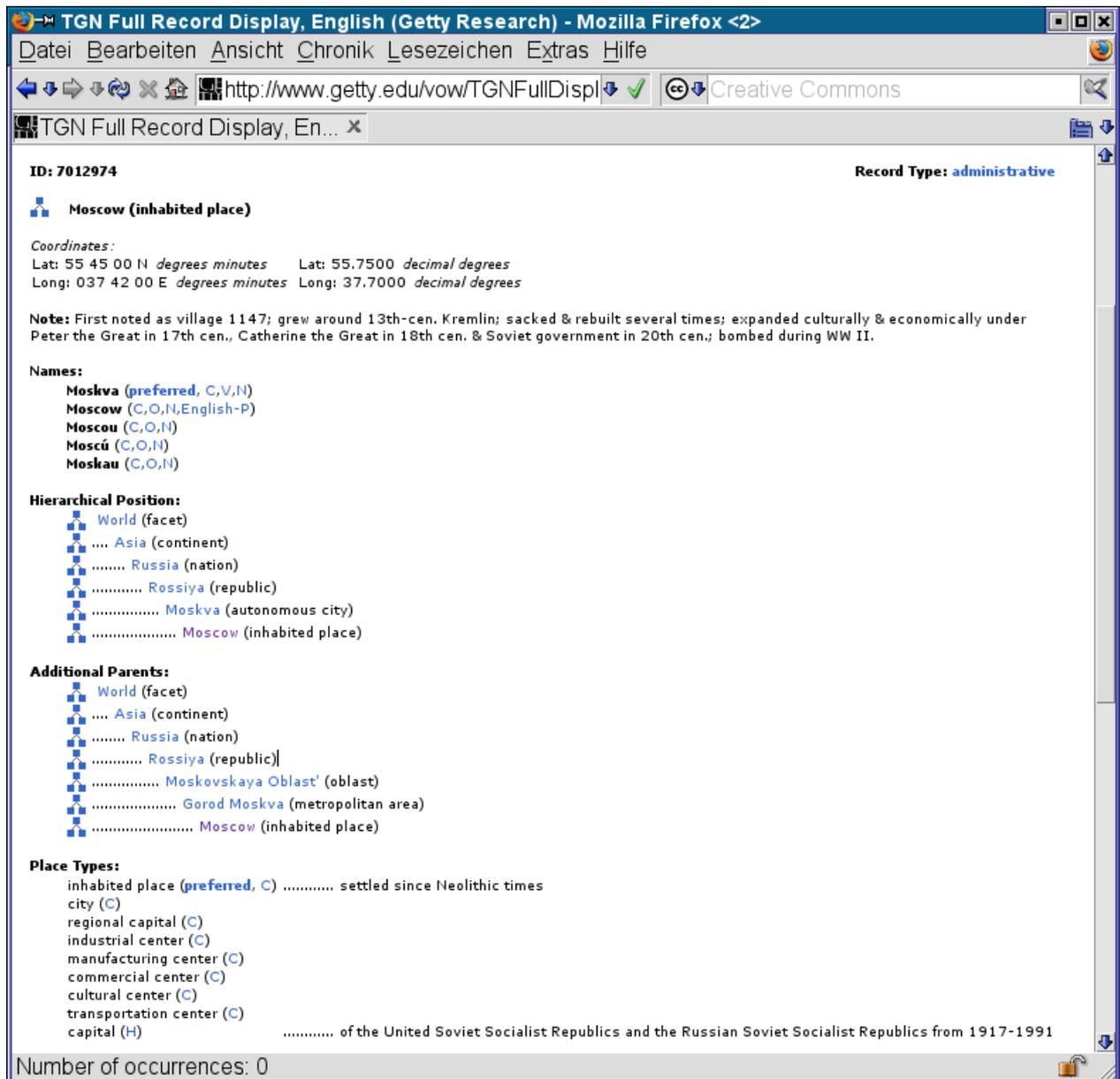


Abbildung 3.2: Die Stadt Moskau im GETTY-Thesaurus

der Flexionsklasse (Kap. 3.3.2.6) repräsentiert sind. Jedes Lexem kann durch eine Reihe von klassifikatorischen Eigenschaften (Genus, Artikelgebrauch, usw.) charakterisiert sein. Diese sind als Key-Value-Paare (Slots, vgl. Abb. 3.3) repräsentiert.

Die Lexeme sind in Wortarten unterteilt, die ihrerseits zur besseren Organisation weiter untergliedert sein können. Die Untergliederung ermöglicht, dass bestimmte Slots per default gefüllt werden können, wie z.B. im Fall des Artikelgebrauchs bei den Toponymen, der für die Untergruppe der Hydronyme obligatorisch zu sein scheint – bei ca. 2000 untersuchten Namen wurde kein einziger mit dem Merkmal +DetZ beobachtet (vgl. Tab. 3.4).

Die Verlinkung zwischen Lemmata – namentlich zwischen Toponym und Derivation, d.h. Adjektiven und Bewohnernamen – erfolgt ebenfalls durch Slots: jedes Nomen hat einen Slot für das abgeleitete Adjektiv, jedes Toponym für dessen Bewohner. Ein weiterer Typ Slots, ist nötig um die historische Dimension der Namen zu kodieren, siehe Kap. 3.4.3.

3.4.1.1 Klassen geographischer Eigennamen

Eine Klassifikation der Toponyme in semantische Klassen wie Siedlungs-, Berg- und Gewässernamen könnte im Prinzip entfallen, da die geographischen Entitäten als die Referenten ohnehin klassifiziert sind (s.u., Kap. 3.4.2.1). Dennoch erscheint auch eine (Grob-)Klassifikation der Namen sinnvoll. Zum einen ist sie in der Onomastik seit jeher üblich, zum anderen besitzt sie z.T. linguistische Relevanz und empirische Evidenz, z.B. bei der Vorhersagbarkeit des Artikelgebrauchs, wo semantische und syntaktische Klassifikation in weiten Teilen zusammenfallen. So werden auf der Datenbasis des hier erarbeiteten Lexikons alle (beobachteten) Gewässernamen mit, die weitaus meisten Siedlungsnamen ohne Artikel verwendet (Kap. 3.3.2.4). Syntaktische Eigenschaften wie Artikelgebrauch und Genus sind Eigenschaften sprachlicher Einheiten – und nicht geographischer.¹²⁶ Trotzdem ist die Klassifikation von Toponymen nicht allein Aufgabe des Linguisten, sondern auch des Geographen, die „Übereinstimmung mit der linguistischen und geographischen Terminologie“ (OsnovenSistem1983: 53) sollte für jedes Klassifikationssystem maßgeblich sein. Die zur Klassifikation der im Lexikon kodierten Toponyme zugrundeliegenden Überlegungen werden deshalb auch in Kap. 3.4.2.1 diskutiert. Als Toponymklassen wurden angesetzt:

- Siedlungsnamen (Oikonyme, +Oikon)
- Raumnamen (Choronyme, +Choron)
- Bergnamen (Oronyme, +Oron)
- Gewässernamen (Hydronyme, +Hydron)
- Namen von Verkehrswegen (Hodonyme, +Hodon)
- Namen für Bauwerke (+Aedif)

¹²⁶ Es wäre absurd, Genus der geographischen Entität als Eigenschaft zuzuschreiben; und abgesehen davon auch unmöglich: So ist die Bezeichnung für die Stadt *Aachen* im Deutschen ein Neutrum, im Russischen (*Ахен*) ein Maskulinum, im Tschechischen (*Cáchy*) ein Plurale tantum (vgl. S. 101). Syntaktische Eigenschaften sind also Merkmale der Bezeichnungen und nicht der geographischen Objekte selbst.

Wenn möglich wurden die in der Onomastik üblichen aus dem Griechischen entlehnten Bezeichnungen auf *-nym* gewählt. Die Gesamtzahl der Toponymklassen ist überschaubar verglichen mit anderen onomastischen Klassifikationssystemen (z.B. Eichler 1971, OsnovenSistem1983, Room 1996, Kamianets 2000, Zelinsky 2002, Walther 2003, Brendler 2004). Eine feinere Untergliederung erfolgt erst auf Ebene der geographischen Entitäten. Für weitere Details und Toponymklassen der Onomastik siehe Tab. A.2 auf S. 301.

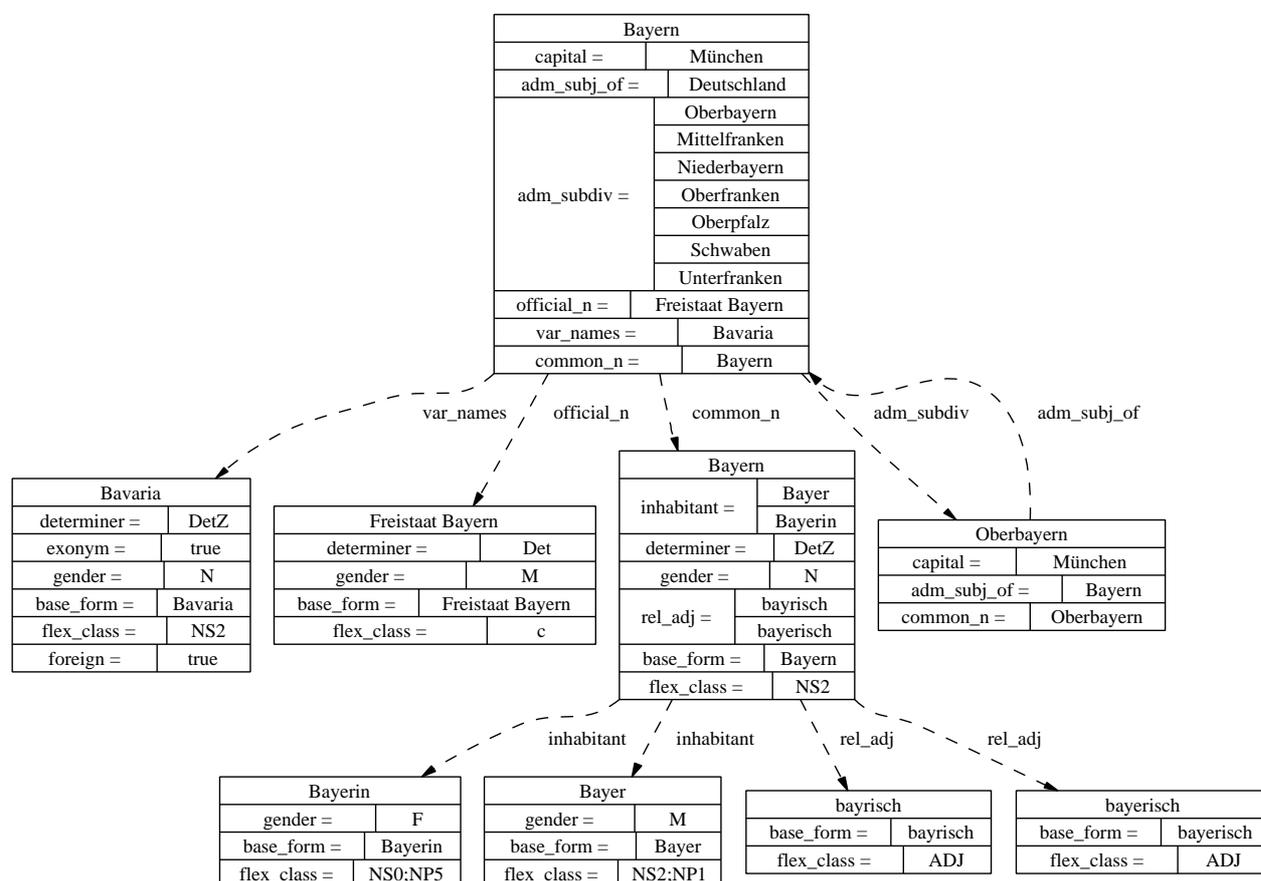


Abbildung 3.3: Die geographische Entität ‚Bayern‘ und andere geographische wie linguistische Entitäten. Jede Box repräsentiert eine Entität, die Zellen enthalten Namen und Werte der Slots. Pfeile stehen für Relationen zwischen den Entitäten.

3.4.2 Geographische Entitäten

Der Zugriff auf die geographischen Entitäten ist möglich über:

- den oder die Namen der Entität
- die Klasse, der die Entität angehört – die Klassen selbst sind hierarchisch organisiert
- Relationen (Links) zwischen den Entitäten selbst, insbesondere über die beiden Relationen ‘administrative Untergliederung’ und ‘räumlich enthalten in’: diese organisieren die Entitäten hierarchisch

3.4.2.1 Klassen geographischer Entitäten

“When you say ‘hill,’” the Queen interrupted, “I could show you hills, in comparison with which you’d call that a valley.”

“No, I shouldn’t,” said Alice, surprised into contradicting her at last: “a hill can’t be a valley, you know. That would be nonsense—” Lewis Carroll, *Through the looking glass*, chap. 2

Helen Kerfoot schreibt in Bezug auf die Entwicklung einer Taxonomie für einen kanadischen nationalen Gazetteer: „On studying existing national gazetteers, we found there to be no generally accepted or standardized approach to the representation of feature types from one country to another. Some national gazetteers use as few as 7 categories, others have many as 60 or more types of features identified, and most fall somewhere in between.“ (UNGEGN 1996) Darin hat sich bis heute nichts geändert. Es existiert keine a-priori-Taxonomie, die sich allgemeiner Akzeptanz erfreut und sich deshalb als Standard durchgesetzt hätte (vgl. Kap. 3.4.2.1.1 und A.1.5). Bereits bei der Wahl des Kriteriums zur Unterscheidung der Klassen in der obersten Ebene der Taxonomie existieren parallel eine Reihe von inkompatiblen Kriterien:

1. Art der räumlichen Repräsentation: Punkt, Linie, Fläche
2. Makro- vs. Mikrotoponyme (,große‘ vs. ,kleine‘ Orte), vgl. Walther (2003), Abb. A.3 auf S. 334
3. ,natürliche‘ vs. vom Menschen geschaffene, ,künstliche‘ Strukturen, z.B. bei Kamianets (2000), Abb. A.2 auf S. 334 oder ADL, S. 135 und Abb. A.17 auf S. 349
4. Beschaffenheit (der Oberfläche)
5. Funktion

Kriterium 1 ist gebunden an die Form der räumlichen Repräsentation im Gazetteer/Ontologie (vgl. die Diskussion ab S. 103). Die Kriterien 2 und 3 werden häufig zur Klassifikation herangezogen, sind aber willkürlich:

Wieder ist eine Untergliederung nach allgemein verbindlichen Ordnungskriterien nicht möglich: in der Art der Einteilung spiegelt sich die Weltansicht des Klassifizierenden. So lässt sich auch hier allenfalls eine Aufzählung möglicher Untergruppen geben, denn schon die erste grobe Aufspaltung in Makro- und Mikrotoponym (etwa Raum- und Punktnamen) erweist sich als relativ, sobald man die Subkategorien Gewässernamen (Hydronyme) [...], Gebirgsnamen (Oronyme) – wozu auch die Namen einzelner Berge gehören [...] und die Gemarkungs- oder Flurnamen (Mikrotoponyme im eigentlichen Sinne) [...] im System sinnvoll unterzubringen versucht. Auch die Gliederung in Namen für naturbedingte Sachverhalte (Naturnamen) und Namen für durch den Zugriff des Menschen zustandgekommene Gegebenheiten der Natur (Kulturnamen) ist letztlich willkürlich, weil vom Standpunkt einer bestimmten menschlichen Entwicklungsstufe aus getroffen. (Bauer 1998: 55–6)

Die verbleibenden beiden Kriterien Oberflächenbeschaffenheit und Funktion werden von allen Klassifikationssystemen zumindest auf untergeordneten Ebenen herangezogen. Aus der Perspektive einer Einteilung in Selektionsklassen von Prädikaten, ist zu erwarten, dass die Selektionsklassen teilweise mit Beschaffenheit und Funktion Hand in Hand gehen. Menschen „tun“ vergleichbare Dinge mit (funktional) gleichartigen Objekten. Allerdings erschwert auch hier die Polyfunktionalität einzelner Objekte eine Zuordnung. Oder Beschaffenheit und Funktion ermöglichen jeweils eine Zuordnung zu einer Klasse, wie z.B. Wasserwege die als Gewässer und Verkehrswege klassifiziert werden können. Einander überlappende Klassen sind die

Folge, und die Gefahr einer inkonsistenten und unlogischen Taxonomie ist groß (vgl. Langer 2003).

Ein weiterer kritischer Punkt sind Klassengrenzen innerhalb von Kontinua, z.B. den Fließgewässern (Bach, Fluss, Strom) oder Siedlungen (Dorf vs. Stadt). Für den Lexikographen bedeutet die Grenzziehung hier eine unsichere, willkürliche und potentiell fehlerträchtige Entscheidung, die auch mit längerem Abwägen und Absprachen mit Kollegen nicht sicher zu lösen ist.

Die Unmöglichkeit oder Schwierigkeit einer multilingualen Klassifikation wird oft erwähnt. Paradebeispiel aus der Domäne der geographischen Entitäten ist das Begriffspaar engl. *river* (‚Fluss‘) und *stream* (‚großer Fluss, Strom‘), das nicht mit franz. *rivière* (‚Fluss, der in einen anderen Fluss mündet‘) und *fleuve* (‚Fluss, der ins Meer mündet‘) in Übereinstimmung zu bringen ist:¹²⁷

Any concept can be refined into more specialized subtypes by making more detailed distinctions. Since different cultures may be sensitive to different features, their languages may have words that have to be translated into other languages either by rough approximations or by clumsy paraphrases. In English, for example, size is the feature that distinguishes *river* from *stream*; in French, a *fleuve* is a river that flows into the sea, and a *rivière* is either a river or a stream that flows into another river. (Sowa 1993: 246)

Damit bleibt es aber weiterhin möglich, gute und universelle Ontologien zu konstruieren.¹²⁸ Es ist ausreichend, Konzepte zu kategorisieren, die nicht unbedingt eine 1:1-Entsprechung in der Sprache haben müssen. Denn nicht jeder in einer Sprache vorhandene Klassenbezeichner (‘X ist ein Y’) muss eine Kategorie der Ontologie konstituieren.

Auch den Geographen ist bewusst, dass sie über keine einheitliche und allgemein akzeptierte und verwendbare Klassifikation geographischer Objekte verfügen. Versuche, die Klassen verschiedener Klassifikationssysteme und Ontologien mittels statistischer Verfahren aufeinander abzubilden (Rodríguez, Egenhofer & Rugg 1999, Rodríguez 2000, Rodríguez & Egenhofer 2003, 2004, Kavouras, Margarita & Tomai 2005, Tomai & Kavouras 2005) konstatieren die Existenz des Problems, liefern aber keine Lösung.

Die Erarbeitung einer Klassifikation deutscher Toponyme war angesichts der oben geschilderten Probleme eher ein Versuch, das beste aus bestehenden Klassifikationssystemen (siehe Kap. 3.4.2.1.1) zu übernehmen, um zu einer brauchbaren, pragmatischen, bei weitem aber nicht perfekten Lösung zu kommen. Die Zuordnung zu einer Klasse erfolgte mit der syntaktischen Beschreibung eines Toponyms bei der Aufnahme ins Lexikon (vgl. Kap. 3.3.1). Erwartungsgemäß sind nur die Klassen geographischer Entitäten vertreten, deren Instanzen auch durch Toponyme benannt sind. Verfolgt wurde schließlich ein synkretistischer Ansatz, der sich primär an der Funktion und Oberflächenbeschaffenheit der Objekte orientiert und mehrfache Superklassen von Klassen (nicht aber Instanzen) erlaubt. Die Taxonomie stellt also einen Graph, keinen Baum dar (siehe Abb. 3.6 und A.7).

127 Vgl. die Diskussion in Janssen (2002: 115–43). Weitere Beispiele sind bei Tichelaar (2003b: 80) zu finden – nicht nur aus dem Finnischen, Arabischen und anderen „exotischen“ Sprachen, sondern auch z.B. dt. *Kloster* vs. engl. *convent* ‚Nonnenkloster‘ vs. *monastery* ‚Kloster für Mönche‘.

128 Für theoretische Überlegungen zur Universalität einer geographischen Ontologie, siehe Mark (1989, 1999), Mark, Smith & Tversky (1999), Mark & Smith (2003), Fisher & Wood (1998) und BUFFALOONTOLOGY.

3.4.2.1.1 Überblick über Klassifikationssysteme Klassifikationssysteme geographischer Entitäten werden hier im Überblick dargestellt. Beispiele für die voneinander verschiedenen Taxonomien sind im Anhang (Kap. A.1.5) zu finden.

SDTS Der Spatial Data Transfer Standard definiert 200 „standard entity types“, verzichtet aber auf eine hierarchische Organisation der Klassen: „This standard specifically avoids adopting any particular hierarchy, in part because the higher level classifications differ from one organization to another. Hierarchical classification, if desired, may be constructed using standard attributes (such as air/land/water, or manmade/natural) or through user-defined non-standard attributes.“¹²⁹

DAML:GEOFILE enthält knapp 60 000 „Locations“ aus verschiedenen Teilen der Welt, die jeweils durch Namen und Koordinatentripel (geogr. Länge und Breite, Höhe über NN) repräsentiert sind. Jede Lokation ist einem „Geographic Area“ zugeordnet, d.h. einer nur durch ihren Namen spezifizierten Entität. Geographic Areas sind 100 größere Gewässer, 250 unabhängige Staaten und die 50 Bundesstaaten der USA. Die Lokationen verteilen sich auf 70 Klassen mit Schwerpunkt auf militärischen und Infrastruktureinrichtungen (siehe Abb. A.9 auf S. 341).

Die Klassifikation folgt pragmatischen Kriterien und berücksichtigt, wofür und wie ein Objekt genutzt wird. Die Taxonomie hat die Form eines Baumes. Das zwingt zu eindeutigen Zuordnungen, die mitunter nicht unmittelbar einsichtig sind. Z.B. ist nicht nachzuvollziehen wieso PORT zu SEAAREA und NAVALBASE zu MILITARYINSTALLATION gerechnet werden, wenn andererseits MILITARYAIRPORT und INTERNATIONALAIRPORT in die gleiche Klasse AIRLANDINGAREA fallen.

NASA:SWEET verknüpft Klassen geographischer Entitäten mit zahlreichen physischen Eigenschaften (Beschaffenheit des Untergrundes, Bodenbedeckung, Erdgeschichte). Das zugrundeliegende wissenschaftliche Weltbild (Geologie, Biologie, Ökologie) macht den Taxonomiebaum (Abb. A.11 auf S. 343) eher ungeeignet für eine linguistisch orientierte Klassifikation.

NGA:GNS stellt mit über 600 Klassen die vermutlich größte Taxonomie auf. Jede Klasse ist kurz beschrieben und im wohl größten Gazetteer (vgl. S. 107) durch zahlreiche Entitäten belegt. Allerdings ist die Taxonomie mit nur zwei Ebenen für diese Anzahl von Klassen zu flach. Die Wahl der neun übergeordneten Klassen erscheint nicht intuitiv:

- administrative boundary features
- hydrographic features
- area features
- populated place features
- road / railroad features
- spot features
- hypsographic features
- undersea features
- vegetation features

Auch die Zuordnung einer Klasse zu einer der übergeordneten Klassen ist nicht immer nachvollziehbar, z.B. findet sich unter der Kategorie ‘populated place features’ eine Klasse ‘religious populated place’ („a populated place whose population is largely engaged in religious occupations“). Die Klasse ‘monastery’ ist aber nicht etwa als Subklasse dieser definiert, sondern der Kategorie ‘spot features’ untergeordnet.

¹²⁹ Siehe <http://thor-f5.er.usgs.gov/sdts/standard/ps/part2.ps>

Langer (1996) stellte im Rahmen des CISLEX-Projektes eine allgemeine Klassifikation einfacher (nicht-präfigierter, nicht-komplexer) Nomina auf. Gut 40 000 Nomina wurden etwa 300 Klassen zugeordnet. Abb. A.5 auf S. 336 zeigt einen Ausschnitt aus der Taxonomie. Bezeichnungen für Orte und Lokationen nehmen nur einen kleinen Teil ein und verteilen sich auf mind. zwei Superklassen, für die übergeordnete Kriterien wie die Unterscheidung Lebewesen – Artefakt – andere höher gewichtet wurden als unmittelbar auf Eigenschaften von Orten bezogene Kriterien. Als klare Taxonomie und durch ihren großen Umfang lieferte sie jedoch gutes Material für das Lexikon der lokativen Nomina (siehe Kap. 3.1.6).

GEOFEATURES.OWL von MINDSWAP:GEOONTOLOGIES ist geprägt durch eine sehr flache Taxonomie, die alle topographischen Objekte auf einer Ebene vereint. Dies und Klassen wie „Pakistan-administered“ empfehlen die Taxonomie nicht zur Nachahmung, siehe Abb. A.12 auf S. 344.

CORINE ist ein Projekt der Europäischen Union mit dem Ziel einheitliche Daten zur Bodenbedeckung bereitzustellen. Die Taxonomie von CORINE liegt in mehreren Sprachen vor, ihrer Aufgabe entsprechend beinhaltet sie nur Klassen der Bodennutzung/-bedeckung. Diese Ausrichtung macht sie jedoch zu einer brauchbaren, wenn auch nicht vollständigen Taxonomie, siehe Abb. A.13 auf S. 345.

SEKINEEXTNER von Satoshi Sekine et al. (Sekine, Sudo & Nobata 2002, Sekine & Nobata 2004, SEKINEEXTNER, Sekine 2004) stellt einen Taxonomiebaum für das Japanische auf, der aber auch in englischer Übersetzung vorliegt. Die Anzahl der lokativen Klassen beträgt etwa 50 (von insgesamt 200 Klassen, vgl. Abb. A.14, A.15 und A.16). Die Lokativa verteilen sich dabei auf die beiden Oberklassen LOCATION und FACILITY.

Die Taxonomie ist funktional ausgerichtet, d.h. sie klassifiziert die Entitäten primär nach ihrer Funktion. Klassen, die regulär polysem sind (Ort und Organisation), werden zu den Klassen GPE (Geographical and Political Entity) und GOE (Geographical and Organizational Entity) zusammengefasst, die jeweils Unterklassen von LOCATION bzw. FACILITY sind. Der Aufbau der Taxonomie erfolgte nach Angaben von Sekine, Sudo & Nobata (2002) einerseits „bottom-up“, ausgehend von in einem Zeitungskorpus vorkommenden Named Entities, andererseits orientiert er sich an den Klassen existierender NER-Systeme und den beiden Thesauri WORDNET und Roget (1852).

Eine Konsequenz aus dem Bemühen, die Klassen von Named Entities zu erweitern ist die Aufweichung der Grenze zwischen Namen, d.h. Referenzen auf einzelne, unique Entitäten und Klassenbezeichnern, z.B. Produktnamen, die auf eine Klasse von Entitäten referieren. Damit werden aber auch gewöhnliche Nomina zu Entitäten. Das äußert sich vor allem in nicht-lokativen Klassen wie einer Klasse BODY_REGION mit *head, heart, rib, muscle, nerves, upper half of the body, gallstone*, oder SCHOOL_AGE (als Unterklasse von NUMEX) mit *first-grade pupil, freshman, sophomore, junior, senior*. Bei den Lokativa finden sich Ausdrücke wie *a subway line, flight route to Las Vegas from New York, heaven, hell, 8th floor, the South, John's kitchen*. Das letzte Beispiel (*John's kitchen*) verfügt zwar über eine eindeutige Referenz, ist aber (a) als Named Entity so gut wie gar nicht relevant, (b) nicht beständig genug, um wirklich informativ zu sein (Personen ziehen zu oft um), und schließlich (c) ist es vermutlich eine der wichtigsten Eigenschaften von Sprache, im Diskurs eindeutige Referenz zu Objekten herzustellen, aber deshalb von „benannten Objekten“ zu sprechen ist gewagt. Es erscheint kaum sinnvoll, den Begriff der Named Entity aufzuweichen in Richtung solcher (im Diskurs) monoreferentieller Ausdrücke (vgl. die Diskussion auf S. 94 und Fußnote 139 auf S. 176). Dennoch: für die Lokativa ist die Taxonomie von

Sekine et al. brauchbar, da pragmatisch und funktional orientiert.

ADL Feature Type Thesaurus (Hill 2000, Hill, Goodchild & Janée 2004) liefert eine gute und umfangreiche Taxonomie mit gut 200 Klassen (Abb. A.17 und A.18). Sie ist geprägt durch die Unterscheidung zwischen natürlichen und vom Menschen geschaffenen Objekten auf oberster Ebene, hinter der funktionale Kriterien zurücktreten. Die Klasse RESERVOIRS ist zusammen mit OFFSHORE PLATFORMS den HYDROGRAPHIC STRUCTURES untergeordnet und steht damit weit entfernt von der Klasse LAKES.

KIM ist eine vollwertige Ontologie mit 250 Klassen und 40 Relationen. Darunter sind gut 100 lokative Klassen, laut Manov et al. (2003) basierend auf der Taxonomie des ADL. Die Taxonomie erlaubt mehrfache Klassenzugehörigkeit, z.B. ist HARBOR den Klassen FACILITY und WATERREGION untergeordnet. Sie ist intuitiv, und nur Kleinigkeiten sind zu beanstanden oder bedürfen einer Diskussion, so die Entscheidung WATERFALLS als LANDREGION und nicht wie RIVER und SPRING als WATERREGION zu klassifizieren. Der Taxonomiebaum ist in Abb. A.20 auf S. 352 abgebildet. Die Definition der Taxonomie enthält ein Mapping auf die Klassen von NGA:GNS (s.o.), womit ein riesiger Gazetteer in die Taxonomie integriert ist.

WORDNET (Fellbaum ed. 1998) ist eines der bekanntesten Klassifikationssysteme, und dazu das mit Abstand umfangreichste. Jedoch erfolgt die Klassifikation inkonsequent und unsauber. So sind beispielsweise die beiden Lesarten von COUNTY „1. (United States) the largest administrative district within a state; 2. (United Kingdom) a region created by territorial division for the purpose of local government“ einmal als (wohl korrekte) Unterklasse von ADMINISTRATIVE DISTRICT, das andere Mal mit INTERSTELLAR SPACE oder DEPTH als Kohyponymen dem Synset REGION, PART untergeordnet (vgl. Abb. A.6 auf S. 337). Dass Kohyponyme eine homogene Menge bilden ist eine der wichtigsten Eigenschaften einer Taxonomie. Als Hyponyme von COUNTRY, STATE, LAND (vgl. Abb. A.6 auf S. 337) sind aber Klassen wie BANANA REPUBLIC, TAX HAVEN ‚Steuerose‘, KINGDOM und AFRICAN COUNTRY gelistet, die sich durch deutlich verschiedene Eigenschaften auszeichnen (vgl. die Kritik von Langer 2003: 139). Auch ist nicht nachvollziehbar wieso zwar KINGDOM und SULTANATE Instanzen eines Landes sein sollen, REPUBLIC dagegen nur Instanz des Synsets POLITICAL SYSTEM, FORM OF GOVERNMENT ist (vgl. Kap. 3.1.4 zum Problem der regulären Polysemie).

Aufgrund dieser chaotischen Klassifikation, ist WORDNET als Ausgangsbasis für eine Taxonomie geographischer Objekte nicht geeignet.

BBNTAXON Die Taxonomie von BBN mit Guidelines zur Annotierung eines vom LDC vertriebenen Korpus¹³⁰ besteht aus 29 Oberklassen (davon 12 NE-Klassen, 9 nominale Entity-Klassen und 7 Klassen numerischer Entitäten), und 64 Unterklassen. Abb. A.19 auf S. 351 zeigt alle fünf lokativen Oberklassen, darunter auch die Klasse FACILITY DESCRIPTOR, die nicht den Eigennamen zuzurechnen ist. Die Klasse NORP (“nationality, religion, organization, political”) enthält Einwohnerbezeichnungen und von Toponymen abgeleitete Adjektive.

Problematisch ist die Unterscheidung zwischen FACILITY und ORGANIZATION. Alle Gebäude oder Einrichtungen, die Organisationen beherbergen (können), wie *White House*, *Kremlin*, sowie Kirchen, Museen usw., werden als ORGANIZATION klassifiziert. Dies ist

130 BBN Pronoun Coreference and Entity Type Corpus (<http://www ldc.upenn.edu/Catalog/docs/LDC2005T33/>), erstellt von Ralph Weischedel und Ada Brunstein.

teilweise durch die häufige metonymische Verwendung gerechtfertigt (vgl. Kap. 3.1.4), aber mangels Flexibilität kaum einer Lösung mit metonymischen Klassen und multiplen Hyperonymen vorzuziehen.

PROLEX mit einer langen Geschichte vom französischen elektronischen Orts- und Eigennamenlexikon bis zur multilingualen Ontologie (Prolintex, Prolex, Belleil 1997, Belleil & Maurel 1997, Maurel et al. 1997, Maurel, Piton & Eggert 2000, Maurel et al. 2004, 2006) organisiert die enthalten Eigennamen aller Typen nach einer Taxonomie die mit Modifikationen auf Bauer (1998) basiert. Für alle Eigennamen sind nur zwei Gliederungsebenen vorgesehen, d.h. alle acht Toponymklassen stehen auf einer Ebene. Die metonymische Verwendung von Eigennamen (vgl. Kap. 3.1.4) findet durch „sekundäre Hyperonyme“ Berücksichtigung: so ist eine Stadt nicht nur ein Toponym, sondern bezeichnet auch deren Bewohner (=Anthroponym); sie ist von Menschen geschaffen und deshalb auch ein Ergonym (siehe Abb. A.4).

GETTY Die 1700 Klassen des GETTY-Thesaurus, sind zu viele, so dass die Taxonomie schwer handzuhaben ist. Die Zuordnung der Klassen erfolgt eher nach Art eines Epithetons, mehrfache Klassenzugehörigkeit von Objekten ist eher die Regel, denn die Ausnahme. Vgl. Moskau als INHABITED PLACE, CITY, REGIONAL CAPITAL, INDUSTRIAL CENTER, MANUFACTURING CENTER, COMMERCIAL CENTER, CULTURAL CENTER, TRANSPORTATION CENTER und CAPITAL (Abb. 3.2 auf S. 128).

3.4.2.2 Relationen zwischen geographischen Entitäten

SCENE III. Bohemia. The sea-coast

ANTIGONUS Thou art perfect then our ship hath touch'd upon The deserts of Bohemia?

William Shakespeare, The Winter's Tale

Wie Böhmen [...] eines schönen Tags ans Meer begnadigt wurde und jetzt am Wasser liegt.

Ingeborg Bachmann: „Böhmen liegt am Meer“

Relationen zwischen geographischen Entitäten ermöglichen einen alternativen Zugriff auf die Entitäten, z.B. auf Paris als Hauptstadt von Frankreich (= die Entität der Klasse STADT, die in der Hauptstadt-Relation zur Entität ‚Frankreich‘ steht). Die formale Modellierung der Relationen innerhalb der Ontologie ist Voraussetzung für die automatische Verarbeitung der Ontologie. Relationen zwischen geographischen Entitäten können sein:

- die administrative Untergliederung: Y ist Bestandteil/Subjekt von X ; X ist in Y , ..., Z untergliedert
- (partielle) Inklusion: Y ist Teil von X ; X enthält Y vollständig oder teilweise (=Überlappung). Diese Relation ist meist Voraussetzung für die administrative Untergliederung und die intuitivste Form zur Navigation innerhalb der Ontologie (s.u.)
- X ist Hauptstadt von Y
- Fluss X fließt in Gewässer Y
- räumlicher Kontakt: X grenzt an Y , Stadt X liegt an Gewässer Y
- räumliche Entfernung oder Nähe (Distanz). Diese lässt sich aus der räumlichen Repräsentation der beteiligten Entitäten errechnen.

Relation	Inverse R.	transitiv	symmetr.	
adjacent to		–	+	(adjacent to a river, ≈borders)
overlaps		–	+	
borders		–	+	
has part	part of	+	–	
part of	has part	+	–	
located in	location of	+	–	(fully located in, ≠ overlaps)
location of	located in	+	–	
has capital		–	–	
has location city		–	–	(e.g. for events)
has location province		–	–	(e.g. for events)
has location country		–	–	(e.g. for events)

Tabelle 3.6: Relationen zwischen Lokationen in ONTOWORLD.ORG (vgl. SEMANTIC-MEDIAWIKI)

Die Zahl möglicher Relation ist praktisch unbegrenzt. So könnte man für einen Fluss X auch noch folgende Relation ansetzen: entspringt an Ort Y , durchfließt See Y , ist Grenzfluss zwischen Y und Z , wird überquert von Brücke/Straße/Bahnlinie/Fährverbindung Y , usw. Die Tendenz ist jedoch der Komplexität der Welt zum Trotz nur wenige Relationen in einer Ontologie zu kodieren (vgl. Tab. 3.6). Die linguistische Relevanz der Relationen entsteht – neben der trivialen Tatsache, dass die Relationen auch sprachlich kodiert werden können – daraus, dass sie zur Modulierung bestimmter metonymischer Verwendungen (z.B. Hauptstadt für Regierung, vgl. Kap. 3.1.4) notwendig sind. Weiter ermöglichen die Relationen einen intuitiven Zugriff auf die Entitäten der Ontologie. Vor allem die Inklusionsrelation ist hier wichtig, sofern die Anzahl der Äste klein bleibt (7±2-Regel von Miller 1956). Die administrative Unterteilung ist meist weniger gut geeignet, da die Anzahl der Äste oft wesentlich größer ist (z.B. die USA mit 50 Bundesstaaten), und die Aufteilung in Verwaltungseinheiten sich häufig ändert. Neben den beiden Meronym-Holonym-Relationen bieten auch die anderen Relationen wichtige Assoziationen zum Zugriff auf die eigentliche geographische Entität.¹³¹ Diese Assoziationen manifestieren sich auch in den Definitionen/Bedeutungsexplikationen geographischer Entitäten (*Paris* = ‚Hauptstadt von Frankreich‘, *Köln* = ‚Stadt am Rhein‘).

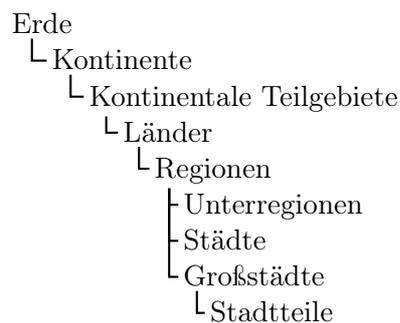


Abbildung 3.4: Geografische Hierarchie in WIKIVOYAGE.ORG/WIKITRAVEL.ORG. Diese wird zur Navigation durch sog. Breadcrumbs verwendet, z.B. Europa > Mitteleuropa > Deutschland > Bayern > Oberbayern > München

¹³¹ Vgl. die ‚relation d’accessibilité‘ bei Maurel et al. (2006).

3.4.3 Zeitliche Dimension: historische Namen und Orte

Die zeitliche Dimension (vgl. Axelrod 2003, Tran, Grass & Maurel 2004) wird in der Ontologie auf zwei Ebenen behandelt:

1. Eigennamen können mit einer Zeitperiode assoziiert werden, innerhalb derer der Name verwendet wird oder der offizielle Name eines Objekts war, so z.B. die Toponyme *Sankt Petersburg* (1701–1914,1991–), *Petrograd* (1914–1924) und *Leningrad* (1924–1991), vgl. Abb. 3.5.
2. geographische Entitäten, die nicht mehr existent sind, werden als ‘historisch’ markiert, ev. mit Angabe der Zeitperiode der Existenz. So existierte der Schweizer Kanton *Ausserschwyz* nur 1831.

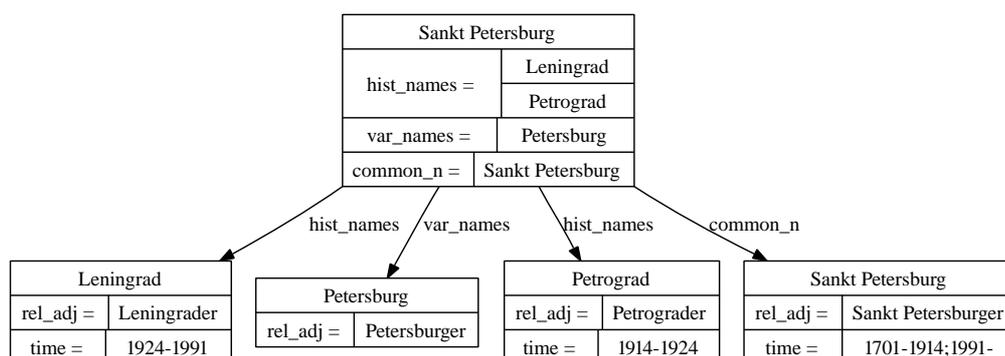


Abbildung 3.5: Die geographische Entität ‚Sankt Petersburg‘ mit ihren Benennungen, historische Namen eingeschlossen, in der Ontologie.

3.5 Phrasengrammatiken zur Erkennung lokativer Argumente

Ziel der Phrasengrammatiken ist die automatische Erkennung von (1.) Argumenten lokativer Prädikate und (2.) Supplementen/Angaben, d.h. Adverbialen des Ortes (Lokalbestimmungen). Als einzelne Phrasentypen kommen in Frage:

- Nominalphrasen, siehe Kap. 3.5.1
- Präpositionalphrasen, siehe Kap. 3.5.2
- und einige weitere: Adverbien und Adverbiale (Kap. 3.5.3), Lokalsätze (Kap. 3.5.4) und Adressen (Kap. 3.5.5)
- sowie Kombinationen aus diesen: quasikoordinierte statische Ortsangaben (Kap. 3.6.1) und Pfadangaben mit mehreren Pfadbestandteilen (Ausgangspunkt, Wegpunkte, Zielpunkt: Kap. 3.6.2)

Beschrieben wurden diese Phrasen mit lokalen Grammatiken und Lexikogrammatiken. Alle syntaktisch relevanten Phänomene wurden behandelt, und zwar (1.) Kongruenz in Kasus, Genus, Numerus und Adjektivdeklinatation (stark/schwach) innerhalb der NP, und (2.) die Kasus- und Numerusreaktion von Präpositionen. Da UNITEX als Engine für lokale Grammatiken über keinen Unifikationsmechanismus verfügt, wurde via Expansion ‚unifiziert‘, d.h. für jede mögliche Kombination aus Kasus, Genus, Numerus und Adjektivdeklinatation führt

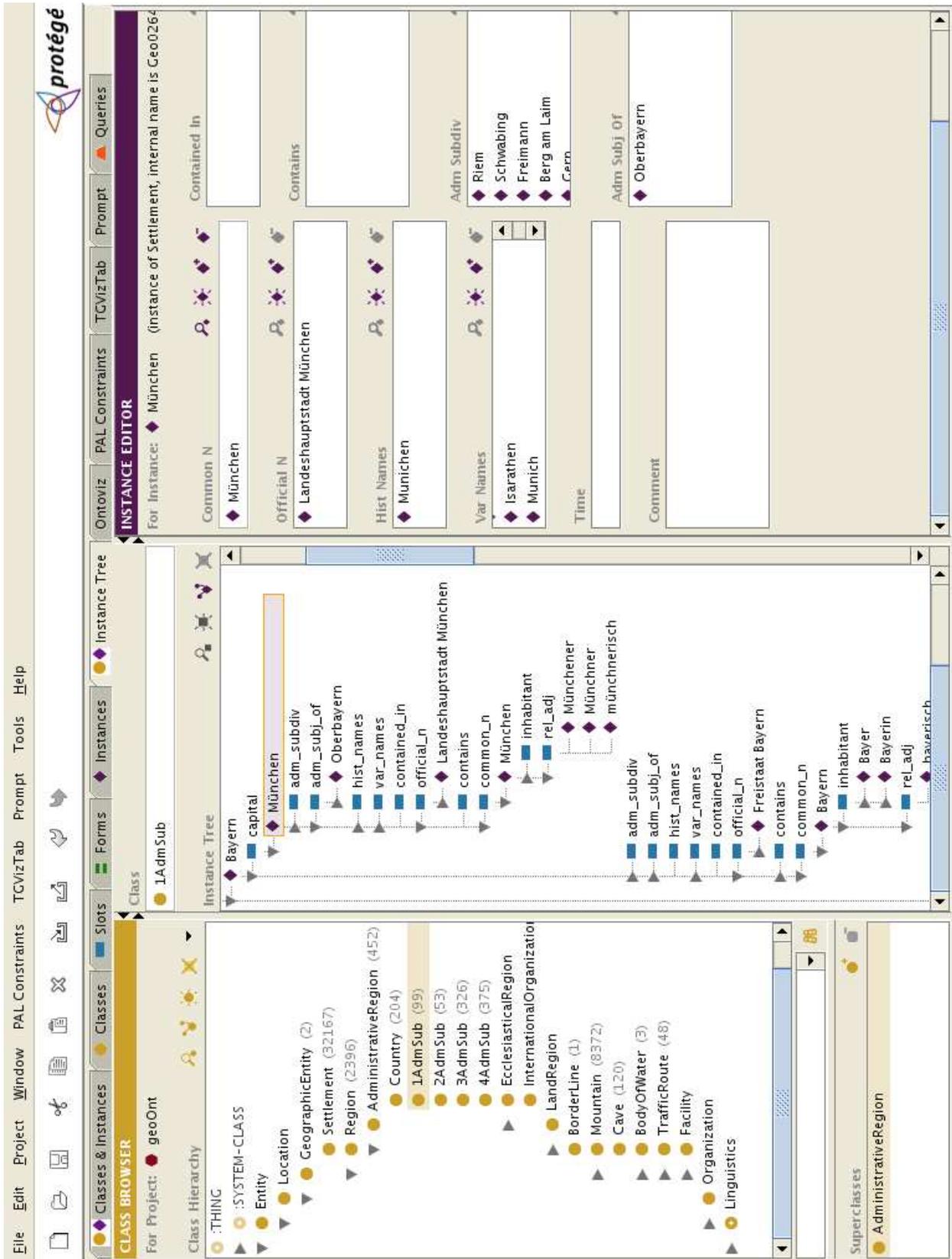


Abbildung 3.6: Screenshot des Ontologiebrowsers PROTÉGÉ. Im linken Fenster ist die Klassenhierarchie zu sehen. Das mittlere Fenster zeigt Relationen zwischen der Entität ‚Bayern‘ und anderen geographischen wie sprachlichen Entitäten. Rechts sind die Slots der gerade aktivierten geographischen Entität ‚München‘ zu sehen.

ein eigener Pfad durch den Automaten. Syntaktische Merkmale von Toponymen wurden, wenn immer möglich im Lexikon kodiert. Der Fokus der lokalen Grammatiken liegt auf spezifischen Pattern, die entweder soviel Variation zeigen, dass sie als Stringrepräsentation im Lexikon zu viel Platz einnehmen würden, oder semantisch klar kompositional aufgebaut sind, und deshalb nicht als lexikalische Einheiten gelten können. Lexikongrammatiken fanden für die Beschreibung der syntaktischen und kombinatorischen Eigenschaften der Präpositionen Anwendung.

Probleme bei der Beschreibung von Nominal- und Präpositionalphrasen resultieren aus:

- der Koordination unterhalb der Wortebene (*Ober- und Unterammergau, zwischen Rotem und Weißem Main*, vgl. Kap. 3.3.2.10). Eine adäquate Beschreibung würde partiellen Lexikon-LookUp erfordern, der in UNITEX nicht möglich ist.
- die Beschreibung von auf verschiedenen Phrasenebenen koordinierten Ausdrücken stellt dagegen kein größeres Problem dar:

(3.77) *in Österreich, Frankreich, Italien und in der Schweiz*

Ursache sind Unterschiede beim Artikelgebrauch und dem Genus der koordinierten Toponyme und resultierenden Kollisionen in den kontrahierten Formen aus Präposition und Artikel (vgl. auch (3.82)):

(3.78) a. *in Griechenland und in der Türkei*

b. *in Griechenland und der Türkei*

(3.79) a. *im Irak und in der Türkei*

b. *im Irak und der Türkei*

c. *in der Türkei und dem Irak*

d. *in der Türkei und im Irak*

e. * *in der Türkei und Irak*

- Phrasen wie *in der Stadt und im Landkreis München* (siehe (3.114) auf S. 144) werden als Koordination zweier Ortsadverbiale erkannt, also in der Bedeutung ‚in einer (beliebigen) Stadt und im Landkreis München‘, und nicht in der wahrscheinlicheren Lesart ‚in der Stadt München und im Landkreis München‘. Eine Beschreibung dieser würde die Anzahl der Subgraphen der Grammatik drastisch erhöhen, da ja die Präpositionalphrasen über eine Lexikongrammatik lexikalisiert sind, also für jede Form einer Präposition ein eigener Subgraph existiert. Auf ein ähnlich gelagertes Problem bei der Kombination von Klassenbezeichner und Toponym ist Constant (2002, 2003b) gestoßen (siehe Kap. 2.5.1.4).

3.5.1 Nominalphrasen

Nominalphrasen können die Position des Aktanten eines lokativen Prädikats unmittelbar (3.80) oder mittels einer vom Prädikat regierten Präposition (3.81) auffüllen.

(3.80) *Der Schornbach durchfließt das Tal.*

(3.81) *Der Fluss Tay mündet in die Nordsee.*

In beiden Fällen bleibt die Struktur der Nominalphrase gleich. In einigen Kasus-Genus-Kombinationen im Singular können Präposition und bestimmter Artikel verschmelzen (PDet: *im, zur*). Dies macht eine besondere NP-Variante erforderlich (siehe auch Beispiel (3.60) auf S. 112 mit Erläuterungen). Schwierig, aber nicht unlösbar ist die Beschreibung von koordinierten Toponymen mit PDet. Alle drei Varianten in (3.82) sind in Texten anzutreffen und dabei etwa gleich häufig.

- (3.82) a. *im Westjordanland und im Gazastreifen*
 b. *im Westjordanland und Gazastreifen*
 c. *im Westjordanland und dem Gazastreifen*

Kopf der Nominalphrase ist entweder ein lokatives Nomen oder ein Toponym. Beide können durch weitere Elemente modifiziert oder spezifiziert werden. Die Modifikation des Kopfes durch Adjektive wird durch die Grammatik beschrieben, weitere Modifikatoren (Genitivattribut, Präpositionalattribute/PP-attachments) dagegen nur rudimentär, da eine adäquate Beschreibung nur möglich ist, wenn die Argumentstruktur des Kopfnomens oder weiterer dependenter Nomina bekannt ist (siehe auch Kap. 4.6). Ausführlich wurden Appositionen von Klassenbezeichnern wie Toponymen behandelt, da sie typisch für Phrasen sind, die einen Eigennamen enthalten.

3.5.1.1 Adjektivmodifikatoren

Spezifische Adjektivmodifikatoren lokativer Nomina und Toponyme, z.B. wurden *baumbe-standen, schiffbar*, in ein Lexikon aufgenommen (Kap. A.3). Darüberhinaus mussten einige typische Adjektivphrasen, d.h. ein Adjektiv mit weiteren modifizierenden Elementen, mittels lokaler Grammatiken beschrieben werden. Diese verteilen sich im Wesentlichen auf folgende Typen:

- Zahl der Einwohner

(3.83) *die **knapp 100.000 Einwohner zählende** Industriestadt*

(3.84) *Belgrad, die **mit knapp 1,7 Millionen Einwohnern drittgrößte** Stadt an der Donau*

- eindimensionale Ausdehnung (Länge, Breite, Höhe, Tiefe)

(3.85) *Die Sempt ist ein **knapp 60 km langer** rechter Nebenfluss der Isar*

(3.86) *ein **sieben Kilometer langer und über einen Kilometer breiter** Stausee*

(3.87) *im **bis 1.537 m hohen** Bergland von Futa Dschalon*

- zweidimensionale Ausdehnung (Fläche)

(3.88) *das **etwa 1854 m² große** Grundstück*

(3.89) *das **13,5 Hektar umfassende** Gebiet*

- dreidimensionale Ausdehnung (Volumen)

(3.90) *der **109 Millionen Kubikmeter fassende** Stausee*

- Entfernung

(3.91) *der etwa 6 bis 9 km entfernt liegende Fluss*

- Adjektive mit temporaler Ergänzung (Bau, Gründung, Alter, usw.)

(3.92) *der von 1960 bis 1971 errichtete Assuan-Damm*

(3.93) *der 2–4 Millionen Jahre alte Kuiseb-Cañon*

(3.94) *der schon um 1200 gegründete Ort Krásna nad Hornádom*

3.5.1.2 Toponyme

Nominalphrasen mit Toponymen sind wesentlich geprägt von den syntaktischen Eigenschaften (Kap. 3.2.1.2) des Toponyms. Die wichtigste Eigenschaft ist dabei die der Verwendung mit und ohne Artikel (Kap. 3.3.2.4), die in folgenden syntaktischen Mustern resultiert:

<Topon+DetZ> Toponym, ohne Artikel gebraucht: *in München*

<Det> <Topon+Det> Toponym mit Artikel: *in den Alpen*

Toponyme können durch Adjektive oder Adjektivphrasen modifiziert werden, siehe Kap. 3.5.1.1. Auch artikellose Toponyme stehen dann mit Artikel, siehe (3.59) auf S. 112. Auch eine Modifikation durch koordinierte pränominalen Phrasen ist möglich. Die resultierende NP erhält dann das Merkmal ‘Plural’, in dem das finite Verb steht:

(3.95) *die alte und die neue Türkei sind ...*

Ein spezifisches syntaktisches Muster (*ganz+halb* <Topon+DetZ>) ist für die beiden Quantoren *ganz* und *halb* in Kombination mit einem artikellosen Toponym nötig, siehe auch S. 112. Artikellose Toponyme stehen hier ohne Artikel:

(3.96) *in ganz Deutschland und halb Polen*

In bestimmten Kontexten (Telegrammstil usw., siehe Kap. 3.3.2.4) können auch mit Artikel gebrauchte Toponyme ohne Artikel verwendet werden. Adjektivmodifikatoren sind dabei möglich (*Medienzensur in Ungarn, DDR und vereintem Deutschland*), das Adjektiv steht in der starken Form. Im Genitiv ist die artikellose Verwendung ebenso wie die Verwendung artikelloser Toponyme nur möglich, wenn das Toponym eine auf [s] endende Genitivendung trägt. Wird das Toponym durch Attribute modifiziert oder steht mit Artikel (oder einigen anderen Determinativen), muss stattdessen eine endungslose Genitivform gebraucht werden.

(3.97) *Türkeis Armee geht gegen Kurden vor*

(3.98) a. *Münchens Residenz*

b. *die Residenz des kurfürstlichen München*

c. * *die Residenz des kurfürstlichen Münchens*

Dieses nicht ausschließlich auf Eigennamen beschränkte Phänomen, ist noch wenig theoretisch beschrieben, vgl. DUDEN2006: §321–6, Spencer (o.J.), Mel’čuk (2006:129). Da es Flexionsklasse und syntaktische Stellung – also die Ebene der morphologischen und syntaktischen Beschreibung – kombiniert, ist es schwer zu modellieren. Hier wurde ein Merkmal ‘S’ der flektierten Form (z.B. *Kölns*, *Köln*.EN+Topon+Oikon+DetZ:geNS) eingeführt, was nicht heißen soll, dass damit eine neue grammatische Kategorie postuliert wird.

Appositive Konstruktionen aus Toponymen

Durch nachgestellte Apposition können Toponyme genauer spezifiziert werden und lassen sich damit von homonymen Toponymen unterscheiden. Für diese spezifischen Konstruktionen sind folgende Muster zu beobachten:

<Topon>/<Topon> Toponym 1 liegt in Top. 2: *im Flugzeug nach Vail/Colorado*

<Topon>, <Topon+Choron> dito, aber mit semantischer Einschränkung des zweiten (und dritten) Toponyms auf eine bestimmte Toponymklasse: *in Sabaneta, Bundesstaat Barinas, Venezuela*

<Topon> (<Topon>) dito: *Landshut (Niederbayern)*

<Topon> (<Class>) Toponym der Klasse <Class>: *München (Stadt) = Stadt München vs. München (Land) = Landkreis München*

<Topon> (<spec> <Topon>) mit <spec> wird entweder die Inklusionsrelation (liegt in) genauer spezifiziert: *in Straßburg (heute Frankreich)*; oder eine Namensvariante angegeben: *der Landkreis Neumarkt in der Oberpfalz (amtlicher Name Landkreis Neumarkt i.d.OPf.)*

<Topon> (<coord>) Angabe der Koordinaten von <Topon>: *Nebelhausen (Koordinaten: 35° 58’ 08" nördliche Breite, 95° 36’ 56" westliche Länge)*

Phrasen mit Klassenbezeichner als Kopf können dem Toponym nachgestellt werden, der Klassenbezeichner ist aber nicht morphologischer Kontaktpunkt (siehe Kap. 3.5.1.3). Eine Kombination mit den oben genannten Pattern ist möglich:

(3.99) *Richmond, die Hauptstadt Virginias, ...*

(3.100) *in Aspen/Colorado (USA), der Partnerstadt von Garmisch-Partenkirchen*

(3.101) *Tempe, Bundesstaat Arizona (Vereinigte Staaten von Amerika)*

Die nachgestellte Klassenbezeichnerphrase kongruiert oft, aber nicht notwendigerweise im Kasus mit dem Toponym:

(3.102) a. *in der Nähe des Niger, des über 4000 Kilometer langen Flusses*

b. *Köln liegt an den Ufern des Rheins, einem der symbolträchtigsten Flüsse Europas.*

Die Angabe von Namensvarianten ist ebenso über nachgestellte Phrasen möglich:

(3.103) *die Inselhauptstadt Skyros, auch Chora genannt*

3.5.1.3 Klassenbezeichner und Toponym

Das Kriterium für diesen Phrasentyp ist, dass der Klassenbezeichner den Kopf der Phrase bildet. Damit lassen sie sich von den oben beschriebenen Phrasen (3.99)–(3.102), die ja ebenso Toponym und Klassenbezeichner enthalten können, abgrenzen. Der Klassenbezeichner ist morphologischer Kontaktpunkt der Phrase und trägt als solcher die Kasusmarkierung (siehe Mel'čuk 2003: 200–201 für eine detaillierte Diskussion). Das Toponym steht meist in Apposition zum Klassenbezeichner,

- (3.104) *die Stadt Köln*
- (3.105) *im Nationalpark Hohe Tauern*
- (3.106) *am Stausee „Lac de Ste. Croix“*

aber auch einige andere Varianten der Anbindung sind möglich:

- (3.107) *in einer Stadt namens Schilda*
- (3.108) *in Staaten wie der DDR*
- (3.109) *der größte Staudamm Südafrikas, der Gariep Dam*

Sowohl Klassenbezeichner als auch Toponyme können mit ihresgleichen koordiniert sein, bei mit *und* koordinierten Toponymen steht der Klassenbezeichner im Plural:

- (3.110) *die Inseln Helgoland und Sylt*
- (3.111) *in der Stadt und Fürstenresidenz Bamberg*
- (3.112) *aus der Stadt und dem Kreis Aussig*
- (3.113) *in den Gemeinden Bad Sassendorf, Lippetal, Welver und der Stadt Hamm*

Solange die Koordination auf Ebene der NP erfolgt, konnten alle koordinierten Phrasen beschrieben werden. Ausgeschlossen blieben deshalb Phrasen wie (3.114a), die in der Dependenzanalyse nicht (oder nur mit Kniffen) korrekt als Baum darstellbar sind.

- (3.114) a. *in der Stadt und im Landkreis München* (nicht beschreibbar)
- b. *in Stadt und Landkreis München* (beschreibbar)

Viele der Klassenbezeichner sind prädikative Nomina und können als solche Aktanten an sich binden. Ohne eine vollständige und adäquate Beschreibung zu beanspruchen – diese war nur für einige wenige Nomina (Kap. 4.6) möglich, wurde versucht gängige Realisierungen schematisch zu beschreiben, um eine hohe Abdeckung der Phrasengrammatik zu erzielen.

- (3.115) *im Kanton Waadt der Schweiz*
- (3.116) *in Japans einstiger Hauptstadt Kyoto*
- (3.117) *in Simbabwe Hauptstadt und Münchens Partnerstadt Harare*

Auch wenn die beschriebenen Muster meist aktantische Klassenbezeichner erkennen, beschreiben sie auch andere Konstellationen. In (3.118) regiert das Superlativgrammem der Wortform *größte* einen Aktanten (vgl. (2.11) auf S. 30), die Vergleichs- oder Referenzmenge, vgl. die Paraphrase *die im Bundesstaat größte Stadt, Milwaukee*. In (3.119) und (3.120) haben die Attribute des Klassenbezeichners nicht den Status eines Aktanten.

(3.118) *die größte Stadt des Bundesstaates, Milwaukee*

(3.119) *in vielen Ländern Afrikas und Asiens*

(3.120) *in den Hauptstädten des Baltikums*

In allen diesen Fällen werden die Nominalphrasen zwar korrekt als solche erkannt, aber aus dem Parsebaum der Grammatik würde eine falsche Interpretation dieser nicht-projektiven (nicht-konfigurativen) Konstruktionen folgen.

Adverbiale Modifikatoren des Nomens (*das Haus dort, ein See in Bayern*) werden nicht von der Grammatik erfasst. Versuche haben gezeigt, dass die Precision der Grammatik leidet, wenn diese als Pfade in der Grammatik ergänzt werden. Möglich wäre das nur mit einer extrem restriktiven Beschreibung, und auch dann wären Ambiguitäten zu quasikoordinierten Ortsangaben/-adverbialen nicht aufzulösen (*an einem See in Bayern = in Bayern an einem See; dort in dem Haus = in dem Haus dort*).

Eigene Pfade innerhalb der Grammatik erfordern zwei Konstruktionen: sog. elektive Konstruktionen (3.121) und solche mit *kein ... außer* (3.122).

(3.121) a. *einige der größten Städte der USA*

b. *eines der 23 Oberzentren des Freistaates Bayern*

c. *auf einer der Jungferninseln*

(3.122) a. *in keinem anderen Land außer Deutschland*

b. *auf keiner anderen Insel außer Kreta*

Wie (3.121c) zeigt, sind elektive Konstruktionen auch für pluralische Toponyme möglich (vgl. Kap. 3.3.2.3).

Kombinationen aus Klassenbezeichner und Toponym lassen sich im Deutschen auch in Form von Nominalkomposita realisieren; auf die zur Bildung notwendigen Fugenformen wurde bereits in Kap. 3.3.2.12 eingegangen. Das Wort *Mittelmeerinsel* ist eine Kombination des Prädikats *Insel* und seines zweiten Aktanten: *eine Insel im Mittelmeer*. Für *Schwarzmeergebiet* ist die Situation nicht mehr so klar. In Kombination mit Gewässern hat der Bestandteil *-gebiet* die Bedeutung ‚das Gebiet oder die Gebiete, die an ein Gewässer angrenzen‘. Paraphrasieren lässt sich das Beispiel mit ‚das Gebiet um das Schwarze Meer herum‘, aber nicht mit ‚das ‚Gebiet des Schwarzen Meeres‘. Diese zusätzliche Bedeutungskomponente muss beschrieben werden, um die Paraphrasen einer Nominalphrase korrekt zu erfassen.

3.5.1.4 Relativsätze

Relativsätze, die den Kopf einer Nominalphrase, hier das lokative Nomen, der Klassenbezeichner oder das Toponym, wurden nicht beschrieben, da der Aufwand zur Beschreibung bedingt durch die hohe Zahl möglicher Prädikate im Relativsatz hoch ist. Eingeleitet werden lokative Relativsätze entweder von einem Relativpronomen (3.123) (ev. mit Präposition) oder durch *wo*, *wohin*, *woher* (3.124). Mit *wo* (*wohin*, *woher*) eingeleitete Relativsätze sind auch mit Ortsadverbien und -adverbialen kombinierbar (3.125) und können ev. selbst Aktantenstatus besitzen, vgl. die Diskussion in Kap. 3.5.4.

(3.123) *Im Gegensatz zu **München**, das in wesentlichen Teilen die historische Stadtsilhouette erhalten hat, ...*

(3.124) *Gerade für **München**, wo die Schlangen vor dem Wohnungsamt immer länger werden und die Grundstückspreise längst bundesweiten Rekord darstellen, ...*

(3.125) ***Südlich von München**, wo das Zeller Tal in die Isarsenke mündet, liegt das kleine Dorf Bairawies.*

3.5.1.5 Paraphrasen einer NP

Bereits auf Ebene einer Nominalphrase sind zahlreiche Möglichkeiten der Paraphrasierung gegeben. Die Möglichkeiten resultieren aus Ersetzungsmöglichkeiten zwischen attributivem Adjektiv, untergeordneter NP oder PP, beigeordneter NP (Apposition) oder einer Fugenform im Nominalkompositum, siehe die Beispiele unten (viele der Beispiele stammen von Sugarewa 1974: 237–49).

Die überwiegende Anzahl der Paraphrasierungen einer NP sind Ergebnis von alternativen Realisierungen der Aktanten, d.h. Variationsmöglichkeiten in der Rektion:

(3.126) a. *Reise nach Afrika*
b. = *afrikanische Reise*
c. = *Afrikareise*
d. = *Die Reise geht nach Afrika.*
(elem. Satz)

(3.127) a. *Europatournee*
b. = *Tournee durch Europa*

(3.128) a. *afrikanische Küste*
b. = *Küste Afrikas*
c. = *Afrikas Küste*
d. = *Küste von Afrika*

(3.129) a. *DDR-Regierung*
b. = *Regierung der DDR*

(3.130) a. *sowjetische Armee*
b. = *Armee der Sowjetunion*
c. = *Sowjetarmee*

(3.131) a. *sowjetische Soldaten*
b. = *Sowjetsoldaten*

(3.132) a. *die deutschen Stellungen (an der Front)*
b. = *die Stellungen der Deutschen (an der Front)*

Andere Paraphrasen basieren auf verschiedenen Möglichkeiten des Ausdrucks der Lage (statische Position, containment) eines Objekts an einem Ort (3.133), oder (3.134)–(3.135) als ein Spezialfall davon, Ortsnamen, die in andern, komplexeren Namen, enthalten sind, und dabei eine diffuse Ortszugehörigkeit ausdrücken (vgl. Kap. 5.1.5.1.2 mit weiteren Beispielen).

(3.133) a. *rumänische Obstplantagen*

- b. = *Obstplantagen in Rumänien*
- c. = *Obstplantagen Rumäniens*
- (3.134) a. *Rostocker Universität*
- b. = *Universität Rostock*
- (3.135) a. *Leipziger Hochbaukombinat*
- b. = *Hochbaukombinat Leipzig*

3.5.2 Präpositionalphrasen

Präpositionalphrasen können verwendet werden

- als Bestandteil eines elementaren Satzes, die Präposition selbst ist dabei das „Kern“-Prädikat des Satzes
- als Komplement zu einem Prädikat, und zwar
 - mit stark regierter, unveränderlicher Präposition
 - mit variabler Präposition, die zur Semantik des Satzes beitragen kann
- als Adverbialsupplement (Kap. 4.2)
- Komplement wie Supplement können als Satzadverbial oder als Attribut eines Nomens, Adjektivs, Adverbs im Satz fungieren

Diese Polyfunktionalität von Präpositionalphrasen führt in der Sprachverarbeitung zum bekannten Problem der Disambiguierung von sog. PP-attachments (vgl. Brill & Resnik 1994). Entscheidend für die Beschreibung ist aber, dass sich die innere Struktur einer PP in allen diesen Funktionen nicht unterscheidet. Die Beschreibung kann deshalb in einem Zug erfolgen, mit zwei Ausnahmen: Falls die Präposition nicht stark regiert ist, können Präpositionalphrasen (1.) mit anderen Adverbialen quasikoordiniert (Kap. 3.6) und (2.) modifiziert werden (*mitten in, quer durch, weit jenseits*). Welche Formen und Ausdrücke als Modifikatoren erlaubt sind, wird durch die Präposition selbst gesteuert. Allgemeine Charakteristika räumlicher Präpositionen wurden bereits in Kap. 2.4 besprochen, der interne Aufbau von Präpositionalphrasen wird zusammen mit präpositionalen elementaren Sätzen (Kap. 4.4) und der detaillierten Beschreibung der syntaktischen Eigenschaften in einer Tabelle der Lexikongrammatik (Kap. C.1) in beschrieben.

3.5.3 Ortsadverbien

Mit Adverbien geht der *native speaker* verschwenderisch und diskriminierend in einem um. Nach München geht's *auffi*, nach Passau *obi*, nach Altötting *eini*. Amerika liegt *drent*, man fährt *auf* New York *ummi* und war in Kalifornien *hint*.

(Richard McCormack, *Tief in Bayern*, S. 55)

Auch Adverbien können Argumentstellen realisieren, zum einen als Substitut von präpositional angeschlossenen Nomina (vgl. die Anmerkungen in Kap. 4.3), also in pronominaler Funktion, zum anderen als Bestandteile mit eigener Semantik – als „location quantifiers“ (*überall, irgendwo, nirgendwo, allerorts*, siehe Kracht 2002, Creary, Gawron & Nerbonne

1987, 1989) oder lokale Deiktika (*hier, gegenüber*, und im weiteren Sinn: *hinten, unten, hinter, heimwärts*). Über Adverbien in der Funktion des Prädikats wurde bereits gesprochen (Kap. 2.8), in Kap. 4.2 werden als Supplemente verwendete Adverbien oder Adverbialphrasen behandelt.

Zu den etwa 200 einfachen räumlichen Adverbien kommen noch Kombinationen aus diesen und komplexe Bildungen. Die Bemerkung von Engel (1996: 760) „im ganzen sind Adverbien nur in geringem Maße erweiterungsfähig“ trifft zwar zu, dafür ist die Kombinatorik von Adverbien mit weiteren Adverbien, Partikeln und Präpositionen stark eingeschränkt und nicht einfach zu beschreiben. Als Muster bei der Bildung sind u.a. belegt:

- dem eigentlichen Adverb vorangestelltes modifizierendes Adverb: *knapp dahinter, genau dort, ungefähr hier, weiter links, ganz unten, schräg gegenüber*
- Präposition und Adverb: *ab da, von oben, nach unten, abseits von hier, von oben her, zwischen hier und ...*
- statisches lokatives Adverb und Direktionaladverb: *dort hinauf, hier hinunter*
- Adverbien verschiedener deiktischer Dimensionen: *hier oben, dort drüben, da hinten*
- diese Kombinationen wiederum kombiniert mit einer Präposition: *von hier oben*

Etwa 300 dieser Kombinationen wurden direkt ins Lexikon (Kap. A.3) aufgenommen.¹³² Waren die kombinatorischen Möglichkeiten zu zahlreich, wurden sie mittels lokaler Grammatiken beschrieben (Beispiele in Kap. B.1). Zu dieser Gruppe zählen komplexe Adverbien wie *mehrere hundert Kilometer weiter nordwestlich*.

Ebenso wie bei Orts- und Richtungsangaben, die ein lokatives Nomen oder Toponym enthalten und als NP oder PP realisiert sind (Kap. 3.5.1.4), kann ein Adverb mit einem durch *wo, wohin*, usw. eingeleiteten Relativsatz ergänzt werden.

3.5.4 Lokalsätze

Lokalsätze werden Nebensätze genannt, die „Ort, Richtung oder Erstreckungsbereich des im Hauptsatz bezeichneten Sachverhalts“ erläutern und „durch Lokaladverbien (*wo, wohin, woher*) eingeleitet“ werden (Bußmann ed. 2002: 415) und alternativ zu Adverbien und Präpositionalphrasen verwendet werden können (Eroms 2000: 234).

(3.136) *Er segelte, wohin ihn der Wind trieb.* (Bußmann ed. 2002: 415)

(3.137) *Ich wohne, wo es mir gefällt.* (Eroms 2000: 234)

(3.138) *Er geht, wohin er will.* (Pittner 1999: 225)

(3.139) *Er wohnt, wo sich Fuchs und Hase gute Nacht sagen.* (Pittner 1999: 225)

Nun ist das von Bußmann angeführte Beispiel nicht glücklich gewählt, da *wohin ihn der Wind trieb* eher als feste Wendung (nicht-standardisierte lexikalische Funktion) mit der Bedeutung ‚irgendwohin‘/‚zum Spaß‘/‚wohin es einem gefällt‘ zu betrachten ist. Ähnliches gilt für die Beispiele von Eroms und Pittner. Insofern ist die Beschreibung dieser Sätze von sekundärer

¹³² Als Quelle dienten Listen aus den Arbeiten von Helbig & Buscha (1991: 343–345), Engel (1996: 751) und Engelke (2003).

Priorität. Nicht zu verwechseln sind die Lokalsätze mit durch die gleichen Lokaladverbien eingeleiteten Relativsätzen (siehe Kap. 3.5.1.4).

3.5.5 Adressen

Adressen (Postanschriften) stellen präzise und eindeutige Ortsangaben dar. Aufbau und Anordnung der Adressbestandteile unterliegen spezifischen Konventionen, die weniger linguistisch motiviert sind, sondern durch Vorgaben der jeweiligen Post oder auch nationale bzw. internationale Normen (z.B. DIN 5008). Hinzu kommen Elemente, die keine Lokativa sind: (1.) Personennamen (Empfänger oder Absender mit Anrede und Titel), und (2.) Organisationen (Firma, Behörde, Institution) inklusive ihrer Untergliederung (Abteilung, Referat). Am CIS haben sich zwei Magisterarbeiten mit der Erkennung und dem Parsen von Adressen beschäftigt (Altfeld 2000, Mederle 2004). Adressen stehen jedoch außerhalb des Fokus dieser Arbeit, da sie, wenn überhaupt, nur sehr begrenzt in Sätze eingebettet werden können. Sieht man von (3.140a) und (3.140b) ab, sind nur zwei Bestandteile/Felder einer Adresse als lokale Argumente in einem elementaren Satz verwendbar: Straße mit Hausnummer (3.140c) und Wohnort mit Postleitzahl (3.140d), oder auch eine Kombination von beiden (3.140e):

- (3.140) a. *Die Adresse von Max lautet: Max Mustermann, Musterstr. 123, 12345 Musterstadt.*
b. *Diese Anmeldung schicken Sie bitte an folgende Adresse: Mustermann KG, ...*
c. *Laras Schwester wohnt in der Romanstraße 16.*
d. *Thomas Maier wohnt in 45149 Essen.*
e. *Der neue Schornsteinfeger wohnt in der Gromergasse 13/2 in 72124 Pliezhausen*

Die Verbindung von Straße oder Ort mit Hausnummer bzw. Postleitzahl verhält sich syntaktisch voraussagbar: Straßen und Hausnummer stehen mit Artikel wie der Straßename allein (ohne Hausnummer), Ort und Postleitzahl stehen immer ohne Artikel und sind Neutra. Sie werden deshalb auch so annotiert, sie stehen jedoch nicht im Lexikon, sondern werden von einer Grammatik erkannt (siehe Kap. 5.1.5.3.1.1, vgl. auch Mederle 2004: 9–11 mit einer Straßennamengrammatik).

3.6 Kombination von Ortsangaben

Ortsangaben (als Komplemente und Supplemente) lassen sich auf zweierlei Weise verknüpfen – per Koordination oder Quasikoordination.¹³³ Aus der Unterscheidung von Koordination und Quasikoordination folgt ein Unterschied in der Bedeutung:¹³⁴

133 Für den Begriff ‚Quasikoordination‘ siehe Mel’čuk (2006: 9) und Mel’čuk (2004c: 250,276): „QUASI-COORD stands for the quasi-coordinative deep syntactic relation, which holds between an lexical unit L and its conjoined dependent that semantically elaborates L, adding more specific information: *John was born in the USA*,–quasi-coord→*in New York*,–quasi-coord→*in Manhattan*,–quasi-coord→*on the 56th Street*.“

134 Zumindest bei Orts- und Temporalangaben. Bei anderen quasikoordinierten Elementen, ist die Situation weniger klar, vgl. das Beispiel aus Rauh (1997): *ein Brot mit Käse ohne Butter* vs. *ein Brot mit Käse und ohne Butter*.

- (3.141) *Nach dem Studium **hier und in München** lebte er für einige Jahre in seiner Heimatstadt.* (koordiniert)
- (3.142) *Sie hat nach dem Studium **hier in München** an der Medizinischen Kleintierklinik der LMU im Bereich der Inneren Medizin promoviert und mehrere Jahre gearbeitet ...* (quasikoordiniert)

Die Aussage in (3.141) lautet, dass die Person *X* an zwei Orten, in München und dem durch *hier* referenzierten Ort, studiert hat. In (3.142) studierte *X* nur in München, auf das deiktisch durch *hier* Bezug genommen wird. Durch Quasikoordination verbundene, nachgestellte Ortsangaben spezifizieren die vorhergehenden Bestandteile. Letzere Interpretation liegt auch deshalb nahe, da es strenge Regularitäten in der Abfolge quasikoordinierter Ortsangaben gibt (**in München hier*), während dies für koordinierte nicht der Fall ist (*in München und hier*). Koordinierte Ortsangaben sind durch allgemeine Regeln zur Phrasenkoordination beschreibbar und werden hier nicht behandelt.

Neben den quasikoordinierten statischen Ortsangaben geben Pfadangaben, die den räumlichen Verlauf einer Bewegung beschreiben, ein weniger klares Bild. Sie lassen sich aber doch als Sonderfall der Quasikoordination behandeln.

3.6.1 Quasikoordinierte statische Ortsangaben

Quasikoordinierte statische Ortsangaben können einfach als eine Region betrachtet werden, die als Schnittmenge aus allen beteiligten Regionen definiert ist: „we allow that multiple locatives (in the same verb phrase) jointly determine a single location argument via the intersection of regions“ (Creary, Gawron & Nerbonne 1989).

- (3.143) *Stark zerstört war auch das Gebiet nördlich der Brunswik zwischen Holtenuer- und Feldstraße.*
- (3.144) *Oberlangenberg ist eine Ortschaft in Hückeswagen im Oberbergischen Kreis im Regierungsbezirk Köln in Nordrhein-Westfalen (Deutschland).*
- (3.145) *Die Region Franche-Comté liegt westlich des Rheins, südlich des Elsasses und nördlich der Alpen, zwischen Visoul und Besançon.*

In (3.144) ist die Ortsangabe nur im weiteren Sinne als Schnittmenge der beteiligten Regionen zu bezeichnen, im engeren Sinne ist es eine Menge ineinander geschachtelter Region: Deutschland > NRW > Reg.Bez. Köln > Oberberg. Kreis > Hückeswagen > Oberlangenberg. Quasikoordinierte Ortsangaben weisen also Parallelen auf zu Toponymen wie *Frankfurt an der Oder* und zu Toponymreihungen wie *Peoria, Illinois (USA)*, vgl. (3.100)–(3.101). Ist es auf der Ebene der Semantik möglich, den Unterschied zwischen Koordination und Quasikoordination zu fassen, schwimmt er auf der Ebene der Syntax oder auf der orthographischen Ebene. In (3.145) stehen Komma und die Konjunktion *und*, aber vielleicht ist das auch ein Reflex davon, dass die Ortsangaben hier nicht nur einander spezifizieren, sondern sich tatsächlich überlappen.

Bestandteile einer Grammatik, die versucht die quasikoordinierten statischen Ortsangaben zu erkennen sind:

- Adverbien, inklusive Adverbialkombinationen, und „frozen adverbials“, siehe Kap. 3.5.3. Das schließt also Ausdrücke wie *einen Steinwurf entfernt* oder Höhenangaben (*auf 600m über dem Meeresspiegel*) ein.
- Präpositionalphrasen mit räumlicher Präposition und Toponym oder lokativem Nomen

Die Anordnung der Bestandteile ist grob ‘Adverb > frozen adverbial > PP (Nloc) > PP (Toponym)’ mit vielen Variationsmöglichkeiten und Wiederholungen:

(3.146) *hier im Süden um den Beinivørð an der steilen Westküste*

(3.147) *im Ultental auf 736 m Höhe*

(3.148) *auf herausragendem Platz hoch über der Isar nahe der Ampereimündung (Gemeinde Wang)*

Quasikoordination ist nicht nur bei statischen Ortsangaben zu beobachten, sondern auch bei direktionalen:

(3.149) *Im Sommer zog ich nach München in die Reisinger Straße.*

Doch liegt die eigentliche Herausforderungen bei der Beschreibung direktonaler Ortsangaben auf der Kombinatorik der einzelnen Pfadbestandteile.

3.6.2 Pfadangaben

(3.150) *I traveled from LA along Rt. 1 northwards 450 miles at 60 all the way to SF on foot.*
 source trajectory direction distance speed totality goal mode

Dieses Beispiel aus Ross (1995) zeigt, wie komplex eine Pfadangabe sein kann. Verschiedene semantische Komponenten einer Bewegung im Raum können beteiligt sein (siehe Kap. 2.1.2 und Fußnote 24 auf S. 24 zum Pfadkonzept).

Von einigen Linguisten werden Pfadangaben als einziges lokatives Argument von Bewegungsverben angenommen (vgl. S. 65–65). Gleiches gilt analog für die anderen Bewegungsprädikate: deverbale Nomina (3.152) und Nomina, die ein Objekt bezeichnen (3.153).

(3.151) a. *Äthiopiens Starläufer Haile Gebrselassie ist aus Addis Abeba in_{dat} Berlin angekommen.*

b. *Sie ist von Paris nach Amsterdam gefahren.*

(3.152) a. *Seit 1940, dem Jahr ihrer Ankunft aus Prag in Palästina, ...*

b. *Nach dem Frühstück, Abfahrt aus Seefeld nach Innsbruck*

(3.153) *Das Ufer entlang des Dorfes zierte eine alte Nussbaumallee.*

In welcher Form eine Pfadangabe realisiert wird, wird dabei wesentlich durch das Prädikat gesteuert. Pfadangaben sind deshalb einfacher zu modellieren, wenn man die einzelnen Pfadbestandteile als eigene Aktanten behandelt. Die Frage ist dann aber, wieviele Typen von Bestandteilen man ansetzt, und ob jeder Typ Aktantenstatus erhalten soll oder nicht. Semantisch, darin besteht kein Zweifel, gehören alle zu einer Bewegung dazu. Üblicherweise werden

drei Typen angenommen: Ausgangspunkt und Endpunkt, und den eigentlichen „Pfad“, die dazwischenliegenden Wegpunkte. Letzteren Typ unterteilen Eschenbach et al. (2000) in die Typen ‘course’ („intermediate place“: *Tim geht durch die Tür*) und ‘shape’ („shape of the path“: *Tim geht den Bach entlang*).

Unabhängig von der angesetzten Anzahl an Typen, ist zu beschreiben, wie diese zur komplexen Pfadangabe zusammengesetzt werden. Hier einige Beispiele:

- (3.154) *Er fährt von Jyväskylä via Tampere nach Helsinki.*
- (3.155) *Mit dem Schiff von Kos, Patmos, über Samothrake nach Kavalla zu den Meteora-Klöstern.*
- (3.156) *Abfahrt aus Swakopmund in südlicher Richtung entlang der Küste zwischen dem Meer und dem Wanderdünen Gürtel bis Walvis Bay*
- (3.157) *entlang der Donau von Ulm bis Regensburg*
- (3.158) *von St. Martin ausgehend den Gosaukamm entlang über Postalm und Osterhorngruppe bis Puch im Salztal*
- (3.154') *Von Jyväskylä fährt er via Tampere nach Helsinki.*

Die natürliche, logische Abfolge Source → Passage → Destination wird meist ((3.154), (3.155), (3.156) und (3.158)), muss aber nicht eingehalten werden (3.157). Aus kommunikativen Gründen (Thema-Rhema-Gliederung) können im Deutschen auch einzelne Pfadbestandteile abgetrennt und verlagert werden, z.B. ins Vorfeld wie in (3.154'). Teilweise entspricht aber auch die Abfolge Passage → Source → Destination der neutralen Wortfolge (3.157). Und schließlich wird aus (3.157) klar, dass die Beschreibung von Pfadangaben sehr komplex ist, und vermutlich weit mehr als 3–4 Typen von Bestandteilen verlangt.

4 Lokative Sätze

Nachdem in den beiden vorigen Kapiteln lokative Prädikate und Argumente besprochen wurden, behandelt dieses Kapitel Sätze, die mit diesen gebildet werden. Dem lexikongrammatischen Ansatz folgend werden nur einfache Sätze beschrieben. Allerdings sind einfache Sätze in realen (journalistischen) Texten außer in Überschriften kaum zu beobachten. Das führt zu einem Widerspruch im Hinblick auf die praktische Ausrichtung der Arbeit – die Grammatiken sollen ja zur Informationsextraktion einsetzbar sein. In zwei Punkten wird deshalb das Konzept des elementaren Satzes erweitert, um den Recall der Grammatiken zu erhöhen. (1.) soweit möglich lassen die Grammatiken adverbiale Einschübe (‘Inserts’, vgl. Kap. 4.2), insbesondere Ortsadverbiale zu. (2.) Elementare Sätze können in Formen transformiert werden, die keine selbständigen Sätze sind. Die Grammatiken versuchen deshalb auch Teilsätze, Chunks, Partizipalphrasen, Relativsätze, im Falle substantivischer Prädikate auch Nominalphrasen zu erkennen. Das erscheint auch deshalb nötig, weil die Möglichkeiten einen Satz allein durch Umstellung der Satzteile zu variieren im Deutschen groß sind (Kap. 4.1).

Mit den zugrundegelegten Klassifikationskriterien (Kap. 4.3) werden dann lokative Prädikate der Wortarten Präposition, Verb, Nomen und Adverb Satztypen zugeordnet und beschrieben. Tab. 4.2 am Ende des Kapitels zeigt die behandelten Satz schemata im Überblick. Die Tabellen der Lexikongrammatik sind mit Erläuterungen im Anhang (Kap. C) zu finden.

4.1 Lokale Grammatiken und deutscher Satzbau

Der deutsche Satzbau ist durch folgende Eigenschaften charakterisiert:

- zwei grundlegende Anordnungsmöglichkeiten der Bestandteile im Satz:
 - Verberststellung, im nicht eingebetteten Deklarativsatz, W-Fragesatz, oder ohne Konjunktion eingebetteten Satz
 - Verbzweitstellung im mit Konjunktion eingebetteten/subordinierten Satz oder im Fragesatz ohne Fragepronomen
- Satzklammer:
 - die linke Satzklammer enthält entweder (1.) ein finites Verb (in Sätzen mit Verbzweitstellung) oder (2.) die Konjunktion in subordinierten Sätzen
 - die rechte Satzklammer enthält in Abhängigkeit von der linken Klammer die weiteren Verbbestandteile, die noch nicht in der linken Klammer realisiert wurden, sowie die Infinitivpartikel und/oder Verbalpräfixe
- Stellungsfelder, topologische Felder: linke und rechte Satzklammer unterteilen die verbleibenden Bestandteile des Satzes in drei Felder – Vor-, Mittel-, Nachfeld. Das Vorfeld links vor der linken Klammer enthält (meist) nur eine Konstituente, das Mittelfeld zwischen linker und rechter Klammer enthält beliebig viele Konstituenten, das Nachfeld rechts von der rechten Klammer kann extraponierte Konstituenten enthalten.

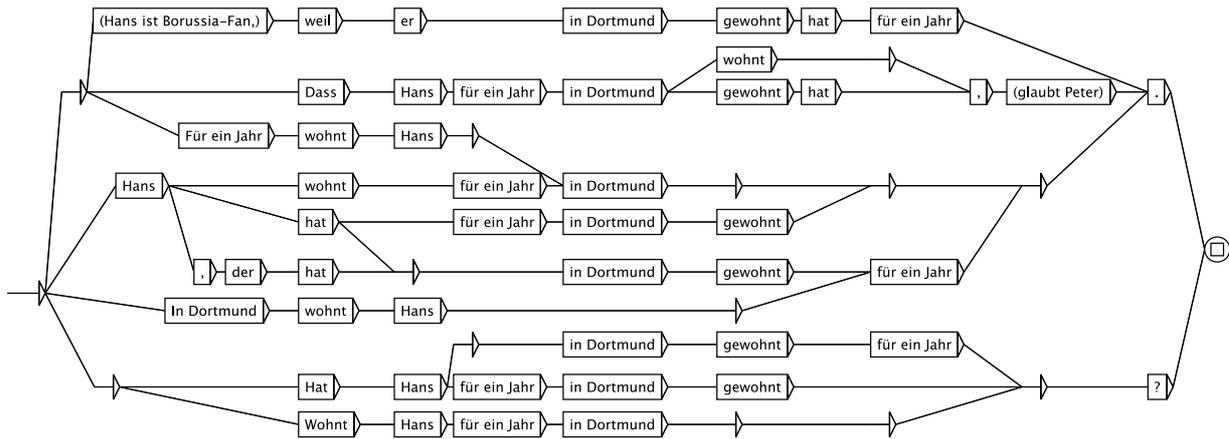


Abbildung 4.1: Alle Sätze aus Tab. 4.1 als lokale Grammatik (Textautomat).

- nur die Verbalklammer ist in ihrer Stellung fix, andere Konstituenten (aktantische NPs wie Adverbialsupplemente) können nahezu beliebig permutiert werden

Diese sehr knapp gehaltene Charakteristik¹³⁵ soll hier genügen. Wichtig sind aber die Konsequenzen für die Beschreibung deutscher Sätze mittels lokaler Grammatiken. Zum einen bedingen topologische Felder nicht-lokale Abhängigkeiten innerhalb des Satzes, zum anderen sind Permutationen mit lokalen Grammatiken nur schwer zu beschreiben (vgl. die Diskussion ab S. 10). Versucht man alle Realisierungsmöglichkeiten eines Satzes (siehe Tab. 4.1) in einer lokalen Grammatik zu erfassen, wird der Graph (Automat) verhältnismäßig groß, obwohl bei weitem noch nicht alle möglichen Varianten realisiert sind (Abb. 4.1). Die einzelnen Pfade lassen sich nicht zusammenfalten. Die beiden Satzklammern sind voneinander abhängig, ohne nebeneinander zu stehen. Im Mittelfeld zwischen den beiden Klammern ist der Großteil des Materials akkumuliert. Immerhin, Stellungsregularitäten von Aktanten und Supplementen im Mittelfeld schränken die Zahl der möglichen Permutation deutlich ein.

4.2 Adverbialsupplemente (Inserts)

Adverbialsupplemente¹³⁶ sind, da sie die Struktur eines elementaren Satzes kaum verändern, leicht zu beschreiben. Voraussetzung ist dabei, dass die interne Struktur der Adverbiale selbst bereits beschrieben ist. Dass sich dafür lokale Grammatiken gut eignen, wurde bereits in Kap. 1 angesprochen. Beschrieben werden nur Ortsadverbiale, Temporaladverbien u.ä. fallen nicht in den Fokus dieser Arbeit. Zugute kommt, dass die interne Struktur von Adverbialkomplementen und -supplementen sich nicht unterscheidet. Beide können also durch ein und dieselben Grammatiken beschrieben werden.

Die Stellung der Adverbialsupplemente wird durch zwei Prinzipien gesteuert:

¹³⁵ Siehe Höhle (1986), Lang & Zifonun eds. (1996), Pasch et al. (2003: 64–75), Berman (2003: Kap. 3) für einen allgemeinen Überblick, Müller (1999), Müller & Ule (2001), Klatt (2004), Hinrichs (2005), Gerdes & Kahane (2001, 2006, 2007), Clément, Gerdes & Kahane (2002), Telljohann et al. (2006) für Beispiele einer formalen Beschreibung und computerlinguistischen Modellierung.

¹³⁶ In der Tradition von Harris mit anderen Arten von Einschüben auch ‚Inserts‘ genannt, siehe S. 14.

Satztyp	Vorvorfeld	Vorfeld	linke Satzkl.	Mittelfeld	rechte Satzkl.	Nachfeld
Verbzweitsatz		Hans	wohnt	für ein Jahr in Dortmund.		
		Hans	hat	für ein Jahr in Dortmund	gewohnt.	
		Hans	hat	in Dortmund	gewohnt	für ein Jahr.
	Hans,	der	hat	in Dortmund	gewohnt	für ein Jahr.
		In Dortmund Für ein Jahr	wohnt wohnt	Hans für ein Jahr. Hans in Dortmund.		
Verberstsatz			Wohnt	Hans für ein Jahr in Dortmund?		
			Hat	Hans für ein Jahr in Dortmund	gewohnt?	
			Hat	Hans in Dortmund	gewohnt	für ein Jahr?
Verbletztsatz			Dass	Hans für ein Jahr in Dortmund	wohnt,	(glaubt Peter).
			Dass	Hans für ein Jahr in Dortmund	gewohnt hat,	(glaubt Peter).
	(Hans ist Borussia-Fan,)		weil	er in Dortmund	gewohnt hat	für ein Jahr.

Tabelle 4.1: Topologische Felder und einige Varianten des Satzes *Hans wohnt für ein Jahr in Dortmund.*

1. es besteht ein Zusammenhang zwischen dem Aktantenstatus eines Adverbials und seiner Stellung: ein valenzgebundenes steht hinter einem freien Adverbial (siehe auch S. 34):

- (4.1) a. *Die Ausbildung zum staatlich geprüften Modegestalter dauert in München zwei Jahre.*
 b. *Max wohnt zwei Jahre in München.*

2. in Abhängigkeit vom semantischen Typ des Adverbialsupplements stehen von links nach rechts: Zeit- < Orts- < modale Angaben (vgl. Haider 2002: 68):

- (4.2) a. *Max hat zwei Jahre in München gearbeitet.*
 b. *Max war heute in München beim Einkaufen.*

! diese Regeln gelten so nicht mehr, wenn einzelne Adverbiale rhematisch hervorgehoben werden, d.h. wenn der Satz eine nicht-neutrale Wortstellung aufweist.

4.3 Kriterienkatalog zur Klassifikation lokativer elementarer Sätze

Die folgenden Prinzipien wurden der Klassifikation der elementaren Sätze zugrundegelegt, Ergebnis sind eine Reihe von Satz schemata, die die Klassifikation unmittelbar widerspiegeln.

0. bestimme Prädikat bzw. die Wortart des Prädikats (vgl. Kap. 2.3)

1. bestimme die Stelligkeit des Prädikats und beschreibe für jeden Aktanten N

a) dessen syntaktische Realisierung, d.h.

i. den (obligatorischen) Kasus (z.B. $N_{x_{akk}}$) oder

ii. die regierte Präposition als

A. *Prep* – stark regierte Präposition

B. *Loc_{in}*, *Loc_{ad}*, *Loc_{ab}* – durch N determinierte Präposition, im Sinne der lexikalischen Funktionen *Loc_{in}* usw. (S. 28)

C. *Loc_{stat}*, *Loc_{dir}*, *Loc_{dst}*, *Loc_{src}*, *Loc_{pass}* – durch das Prädikat im semantischen Typ bestimmte Präposition (statisch, direktional: auf Ausgangs-, Ziel-, Wegpunkt fokussiert)¹³⁷

* in den Fällen B und C kann die Präpositionalphrase durch Adverbien oder Adverbialkomplexe substituiert werden: *Max wohnt hier*, *Max fährt dorthin*

137 Ein kurzer Kommentar zur Realisierung der Präposition als *Loc*: Sowohl Lexikongrammatiker (siehe S. 46), aber auch andere vgl. Schnorbusch 1999: 165) haben erkannt, dass für die meisten lokativen Verben die Präposition, die das lokative Argument markiert, nicht stark regiert ist, sondern variiert werden kann und auch mit zur Semantik des Satzes beiträgt. Alle lösen dieses Problem, indem die Präposition in der Beschreibung als Typ 'räumliche Präposition' (hier *Loc*) angegeben wird. Diese muss (im Deutschen) weiter spezifiziert werden in *Loc_{stat}*, *Loc_{dst}*, *Loc_{src}*, *Loc_{pass}*. Eine Kombination wie *Loc N₁dst* bei den französischen Lexikongrammatikern ist nicht fein genug, da (1) im Kontrast zu (2) beschrieben werden muss:

(1) *Er kommt in München an.*
 N_0 V *Loc_{stat}* *Ndest* *VPart*

(2) *Er fährt nach München.*
 N_0 V *Loc_{dst}* *Ndest*

** in B kann sie nur durch ein Präpositionaladverb ersetzt werden, das dieselbe Präposition enthält: *Max steuert (auf die Insel + darauf) zu*

b) die Objektklasse des Arguments, z.B. N_{loc} =Ort, N_{dest} =Zielort, N_{hum} =Mensch

Die folgenden Kriterien können klassenbildend sein, also ins Satzschema einfließen, oder ein zu beschreibendes Merkmal sein, d.h. sie füllen eine Spalte in der Grammatik:

c) ein trennbares Verbpräfix: $N_0 V Loc_{ad} VPart$: *ankommen*

d) ein reflexives Verb: $N_0 V (Refl) Loc N_1 dest$: *sich begeben*

e) expletives *es*: *es geht nach Frankreich*

f) weitere, nicht fokussierte Pfadb Bestandteile (Wegpunkte) und ihre syntaktische Realisierung

2. alle weiteren Eigenschaften werden als Merkmale in den Spalten der Tabelle beschrieben

Dieser Kriterienkatalog beschreibt in erster Linie die Klassifikation lokativer Verben, für andere Wortarten sind Abweichungen unvermeidbar, z.B. wird die Kasusreaktion als Merkmal in der Tabelle der Präpositionen geführt, als Kriterium würde sie die Präpositionen auf viele verschiedene Satzschemata aufteilen, was nicht erwünscht ist.

4.4 Präpositionen

(Räumliche) Präpositionen wurden bereits mehrfach in dieser Arbeit diskutiert (Kap. 2.4 und 3.5.2). Thema sind hier elementare Sätze mit einer Präposition als „Kern“-Prädikat, das eine statische räumliche Relation zwischen N_0 und N_1 ausdrückt. Das zugrundeliegende Satzschema ist deshalb:

$N_0 V_{sup} LOC_{stat} N_1 loc$

(4.3) *Max ist in Budapest.*

(4.4) *München liegt an der Isar.*

Das Stützverb hat den Charakter einer Kopula. Die Auswahl von V_{sup} ist durch N_0 (bzw. dessen Klasse) determiniert. N_0 ist zwar semantischer Aktant der Präposition, kann aber nie als syntaktischer Aktant der Präposition realisiert sein. Deshalb ist N_0 auch nicht Gegenstand der Lexikontabelle.

Nach ihrem semantischen Typ werden die Präposition auf statische beschränkt. D.h. Verben wie *sich begeben*, *fahren*, *gehen* in Kombination mit einer direktionalen Präpositionen werden als Voll- und nicht als Stützverben beschrieben. Allerdings können direktionale Präpositionen mit Fokus auf dem Wegverlauf häufig auch statisch verwendet werden – die „statische Lesart“ fand dann Aufnahme in die Lexikontabelle:

(4.5) *Die Flussfahrt abwärts des großen Stroms Ayeyarwaddy beginnt abenteuerlich und unvergesslich romantisch gegen 5 Uhr morgens.* (direktional)

(4.6) *Die Stadt Dnepropetrovsk liegt in der östlichen Ukraine, von Kiew aus 500 km abwärts des Flusses Dnepr.* (statisch)

Im elementaren Satz kann die Präpositionalphrase durch ein Ortsadverb ersetzt werden (*Max ist hier*). Oder sie kann um weitere Präpositionalphrasen bzw. Ortsadverbiale ergänzt werden (quasikoordiniert, siehe Kap. 3.6.1). Deshalb wurde folgendes Satzschema auch als lokale Grammatik beschrieben:

N_0 V_{sup} $Advloc_{stat}$

(4.7) *Das Haus steht dort drüben.*

(4.8) *Die Frauenkirche steht hier im Zentrum der Münchner Altstadt.*

(4.9) *Die Flussmündung der Isar liegt auf 312 Metern über Normalnull.*

(4.10) *Kremmen liegt am südlichen Rand des Eberswalder Urstromtales auf dem Glin.*

Zwischen einfachen und komplexen Präpositionen wird in der Tabelle nicht unterschieden. Präpositionen wie *nahe*, *gegenüber*, *fernab*, *flussaufwärts*, die auch adverbial, d.h. ohne regiertes Nomen, gebraucht werden können, wurden in die Tabelle mit aufgenommen und durch ein Merkmal ‘ADV’ markiert (vgl. die Diskussion ab S. 38 zur Unterscheidung zwischen Präposition und Adverb).

Die Lexikontabellen mit einer detaillierten Erklärung zu den beschriebenen Merkmalen sind in Kap. C.1 auf S. 361 zu finden.

4.5 Verben

Drei große Gruppen von Verben werden in diesem Kapitel behandelt: Bewegungsverben und Verben, die entweder eine Relation Mensch – Ort oder eine Relation zwischen zwei Orten zum Ausdruck bringen.

4.5.1 Bewegungsverben

Bewegungsverben sind die mit Abstand größte Gruppe lokativer Verben. Sie werden als eine Gruppe behandelt. Zwischen ‘verbs of locomotion’ und ‘verbs of change of position’ (vgl. Eschenbach et al. 2000) wird nicht unterschieden. Da die Bewegungsverben insgesamt sehr viele Gemeinsamkeiten aufweisen, wurden alle untenstehenden Satz schemata in einer Tabelle (Tab. C.3 auf S. 395) zusammengefasst. Kriterien zur Einteilung und Merkmale bei der Beschreibung sind:

- für das bewegte Objekt (Figure):
 - In welcher Position steht das bewegte Objekt? Dies führt zur Unterteilung in zwei große Gruppen mit Figure
 - * als syntaktischem Subjekt
 - * in der Position des syntaktischen Objekts
 - Welche Objektklassen sind als Figure möglich? Z.B. LEBEWESEN, VERKEHRSMITTEL, oder auch ganz spezifische Klassen wie SCHIFFE.
- die deutschen Bewegungsverben unterscheiden sich vor allem darin, wie sie den Pfad der Bewegung (siehe Kap. 2.5.2.3 und 3.6.2) kodieren.

- Welchen Pfadb Bestandteil fokussiert ein Verb?
- Falls ja, welche Pfadbbestandteile sind blockiert?
- Die syntaktische Realisierung des fokussierten Bestandteils ist durch das Verb determiniert.
- Mit deiktischen Adverbien präfigierte Verben können auch den fokussierten Pfadbbestandteil „inkorporieren“ (z.B. *heimgehen*).
- Wenn nicht explizit durch das Merkmal *Pfad=∅* ausgeschlossen, wird angenommen, dass die nicht fokussierten Pfadbbestandteile immer über Präpositionen oder als Adverbiale angeschlossen werden können. Nicht möglich ist die syntaktische Realisierung z.B. bei *stranden*.
- Die Entscheidung, welcher Pfadbbestandteil fokussiert ist, ist allerdings oft schwer zu treffen und zwischen Sprechern umstritten.

4.5.1.1 Figure (bewegtes Objekt) als Subjekt

Bewegungsverben, die das bewegte Objekt als Subjekt realisieren, bilden die weitaus größte Gruppe. Die Vielzahl an möglichen Satz schemata resultiert aus den zahlreichen Realisierungsmöglichkeiten des oder der lokativen Aktanten.

$N_0 \text{ V } N_1 \text{src}_{\text{akk}}$

(4.11) *Max verlässt das Zimmer.*

(4.12) *Wir ließen die Insel zurück.*

$N_0 \text{ V } N_1 \text{dest}_{\text{akk}}$

(4.13) *Max betritt das Zimmer.*

(4.14) *Max erklimmt den Ortler.*

(4.15) *Das Boot erreicht die Insel.*

Diesen beiden Gruppen von Bewegungsverben stehen einige Inchoativa nahe, die semantisch eine Bewegung implizieren. So fällt die Grenzziehung oft schwer, z.B. für (ein Land) *besiedeln*, (einen Platz) *besetzen*, (ein Zimmer) *beziehen*.

$N_0 \text{ V } N_1 \text{pass}_{\text{akk}}$

(4.16) *Max quert ein Geröllfeld.*

(4.17) *Max durchwandert den Wald.*

Vor allem präfigierte Verben sind in dieser Gruppe enthalten, und zwar mit folgenden Präfixen: *ab-* (*die Strecke abschreiten*), *be-* (*eine Straße befahren*), *durch-* (s.o.), *entlang-* (*die Küste entlangsegeln*), *über-* (*die Straße überqueren*) und *um-* (*die Baustelle umfahren*).

$N_0 \text{ V (Refl) } N_1 \text{dest}_{\text{dat}}$

(4.18) *Ein Taifun nähert sich der Insel Kyushu.*

(4.19) *Der Asteroid kommt der Erde ungemütlich nahe.*

Zu dieser Gruppe gehört vor allem das Verb *sich nähern*. Auch *naherkommen* und *näherkommen* können dazugerechnet werden, drücken aber häufiger eine nicht-räumliche Nähe aus.

N_0 V N_1 pass_{dat}

(4.20) *Die Wanderer folgten dem Tal.*

(4.21) *Die Forscher folgten dem Fluss bis zur Quelle.*

Nur das Verb *folgen* gehört zu dieser Gruppe. Ungeachtet der Nähe zu anderen Bedeutungen („Lesarten“) muss es hier als Bewegungsverb betrachtet mit einem obligatorischen Objekt, das den Wegverlauf angibt. Weitere lokative Aktanten sind fakultativ zur Angabe von Ausgangs-, Ziel- und Wegpunkten. Sie werden durch Präpositionen angeschlossen (4.21).

N_0 V Prep N_1 dest (VPart)

(4.22) *Das Auto rast auf die Ampel zu.*

(4.23) *Das Schiff steuert auf den Leuchtturm zu.*

(4.24) *Anna kommt auf Max zu.*

Die Gruppe ist dadurch gekennzeichnet, dass die Präposition allein vom Verb bestimmt ist und nicht variiert werden kann. Dies kann mit dem Fragetest überprüft werden. Die Abfrage ist in Abhängigkeit von N_1 entweder durch die Kombination aus Präposition und Fragepronomen möglich, oder auch durch das Präpositionaladverb, das die Präposition inkorporiert. Durch ein Adverb (*wohin*) kann N_1 nicht abgefragt werden:

(4.25) * *Wohin kommt Anna zu?*

(4.26) *Auf wen kommt Anna zu? – Auf Max.*

(4.27) * *Worauf kommt Anna zu?*

(4.28) *Worauf steuert das Schiff zu? – Auf den Leuchtturm.*

(4.29) ? *Auf was steuert das Schiff zu?*

Dabei scheint das Verbpräfix die Präposition zu bestimmen. Bei Präpositionen mit doppeltem Kasus wird die direktionale Variante (mit Akkusativ) gewählt. Es sei denn, die Präposition lautet *in*. Hier ist auch der Dativ möglich (4.30).

(4.30) *Die Kompanie rückte in die+der Stadt ein.*

(4.31) *Die Brüder reisten nach+in Deutschland ein.*

Ist N_1 ein Ortsname, der ohne Artikel steht, steht statt direktonalem *in* die Präposition *nach* (4.31).

N_0 V Prep N_1 src (VPart)

(4.32) *Ein ICE-T fährt aus dem Augsburger Hbf aus.*

Nur *ausfahren* kann ausschließlich mit einer Präposition gebraucht werden. Andere Verben mit gleichem Präfix (*auslaufen*) lassen auch andere Präpositionen zu und werden deshalb unter $N_0 V Loc_{src} N_1src$ klassifiziert (4.43).

$N_0 V Prep N_1pass$ (VPart)

- (4.33) *Shermans Divisionen setzten über den Fluss.*
(4.34) *Die 184. sibirische Schützendivision unter Oberst Batjuk setzt über die Wolga über und geht unterhalb des Mamaihügels in Stellung.*
(4.35) *Wir fahren an Weißenkirchen vorbei, passieren Dürnstein ...*
(4.36) *Wir segelten an Saßnitz vorbei durch die Prorer Wiek.*

Der Fokus liegt bei diesen Verben klar auf dem passiertem Wegpunkt (N_1pass), der deshalb als maßgeblicher Aktant angenommen wird. Die Präposition für N_1pass ist dabei nicht variabel, sondern durch das Verb vorgegeben.

Die folgenden drei Schemata sind Varianten von $N_0 V Loc N_1$ und unterscheiden sich nur darin, welchen Teil des Pfades sie fokussieren: Ausgangspunkt (*src*), Wegpunkt (*pass*) oder Zielpunkt (*dest*). Die Abgrenzung ist mitunter schwierig, im Zweifelsfall wurde das Kriterium von Panevová (S. 32) herangezogen. Für das Paar *einwandern/auswandern* beispielsweise liefert es eine klare Entscheidung: gebraucht der Sprecher *einwandern*, muss er wissen, welches das Zielland der Migration ist. Im Fall von *auswandern* kann das Zielland unbekannt sein, nicht jedoch das Ausgangsland.

$N_0 V (Refl) Loc_{dst} N_1dest$ (VPart)

- (4.37) *Max fährt an die Ostsee.*
(4.38) *Max begibt sich ins Haus.*
(4.39) *Alexander bricht nach Persien auf.*

Hier wie auch bei den folgenden Verbgruppen ist die Präposition variabel und kann durch andere Präpositionen des gleichen Typs ersetzt werden, womit meist eine Bedeutungsveränderung verbunden ist.

$N_0 V (Refl) Loc_{src} N_1src$ (VPart)

- (4.40) *Max stieg vom Baum herab.*
(4.41) *Max stahl sich aus dem Saal davon.*
(4.42) *die Großeltern, die in den dreißiger Jahren aus Deutschland auswanderten, ...*
(4.43) *Die Fregatte läuft von Wilhelmshaven Richtung Libanon aus.*

Die Variabilität direktonaler Präpositionen, die den Ausgangspunkt der Bewegung fokussieren, ist dadurch eingeschränkt, dass es nur die zwei einfachen Präpositionen *aus* und *von* gibt. Alle komplexen Präpositionen dieses Typs enthalten außerdem eine der beiden (*aus Richtung, von ... her*).

N_0 V (Refl) Loc_{pass} N_1pass (VPart)

(4.44) *Max flanierte durch die Boulevards von Paris.*

(4.45) *Er rast unter der Brücke durch.*

N_0 V Loc_{stat} N_1dest

(4.46) *Max kommt in München an.*

(4.47) *Das Schiff legt am Kai an.*

(4.48) *Die Reisegruppe traf im Hotel ein.*

Diese Verben schließen den Zielpunkt über eine statische Präposition an, und legen dabei einen starken Fokus auf den Zielpunkt. Die Realisierung von Ausgangs- oder Wegpunkt ist aber möglich:

(4.49) *..., die aus den Vereinigten Staaten in Europa ankamen.*

(4.50) *Der FC Nania wird am Mittwoch aus Ghana in Zürich eintreffen ...*

N_0 V $Loc_{src/stat}$ N_1src

(4.51) *Max fährt vom Bahnhof ab.*

(4.52) *Der Bus fährt an der Brücke ab.*

Auch der Ausgangspunkt der Bewegung wird von einigen Verben über eine statische Präposition angebunden, diese kann aber mit der direktionalen alternieren. Entsprechend ist auch die Situation beim Fragetest, hier in Gegenüberstellung zur Gruppe der *ankommen*-Verben (4.48)–(4.48):

(4.53) a. *Wo fährt der Zug ab?*

b. *Von wo fährt der Zug ab?*

(4.54) a. *Wo kommt der Zug an?*

b. * *Wohin kommt der Zug an?*

Aus dieser Alternation resultiert die Beschreibung durch ein Satzschema mit variablem Typ der Präposition ($Loc_{src/stat}$).

N_0 V (Loc N_1dest) V_{pp}

(4.55) *Max kommt angelaufen.*

(4.56) *Peter kommt nach Hause gehinkt.*

Konstruktionen mit *kommen* und dem Partizip II von oft mit *an-* präfigierten Bewegungsverben werden hier nicht behandelt. *Kommen* kann auch als Modal- oder Auxiliärverb betrachtet werden, Duden2006: §577 gebraucht als Terminus ‚infinittregierendes Verb‘.

4.5.1.2 Figure als syntaktisches Objekt

$N_0/loc V N_1$

(4.57) *Der Wald verschluckt Max.*

Verben dieser Klasse mit dem bewegten Objekt (Figure) in der Position des syntaktischen Objekts (vgl. Tabelle 38L0 der französischen Lexikongrammatik) sind in Texten eher selten und wurden deshalb hier nicht beschrieben.

4.5.1.3 Subjektlose Sätze

$Es V LocDst N_0dest$

(4.58) *Es geht nach Frankreich!*

Auch dieser wie der folgende elementare Satz stellt eher ein Randphänomen dar.

$Es V LocSrc N_0src$

(4.59) *Mit dem Zug geht's hinaus aus Passau.*

4.5.2 Transportverben und 'local causatives'

Transportverben und 'local causatives', die dadurch charakterisiert sind, dass sie neben Figure und Ground auch noch ein Agens realisieren, dass Verursacher der Bewegung ist, aber selbst nicht an der Bewegung beteiligt sein muss, werden hier nicht behandelt. Die Erweiterung um einen Aktanten bedeutet zusätzlichen Aufwand beim Erstellen der Grammatiken. Mit den Arbeiten von Caroli (siehe Kap. 2.5.1.2) existiert bereits eine Beschreibung.

4.5.3 Verben mit Infinitivergänzung

$N_0 V (Loc N_1) V_{inf} W$

(4.60) *Max geht (in den Supermarkt) Milch kaufen.*

(4.61) *Max kommt Peter besuchen.*

Verben mit Infinitivergänzung und zusätzlichem lokativem Komplement wurden von Gross (1975:165–8) in Tab. 2 beschrieben. Im Deutschen ist diese Gruppe Verben kleiner als im Französischen, es gelten außerdem deutliche Beschränkungen welche Propositionen in der Infinitivkonstruktion formuliert werden können. In den meisten Fällen muss der Infinitiv um die Partikel *zu* erweitert werden und mit einer Konjunktion angeschlossen werden (vgl. auch die Übersetzungen von (2.96) und (2.97)):

- (4.62) *Max traverse la rivière retrouver ses copains.*
 ‚Max überquert den Fluss, um seine Schulkameraden wiederzufinden.‘

Um die Infinitivergänzungen von anderen Lokalgänzungen zu unterscheiden, erweitert Caroli (1984a: 2) den Fragetest in (4.63) zu (4.64). Er zeigt damit auch, dass durch die Infinitivergänzung die Intention des Subjekts und nicht etwa eine Lokalgänzung ausgedrückt wird.

- (4.63) *Hans läuft den Arzt holen.*
Wohin läuft Hans?
Den Arzt holen.

- (4.64) *Wohin läuft Max?*
Ins Dorf.
Warum? / Wozu?
Den Arzt holen.

Wirklich üblich sind nur wenige Kombinationen von Bewegungsverb (meist *gehen*) und Infinitiv, z.B. *einkaufen gehen* und Synonyme (*Milch holen gehen*), ‚zum Sport gehen‘ (*joggen gehen*, *schwimmen gehen*, *radfahren gehen*), usw. *Gehen* hat hier eher eine modale Funktion, und nicht die eines echten Bewegungsverbs, es kann auch kaum durch andere Bewegungsverben ersetzt werden (*?joggen fahren*, *?Max radelt eine Milch holen*).

Für den Linguisten ist diese Gruppe der Verben sicher eine der interessantesten. Aus praktischen Erwägungen werden sie jedoch in dieser Arbeit tabellarisch nicht erfasst.

4.5.4 Mensch – Ort statisch

Eine statische Relation zwischen Mensch und Ort wird durch eine verhältnismäßig kleine Gruppe von Verben ausgedrückt. Zu ihnen gehören aber prototypisch lokative Verben (*wohnen*), weshalb der Gruppe eine eigene Tabelle gewidmet ist (Tab. C.4 auf S. 409). Die Verben verteilen sich dabei auf die folgenden Satzchemata:

N_0 V N_1 *loc*_{akk}

- (4.65) *Die Meyers bewohnten ein herrschaftliches Haus in Washington.*
 (4.66) *Zehntausende Touristen bevölkern Hotels und Campingplätze.*

Diese mit *be-* präfigierten Verben werden tendenziell eher mit einer Gruppe von Menschen (*NhumColl*) als erstem Aktanten verwendet. Einige der Verben (*besiedeln*) können auch inchoativ verwendet werden und kommen dann in der Bedeutung den Bewegungsverben, vgl. S. 159, nahe.

N_0 V *Loc*_{stat} N_1 *loc*

- (4.67) *Max wohnt in Hamburg.*
 (4.68) *Reinhold Messmer biwakiert auf dem Gipfel des K2.*
 (4.69) *In Frankfurt findet heute die Hauptversammlung der Dresdner Bank statt.*

Satz (4.67) stellt den Prototyp eines lokativen Verbs dar, vgl. Kap. 2.5.3.1. Das Ortsargument wird bei dieser Gruppe über eine schwach regierte Präposition angeschlossen. Die Präposition kann deshalb variieren. Sie drückt einen Teil der räumlichen Relation aus, nämlich die räumliche Konstellation. Falls das Verb selbst nicht zur Bedeutung beiträgt, muss es als Kopula oder Stützverb betrachtet werden. Die Abgrenzung ist aber nicht immer eindeutig: so unterscheiden sich die Verben *sich befinden* und *bleiben* nur in ihrer aspektuellen Komponente. Der Unterschied zu Verben wie *wohnen*, deren Bedeutung klar mit Konzepten wie ‚Wohnung‘ und ‚Wohnsitz‘ verbunden ist, bleibt aber deutlich. Die möglichen Stützverben wurden aber mit in Tab. C.4 aufgenommen und durch V_{sup} gekennzeichnet.

$N_0 V Loc_{src} N_1/loc$

(4.70) *Max stammt aus Hamburg.*

(4.71) *Die Sängerin Lura stammt von den kapverdischen Inseln vor Afrikas Westküste.*

Die Präposition kann hier nur zwischen *aus* und *von* variieren, vgl. die Bewegungsverben des Schemas $N_0 V Loc_{src} N_1/src$ (161).

$N_0 V Loc_{stat} N_1/stat) V_{pp}$

(4.72) *Max liegt in Adelshausen begraben.*

Diese Konstruktion ähnelt sehr den Verben des Typs *angelaufen kommen* (siehe S. 162). Der modale Charakter von *liegen* ist aber noch offensichtlicher, da das Verb durch *sein* substituiert werden kann:

(4.72') *Max ist in Adelshausen begraben.*

Liegen erscheint damit eher als Kopula des Adjektivs/Partizips, *begraben liegen* wurde deshalb auch in der Tabelle der Adjektive (Tab. C.7) behandelt.

4.5.5 Relationen zwischen Orten (Verben)

Relationen zwischen Orten werden durch eine syntaktisch erstaunlich vielfältige Menge von Verben ausgedrückt. Auch einige dreistellige Relationen sind darunter. Semantisch drücken diese Verben (1.) eine räumliche Relation aus, d.h. sie beschreiben die räumliche Lage, Ausdehnung, oder den Verlauf einer geographischen Entität, oder (2.) sie sagen etwas über die Beziehung zwischen (Fließ)Gewässern aus, d.h. welches Gewässer Zu- bzw. Abfluss eines zweiten Gewässers ist. Beispiele sind auch in Tab. C.5 zu finden, die versucht, etwa 40 dieser Verben als Lexikontabelle zu beschreiben.

$N_0/loc V N_1/loc_{akk}$

(4.73) *Die Isar durchfließt das Stadtgebiet von München.*

(4.74) *Das Königsbrunnchen speist einen der Weiher im Stadtwald.*

(4.75) *Die Brücke von Andau quert den Einser Kanal.*

- (4.76) *Die Freie Hansestadt Bremen ist mit 404 km² das kleinste Land Deutschlands und umfasst die beiden 60 km voneinander entfernten Städte Bremen und Bremerhaven.*

N₀loc V N₁loc_{dat} (VPart)

- (4.77) *Hier fließt die Samara der Wolga von links zu.*
(4.78) *Die Straße folgt der Donau.*

N₀loc V (Refl) Prep N₁loc (VPart)

- (4.79) *Die Isar fließt in die Donau.*
(4.80) *Die Weser speist sich aus den Flüssen Werra und Fulda.*
(4.81) *Der Kanton Schwyz grenzt an die Kantone Uri, Glarus, St. Gallen, Zürich, Zug und Luzern.*
(4.82) *Die Picardie schließt sich nordöstlich an die Normandie an.*
(4.83) *Der Schwarze Drin entspringt aus dem Ohridsee.*

N₀ V Loc_{stat} N₁loc

- (4.84) *Die Elbe entspringt im Riesengebirge.*

N₀loc V (Refl) Loc_{dst} N₁loc

- (4.85) *Das Weideland erstreckt sich bis zum Flussufer.*
(4.86) *Das Grundstück reicht bis zur Straße. (aus Pittner 1999: 73)*
(4.87) *Das französische Departement Alpes Maritimes reicht von der Côte d'Azur ins alpine Hinterland.*

Dieses Satzschema, wie auch das folgende und einige der Beispiele oben, ist der 'fictive motion' (Kap. 2.5.2.4) geschuldet und ähnelt den Schemata der Bewegungsverben. Der Ausdruck des „Pfades“, also der als 'fictive motion' beschriebene Verlauf oder die Ausdehnung des räumlichen Objektes kann mehr als nur einen Bestandteil aufweisen. Wie auch bei der Beschreibung der Bewegungsverben wird nur der markierte und/oder obligatorische Pfadbbestandteil als Aktant im Satzschema angegeben.

N₀loc V (Refl) Loc_{pass} N₁loc

- (4.88) *Der Radweg verläuft vorbei am Ameisberg, ...*
(4.89) *Von der Quelle am Rothaarkamm im Siegerland schlängelt sich die Sieg rund 140km zwischen dem Bergischen Land und dem Westerwald entlang bis zu ihrer Mündung in den Rhein bei Niederkassel-Mondorf.*

N₀loc V N₁loc Prep N₂loc

- (4.90) *Die kurze Buna entwässert den Shkodrasee in die Adria.*

(4.91) *Durch den Sankt-Lorenz-Strom werden die Großen Seen in den Atlantik entwässert.*

N_0loc und $V Prep N_1loc Prep N_2loc$

(4.92) *Der Tigris vereinigt sich mit dem Euphrat zum Schatt al-Arab, der in den Persischen Golf mündet.*

- (4.93) a. *Die Zwickauer Mulde vereinigt sich mit der Freiburger Mulde zur Mulde.*
b. *Die Zwickauer Mulde und die Freiburger Mulde vereinigen sich zur Mulde.*

Eine Besonderheit dieses Schemas ist, dass der zweite Aktant (N_1) auch in die syntaktische Subjektposition verschoben werden kann (4.93b).

4.6 Nomina

Als Nomina, die als lokative Prädikate in Frage kommen, wurden in Kap. 2.6 (deverbale) Nomina der Fortbewegung und des Aufenthalts und eine Reihe quasiprädikativer Nomina genannt. Letztere drücken entweder eine Relation zwischen Mensch und Ort, oder zwischen zwei Orten aus. Auch Verkehrsmittel und -wege wurden als Quasiprädikate genannt, sie werden hier aber nicht durch eine Lexikongrammatik beschrieben.

Auch wenn auch für die Nomina der elementare Satz Ausgangsbasis der Beschreibung ist, so liegt der Fokus der Grammatiken auf der Nominalphrase und ihrer Struktur. Gemeinsam ist allen Tabellen der Nomina, dass die Anschlussmöglichkeiten von Nomen und zweitem Aktanten (N_1) zahlreich sind (vgl. Kap. 3.5.1.5). Prinzipiell existieren folgende Varianten (§ bezeichnet die Position des prädikativen Nomens):

§ N_{1gen} Kasusreaktion Genitiv, in Prä- und Poststellung: *die Hauptstadt Frankreichs, Frankreichs Hauptstadt*

§ von N_1 Präposition *von* mit Dativ: *die Hauptstadt von Frankreich,*

§ $Prep N_1$ Anschluss über eine zu spezifizierende Präposition: *Die Müglitz ist ein linker Nebenfluss zur Elbe*

N_{1adj} § Realisierung von N_1 als abgeleitetes Adjektiv: *die französische Hauptstadt, der französische Präsident*

$N_1\#\S$ Realisierung von N_1 als Kompositaerstglied (Fugenform): *am Ufer des Amazonasflusses Huallango, Usedom ist eine Ostseeinsel*

N_1 -§ Anschluss von N_1 durch Bindestrich: *DDR-Hauptstadt, die Themse-Mündung*

Der erste Aktant (N_0) wird entweder über ein Stützverb/Kopula angebunden, bei quasiprädikativen Nomina kann er auch in appositiver Stellung realisiert sein, vgl. (4.101).

4.6.1 Nomina der Fortbewegung und des Aufenthalts

Die Mehrzahl der Nomina, die Bewegungen oder den Aufenthalt an einem Ort ausdrücken, sind deverbale, von den entsprechenden Bewegungsverben abgeleitete Nomina. Dazu kommen doch einige Nomina wie *Ausflug* und *Exkursion*, für die keine verbalen Pendant existieren. Naheliegender ist, die deverbalen Nomina ausgehend von der Beschreibung des Verbs zu behandeln. Die semantischen Argumentrahmen sind, da die gleiche Bedeutung vorausgesetzt wird, die gleichen. Beim Anschluss der präpositional kodierten Aktanten, also insbesondere dem Pfad, sollten sich Verb und Nomen auch auf der syntaktischen Ebene kaum unterscheiden.

- (4.94) a. *Ein Viertel der Bevölkerung ist **ins** Ausland abgewandert.*
b. *die Abwanderung (von einem Viertel der Bevölkerung) **ins** Ausland*
- (4.95) a. *Max ist **in** München angekommen.*
b. *seine Ankunft **in** München / die Ankunft von Max in München*

Allerdings gilt dies nicht ohne Ausnahmen. So kann ein Aktant blockiert sein (4.96) oder syntaktisch anders angeschlossen werden (4.97). Möglicherweise sind diese Unterscheide aber auch Resultat einer wenn auch minimalen Änderung in der Bedeutung.

- (4.96) a. *Max fuhr (?nach München) **hin**.* (Source geblockt)
b. *die Hinfahrt nach München*
- (4.97) a. *Die Boeing fliegt **den** Flughafen an.* (Akk.)
b. *Das Anfliegen **des** Flughafens ist aus Lärmschutzgründen für Hubschrauber verboten.* (Gen.)
c. *Die Boeing stürzte beim Anflug **auf** den Flughafen ab.* (Präp.)

Neben der syntaktischen Realisierung der Aktanten besteht die zweite große Aufgabe in der Beschreibung der Stützverben, mittels derer die Nomina als Kernprädikat eines elementaren Satzes verwendet werden können.

- (4.98) a. *Max reist nach Helgoland.*
b. = *Max unternimmt eine Reise nach Helgoland.*
c. = *Die Reise (von Max) geht nach Helgoland.*
- (4.99) a. $\text{Oper}_1(\text{Reise}) = \text{machen, unternehmen, auf} \sim \text{en gehen, auf} \sim \text{en sein}$
b. $\text{Func}_2(\text{Reise}) = \text{gehen}$

Stützverben lassen sich optimal durch lexikalische Funktionen beschreiben (4.99), vgl. S. 29. Eine Behandlung und Beschreibung der französischen Äquivalente als Lexikogrammatik ist in G. Gross (1989) zu finden.

4.6.2 Relation Mensch – Ort

Nomina, die Relationen zwischen einem Mensch und einem Ort ausdrücken, sind durchweg Quasiprädikate, d.h. das Nomen bezeichnet entweder den Menschen oder den Ort, der an der Relation beteiligt ist.

$N_0hum V_{sup} Nhum N_1loc_{gen}$

(4.100) *Christian Ude ist Oberbürgermeister von München.*

(4.101) *der Münchner Oberbürgermeister Christian Ude*

$N_0hum V_{sup} Nhum N_1loc_{gen} Prep N_2loc$

(4.102) *Carlos Alberto Vargas Solis ist Botschafter von Costa Rica in Polen.*

(4.103) *der Botschafter von Costa Rica in Polen, Carlos Alberto Vargas Solis*

$N_0hum V_{sup} Nhum Prep N_1loc (Prep N_2loc)$

(4.104) *Die Meyers sind Emigranten aus Deutschland.*

Nach Mel'čuk (2004b: 41) kann der dritte Aktant (N_2loc) syntaktisch nicht als Dependent des Nomens realisiert werden. Sätze wie

(4.105) *Der Konflikt um die Emigranten **aus** Bangladesh **nach** Indien ist nicht neu.*

sind aber anzutreffen.

$N_0loc V_{sup} Nloc N_1hum$

(4.106) *Grevesmöhlen ist Wohnort von vier jungen Männern, die ...*

(4.107) *Ansbach ist Sitz der Regierung von Mittelfranken.*

(4.108) *Sitz der Landesgeschäftsstelle ist Passau.*

Wie die letzten Beispiele zeigen, können auch Organisationen ($NhumColl$) die Rolle von N_1 ausfüllen. Häufiger ist bei diesem Schema eines elementaren Satzes, aber die Konverse ($Oper_2 \leftrightarrow Oper_1$), die sinnvollerweise in der gleichen Tabelle beschrieben wird.

(4.109) *Die WIBV hat ihren Hauptsitz in Eindhoven.*

(4.110) *Peter Müller hat seinen Wohnsitz in Zürich.*

4.6.3 Relationen zwischen Orten

$N_0/loc V_{sup} Nloc N_1/loc$

- (4.111) *Paris ist die Hauptstadt von Frankreich.*
(4.112) *Piene ist ein Stadtteil von Gummersbach.*
(4.113) *Die Wörnitz ist ein linker Nebenfluss der Donau.*
(4.114) *Münchwilen ist der Nachbarort der sanktgallischen Stadt Wil.*
(4.115) *Guadalcanal ist eine pazifische Insel.*

Nomina, die binäre Relationen zwischen Orten ausgedrückt, sind sehr zahlreich, zumal mit den Affixen *Nachbar-*, *Neben-*, *Haupt-*, *Vor-* und *-teil* auch produktive Bildungen möglich sind. Nur ein Ausschnitt aus diesen Nomina wurde in Tab. C.6 auf S. 415 beschrieben, vor allem aus dem Bereich administrative Untergliederung und Gewässerzu- und -abflüsse. Wäge Relationen wie *Nachbarstadt* blieben außen vor.

$N_0/loc V_{sup} Nloc N_1/loc Prep N_2/loc$

- (4.116) *Bremerhaven ist eine Exklave des Bundeslandes Bremen in Niedersachsen.*

Nomina, die ternäre Relationen zwischen Orten ausdrücken, sind selten, die wenigen Nomina dieser Klasse werden deshalb mit in Tab. C.6 beschrieben. Zwei Beispiele – *Exklave* und *Mündung* – wurden bereits in Kap. 2.6 angesprochen.

4.7 Adverbien und Adjektive

Die Adverbien und prädikativen Formen von Adjektiven, die einen lokativen Aktanten syntaktisch realisieren, wurden wegen ihrer geringen Zahl in einer Tabelle zusammengefasst (Tab. C.7 auf S. 419). Sie verteilen sich auf folgende Satzchemata:

$N_0 V_{sup} Loc_{stat} N_1/loc Adv$

- (4.117) *Der Künstler Tróndur Patursson ist in Kirkjubøur ansässig.*
(4.118) *Peter Müller ist in München (polizeilich) gemeldet.*
(4.119) *Das Industrieunternehmen Precise Präzisionsspindeln GmbH ist in Leichlingen (Rheinland) geschäftsansässig gemeldet.*
(4.120) *Die Erle ist an Fließgewässern oder Mooren zuhause.*
(4.121) *Die Chilischote ist in Südamerika beheimatet.*

$N_0 V_{sup} Prep N_1/loc N_2/mess Adv$

- (4.122) *Der Strand ist nur 50m vom Hotel entfernt.*
(4.123) *Eure Eminenz, der Bischof von Durango ist ein alter Mann; sein Sitz ist fünfzehnhundert englische Meilen von Santa Fe entfernt.*

(4.124) *Die Großstadt Bialystok im Nordosten Polens liegt nur 30 Kilometer von der Grenze entfernt.*

Dimensionsadjektive (*hoch, tief, lang*, usw., siehe auch (2.86) auf S. 44), die einem ähnlichen Schema folgen ($N_0 V_{sup} N_1 mess Adv$), wurden nicht beschrieben, da ihnen der lokative zweite Aktant fehlt. Adverbien/Adjektive wie *entfernt* binden dagegen ein Lokativum als (fakultativen) Aktanten, was aus Paraphrasen mit einem Derivat deutlich wird (vgl. die Beispiele in Kap. 2.3.5):

(4.124') *Die Entfernung zwischen Bialystok und der Grenze beträgt 30 Kilometer.*

Der dritte Aktant dieser Gruppe von Adverbien wurde als $N_2 mess$ beschrieben. In den meisten Fällen ist seine Position durch eine nominale Maßphrase im Akkusativ gefüllt. Der adverbiale Charakter der Maßphrase äußert sich aber darin, dass auch *weit* und dessen Varianten (*unglaublich weit, nicht weit*, usw.) die Rolle übernehmen können. Auch Kombinationen (*einen Kilometer weit*) sind möglich.

$N_0 V_{sup} Loc_{dst} N_1 dest Adv$

(4.125) *Nierwetberg, Chef des Bonner Hilfswerks Help, ist unterwegs nach Goma.*

Manche dieser Adverbien lassen sich auch mit statischen Ortsangaben kombinieren, und könnten deshalb auch durch das Schema $N_0 V_{sup} Loc_{stat} N_1 loc Adv$ (s.o.) klassifiziert werden:

(4.126) *Peter Müller ist unterwegs nach Afrika.* $N_0 V_{sup} Loc_{dst} N_1 dest Adv$

(4.127) *Peter Müller ist unterwegs in Afrika.* $N_0 V_{sup} Loc_{stat} N_1 loc Adv$

Neben *unterwegs* gehören zu dieser Gruppe Adverbien einige komplexe mit einem Substantiv als Kern:

(4.128) *Er ist auf dem Weg nach Bad Villingen, um sich bei einer Kur zu erholen und Pfunde abzuspecken.*

(4.129) *James Garner ist auf der Durchreise nach Australien, als er ...*

Um diese von den Bewegungsnomina (Kap. 4.6.1) abzugrenzen, gilt das Kriterium, dass das Substantiv in dieser Lesart nicht außerhalb des Adverbs verwendet werden darf:

(4.130) *Der Flug (?der Weg) nach New York dauert acht Stunden.*

(4.131) *Die Reise (?der Weg) geht nach Frankreich.*

(4.132) a. *Die Familie besucht Berliner Verwandte und ist **auf der Durchreise nach Brandenburg**, wo sie zu einer Jugendweihe eingeladen ist.*

b. *? die Durchreise nach Brandenburg*

Allen Adverbien der drei Schemata ist gemeinsam, dass die Position des (Stütz)Verbs im Satz durch die Kopula *sein* ausgefüllt werden kann. Andere Stützverben (*liegen, sich befinden, gehen*, usw.) sind möglich. Semantisch drücken die Adverbien teils sehr verschiedene Relationen aus, was sich auch darin äußert, dass sich die möglichen semantischen Typen des ersten Aktanten deutlich variieren.

N_0 loc V_{sup} Adv

(4.133) *Memmert, Mellum und Minsener Oog sind unbewohnt.*

(4.134) *Das Knüllgebirge ist dünn besiedelt und stark bewaldet.*

(4.135) *Der Rhein ist von der Mündung flussaufwärts bis Basel schiffbar.*

Diese Adjektive realisieren als ersten semantischen realisieren ein Lokativum. Als semantisch sehr inhomogene Gruppe zeichnen sich die einzelnen Adjektive durch spezifische Modifikatoren oder auch weitere Aktanten aus. Eine Beschreibung als Lexikontabelle erfolgt nicht.

Weitere lokative Adverbien und Adverbiale

Adverbien und Adverbiale die lediglich eine Lokalisierung ausdrücken, oder solche, die sich wie eine Präposition verhalten wurden bereits unter dem Schema $N_0 V_{sup} Adv_{loc_{stat}}$ (S. 158) klassifiziert.

POS	Satzschema	Erl.	Tabelle	Beispiel
Prep	$N_0 V_{sup} Loc_{stat} N_1 loc$	157	C.1, S. 363	<i>Max ist in Budapest.</i>
Adv	$N_0 V_{sup} Advloc_{stat}$	158		<i>Das Haus steht dort drüben.</i>
V	$N_0 V N_1 src_{akk}$	159	C.3, S. 395	<i>Max verlässt das Zimmer.</i>
V	$N_0 V N_1 dest_{akk}$	159	C.3, S. 395	<i>Max betritt das Zimmer.</i>
V	$N_0 V N_1 pass_{akk}$	159	C.3, S. 395	<i>Max quert ein Geröllfeld.</i>
V	$N_0 V N_1 dest_{dat}$	159	C.3, S. 395	<i>Ein Taifun nähert sich der Insel Kyushu.</i>
V	$N_0 V N_1 pass_{dat}$	160	C.3, S. 398	<i>Die Wanderer folgten dem Tal.</i>
V	$N_0 V Prep N_1 dest$	160	C.3, S. 398	<i>Das Auto rast auf die Ampel zu.</i>
V	$N_0 V Prep N_1 src$	160	C.3, S. 398	<i>Ein ICE-T fährt aus dem Augsburger Hbf aus.</i>
V	$N_0 V Prep N_1 pass$	161	C.3, S. 399	<i>Shermans Divisionen setzten über den Fluss.</i>
V	$N_0 V Loc_{dst} N_1 dest$	161	C.3, S. 402	<i>Max fährt an die Ostsee.</i>
V	$N_0 V Loc_{src} N_1 src$	161	C.3, S. 400	<i>Max stieg vom Baum herab.</i>
V	$N_0 V Loc_{pass} N_1 pass$	162	C.3, S. 406	<i>Max flanierete durch die Boulevards von Paris.</i>
V	$N_0 V Loc_{stat} N_1 dest$	162	C.3, S. 408	<i>Max kommt in München an.</i>
V	$N_0 V Loc_{src/stat} N_1 src$	162	C.3, S. 408	<i>Max fährt vom Bahnhof ab.</i>
V	$N_0 V (Loc N_1 dest) V_{pp}$	162		<i>Max kommt angelaufen.</i>
V	$N_0 loc V N_1$	163		<i>Der Wald verschluckt Max.</i>
V	$Es V LocDst N_0 dest$	163		<i>Es geht nach Frankreich!</i>
V	$Es V LocSrc N_0 src$	163		<i>Mit dem Zug geht's hinaus aus Passau.</i>
V	$N_0 V (Loc N_1) V_{inf} W$	163		<i>Max geht (in den Supermarkt) Milch kaufen.</i>
V	$N_0 V N_1 loc_{akk}$	164	C.4, S. 409	<i>Die Meyers bewohnten ein herrschaftliches Haus in Washington.</i>
V	$N_0 V Loc_{stat} N_1 loc$	164	C.4, S. 409	<i>Max wohnt in Hamburg.</i>
V	$N_0 V Loc_{src} N_1 loc$	165	C.4, S. 409	<i>Max stammt aus Hamburg.</i>
V	$N_0 V Loc_{stat} N_1 stat) V_{pp}$	165		<i>Max liegt in Adelshausen begraben.</i>
V	$N_0 loc V N_1 loc_{akk}$	165	C.5, S. 411	<i>Die Isar durchfließt das Stadtgebiet von München.</i>
V	$N_0 loc V N_1 loc_{dat}$	166	C.5, S. 411	<i>Hier fließt die Samara der Wolga von links zu.</i>
V	$N_0 loc V Prep N_1 loc$	166	C.5, S. 411	<i>Die Isar fließt in die Donau.</i>
V	$N_0 V Loc_{stat} N_1 loc$	166	C.5, S. 413	<i>Die Elbe entspringt im Riesengebirge.</i>
V	$N_0 loc V Loc_{dst} N_1 loc$	166	C.5, S. 412	<i>Das Weideland erstreckt sich bis zum Flussufer.</i>
V	$N_0 loc V Loc_{pass} N_1 loc$	166	C.5, S. 413	<i>Der Radweg verläuft vorbei am Ameisberg, ...</i>
V	$N_0 loc V N_1 loc Prep N_2 loc$	166	C.5, S. 413	<i>Die kurze Buna entwässert den Shkodrasee in die Adria.</i>
V	$N_0 loc \text{ und } V Prep N_1 loc Prep N_2 loc$	167		<i>Der Tigris vereinigt sich mit dem Euphrat zum Schatt al-Arab, der in den Persischen Golf mündet.</i>
N	$N_0 hum V_{sup} N hum N_1 loc_{gen}$	169		<i>Christian Ude ist Oberbürgermeister von München.</i>
N	$N_0 hum V_{sup} N hum N_1 loc_{gen} Prep N_2 loc$	169		<i>Carlos Alberto Vargas Solis ist Botschafter von Costa Rica in Polen.</i>
N	$N_0 hum V_{sup} N hum Prep N_1 loc (Prep N_2 loc)$	169		<i>Die Meyers sind Emigranten aus Deutschland.</i>
N	$N_0 loc V_{sup} N loc N_1 hum$	169		<i>Grevesmühlen ist Wohnort von vier jungen Männern, die ...</i>
N	$N_0 loc V_{sup} N loc N_1 loc$	170	C.6, S. 415	<i>Paris ist die Hauptstadt von Frankreich.</i>
N	$N_0 loc V_{sup} N loc N_1 loc Prep N_2 loc$	170		<i>Bremerhaven ist eine Exklave des Bundeslandes Bremen in Niedersachsen.</i>
Adv	$N_0 V_{sup} Loc_{stat} N_1 loc Adv$	170	C.7, S. 419	<i>Der Künstler Tróndur Patursson ist in Kirkjubøur ansässig.</i>
Adv	$N_0 V_{sup} Prep N_1 loc N_2 mess Adv$	170	C.7, S. 419	<i>Der Strand ist nur 50m vom Hotel entfernt.</i>
Adv	$N_0 V_{sup} Loc_{dst} N_1 dest Adv$	171	C.7, S. 420	<i>Nierwetberg, Chef des Bonner Hilfswerks Help, ist unterwegs nach Goma.</i>
Adv	$N_0 loc V_{sup} Adv$	172		<i>Memmert, Mellum und Minsener Oog sind unbewohnt.</i>

Tabelle 4.2: Überblick über die behandelten Typen lokativer Sätze. Die Verweise beziehen sich zum einen auf die Erläuterungen in Kap. 4, zum anderen auf die Lexikontabellen, die die Lemmata des jeweiligen Typs behandelt. Wird auf keine Tabelle verwiesen, wurde dieser Typ nicht tabellarisch beschrieben.

5 Lokativa im Information Retrieval

Das ‚Information Retrieval‘ (IR)¹³⁸ befasst sich mit dem Auffinden von Information in Form relevanter Dokumente innerhalb einer Dokumentensammlung. Lokativa und Toponyme werden dabei gewöhnlich nicht anders als andere Wörter/Terme behandelt: das IR betrachtet ein Dokument als Menge von Termen. Nun kann man Dokumente (zumindest die meisten davon) aber auch als Menge von Lokationen in einer räumlichen Repräsentation, z.B. in Form von Koordinaten, betrachten. Das Ziel des ‚Geographic Information Retrieval‘ (GIR) ist, die beiden Repräsentationen zusammenzubringen, oder anders ausgedrückt, die präzisen Methoden der Geographen mit der alltäglichen Sprache des Menschen zu verbinden, wobei sich eine Reihe von Teilaufgaben ergeben, siehe auch GEOCLEF, SPIRIT, Sester & Heinzle (2004), Heinzle, Kopczynski & Sester (2003), Gey et al. (2005, 2006, 2007):

- der Aufbau von Gazetteers und Ontologien
- die Erkennung von Referenzen auf Lokationen in Texten, nämlich
 - die Erkennung von Ortsnamen und Ortsangaben im Text („Named Entity Recognition“ oder „geo-parsing“), siehe Kap. 5.1
 - die korrekte Zuordnung eines Ortsnamens zu einem bestimmten Ort („grounding“/„Geokodierung“, siehe Kap. 5.2). Diese stellt insbesondere dann ein Problem dar, wenn verschiedene Orte übereinstimmende Namen tragen.
- die Zuordnung eines Dokuments zu einem, oder mehreren Orten oder Regionen („geographic scope“)
- im Zusammenhang mit Anfragen/Queries (siehe Kap. 5.4.3):
 - Queryanalyse und -expansion im Hinblick auf räumliche Information
 - Ranking nach räumlichen Kriterien wie Nähe (geo-relevance ranking)
 - und die Benutzerführung (auch) über Landkarten und Ontologien (Kap. 5.4)
- und schließlich die Abbildung zwischen sprachlicher und nicht-sprachlicher Form der Ortsangabe (Kap. 5.5), die im Unterschied zur Geokodierung nicht nur Ortsnamen betrifft, sondern auch komplexere Ausdrücke, die komplexe Raumkonstellationen beschreiben.

5.1 Named Entity Recognition

Unter ‚(Named) Entity Recognition‘ (NER) versteht man die Erkennung der Benennungen von atomaren Entitäten wie z.B. Personen, Organisationen, Zeit-, Maß-, und nicht zuletzt Ortsangaben und Orten in natürlichsprachigen Texten. Zu unterscheiden ist zwischen ‚Named Entities‘, die Eigennamen sind oder aus solchen zusammengesetzt sind, und „unbenannten“

138 Viele der hier vorgestellten Aufgaben lassen sich auch der ‚Informationsextraktion‘ (IE) zuordnen. Diese steht dem IR nahe und wird oft auch als eines ihrer Teilgebiete betrachtet.

Entitäten, wie z.B. Datumsangaben.¹³⁹ Die NER gehört zum Bereich der Informationsextraktion (IE). Aber auch in anderen Gebieten der Sprachverarbeitung ist die NER relevant, überall dort, wo Named Entities als Informationseinheiten behandelt werden, z.B. im Parsing oder in der maschinellen Übersetzung. Dort müssen Eigennamen anders behandelt werden, sie werden z.B. nicht „übersetzt“ (vgl. S. 95). Die gesonderte Behandlung setzt ihre Erkennung voraus.

In der Theorie lässt sich die NER in zwei Teilaufgaben aufteilen, die aber in der Praxis kaum zu trennen sind: (1.) die eigentliche Erkennung und (2.) die Klassifikation, d.h. die Zuordnung zu einer vordefinierten Klasse einer Eigennamen- oder NE-Taxonomie. In Abhängigkeit von der Feingliedrigkeit der Taxonomie kann auch die zweite Teilaufgabe einen hohen Schwierigkeitsgrad aufweisen.

Systeme zur NER existieren zum einen als regelbasierte Systeme mit manuell erstellten, linguistisch motivierten oder pragmatischen Regeln und Heuristiken, zum anderen als statistische Systeme, die ihre Modelle an annotierten Korpora trainieren.

Dieses Kapitel hat vornehmlich ein Ziel: ein NER-System zur Erkennung geographischer Entitäten in deutschen Texten soll vorgestellt werden. Vergleichbare Systeme, die wie dieses auf lokalen Grammatiken basieren, behandeln andere Typen von Entitäten und wurden für andere Sprachen als das Deutsche entwickelt. Im Vergleich werden außerdem verschiedene in der Literatur beschriebene Systeme, die zwar nicht mit lokalen Grammatiken arbeiten, aber die auch die Erkennung deutscher geographischer Namen zum Ziel haben, angesprochen. Desweiteren wird kurz auf Standards zur Annotierung geographischer Entitäten eingegangen, vor allem auf die Kriterien für die Abgrenzung geographischer Entitäten von anderen Typen wie z.B. Organisationen aus der Domäne Sport. Den Anfang machen jedoch kurze Statements, die einige wichtige Fragen und Probleme der NER anreißen.

5.1.1 Anteil der (geographischen) Eigennamen in (deutschen) Texten

Als Argument, wieso die Erkennung von Eigennamen von Bedeutung sei, wird immer wieder angeführt, dass Eigennamen einen erheblichen Anteil der Wörter eines Textes ausmachen, insbesondere von Nachrichtentexten. Laut Friburger & Maurel (2002), Maurel (2004) machen Eigennamen ungefähr 10% eines französischen Zeitungstextes aus. Nach Gey (2001: 85) sind sogar mehr als 30% der „informationstragenden“ Wörter eines Nachrichtentextes Eigennamen, wobei er eine genauere Bestimmung unterlässt, welche Wörter denn nun informationstragende sind. Die Zahlen erscheinen eher hochgegriffen. Eigene Messungen ergaben, dass im englischen Brownkorpus lediglich 4,3% der Worttoken (vgl. Fußnote 81 auf S. 95) Eigennamen oder Teil eines solchen sind. In einem NE-annotierten Ausschnitt der deutschen „Computer Zeitung“ (siehe auf S. 205) gehören von 110 000 Wort-Token 7 500, und damit knapp 7% zu den

139 Aktuell ist eine Aufweichung des Konzeptes der ‚Named Entity‘ zu beobachten. In Sekines ‚Extended Hierarchy‘ (SEKINEEXTNER) sind zahlreiche Nicht-Eigennamen zu finden (vgl. S. 134). Toral, Muñoz & Montoyo (2005) behandeln als ‚Weak Named Entities‘ verschiedene Typen monoreferentieller Syntagmen. Der Begriff der ‚Named Entity‘ ist relativ jung, er wurde erst bei der Vorbereitung der MUC-6 (Sundheim 1995, Grishman & Sundheim 1996) eingeführt. Das zugrundeliegende Konzept (der ‚Eigennamen‘!) ist selbstverständlich älter. Für frühe Anwendungen der ‚Eigennamenerkennung‘ vgl. Rau (1991), Coates-Stephens (1991, 1992) und die ‚Proper Name Facility‘ in McDonald (1993). Nadeau & Sekine (2007) geben einen Überblick über die Entwicklung der NER in den vergangenen 15 Jahren.

Eigennamen. Im deutschen Korpus des CoNLL-2003-Shared-Task sind 7,5%¹⁴⁰ der Worttoken Named Entities oder Bestandteile davon.

Kalkulationen, welchen Anteil geographische Entitäten innerhalb aller Named Entities einnehmen, variieren zwischen 49% (Schnetzler 2004: 29 bzgl. der CLEF-Topics 2000–2003), 44% (Friburger & Maurel 2002) für französische Nachrichtentexte, 36%¹⁴⁰ (deutsches Korpus CoNLL-2003-Shared-Task), und 20% in der „Computer Zeitung“, wobei letztere sehr spezifisch im Nachrichtengenre ist. Zählt man dort die annotierten Adjektive und Wohnernamen zu den Named Entities hinzu, erhöht sich der Anteil aller lokativen Entities auf 30%.

Der Anteil geographischer Entitäten an den Worttoken eines Textes variiert für die vorliegenden deutschen NE-annotierten Korpora zwischen 2,7% (SZ), 2,3% (CoNLL-2003-Shared-Task) und 0,9% (Computer Zeitung). Toponyme, die aus mehreren Token bestehen wurden auch mehrfach gezählt, Adjektive und Wohnernamen nicht mithinzugerechnet.

5.1.2 Strategien zur Erkennung von Eigennamen

Strategien zur Erkennung von Eigennamen können auf verschiedenen sprachlichen Ebenen ansetzen:

Wort/Entität: auf der Wortebene erfolgt die Erkennung mittels einer Eigennamenliste (Gazetteer) oder anhand des internen Kontexts.

lokaler externer Kontext : der Kontext rechts und links von der Entität im Umfang einiger weniger Wörter bis zur Länge eines Satzes dient zur Erkennung wie auch Disambiguierung von Entitäten.

Diskurs: die Annahme, das auf der Diskurs- oder Textebene ein Wort/Entität einheitlich verwendet wird (one-sense-per-document), kann zur Erkennung (Learn-Apply-Forget) wie Disambiguierung von homonymen (Wörtern/Entitäten) verwendet werden.

Korpus: Korpora werden zur Lexikonaquisition und zum „Lernen“ von Kontexten oder statistischen Gewichten verwendet.

Die Verwendung einer dieser Strategien ist unabhängig davon, welchem Paradigma (statistisch vs. regelbasiert, vgl. Kap. 5.1.3) ein System verpflichtet ist. So können sowohl regelbasierte wie statistische Systeme auf Gazetteers zurückgreifen. Eventuell unterscheiden sich die Systeme bei der Größe und Qualität des Gazetteers und möglicherweise darin, ob der Gazetteer unabhängig von einem Korpus erstellt, oder aus einem annotierten Trainingskorpus extrahiert wird.

5.1.2.1 Listenbasierte Erkennung: Gazetteers

Die Erkennung von Named Entities anhand von Eigennamenlisten (Gazetteers) ist eine der simpelsten und zugleich wirkungsvollsten Strategien. Allerdings sind eine Reihe von Problemen unvermeidlich, wenn sich ein System ausschließlich auf Gazetteers verlässt:

140 Ohne die problematische Kategorie 'MISC'. Mit dieser beträgt (a) der Anteil der NE-Token an allen Wort-Token gut 9%, (b) der Anteil der Kategorie 'LOC' an allen NE-Kategorien nur 27%.

English		French		German	
And	(Iran)	De	(Burkina Faso)	Die	(France)
To	(Ghana)	Du	(Ghana)	Den	(Ethiopia)
Be	(India)	Un	(Russia)	Zu	(Zaire)
By	(Sweden)	Une	(Colombia)	Ist	(Hungary)
Are	(Nigeria)	Est	(Netherlands)	Im	(Russia)
This	(France)	Il	(Iran)	Dem	(Cameroon)
But	(Afghanistan)	Au	(Austria)	Als	(Denmark)
Had	(Oman)	Par	(U.K.)	Auch	(France)
She	(India)	Sur	(Oman)	An	(Mexico)
We	(Zaire)	Pas	(Turkey)	Aus	(Namibia)

Tabelle 5.1: GEO-Stopwörter in drei Sprachen (aus Kimler 2004, Pouliquen et al. 2006, vgl. S. 108). Obwohl die GEO-Stopwörter Orte bezeichnen und auch in einem Gazetteer enthalten sind, ist es ratsam, diese nicht in einem Text der jeweiligen Sprache als Entitäten zu annotieren: die Wortform *And* bezeichnet in einem englischen Text mit sehr hoher Wahrscheinlichkeit nicht den im Iran gelegenen Ort.

1. im Gazetteer fehlende und damit dem System unbekannte Eigennamen (sog. ‚lexical lacunae‘)
2. Homonymien zwischen Eigennamen und Nicht-Eigennamen oder verschiedenen Klassen von Eigennamen (*Essen* bzw. *Washington*, vgl. S. 108)
3. in komplexere Named Entities eingebettete Eigennamen (‚embedded entities‘, siehe auch Kap. 5.1.5.1.2)

Die Punkte 1. und 2. stehen dabei in Konkurrenz zueinander. Ist der Gazetteer klein, sind Lücken im Lexikon und damit unerkannte Namen im Text die Folge. Ist er groß, enthält er viele Namen, die zu Fehlerkennungen führen, da sie homonym zu Nicht-Eigennamen oder anderen Namensklassen sind. Eine optimale Größe für den Gazetteer zu finden, ist schwierig. Pouliquen et al. (2004, 2006), die mit einem großen Gazetteer arbeiten, nutzen deshalb sprachspezifische Geo-Stopwort-Listen (Tab. 5.1), um die häufigsten und schlimmsten Fehler zu vermeiden. Die zweite, häufig genutzte Strategie ist, anhand der Kontexte zusätzliche Evidenz zu sammeln, und so die falschen Namen auszusondern.

Aber ist die Beschreibung von Kontexten allein ausreichend als Strategie zur Erkennung von Named Entities? Der Titel eines Artikels von Mikheev, Moens & Grover 1999 „Named entity recognition without gazetteers“ lässt dies vermuten und ist doch zugleich irreführend. Denn das dort beschriebene System zeigt mit zunehmender Größe des Gazetteers auch eine bessere Performance. Dass Precision und Recall auch ohne Gazetteer bei etwa 85% liegen, spricht für die Qualität des Systems, aber nicht gegen den Einsatz von Namenslisten und -lexika. Mit großem Gazetteer sind Performance-Werte um 95% möglich. Dass der Sprung von 85% auf 95% erheblichen Aufwand in der Erstellung der lexikalischen Ressourcen bedeutet – der von Mikheev, Moens & Grover verwendet Gazetteer umfasst 5 000 Orts-, 30 000 Organisations- und 10 000 Vornamen –, hat Parallelen in anderen Bereichen der Sprachverarbeitung (z.B. Tagging). Der Sprung bedeutet aber auch den Schritt von einem System an der Grenze der Verwendbarkeit zu einem voll einsatzfähigen System.

Ganz besonders gilt das für Entitäten der Klasse LOCATION: hier steigen die Werte (Recall bzw. Precision) von 46/59% ohne Gazetteer auf 95/94% mit großem Gazetteer, mit einer kleinen Liste von 200 Ortsnamen bereits auf 90/85%. Die Behauptung der Autoren „[...] it is sufficient to use relatively small gazetteers of well-known names, rather than large gazetteers of low-frequency names.“, ist richtig und falsch zugleich. Natürlich ist es notwendig, dass relevante, allgemein bekannte Namen im Gazetteer enthalten sind. Die Ergebnisse lassen sich mit größeren Gazetteers aber weiter verbessern. So verwundert es nicht, dass alle Teilnehmer am CONLL-2003-SHARED-TASK Gazetteers und damit von Hand erstellte Ressourcen benutzen, obwohl die Konferenz ja dem Ansatz des maschinellen Lernens verpflichtet ist.

Fraglich ist auch, ob sich die Beobachtungen, die an englischem Material gemacht wurden, so einfach auf andere Sprachen übertragen lassen:

Mikheev et al. (1999) investigated the role of lists of named entities and showed that reasonable results are possible with small or even no lists. We believe that such results are only possible for languages where capitalization is sufficient to detect named entities and where lists are only needed to support the categorization of the names.

For languages with other characteristics, such as German, named entity recognition is heavily dependent on substantial and reliable lists. A lack of coverage lowers recall while unreliable entries especially frequent ones, dramatically degrade precision. (Rössler 2004)

5.1.2.2 Interner und externer Kontext

Den Unterschied zwischen internem und externem Kontext beschreibt McDonald (1993: 32; 1996: 22) folgendermaßen:

The requirement that a grammar of proper names must be context-sensitive derives from the fact that the classification of a name involves two complementary kinds of evidence, which we will term ‘internal’ and ‘external’. Internal evidence is taken from within the sequence of words that comprise the name. This can be a definitive criteria, such as the presence of known ‘incorporation terms’ that indicate companies (*Ltd.*, *G.m.b.H.*); or it can be heuristic criteria such as abbreviations or known first names, which often indicate people. [...]

By contrast, external evidence is provided by the context in which a name appears. The basis for this evidence is the obvious observation that names are just ways to refer to individuals of specific types (people, churches, rock groups, etc.), and that these types have characteristic properties and participate in characteristic events. The presence of these properties or events in syntactic relation with a proper name can be used to provide confirming or criterial evidence for a name’s category.

External evidence is a necessity for high accuracy performance. One obvious reason is that the predefined word lists so often used as internal evidence can never be complete. Another is that in many instances, especially those involving subsequent references, external evidence will override internal evidence. In the final analysis it is always the way a phrase is used—the attributions and predications it is part of—that make it a proper name of a given sort. Without the consideration of external evidence this definitive criteria is missed, resulting in mistakes and confusion in the state of the parser. (McDonald 1996: 22–3)

Während der interne Kontext praktisch ausschließlich zur Erkennung von (unbekannten) Entitäten Verwendung findet, dient der externe Kontext darüberhinaus zur Disambiguierung zwischen (a) Eigennamen und Nicht-Eigennamen, und (b) verschiedenen Klassen von Eigennamen. Ohne externen Kontext ist keine Erkennung mit hoher Precision möglich.

Als interne oder externe Evidenz kommen eine Reihe von Eigenschaften einzelner Wörter oder Wortfolgen in Betracht:

1. Orthographische Eigenschaften und Groß-/Kleinschreibung
 - die wichtigste interne Evidenz ist in vielen Sprachen mit Sicherheit die Großschreibung von Eigennamen¹⁴¹
 - bestimmte Muster innerhalb eines Wortes (z.B. durchgehend groß, Großbuchstaben im Wortinneren, Ziffer im Wort)
 - orthographische Muster von Wortfolgen (z.B. zwei aufeinanderfolgende großgeschriebene Wörter)
 - Wort(folge) in Anführungszeichen, etc.
2. Affixe von Wörtern
3. lexikalische Einheiten (oder spezifische Wortformen)
 - intern, innerhalb der Entität: Verwendung von *Gazetteers*
 - extern, vor und nach der Entität: sog. ‚trigger words‘
4. Wortartenannotationen (Tags) und charakteristische Abfolgen von diesen
5. andere Eigennamen im Kontext
6. die syntaktische Struktur, in die die Entität eingebettet ist

Zumindest einige dieser Eigenschaften werden von allen Systemen verwendet, von handgemachten regelbasierten wie von statistischen. Unterschiede bestehen, neben der Methode zur Gewinnung der Kontexte und ihrer linguistischen Motivation, darin, wie groß die Kontexte gewählt werden. Statistische Systeme wählen üblicherweise Fenster von einigen (bis zu 5) Wörtern/Token vor und nach der Entität. Bei regelbasierten Systemen erstrecken sich die Kontexte auch bis auf die Längen einer Phrase oder eines Satzes.

Eine allgemeine, und immer wieder zitierte These ist, dass Entitäten der Klasse *LOCATION* weniger über interne oder externe Kontexte verfügen als anderen Klassen:

The hypothesis was that names of organisations and names of people should still be handled relatively well by the system, since they have much internal and external evidence, whereas names of locations have fewer reliable contextual clues. For example, expressions such as “XXX is based in YYY” is not sure-fire evidence that YYY is a location – it could also be an organisation. And since many locations are so well-known, they receive very little extra context (“in China”, “in Paris”, vs “in the small town of Ekeren”).

(Mikheev, Moens & Grover 1999)

Diese These erscheint intuitiv logisch, richtig ist sie nicht. Auch Lokationen stehen in Kontexten, die Evidenz liefern, nur sind diese weniger lokal und auch vielfältiger. „Bequeme“ lokale Kontexte wie *Herr/Mr.* oder *GmbH* sind selten. Aber rund ein Drittel aller Ortsnamen werden zusammen mit einer räumlichen Präpositionen verwendet (vgl. Tab. 5.6), welche zugegebenermaßen nur mit einer detaillierteren syntaktischen Beschreibung (siehe Kap. 3.2.1.2 und 5.1.5.3.1) eindeutig genug sind. Die Satzebene bzw. die Ebene von Prädikat-Argument-Strukturen bietet in jedem Fall genügend Evidenz. Was den internen Kontext betrifft, so

141 Großschreibung stellt aber keine 100% Sicherheit dar: Maurel et al. (2004) berichten, dass sie nur in jedem zweiten Fall erfolgreich ist, wegen Ambiguitäten und komplexer Namen.

unterscheiden sich die verschiedenen Klassen von Ortsnamen deutlich. Länder- und Siedlungsnamen weisen wenig interne Evidenz auf, Straßennamen dagegen sehr viel, vgl. Kap. 3.2.1.1, 5.1.5.3.1.1 und B.2.

5.1.2.3 Die Diskursebene: One-Sense-per-Document und Learn-Apply-Forget

Die Annahme, dass die Bedeutung eines Wortes innerhalb eines Diskurses oder Textes konstant ist (‘one-sense-per-document’) wurde von (Gale, Church & Yarowsky 1992) untersucht und weitgehend bestätigt: „We conclude that with probability about 94% (51/54), two polysemous nouns drawn from the same article will have the same sense.“ Auch Rössler (2004) hat festgestellt, dass im CONLL-2003-SHARED-TASK-Korpus 93,5% aller Eigennamen(-bestandteile) innerhalb eines Dokuments eindeutig sind. Ambiguitäten innerhalb eines Textes resultieren hier aus Namen wie *Deutsche Bank*, *Philip Morris*.

Das One-Sense-Per-Document-Prinzip wird von der Methode ‚Learn-Apply-Forget‘ (LAF) genutzt. Dabei bleibt ein Name, der mit ausreichendem Kontext vom System „gelernt“ wurde, noch eine Weile in dessen „Gedächtnis“, bis er schließlich „vergessen“ wird. Das System verwendet also ein dynamisches Lexikon. Als geeignete Parameter wählten Volk & Clemenide (2001) eine Verweildauer von 15 Sätzen im Gedächtnis, die mit jeder neuerlichen Erwähnung um 5 Sätze verlängert wird. Das System erlaubt homonyme Namen, die verschiedenen Klassen angehören. Dies widerspricht der „one-sense-per-document“-Hypothese: „Note that this algorithm allows a name to belong to different classes within the same document. We have observed this in our corpus especially when a company name is derived from its founder’s name and both are mentioned in the same document. [...] These findings contradict the one-sense-per-document hypothesis [...]“ (Volk & Clemenide 2001).¹⁴²

One-Sense-Per-Document bzw. Learn-Apply-Forget sind mit Sicherheit brauchbare Strategien. Eine Vielzahl von Systemen basierend auf lokalen Grammatiken (Friburger 2002, Mallchok 2005) und zur Erkennung deutscher Toponyme (Piskorski & Neumann 2000, Volk & Clemenide 2001, Rössler 2002) verwenden diese Strategie. Gerade für Toponyme sollte die Strategie aber nicht überbewertet werden: ein guter Gazetteer ist wichtiger und auf die Beschreibung der externen Kontexte ist Learn-Apply-Forget ohnehin angewiesen. Gut geeignet ist Learn-Apply-Forget mit Sicherheit zur Disambiguierung von ambigen Toponymen (Geokodierung, siehe Kap. 5.2).

5.1.3 Ansätze und Systemarchitekturen

Neben den im vorigen Kapitel beschriebenen Strategien lassen sich Systeme zur Entitätenerkennung auch über deren grundsätzliche Systemarchitektur und den gewählten Ansatz beschreiben. Meist werden dabei wie in anderen Anwendungsbereichen von Sprachtechnologie zwei Paradigmen – regelbasierte (oder symbolische) vs. statistische Systeme – angenommen

¹⁴² Diese Beobachtung von Volk & Clemenide (2001) deckt sich mit der Aussage von Krovetz (1998): die Annahme ‚one-sense-per-document‘ gilt nur, wenn man einen sehr weiten Begriff von Polysemie voraussetzt, d.h. Mehrdeutigkeit mit Homonymie gleichsetzt und typische Fälle regulärer Polysemie (siehe Kap. 3.1.4) ausklammert. Wertet man dagegen Fälle regulärer Polysemie als ‚multiple-senses per discourse‘, weisen 33% (so Krovetz 1998) der potentiell mehrdeutigen Wörter auch innerhalb eines Textes mehr als eine Bedeutung auf.

und einander gegenübergestellt. Eine solche binäre Einteilung bedeutet eine Vereinfachung, eine genauere Beschreibung bedarf vielmehr einer Reihe von Parametern wie die Regeln und Strategien, auf denen das System basiert, erworben, repräsentiert und angewandt werden:

Aneignung: die Eingabe der Regeln in das System kann entweder durch den menschlichen Experten erfolgen, oder die Regeln werden aus geeigneten Daten automatisch mittels geeigneter Algorithmen abgeleitet („erlernt“ oder „trainiert“).

Repräsentation: die Regeln sind explizit repräsentiert (symbolische Systeme) oder nur implizit (subsymbolische Systeme). Eine Eingabe der Regeln durch den Menschen setzt eine symbolische Repräsentation voraus. Statistische System halten die Regeln oft in subsymbolischer Repräsentation vor, aber nicht immer (z.B. Brill 1993).

deterministisch vs. stochastisch: ein stochastisches System erlaubt, dass Wahrscheinlichkeiten (oder ‚Gewichte‘) bei der Lösung/Erkennung berücksichtigt werden, ein deterministisches System kennt nur binäre Entscheidungen.

supervised – unsupervised: müssen die Daten zum „Trainieren“ eines Systems neben der Eingabeseite auch die Ausgabeseite (in der NER die Annotationen) enthalten, spricht man von einem „überwachten“ (supervised) System. „Unüberwachten“ Systemen (unsupervised) „genügt“ dagegen nicht annotiertes Sprachmaterial. Daneben gibt Ansätze – „semi-supervised“ genannt, die sowohl annotierte als auch unannotierte Daten nutzen. Vgl. Nadeau & Sekine (2007: 7–11) für eine Diskussion der drei Begriffe supervised, semi-supervised und unsupervised.

Im Folgenden werden diejenigen Systeme „statistische“ genannt, die an geeigneten Daten trainiert werden, und sich dabei statistischen Modellen oder Verfahren des maschinellen Lernens bedienen. Bei „regelbasierten“ Systemen erfolgt die Kodierung der Regeln dagegen durch den Menschen.

5.1.3.1 Statistische Systeme

Statistische Systeme versuchen die Regeln zur Erkennung von Eigennamen aus einem annotierten Korpus abzuleiten. Die Merkmale, die für die Regeln herangezogen werden, unterscheiden sich bei aktuellen Systemen kaum von den in lokalen Grammatiken verwendeten (siehe S. 18). Florian et al. (2003) berücksichtigen z.B. neben den Wortformen die Grundformen, Wortarteninformation, Präfixe und Suffixe, Merkmale wie ‚erster Buchstabe groß‘, ‚alle Buchstaben groß‘, ‚zweistellige Zahl‘ und nutzen zusätzlich große Listen mit Orts-, Personen- und Organisationsnamen. Die Regeln des Systems sind dann Muster, aus den oben genannten Merkmalen kombiniert im Kontext von bis zu fünf Wörtern.

Ein Vorzug der statistischen Systeme und eine übliche Kritik an regelbasierten Systemen lautet, dass die Kosten für das monatelange Schreiben und Testen der Regeln durch Experten zu hoch seien, zumal sich der Aufwand bei jeder Anpassung des Systems an eine neue Sprache oder ein weiteres Genre wiederholt: „Die Entwicklung rein lernbasierter NLP-Komponenten verspricht robuste und schnell entwickelbare Systeme, die den Entwickler vom mühsamen Handwerk des Regelschreibens befreien.“ (Rössler 2007: 61)

Allerdings ist der Aufbau von Ressourcen, und zwar die Bereitstellung ausreichend großer annotierter Korpora, aus denen sich Regeln ableiten lassen, einer der kritischen Punkte beim Bau eines statistischen Systems. Roth (2002: 106) formuliert es folgendermaßen: „Um

also ein *ausgefeiltes* statistisches System zu bauen, werden wahrscheinlich ebenso Ressourcen an Zeit und Geld benötigt wie beim Bau eines leistungsfähigen regelbasierten Systems.“ Die Annotation der Korpora ist „ein nicht zu vernachlässigender“ Aufwand (107). Diese Aussage deckt sich mit meinen eigenen Erfahrungen bei der Annotierung eines Korpus oder den Versuchen, gute annotierte Korpora zu finden (vgl. Kap. 5.1.5.2.2). Rössler (2007: 90) schätzt den Aufwand zur Korpusannotation für ein „lernbasiertes“ System etwa gleich hoch ein wie das Schreiben von Regeln für ein regelbasiertes System. Um die Abhängigkeit von ausreichend großen und gut annotierten Korpora abzumildern, greifen statistische Systeme

- entweder auf Gazetteers zurück, nutzen also eine manuell erstellte Resource. So nutzen z.B. alle dokumentierten statistischen Systeme zur Erkennung deutscher geographischer Entitäten (vgl. Tab. 5.5) und alle Systeme, die am CoNLL-2003-Shared-Task teilnahmen, Gazetteers.
- oder es wird versucht, die Systeme (teilweise) mit unannotierten Daten zu trainieren („unsupervised“ oder „semi-supervised“), oder Ressourcen, z.B. Gazetteers, aus unannotiertem Textmaterial zu extrahieren (vgl. Riloff & Jones 1999, Uryupina 2003, Lee, Lee & Myaeng ed. 2005, Lee & Lee 2005, Erp 2006, Rössler & Morik 2005, Rössler 2007 und Kap. 5.3.1).

Durch den Rückgriff auf Gazetteers sind statistische Systeme dann aber auf gute, manuell erstellte Entitätenlisten angewiesen. Allerdings müssen bei der Erörterung der Frage „regelbasiert oder statistisch?“ auch weitere Faktoren Berücksichtigung finden, vgl. die Diskussion unter dem Stichwort „Korpusadaptivität“ in Rössler (2007: 84–97). Ein Vorteil statistischer Systeme ist, dass sich ihre Entwickler nicht um Standards zur Annotierung (Kap. 5.1.5.1) kümmern müssen: Trainings- und Testkorpora werden als bereits vorhanden vorausgesetzt. Damit reduziert sich aber auch die gemeinsame Basis von Entwicklern und Nutzern des Systems auf eben diese Korpora. Dass auf einem Korpus trainierte Systeme auf einem anderen Korpus versagen können, berichtet Leidner (2007: 179–80). Der verwendete NE-Tagger (Curran & Clark 2003) erreichte beim CoNLL-2003-Shared-Task einen F_1 -Wert von 85% (LOC 88%), wird aber mit einem am MUC-7-Korpus trainierten NE-Sprachmodell ausgeliefert. Unverändert angewendet auf die CoNLL-2003-Shared-Task-Daten erreicht das System nur einen F_1 -Wert von 53% (LOC 59%), obwohl beide Korpora ausschließlich Texte des Nachrichtengenres enthalten, also durchaus vergleichbar sind. Nicht zuletzt ist die Antwort abhängig von der Klasse der Entitäten: „Für den Einsatz der MUC Kategorien Zeit-, Datums- und Mengenangaben etwa, steht der erforderliche Zeitaufwand zur Annotation in keinem Verhältnis zur Effizienz der Entwicklung von Regeln.“ (Rössler 2007: 90)

Regelbasierte Systeme erreichen bislang üblicherweise bessere Ergebnisse als statistische Systeme. So liegen die Ergebnisse der besten regelbasierten Systeme, die an der MUC-6 und MUC-7 (Sundheim 1995, 1996, Marsh & Perzanowski 1998) teilnahmen, mit einem um 5–7% höheren F-Wert (siehe Kap. 5.1.4.1) deutlich vor den statistischen Systemen, die am CoNLL-2003-Shared-Task (Sang & De Meulder 2003) partizipierten. Dennoch stellen statistische Systeme seit Mitte/Ende der 90er Jahre das vorherrschende Paradigma in der NER und haben, zumindest was die Zahl der Publikationen angeht, die regelbasierten Architekturen an den Rand gedrängt.

5.1.3.2 Regelbasierte Systeme

Regelbasierte Systeme bedienen sich der in Kap. 5.1.2 beschriebenen Strategien. Die Regeln, wie mittels der Strategien Named Entities erkannt und klassifiziert werden sollen, sind expliziert („symbolisch“) repräsentiert. Sie sind von Menschen, meist Experten, die mit den sprachlichen Phänomenen und den Regelformaten vertraut sind, erstellt. Das schließt aber nicht aus, das die Regeln mit und anhand von Sprachmaterial entwickelt werden („data-driven development“). Das gilt insbesondere für die Erarbeitung der notwendigen Listen und Lexika (Eigennamen und Trigger-Words).

Auf eine vollständige linguistische (syntaktische) Analyse verzichteten regelbasierte Systeme für die NER praktisch von Anfang an und stützen sich stattdessen auf Finite-State-Techniken, die bei der Anwendung auf große Textmengen mehr Effizienz versprechen.

Selbst für eine kurze Charakteristik verschiedener regelbasierter Systeme fehlt hier der Platz. Friburger (2002: 30–1), Mallchok (2005: 38–48) und Rössler (2007: 54–60) vermitteln einen guten Überblick. Stattdessen werden an dieser Stelle ausschließlich einige regelbasierte Systeme, die auf lokale Grammatiken als zentrale Resource zurückgreifen, vorgestellt.

5.1.3.2.1 NER mit lokalen Grammatiken Dass sich lokale Grammatiken als Mittel zur Named Entity Recognition eignen, demonstrieren die Systeme von Friburger und Mallchok.¹⁴³ Beide nutzen lokale Grammatiken in Kombination mit einer Learn-Apply-Forget-Komponente (Kap. 5.1.2.3). Lokale Grammatiken beschreiben einerseits Pattern für Kandidaten von Named Entities und andererseits Kontexte, die die Kandidaten bestätigen, ausschließen oder bzgl. einer NE-Klasse disambiguieren. Die Learn-Apply-Forget-Komponente dient dazu, alle Vorkommen der bestätigten Kandidaten inklusive ihrer Varianten innerhalb eines Textes/Diskurses zu markieren.

Das von Friburger (2002, 2006; Friburger & Maurel 2001a,b, 2004) beschriebene System zur Erkennung von Named Entities und der Klassifikation in Personen, Organisation und Lokationen bedient sich der Engine INTEX zur Anwendung lokaler Grammatiken als Transduktoren. Folgende Schritte werden dabei ausgeführt:

0. Vorverarbeitung

- Satzsegmentierung
- Tokenisierung
- Lexikon-Lookup (inkl. Eigennamenwörterbüchern)
- Markierung von Eigennamenkandidaten (Friburger 2002: 74)
 - <CNP> („candidats noms propres“): Token, bestehend aus mind. zwei Buchstaben, der erste Buchstabe ist groß-, die folgenden können groß- oder kleingeschrieben sein
 - <CNPCA>: Kandidaten aus mehreren Token werden anhand von orthographischen Pattern erkannt (*O’Reilly, Mac Donald*)

1. Transduktorenkaskade 1 – lokale Grammatiken beschreiben Kontexte für

¹⁴³ Daneben seien noch erwähnt: Garrigues (1993, 1995), die zwar Kontexte lokativer Named Entities beschreibt, aber von einem vollständigen NER-System weit entfernt ist (siehe S. 58), sowie Senellart (1998a,b) Personennamen und die Funktion der Person extrahiert. Senellart wurde bereits im Zusammenhang mit dem Ansatz des Bootstrapping (S. 9) besprochen.

- Bestätigung oder Ausschluss von NE-Kandidaten
 - Erkennung des NE-Typs (Person, Organisation, Ort)
2. Transduktorenkaskade 2 – basierend auf der One-Sense-Per-Document-Annahme („Learn-Apply-Forget“) werden alle Vorkommen der in Schritt 1 erkannten Entitäten markiert, inklusive von generierten Varianten

Bei der Evaluierung anhand zweier Korpora erreicht das System einen F-Wert von über 93,5%. Die Ergebnisse für Lokationen sind dabei tendenziell besser als für die anderen beiden Typen. Insbesondere der Recall ist mit über 96% sehr hoch und resultiert aus dem verwendeten, sehr vollständigen Toponymlexikon (PROLINTEX). Die beschriebenen Kontexte für Toponyme beschränken sich auf Klassenbezeichner (*la rivière de Tréguier, l'estuaire de la Seine*) und Kombination mit Himmelsrichtungen (*Afrique du Sud*) und decken nur ein Fünftel aller Toponymvorkommen ab: „Seuls 20% des noms de lieux ont un contexte gauche et quelques uns une preuve interne.“ (Friburger 2002: 94) Die meisten Toponyme werden deshalb allein durch das Lexikon erkannt. Als Konsequenz bleibt Schritt 2 („Learn-Apply-Forget“) für Lokationen (fast) wirkungslos. In den zwei ausgewerteten Korpora führt er einmal (Le Monde) zu einer Verbesserung des F-Werts um 0,4%, das andere Mal (Ouest France) zu einer Verschlechterung um 0,1%. Für die beiden anderen NE-Klassen führt Schritt 2 zu einer Verbesserung im F-Wert zwischen 0,8% (Ouest France) und 2,7% (Le Monde), und zwar durch eine deutliche Verbesserung des Recalls und nur eine geringfügige Verschlechterung bei der Precision. Der zentrale Bestandteil des Systems von Friburger bleiben damit die lokalen Grammatiken in Kombination mit einem guten Toponymlexikon.

Das von Mallchok (2005) zur Erkennung von Organisationsnamen in englischen Wirtschaftsnachrichten entwickelte System nutzt lokale Grammatiken mit UNITEX als Engine zum „Lernen“ der Organisationsnamen. Der Workflow, die eigentliche Annotierung und Bearbeitung der gewonnenen Organisationsnamen (Säuberung, Grundformreduktion und Generierung von Akronymen) sind in der Skriptsprache PERL implementiert. Die einzelnen Arbeitsschritte gestalten sich folgendermaßen:

0. Der zu bearbeitende Text (ein Zeitungsartikel) wird mit UNITEX vorverarbeitet (Satzsegmentierung, Tokenisierung). Beim Lexikon-Lookup werden neben den Standardwörterbüchern auch Spezialwörterbücher mit Organisationsnamen, Städten und Ländern, Berufsbezeichnern etc. verwendet.
1. Erkennung der Organisationsnamen anhand von externen wie internen Kontexten und einem Organisationsnamenslexikon mit 80 000 Einträgen. 113 UNITEX-Graphen und ein Lexikon mit 21 000 Organisationsdeskriptoren beschreiben interne Kontexte, weitere 100 Graphen externe Kontexte wie Joint Ventures, Fusionen etc. Kontexte (85–100, 109–150) sind z.B.: ACRONYM LAST_NAME & DESCRIPTOR DESCRIPTOR (*MH Meyerson & Co Inc*), LOCATION's ORG_NAME (*Singapore's Changi Airport*), PERSON of ORG_NAME (*Rand V. Araskog of ITT*), ORG EXPRESSES (TIME)? JOINT-VENTURE with ORG (*Advanced Tissue Sciences also said it has signed a joint venture agreement with Smith & Nephew Plc*).
2. Nachbearbeitung der ermittelten Organisationsnamen
 - a) Für gefundene Organisationsnamen werden Initialabkürzungen generiert
 - b) Passende Silbenabkürzungen werden aus der Liste der potentiellen Organisationsnamen ermittelt

- c) Die gefundenen Organisationsnamen werden „gesäubert“. Beseitigt werden vor allem typische Übergenerierungen lokaler Grammatiken, z.B. durch <PRE> erkannte großgeschriebene Wörter am Satzanfang, die als Artikel, Konjunktion usw. eindeutig nicht Teil des Organisationsnamens sein können.
 - d) Die Grundformreduktion für das Englische ist vergleichsweise simpel: ein apostrophiertes 's wird entfernt, aber nur wenn der Organisationsname im Text auch ohne 's vorkommt. Damit wird vermieden, dass z.B. *Hershey's* fälschlich zu *Hershey* lemmatisiert wird.
3. Markiere die gefundenen Organisationen sowie deren Akronyme und Abkürzungen im Text, sofern nicht negative Kontexte für ein konkretes Vorkommen vorliegen. Die Markierung erfolgt durch einen Automaten nach dem leftmost-longest-Prinzip.

Mallchok (2005) evaluiert ihr System an einem Korpus verschiedener Zeitungen, das insgesamt aus 125 000 Token besteht und gut 4 000 Vorkommen von Organisationsnamen enthält. Precision, Recall und F-score liegen durchschnittlich etwa bei 94%, mit abweichenden Werten für einzelne Zeitungen.

5.1.4 Evaluation von Systemen zur NER

Die Evaluierung von computerlinguistischen Systemen ist eine der schwierigsten Aufgaben überhaupt. Bei NER-Systemen basiert sie auf einem Testkorpus, einem Gold-Standard, in dem Named Entities (in der Regel manuell) annotiert sind. Eine Evaluierung ist deshalb abhängig vom verwendeten Korpus und dem Standard zur Annotation (siehe Kap. 5.1.5.1.3), d.h. um NER-Systeme vergleichen zu können müssen diese an ein und demselben Testkorpus evaluiert werden. Das setzt selbstverständlich voraus, dass alle Systeme an das Testkorpus adaptiert sind, und umgekehrt das Testkorpus und seine Annotationen auch für eine Adaption geeignet sind. Vergleichbarkeit in diesem Sinne ist nur bei den Shared Tasks einiger Konferenzen (MUC-6, MUC-7, CoNLL-2003-Shared-Task) gegeben.

5.1.4.1 Precision, Recall, F-Wert

Die Evaluierung von Systemen zur Informationsextraktion im allgemeinen und zur Named Entity Erkennung im speziellen erfolgt durch die drei Maße ‚Precision‘, ‚Recall‘ und ‚F-Wert‘ (‚F-score‘, ‚F-measure‘). Diese sind im Information Retrieval allgemein verbreitet und akzeptiert (siehe Rijsbergen 1979: 134). Sie verwenden die beiden Parameter ‚gefundene Dokumente‘ und ‚relevante Dokumente‘ und setzen diese zueinander in Bezug. Übertragen auf die NER sind die beiden Parameter: n als Anzahl aller Eigennamen im Text und q als Anzahl der vom System als Eigennamen klassifizierten Sequenzen – Fehlklassifikationen eingeschlossen (siehe Makhoul et al. 1999, Nadeau & Sekine 2007). ‚Precision‘ und ‚Recall‘ sind dann:

$$recall\ r = \frac{n \cap q}{n}$$

$$precision\ p = \frac{n \cap q}{q}$$

Der F-Wert mittelt als gewichtetes harmonisches Mittel zwischen den Werten Precision und Recall:

$$F = \frac{1}{\alpha \frac{1}{p} + (1 - \alpha) \frac{1}{r}}$$

Falls Precision und Recall gleich gewichtet werden – dann auch F₁-Wert genannt, lässt sich die Formel mit $\alpha = 0.5$ vereinfachen zu:

$$F_1 = \frac{2pr}{(p + r)}$$

Ein Beispiel: Wenn in einem Dokument 100 Toponyme vorkommen, und das System 80 davon korrekt erkennt und außerdem 5 andere Einheiten (Wörter oder Wortfolgen) fälschlich als Toponyme annotiert, ergeben sich:

$$\begin{aligned} n &= 100 \\ q &= 85 \\ n \cap q &= 80 \\ \text{recall } r &= \frac{80}{100} = 0.8 \\ \text{precision } p &= \frac{80}{85} = 0.94 \\ F_1 &= \frac{2 \cdot 0.8 \cdot 0.94}{(0.8 + 0.94)} = 0.86 \end{aligned}$$

Als harmonisches Mittel ist der F₁-Wert nie höher als das arithmetische Mittel und tendiert in Richtung des kleineren der beiden Werte. Das bedeutet, dass der F₁-Wert Systeme, die Recall und Precision ausbalancieren gegenüber Systemen, die einen der beiden Werte optimieren, bevorzugt.

Üblicherweise werden Recall, Precision und F₍₁₎-Wert auf Ebene der Token berechnet. D.h. jedes Vorkommen einer Entität fließt in die Parameter n und q mit ein. Für Anwendungen wie Dokumentenklassifikation oder zur Benutzerführung in Suchmaschinen, wo es nur entscheidend ist, welche Entitäten einem Dokument zugeordnet sind, kann die Berechnung auch auf Type-Ebene erfolgen. D.h. jede Entität fließt nur einmal in die Berechnung ein, unabhängig davon, wie oft sie im Dokument vorkommt. Die erreichten Werte liegen bei einer Berechnung auf Type-Ebene in der Regel deutlich unter denen auf Token-Ebene, da seltenere Entitäten gleich wie häufige gewichtet werden, erstere aber meist schlechter von dem System erkannt werden (vgl. die Ergebnisse bei Sonntag 2007).

Auch in der Frage, wie partiell erkannte Named Entities zu werten sind, unterscheiden sich verschiedene Richtlinien zur Evaluierung (vgl. Nadeau & Sekine 2007: 19). Die einfachste Lösung ist, nicht exakte Matches, d.h. partielle oder überlappende als falsche Matches zu behandeln, und wird z.B. von den CoNLL-Konferenzen (siehe Sang & De Meulder 2003) so gehandhabt. Problematisch ist dabei, dass partielle Matches dann zweimal negativ zu Buche schlagen. Dies ist einerseits gerechtfertigt: wenn z.B. in der Phrase *in der Holsteinischen Schweiz* nur *Schweiz* als Named Entity annotiert wird, liegen auch zwei Fehler vor: die Phrase bzw. der Text dahinter enthält (1.) keine Referenz auf die Entität *Schweiz* und (2.)

die Entität *Holsteinische Schweiz* wurde nicht erkannt. Schwierig wird es bei Grenzfällen der Annotierung: Klassenbezeichner und Named Entity (*city of New York, New York City*), komplexe Entitäten, ev. mit eingebetteten Entitäten (*Frankfurt am Main*), wo ein partieller Match als besser gewertet werden sollte als gar kein Match.¹⁴⁴ Die Konferenzreihen MUC und insbesondere ACE verwenden deshalb komplexere Evaluationsverfahren, die hier feinere Gewichtungen vornehmen. Das gilt weiter auch für die Unter- und Feinklassifikation der Named Entities. Hier unterschieden MUC und ACE zwischen (a.) als NE erkannt und (b.) richtig klassifiziert.

5.1.4.2 Wann ist ein System gut?

Nun stellt sich auf dem Hintergrund der eben genannten Maße die Frage, wann ein NER-System als gut zu bezeichnen ist. Zwei Antworten sind möglich: (1.) Ein System ist gut, wenn es besser ist als vergleichbare Systeme. (2.) Ein System ist gut, wenn es die Qualität menschlicher Annotatoren erreicht.

Untersuchungen, um die Qualität menschlicher Annotatoren abzuschätzen, wurden bereits für die MUC-6 durchgeführt:

It was also unexpected that one of the systems would match human performance on the task. Human performance was measured by comparing the 30 draft answer keys produced by the annotator at NRaD with those produced by the annotator at SAIC. This test measures the amount of variability between the annotators. When the outputs are scored in “key-to-response” mode, as though one annotator’s output represented the “key” and the other the “response,” the humans achieved an overall F-measure of 96.68 and a corresponding error per response fill (ERR) score of 6%. The top-scoring system, the baseline configuration of the SRA system, achieved an F-measure of 96.42 and a corresponding error score of 5%.

(Sundheim 1996: 425)

Auch Roth (2002: 138) hat in ihrer Evaluation verschiedener Systeme zur Produktnamenerkennung als Vergleichsgruppe zwei Menschen die Aufgabe „von Hand“ erledigen lassen. Die beiden Annotatoren waren nicht perfekt, lagen aber mit 97% und 98% F₁-Wert deutlich vor dem besten maschinellen System (93,5%).

Damit lässt sich feststellen: ein NER-System, das nicht mindestens 96% F₁-Wert erreicht, mag zwar besser als andere sein, ist aber noch nicht gut genug. Erreicht werden solche Werte von den regelbasierten Systemen der MUC-6 für das Englische. Statistische Systeme sind nach wie vor etwas schlechter. Und die Werte für viele andere Sprachen bleiben weit hinter den 96% zurück.

5.1.5 Lokative Named Entities

Ortsangaben (Lokationen) bilden eine Teilmenge der (Named) Entities. Bevor tiefer auf die Erkennung lokativer Named Entities eingegangen wird (Kap. 5.1.5.2), ist die Frage zu klären, welche Einheiten in Texten als lokative Named Entities (kurz LOC oder GEO) gelten sollen und welche Richtlinien und Annotierungsstandards diesbezüglich existieren.

144 Allerdings besteht das Problem in erster Linie in der Notwendigkeit konsistenter, gut dokumentierter Annotierungsstandards und Korpora, die nach diesen Standards konsequent annotiert sind (vgl. Kap. 5.1.5.1.3).

5.1.5.1 Annotierung und Klassifikation

5.1.5.1.1 Probleme der Annotierung Bei der Annotierung von (lokativen) Named Entities (NEs) bestehen in einer Reihe von Fragen Uneinigigkeiten zwischen Annotatoren bzw. den Autoren von Standards und Richtlinien zur Annotierung. Die Problemfälle lassen sich in zwei Gruppen aufteilen, gemäß der paradigmatischen und syntagmatischen Ebene: (1.) Welche Einheiten eines Textes sollen als NEs gelten und welcher NE-Klasse sollen sie zugeordnet werden und (2.) wie wird in der „horizontalen“ Ebene annotiert, d.h. welche Elemente gehören zu einer NE und welche nicht. Die Probleme im Einzelnen sind:

1. Klassen von NEs und Zuordnung zu den Klassen (paradigmatische Ebene)
 - a) Was ist zu markieren, d.h. welche Ausdrücke sind (Named) Entities, welche nicht? Grenzfälle sind insbesondere:
 - i. mythische und fiktionale Namen: *Atlantis*
 - ii. Nicknames, siehe Kap. 3.3.2.9: *Bankfurt*
 - iii. monoreferentielle Ausdrücke, siehe Fußnote 139 auf S. 176: *die französische Hauptstadt*
 - iv. Derivationen, also Variationen in der Form der NEs
 - A. Adjektive: *französisch*
 - B. Bewohnernamen: *Berliner*
 - C. Eigennamen in deutschen Nominalkomposita (*Irakkrieg, Europareise*, vgl. Kap. 3.3.2.12 und 3.5.1.5), oder das Pendant dazu im Englischen – Eigennamen in attributiver Stellung (*the Clinton government, U.S. exporters*)
 - b) Wie sind (Named) Entities zu klassifizieren? Welche Kriterien werden zur Abgrenzung der verschiedenen Klassen von Entitäten herangezogen, insbesondere für die bereits diskutierten Metonymien (Kap. 3.1.4) und die davon betroffene lokative Klasse der FACILITIES (Krankenhäuser, Museen, Theater).
 - c) Granularität der Klassifikation: entweder werden nur wenige Klassen unterschieden („coarse-grained“), d.h. im Wesentlichen die drei Klassen der MUC-6 (PERS, ORG, GEO), oder die Unterteilung ist deutlich feiner („fine-grained“). Die NER entwickelt sich tendenziell hin zu einer feineren Unterteilung, siehe auch Ferrés et al. (2004), und die NE-Klassen von Sekine (S. 134).
2. Extension und interne Struktur der NEs (syntagmatische Ebene)
 - a) Tokenisierung des Textes und Annotierungsrichtlinien müssen aufeinander abgestimmt sein. Werden z.B. mit Bindestrich verbundene Wortformen als ein Token betrachtet, wird es unmöglich NEs korrekt zu annotieren: *a second daily Chicago-Paris flight*. Eine Annotierung unterhalb der Tokenebene ist ebenfalls nicht unproblematisch, vgl. Punkt 1(a)ivC.
 - b) Wie werden Namenszusätze, die der Disambiguierung und Präzisierung dienen, behandelt, z.B. *University of California, Los Angeles*? Unter den Lokativa sind insbesondere Namen wie *Frankfurt am Main, Washington, D.C.* betroffen.
 - c) Entitäten können ineinander verschachtelt sein (eingebettet, embedded, nested), z.B. ist der Standort/Sitz einer Organisation oft im Namen der Organisation enthalten: *Verwaltungsgericht München*. Ohne den Bestandteil *München* ist der Name nicht vollständig. Wegen ihrer Komplexität wird diese Problematik in Kap. 5.1.5.1.2 behandelt.

- d) Gehört der Klassenbezeichner in Kombinationen wie *Berg Sinai*, *Mississippi River*, *Himalaya Mountains* zum Namen oder nicht?
 - i. Gelten die Regeln ebenso für fremdsprachige Klassenbezeichner, die dem durchschnittlichen Sprecher der Sprache der annotierten Texte nicht verständlich sind? Z.B. in *Lairig Gartain* (*Lairig* gälisch ‚Pass‘).
 - ii. Wie werden Kombinationen aus Adjektivderivat und Klassenbezeichner annotiert? Als NE oder Kombination Adjektiv und Nomen, z.B. in *Dominikanische Republik* oder *Deutsche Demokratische Republik*, siehe S. 192.
- e) Wird der Artikel als mit zum Namen gehörig betrachtet in NEs wie *the Walt Whitman Bridge* oder *the Guggenheim Museum*? Wenn ja, gilt das auch für den Artikel in fremdsprachigen Eigennamen, z.B. *Le Havre* oder *El Salvador*?
- f) Wie werden Adjektivmodifikatoren gehandhabt, insbesondere solche wie *lower*, *upper*, z.B. in *lower Manhattan*?
- g) Wie werden Namensteile in Koordinationsellipsen annotiert? *North and South America*, *Nord- und Südamerika*, vgl. Kap. 3.3.2.10.

5.1.5.1.2 Eingebettete (lokative) Entitäten Lokative Entitäten können in andere Entitäten eingebettet sein und sind (notwendiger) Bestandteil eines komplexeren Names, z.B. einer Organisation wie *Verwaltungsgericht München* oder *Stadtsparkasse München*.¹⁴⁵ In der Regel besteht eine metonymische Beziehung zwischen Organisation und Ort, meist ist der Ort der Standort oder Sitz der Organisation. Ohne den Ortsnamen ist der Organisationsname nicht vollständig. Die metonymische Beziehung kann aber andererseits keinen Anspruch auf Ausschließlichkeit erheben: „the relation between The University of Chicago and the city of Chicago is about the same as that of the DePaul University and the city of Chicago.“ (Zaenen 2006: 577–8)

Es ist klar, dass primär der vollständige Name (der Organisation) zu annotieren ist, d.h. die engere und spezifischere Interpretation erhält den Vorzug. Erlauben die Richtlinien aber eine verschachtelte Annotierung, muss auch festgelegt werden, wie mit den eingebetteten Namen zu verfahren ist, d.h. wann diese selbst als NEs zu annotieren sind, und wann nicht. Voraussetzung ist die Annahme, dass eingebettete Elemente auch ihre ursprüngliche Bedeutung/Referenz erhalten, bzw. auf dem Hintergrund des IR, dass User eines Systems die eingebettete Entität als separierbares Element betrachten, das auch einzeln gesucht werden soll.

Das einzig sinnvolle Kriterium zur Unterscheidung zwischen eingebetteten Toponymen und Toponymen als „leeren“ Namensbestandteilen ist die Paraphrase. Ein Toponym innerhalb eines Organisationsnamens ist ein eingebettetes Toponym, wenn eine oder mehrere Paraphrasen möglich sind, d.h. wenn es in Kombination mit dem Einfügen weiterer Elemente wie Präpositionen und Veränderungen der morphologischen Form in andere Positionen verschoben werden kann. Wenn nicht, hat es seinen Ortsbezug (weitgehend) verloren.

- (5.1) a. *(die) Dresdner Bank*
- b. = *(die) Dresdner*
- c. ≠ *(die) Bank in Dresden*

¹⁴⁵ Daneben ist auch die Einbettung in Personennamen wie *Prinz Ferdinand von Bayern*, *Earl of Argyll* (vgl. Byrne 2007, Pouliquen et al. 2006) möglich.

- d. \neq^* (die) *Bank Dresden*
- (5.2) a. (die) *Stadtsparkasse München*
 b. $*$ (die) *Münchner*
 c. \approx (die) *Stadtsparkasse in München*
 d. $=$ (die) *Münchner Stadtsparkasse*
- (5.3) a. (die) *Technische Universität München*
 b. $=$ (die) *Technische Universität in München*
 c. $=$ (die) *Technische Universität zu München*
 d. $=$ (die) *Münchner Technische Universität*
 e. $*$ (die) *Münchner*

Während in (5.1) bis auf die Ellipse des Klassenbezeichners (5.1b) keine Paraphrasen möglich sind, sind (5.2) und (5.3) mehrfach paraphrasierbar, die Ellipse ist aber nicht möglich. In der Praxis mag die Entscheidung oft diffizil und schwierig sein, da sie an eine Beurteilung der Grammatikalität gebunden ist.

Erlauben die Richtlinien eine verschachtelte Annotierung nicht, muss eine Entscheidung getroffen werden, ob die aufeinanderfolgenden Entitäten als eine Einheit oder Entitätenfolge annotiert werden soll. Hier greifen verschiedene Standards zur Annotierung seit Chinchor, Robinson & Brown (1998) (nicht aber Chinchor 1997) auf einen operationalisierbaren, aber zweifelhaften Test zurück: So ist *University of California, Los Angeles* vollständig als Organisation zu annotieren – und nicht als Folge ‘ORG, LOC’ –, da eine Abkürzung (‘Initialismus’) *UCLA* für den gesamten Ausdruck existiert. Als weiteren Test – eine Konsequenz der Paraphrase – nennen Lee & Lee (2005) die Verdoppelung des eingebetteten Ortsnamens nach dem Schema ‘X Y in X’. Ist die resultierende Phrase (z.B. *Beijing University in Beijing*) in Korpora bzw. durch die Abfrage einer Internetsuchmaschine belegt, gilt das fragliche Element ‘X Y’ (hier *Beijing University*) als eine zusammenhängende Entität und X (*Beijing*) nicht als Entität des Typs LOC.

Die Schachtelung kann auch mehr als zwei Ebenen umfassen: [[[*Edinburgh*]_{PLACE} *University*]_{ORG} *Library*]_{ORG} (Byrne 2007). In äquivalenten deutschen Beispielen erhöht sich dann die Anzahl der möglichen Paraphrasen (siehe auch Kap. 3.5.1.5) deutlich: *Universitätsbibliothek München*, *Münchner Universitätsbibliothek*, *Bibliothek der Universität München*, usw., vgl. (5.1), (5.2) und (5.3).

5.1.5.1.3 Standards zur Annotierung Zur Annotierung (lokativer) Named Entities existieren diverse Standards, die sich außer bei der Behandlung der Problemfälle (Kap. 5.1.5.1.1) noch in den folgenden Punkten unterscheiden:

1. Welche Informationen können neben der bloßen Markierung der Named Entity annotiert werden, z.B. Grundform, normalisierte Namen, Koordinaten, Größe, historische Informationen, usw. Nur wenige Standards wie die TEI:P5 (s.u.) erlauben eine tiefe und reichhaltige Annotierung.
2. Wie erfolgt die Markierung im Text:
 - zeilen/tabellenbasiert (z.B. CONLL-2003-SHARED-TASK, Sang & De Meulder 2003),
 - mittels einer Auszeichnungssprache (SGML/XML),

- oder außerhalb des Textes mit Pointern, die auf Positionen im Text verweisen (z.B. Collier et al. 2002).

Von allgemeinen Standards zur Annotierung linguistischer Korpora unterscheiden sich die Standards zur Annotierung von Named Entities darin, dass sie (1.) Eigennamen explizit als Wortart/Klasse vorsehen und klar definieren,¹⁴⁶ (2.) Named Entities weiter in Klassen – mind. Personen, Organisationen und Lokationen – unterteilen und (3.) Eigennamen als atomare Einheiten ansehen und nicht oder nur sekundär in Bestandteile zerlegen. Vor allem der letzte Punkt ist es, der allgemeine linguistisch annotierte Korpora weitgehend unbrauchbar im Zusammenhang mit der NE-Erkennung macht. So werden im Brown-Korpus (Francis & Kučera 1970, 1982) Toponyme wie *Blue Ridge* oder *Central Falls* als Folge Adjektiv–Nomen annotiert. Gleiches gilt für deutsche Korpora (hier das TIGER-CORPUS): *Genfer See*, *Rotes Meer*, *Deutsche Demokratische Republik*. Eine Annotierung der gesamten Phrase als Eigenname – und sei es oberhalb der Tokenebene – erfolgt nicht.

Die Tendenz bei der Annotierung von Named Entities und in der Folge bei der Entwicklung von Systemen, die im Rahmen der Shared Tasks verschiedener Konferenzen anhand der annotierten Korpora trainiert und/oder evaluiert werden, geht von der „bloßen“ Markierung von Eigennamen im Text hin zu komplexeren, und nur teilweise verstandenen Aufgaben, wie Erkennung von (räumlichen) Relationen, Anapherauflösung, usw. Dies macht es – zusammen mit der Tatsache, dass die Arbeit des Annotierens als weniger privilegierte Aufgabe angesehen wird, nicht einfacher, Annotierungsstandards zu entwickeln und zu etablieren (vgl. kritisch Zaenen 2006).

5.1.5.1.3.1 MUC-6 und Nachfolger Die ersten Standards und Korpora für die NER starten mit der MUC-6, die ja auch die Entitätenerkennung an sich als Einzelaufgabe und Shared-Task formulierte (MUC-6-NE-TASK). Die folgenden Konferenzen (MUC-7: Chinchor 1997; Hub-4: Chinchor, Robinson & Brown 1998; ACE: Chinchor et al. 1999, Grishman 2000, LDC 2008a) und die LDC:LCTL:NE-GUIDELINES (siehe auch LDC:LCTL) folgen den Richtlinien der MUC-6 in weiten Teilen, treffen mitunter aber auch andere Entscheidungen für die gleichen Beispiele. Einen Überblick über die Gruppe der auf den MUC-6-Richtlinien aufbauenden Standards gibt Tab. 5.2. Alle versuchen Named Entities auf pragmatische Weise mit einfachen Tokenisierungsregeln unter Verwendung von SGML zu annotieren. Kritik an den Richtlinien ist bei der Behandlung von Metonymien und verschachtelten Named Entities angebracht, vgl. Rössler (2007: 13–19). Die Richtlinien vermitteln mitunter den Eindruck, es handle sich um eine Aneinanderreihung von während des Annotierens getroffenen ad-hoc-Entscheidungen, die sich z.T. widersprechen. Ein Beispiel: Adjektivmodifikatoren werden ausdrücklich als nicht zur Named Entity gehörig bezeichnet (Chinchor 1997, Chinchor, Robinson & Brown 1998, Chinchor et al. 1999). In *lower Manhattan* ist demnach nur *Manhattan* zu annotieren. In Chinchor (1997) wird aber an anderer Stelle *Northern California* inklusive dem Adjektiv als NE annotiert, was nicht nachvollziehbar ist, da sowohl *Northern California* als auch *lower Manhattan* nicht-offizielle, unscharfe Regionen darstellen, die aber beide so prominent sind, dass ihnen in EN.WIKIPEDIA.ORG ein eigener Eintrag gewidmet ist. Immerhin, Chinchor, Robinson & Brown (1998), Chinchor et al. (1999) schließen dann *Northern California* explizit aus, die LDC:LCTL:NE-guidelines versuchen hier eine Regelung über offizielle Bezeichnungen, Häufigkeit des Gebrauchs und „agreed-upon definition“ zu finden.

146 Ein Blick in das Brown-Korpus (Francis & Kučera 1970, 1982) zeigt z.B., dass *German* teils als Eigenname ('np'), teils als Adjektiv ('jj') annotiert ist.

Detailfrage	siehe	MUC-6	MUC-7	HUB-4	ACENER (1999)	ACE EDT (2000)	ACE EDT (LDC)	LDC:LCTL
		MUC-6-NE-TASK	Chinchor (1997)	Chinchor, Robinson & Brown (1998)	Chinchor et al. (1999)	Grishman (2000)	LDC (2008a)	LDC:LCTL:NE- GUIDELINES
mythische und fiktionale Namen	1(a)i		+	+	+		+	+
Nicknames	1(a)ii	+	+	+	+			+
Adjektivderivate	1(a)ivA				–	+	(+)	+
Bewohnernamen	1(a)ivB					– (als PERS)	+ (z.B. GPE.PER)	
Namen in attrib. Verw.	1(a)ivC		+	+	+		+	+
metonymische Verwendung			v	v	v	v	v ^a	v
Gesamtzahl Klassen	1c	3	3	3	3	5	5	4 ^b
Anzahl lokativer Klassen		1	1	1	1	3 ^c	3 ^d	1
Facilities (mit mgl. Ausnahmen) als		ORG	ORG	ORG	ORG/LOC	(eigene Klasse)	(eigene Klasse)	ORG/LOC (Metonym)
Spezifikatoren (<i>Washington, D.C.</i>)	2b ^e	(–)	(–)	–	–	–	±/v	–
eingebettete NEs werden annotiert	2c ^f	–	–	–	–		+/v	–
Klassenbez. als NE-Bestandteil	2d		v	v	v		+/v	+
Artikel als Namensbestandteil	2e		–/v	+/–	+		+	+
Adjektivmodifikatoren	2f ^g	+	+/–	+/–	–		+/v	–/v
ellipt. Bestandteile in Koord.-ell.	2g ^h	+	–	+	+			+
Possessivmarker 's als Teil der NE		–	–	–	(–)		–	+

a Annotierung der Metonymien, falls „Namen“ zusammenfallen (z.B. GPE.ORG: *France signed a treaty*), sonst („Nickname Metonymies“) unter der Klasse des realen Referenten (ORG: *New York defeated Boston 99-97 in overtime.*)

b Vierte NE-Klasse: Personal titles or roles (‘TTL’)

c FACILITY, GSP (geographical-social-political entity), LOCATION (geographische Regionen und Formationen, Gewässer).

d Siehe Fußnote *c*, weitere Unterteilung in viele Subklassen.

e falls –, dann als separate Entität annotiert: <LOC>Washington</LOC>, <LOC>D.C.</LOC>

f falls –, wird längstmögliche Sequenz annotiert

g falls +, als NE-Bestandteil: <LOC>Northern California</LOC>

h falls +, werden elliptische Bestandteile separat annotiert: <LOC>North</LOC> and <LOC>South America</LOC>; falls –, dann <LOC>North and South America</LOC>

Tabelle 5.2: Detailfragen in den Richtlinien der MUC-6 und davon abgeleiteten Standards zur Annotierung von Named Entities, siehe Erläuterungen in Kap. 5.1.5.1.1 ab S. 189. Zeichenklärung: [+] Element wird als NE (oder Bestandteil) annotiert; [–] wird nicht (als NE) annotiert; [v] spezifische Regeln. Angaben in Klammern sind aus Beispielen erschlossen.

Zwei Tendenzen sind bei den in Tab. 5.2 beschriebenen Systemen auszumachen: (1.) Metonymien werden als solche behandelt und kontextabhängig annotiert: „we tag for meaning“ (LDC:LCTL:NE-GUIDELINES, v6.5). In die gleiche Richtung geht, Derivate, vor allem Adjektive als Instanzen von Named Entities, zu annotieren, da sie ja die gleiche Referenz haben („if the entity name occurs in the form of an adjective you should also tag it.“)

Dies ist auch auf dem Hintergrund des Paradigmenwechsels weg von regelbasierten hin zu statistischen NER-Systemen zu sehen. Die Richtlinien der MUC-6 und MUC-7 sind z.B. bei der Behandlung von Metonymien eindeutig zugunsten von Systemen formuliert, die sich auf Gazetteers und interne Kontexte stützen. Lernbasierte Systeme greifen dagegen verstärkt auf externe Kontexte zurück. Was soll ein solches System lernen, wenn in ‘Spiel X gegen Y’ X und Y in Abhängigkeit von der internen Struktur mal als LOC mal als ORG annotiert wird. Der sicher richtige Hinweis in Chinchor (1997), dass die Auflösung von Metonymien eine Aufgabe auf höherer Ebene sei – zusammen mit der Auflösung von Anaphern, hilft dem Entwickler eines NER-Systems hier nicht weiter.

5.1.5.1.3.2 Auflösung von Metonymien Genau diesen Widerspruch versuchen Markert & Nissim (2002, 2007b) zu lösen, indem sie die metonymische Verwendung (siehe auch Kap. 3.1.4) von Named Entities als zusätzliche Ebene der Annotation einführen:

(5.4) `<LOCATION reading="metonymic" metotype="place-for-people">Hungary</LOCATION> took similar actions ...`

Markert & Nissim (2007b) unterscheiden mehrere metonymische Verwendungen, bei den Toponymen sind es: `place-for-people`, `Cap-Gov`, `place-for-event`, `place-for-product`. Die geographische und politische Lesart von Gebietskörperschaften wird als „literale“, nicht-metonymische Verwendung annotiert. Insgesamt geben Markert & Nissim (2007b) ausführliche Regeln an, welche Verwendung nun als „metonymisch“ zu annotieren ist, treffen dabei aber z.T. fragwürdige Entscheidungen, wie ein Blick auf die Daten der Konferenz SemEval-2007¹⁴⁷ zeigt:

(5.5) `Perceptions of product parity are highest in markets like <annot><location reading="metonymic" metotype="othermet" notes="Japan as market"> Japan </location></annot>, where product proliferation is wildest.`

Japan wird hier ganz klar als eine Referenz auf die geopolitische Einheit verwendet, wäre also als ‘literal’ zu annotieren.¹⁴⁸

Von metonymischer Verwendung zu trennen sind auch Ellipsen. ‚Schulische Ausbildung in X‘ kann kaum als metonymische Lexart von *X* bezeichnet werden, die Anzahl der (metonymischen) Lesarten für ein Lexem wäre sonst sehr hoch. In (5.6) liegt eine eliptische Konstruktion vor!

147 <http://nlp.cs.swarthmore.edu/semEval/tasks/task08/data/trial.tar.gz> (Januar 2008)

148 *Market* in der Bedeutung ‚Absatzmarkt‘ [*X ist Absatzmarkt für Y = X ist eine Region, in der die Güter Y (oder die Güter, die von der Organisation Y hergestellt werden) verkauft („abgesetzt“) werden*] ist ein Prädikat mit *Japan* als (semantisches) Argument.

- (5.6) Without urgent action, he added, British education would fall behind Spain, `<annot><location reading="metonymic" metotype="othermet" notes="Education in Italy"> Italy </location></annot>`, and Greece.

5.1.5.1.3.3 Tiefe Annotation: TEI und Spatial ML Der Annotierungsstandard der Text Encoding Initiative (TEI, TEI:P4, TEI:P5, vgl. Southall 2003) versucht im Gegensatz zu den oben besprochen Standards nicht Anweisungen für das Wie der Annotation zu geben, sondern überlässt die Detailentscheidungen dem Annotator. Stattdessen ermöglicht er die tiefe Annotation von Entitäten und versucht diese in weiten Teilen zu vereinheitlichen. Um Toponyme und allgemein Ortsangaben in Texten zu markieren und mit Metainformation anzureichern, stehen Auszeichnungen/Tags zur Verfügung, die den Standard auch als mögliches Format für einen Gazetteer qualifizieren. Einige Beispiele sollen hier genügen:

- Angabe einer Normalform, hier ein Beispiel aus Southall (2003) in einem historischen Text

(5.7) Thence 6 miles to `<placeName reg="Blandford Forum">Blandford</placeName>`,
thence 18 to `<placeName reg="Salisbury">Salsebury</placeName>`.

- Auszeichnung von komplexen Ortsangaben (vgl. Kap. 3.6.1)

(5.8) `<placeName key="Duncan">
<measure unit="km" quantity="17.7">11 miles</measure>
<offset>Northwest of</offset>
<settlement type="city">Providence</settlement>,
<region type="state">RI</region>
</placeName>`

- Annotation von in Toponymen enthaltenen Klassenbezeichnern, inkl. fremdsprachigen Klassenbezeichnern (hier *Lairig*)

(5.9) The isolated ridge separates two great corridors which run from
`<name key="GLC01" type="place">Glencoe</name>` into
`<geogName key="GLET1" type="glen">
<geog reg="glen">Glen</geog>
<name>Etive</name>
</geogName>`, the
`<geogName key="LAGA1" type="hill">
<geog lang="gaelic" reg="sloping hill face">Lairig</geog>
<name>Gartain</name>
</geogName>` and the
`<geogName key="LAEI1" type="hill">
<geog lang="gaelic" reg="sloping hill face">Lairig</geog>
<name>Eilde</name>
</geogName>`

- Metainformationen zu Ortsangaben und -namen

(5.10) `<place type="cemetery">
<placeName>Protestant Cemetery</placeName>
<placeName type="official" xml:lang="it">Cimitero Acattolico</placeName>
<location type="geopolitical">
<country>Italy</country>
<settlement>Rome</settlement>`

```

    <district>Testaccio</district>
  </location>
  <location type="address">
    <address>
      <addrLine>Via Caio Cestio, 6</addrLine>
      <addrLine>00153 Roma</addrLine>
    </address>
  </location>
</place>
(5.11) <place type="city">
  <placeName notBefore="1400">Lyon</placeName>
  <placeName notAfter="0056">Lugdunum</placeName>
  <location>
    <gml:Polygon>
      <gml:exterior>
        <gml:LinearRing> 45.256 -110.45 46.46 -109.48 43.84 -109.86 45.8 -109.2
          45.256 -110.45 </gml:LinearRing>
      </gml:exterior>
    </gml:Polygon>
  </location>
</place>

```

- zu einer Taxonomie geschachtelte Elemente, bis hin zu einer multilingualen Taxonomie (5.13)

```

(5.12) <place>
  <country>Lithuania</country>
  <country xml:lang="lt">Lietuva</country>
</place>
  <place>
    <settlement>Vilnius</settlement>
  </place>
  <place>
    <settlement>Kaunas</settlement>
  </place>
</place>
(5.13) <place xml:id="wales" type="country">
  <placeName xml:lang="cy">Cymru</placeName>
  <placeName xml:lang="en">Wales</placeName>
  <placeName xml:lang="la">Wallie</placeName>
  <placeName xml:lang="la">Wallia</placeName>
  <placeName xml:lang="fro">Le Waleis</placeName>
  ...
</place>
</place>

```

- explizite Angabe von Relationen zwischen Orten

```

(5.14) <listPlace>
  <place type="country" xml:id="FRA">
    <placeName>France</placeName>
  </place>
  <place xml:id="MASC">
    <placeName>Mascarene islands</placeName>
    <placeName>Mascarenhas Archipelago</placeName>
  </place>
</listPlace>

```

```

    <place xml:id="REN">
      <placeName>Réunion</placeName>
    </place>
    <relation name="partOf" active="#REN" passive="#FRA #MASC"/>
  </listPlace>

```

SPATIALML, ein möglicher, aber noch im Entstehen begriffener Standard, stammt anderes als TEI aus dem Bereich der NE-Erkennung, zeigt aber große Ähnlichkeit. Ziel ist einerseits zusätzlich zur bloßen Markierung der Ortsnamen eine tiefere Annotierung inklusive Disambiguierung und Geokodierung zu ermöglichen, andererseits auch komplexe Ortsangaben für die maschinelle Verarbeitung im Sinne des Spatial Reasoning (Kap. 5.5) aufzubereiten:

(5.15) *He arrived, in a disturbed state, in [Rochester], [Illinois].*

```

<PLACE id=1 state="IL" country="US" form="NAM">Rochester</PLACE>
<PLACE id=2 state="IL" country="US" type="STATE" form="NAM"
  gazref="IGDB:30125575">Illinois</PLACE>
<LINK source=1 target=2 linkType="IN"/>

```

Zur automatischen Auflösung der komplexen Ortsangaben werden die einzelnen Bestandteile explizit über topologische Relationen (LINK) und Vektoren (PATH) verknüpft:

(5.16) *a [town]₁ some [50 miles]₂ [south]₃ of [Salzburg]₄ in the central [Austrian]₅ [Alps]₆*

```

<PLACE type="PPL" id=1 form="NOM" ctv="TOWN">town</PLACE>
<SIGNAL id=2>50 miles</SIGNAL>
<SIGNAL id=3>south</SIGNAL>
<PLACE id=4 type="PPLA" country="AU" form="NAM">Salzburg</PLACE>
<PLACE id=5 type="COUNTRY" country="AT" mod="C">Austrian</PLACE>
<PLACE id=6 type="MTS" >Alps</PLACE>
<PATH id=7 distance="50:mi" direction="S" source= 4 destination=1 signals="2 3"/>
<LINK id=8 source=1 target=6 linkType="IN"/>
<LINK id=9 source=6 target=5 linkType="IN"/>

```

Die Verlinkung von Entitäten – und vieles andere – übernimmt SPATIALML von den Annotierungsrichtlinien der ACE, genauer der Teilaufgabe des „Relation Detection and Characterization“. Die aktuellen Richtlinien dazu (LDC 2008b) sehen für die Relationen typisierte Argumente und Selektionsrestriktionen vor. Für die Lokativa relevant sind die drei Relationen `Physical.Located`, `Physical.Near` und `Part-Whole.Geo`.

5.1.5.1.3.4 Geographical Markup Ziel von Standards zum ‚Geographical Markup‘ ist die Verknüpfung räumlicher und sprachlicher Information, bzw. die Einbettung von durch Koordinaten explizit kodierten Ortsangaben in Texte (siehe auch Kap. 5.2). Drei Standards zur Auszeichnung seien hier erwähnt:

- Geographical Markup Language (GML), siehe Smart, Abdelmoty & Jones (2004) und Beispiel (5.11),
- Toponym Resolution Markup Language (TRML, Leidner 2004a, 2007),
- Keyhole Markup Language (KML), die von Google in verschiedenen Projekten verwendet wird.

5.1.5.2 Erkennung geographischer Entitäten in deutschen Texten

Eine zentrales Problem dieser Arbeit ist die Erkennung geographischer Entitäten in deutschen Texten. Aus der Beschränkung auf Orte und das Deutsche als Objektsprache folgen die Fragen: (1.) Was unterscheidet geographische Entitäten von den anderen Typen von Named Entities? (2.) Welche Besonderheiten ergeben sich bei der Anwendung auf deutschsprachige Texte im Gegensatz zu anderen Sprachen, namentlich dem Englischen?

Im Vergleich zu den beiden NE-Klassen Person und Organisation sind die Lokationen durch folgende Eigenschaften ausgezeichnet:

- geographische Entitäten sind häufig Teil anderer Entitäten, und zwar von
 - Personennamen, siehe Fußnote 145 auf S. 190
 - Organisationsnamen, siehe Kap. 5.1.5.1.2
 - komplexen Ortsnamen (*Nußdorf ob der Traisen*)
- ein Gazetteer/Lexikon ist notwendig, um brauchbare Ergebnisse zu erzielen, vgl. die Diskussion des Aufsatzes von Mikheev, Moens & Grover (1999) auf S. 178
- die Performance der Systeme liegt für Lokationen meist im oberen Bereich, deutlich vor den Organisationen, aber meist hinter den Personen¹⁴⁹

Besonderheiten für das Deutsche im Kontrast zum Englischen und anderen europäischen Sprachen sind:

- Großschreibung der Nomina. Während es für das Englische ausreichend ist, alle großgeschriebenen Wörter (außer am Satzanfang) einer Eigennamenklasse zuzuordnen, muss im Deutschen zuerst zwischen Nomen und Eigennamen unterschieden werden.¹⁵⁰ Damit entfällt der vermutlich wichtigste Anhaltspunkt bei der Erkennung von Named Entities.¹⁵¹
- Flexion der Eigennamen, vgl. Kap. 3.3.2.6

Diese Eigenschaften führen dazu, dass die Baseline – definiert als der Wert, den ein System erreicht, das nur im Trainingskorpus eindeutige Entitäten annotiert – für deutsche Texte verglichen mit englischen deutlich schlechter ist. Die Baseline des CONLL-2003-SHARED-TASK liegt für den deutschen Part bei 30% F₁-Wert, während das Englische 60% erreicht. Dieser Zustand schleppt sich weiter, so dass auch die Endergebnisse für das Deutsche gut 15% hinter denen des Englischen zurückbleiben (Sang & De Meulder 2003).

149 So beim CoNLL-2003-Shared-Task für alle drei Werte (Precision, Recall, F₁-Wert): Personen > Lokationen > Organisationen.

Schnetzler (2004) sieht für die Systeme KIM, GATE und LINGPIPE die Lokationen sogar vor allen anderen Typen.

150 Außerdem werden viele von geographischen Namen abgeleitete Adjektive (*deutsch, schwedisch*) kleingeschrieben.

151 Ein Blick auf den Einfluss dieser Eigenschaft beim Tagging ermöglicht eine grobe Abschätzung: Versucht ein Part-of-Speech-Tagger (hier der TNT-TAGGER) zwischen Nomina/Appellativa und Eigennamen zu unterscheiden, entfallen 25% aller Fehler auf diese Unterscheidung (Müller & Ule 2001).

5.1.5.2.1 Überblick über implementierte Systeme Christine Thielen (1995) implementiert ein inkrementelles, bootstrapping-artiges Verfahren: unbekannte Eigennamen werden aufgrund des internen und externen Kontexts erkannt, das Eigennamenlexikon so erweitert. Interessant ist das System durch die Anpassungen der manuell erarbeiteten Heuristiken an deutsche Besonderheiten. So nutzt es vor allem den Null-Artikel in Kombination mit markierten Genitivformen als positiven Kontext. In Bezug auf Toponyme verwendet Thielen Suffixlisten von Siedlungsnamen (siehe Fußnote 156 auf S. 208) und Derivationen (*Münsteraner* etc.). Das Verfahren erreicht (für alle Typen von Eigennamen) eine hohe Precision bei relativ niedrigem Recall ($p = .98$, $r = .65$, $f = .78$).

Das System von Volk & Clemenide (2001) arbeitet im Prinzip regel- und lexikonbasiert, es nutzt Regeln z.B. zur Erkennung unbekannter Familiennamen nach einem bekannten Vornamen und das „learn-apply-forget“-Prinzip (siehe Kap. 5.1.2.3). Die Erkennung von geographischen Namen geschieht allerdings ausschließlich listen- bzw. lexikonbasiert. Verwendet werden Listen mit etwa 1000 Städte- und 250 Ländernamen inklusive der Genitivform (immer auf -s), außerdem Derivationen (Bewohner, Adjektive) aus der Analyse des Systems GERTWOL. Für ein Testkorpus mit 990 Sätzen und 166 enthaltenen geographischen Namen berichten Volk & Clemenide (2001) eine Precision von 81% und einen Recall von 91%.

Piskorski & Neumann (2000), Neumann & Piskorski (2002) beschreiben ein auf endlichen Automaten und „shallow text processing“ basierendes System, das kurze Namenslisten und ein dynamisches Lexikon (learn-apply-forget, vgl. Kap. 5.1.2.3) verwendet. Bei der Erkennung von Lokationen erreichen Neumann & Piskorski (2002: 32) 75,11% Recall und 88,20% Precision auf einem 20 000 Worte-Testkorpus aus der Zeitung „Wirtschaftswoche“.

Das NER-System von Marc Rössler (2002) basiert auf einem Trigramm-Markov-Modell, das TNT-TAGGER getaggten Korpus deutscher Texte trainiert wird. Rössler trainiert das Markov-Modell in zwei verschiedenen Konfigurationen – (a) nur auf den Annotationen, (b) auf den Annotationen und Wortformen, die im Korpus eindeutig als positiver externer oder interner Kontext auftreten. Alle Konfigurationen werden entweder durch einen learn-apply-forget-Filter, der alle an anderer Stelle innerhalb eines Dokuments gefundenen Entitäten annotiert, oder durch einen vollständigen, aus dem Korpus selbst gewonnenen Gazetteer unterstützt. Für Lokationen sind nur die mit Hilfe des Gazetteers erzielten Ergebnisse akzeptabel mit 77% Recall und 86% Precision: „The use of gazetteers seems inevitable, especially for the category LOCATION. Any further improvement seems hard to achieve without it, even when they have just been extracted from an annotated corpus.“ (Rössler 2002: 36) Wird das Markov-Modell nur auf den POS-Tags trainiert, werden für Lokationen gerade einmal 42% Precision und 5% Recall erreicht, was (Rössler 2002: 35) mit den wenig spezifischen Kontexten der Lokationen begründet: „LOCATIONS are often surrounded by prepositions like *in* or *aus* which are very frequent in German and not specific for LOCATIONS.“ Die Evaluation erfolgt anhand eines 100 000 Token umfassenden Korpus der Computer Zeitung 1998. Das Korpus ist beschrieben bei Rössler 2002: 31 und Kühnlein 2003: 138–41, und wird in einer adaptierten Version auch für diese Arbeit verwendet, siehe S. 205.

Rössler & Morik (2005), Rössler (2007) trainieren eine Support Vector Machine (SVM) auf den deutschen Korpora des CoNLL-2003-Shared-Task. Im Anschluss werden durch Bootstrapping auf einem nicht-annotierten Korpus mit 40 Mio. Wortformen (Frankfurter Rundschau) die gelernten Pattern erweitert. Während die Ergebnisse für Personennamen beeindruckend sind und sich bereits nach wenigen Bootstrappingläufen bei 91% F-Wert stabilisieren,

können sie für Lokationen und Organisation mit erreichten 76% bzw. 65% nach acht Läufen nicht überzeugen. Sie liegen damit hinter den besten des CONLL-2003-SHARED-TASK.

Der CONLL-2003-SHARED-TASK hatte wie der des Vorjahres das Ziel, die Entwicklung sprachunabhängiger NER-Systeme zu fördern, und zwar in Form eines Wettbewerbs zwischen den teilnehmenden Systemen. Dazu werden vom Veranstalter annotierte Trainings- und Testkorpora gestellt, anhand derer die Systeme trainiert und evaluiert werden. Zu erkennen sind vier NE-Kategorien (PERSON, ORGANIZATION, LOCATION, MISCELLANEOUS). Um die Sprachunabhängigkeit der Systeme zu gewährleisten werden diese auf zwei Sprachen evaluiert. 2002 waren es die Sprachen Spanisch und Holländisch, 2003 Englisch und Deutsch. Die Korpora für jede Sprache umfassen jeweils etwa 300 000 Token. Zur Qualität der deutschen Korpora, siehe Kap. 5.1.5.2.2 auf S. 204.

2003 nahmen sechzehn Systeme teil, die sich wie folgt charakterisieren lassen:

- alle Systeme benutzen Gazetteers
- zahlreiche Algorithmen des Maschinellen Lernens fanden Anwendung (AdaBoost, Conditional Random Fields, Hidden Markov Models, Maximum Entropy Models, Memory-Based Learning, Recurrent Neural Networks, Robust Risk Minimization, Support Vector Machines, System Combination, Transformation-Based Learning, Voted Perceptrons)
- manche Systeme benutzen weiter unannotierte Korpora
- das beste System (Florian et al. 2003) benutzt zusätzlich zwei externe NER-Systeme
- (Florian et al. 2003) erreichen als bestes System für deutsche Lokationen (LOC): 71,59% Recall, 80,19% Prec. und 75,65% F₁-Wert

Carola Kühnlein (2003) übernimmt die Regeln von Volk (2002) und überträgt sie in die „Konzeptsensor“-Sprache der Firma Eurospider, eine Sprache zur Mustererkennung. Die Eigenamen- und Triggerlisten von Volk (2002) erweitert sie. Die resultierenden Listen deutscher Ortsnamen enthalten 18 552 Einträge, davon 15 192 Städte und 544 „Nationen“. Darin sind 1 839 bzw. 186 Genitivformen enthalten – eine Zahl die viel zu klein ist! Weitere Kasusformen sind nicht vorgesehen. Das Lexikon verteilt sich, bedingt durch die Eigenheiten und Beschränkungen der „Konzeptsensor“-Sprache, auf über 30 verschiedene Listen (Ein- und Mehr-Wort-Städtenamen, Städtenamen mit Slash, Genitivformen, auf sehr einfache Weise gewonnene Adjektivderivationen, u.a.). „Bei der Erkennung von geographischen Eigennamen werden keine Regeln angewendet. Die Erkennung basiert lediglich auf den Ressourcen“, d.h. den Lexikonlisten (63). Sie ist allerdings abhängig von Personen- und Firmennamenerkennung. Kühnlein verwendet deshalb folgende Entscheidungskriterien (103–6):

1. ein Toponym-Kandidat nach einer Personenanrede wird als Personennamenname klassifiziert (*Herr Rombach*)
2. ein Mehr-Wort-Toponym-Kandidat, der auch eine Person bezeichnen könnte, wird als Toponym klassifiziert (*St. Petersburg*)
3. ein Kandidat, ambig zwischen Toponym und Vorname, wird als Toponym klassifiziert (*Austin*)
4. steht ein toponymisches Adjektiv vor einem Firmennamen gehört dieses nicht zum Firmennamen (*die Düsseldorfer Mannesmann AG*)

Resultate für	Recall	Precision
Personennamen	78.46	84.14
geographische Namen	84.34	72.61
Firmennamen	14.39	58.62
deutsch (alle)	44.68	73.80

Tabelle 5.3: Performance Kühnlein (2003)

Die Evaluation erfolgt anhand der von Marc Rössler annotierten Texte (vgl. auf S. 199). Recall und Precision liegen bei 84 bzw. 73% (siehe Tab. 5.3).

Pouliquen et al. (2004, 2006), Steinberger & Pouliquen (2007) verwenden einen Gazetteer-Lookup kombiniert mit Geo-Stopwort-Listen (siehe S. 178). Externen Kontext berücksichtigen Pouliquen et al. nicht – Versuche zeigten angeblich keine Verbesserungen. Lediglich (ambigue) geographische Namen, die Teil eines Personennamens sind werden als Ortsnamen ausgeschlossen. Die Ergebnisse einer Evaluation liegen nur für NE-Erkennung in Kombination mit Geokodierung vor ($f = .73$, $p = .80$, $r = .68$ für deutsche Toponyme), sind also nicht vergleichbar, da die Werte für die reine Erkennung besser sein müssen.

In seiner Diplomarbeit erweitert Kimler (2004) die Methoden von Pouliquen et al. (2004, 2006) um folgende Schritte:

- der Gazetteer-Lookup erfolgt in zwei Schritten
 1. nur Namen der Klassen COUNTRY, REGION, CAPITAL, MAJOR CITY werden nachgeschlagen. Wenn mehr als drei Orte oder mind. 50% aller Orte im Text in einem Land (COUNTRY) liegen, wird der Text diesem „Geo-Kontext“ zugeordnet
 2. anschließend erfolgt ein Lookup im vollständigen Gazetteer, allerdings werden nur Orte, die in dem in 1 ermittelten Geo-Kontext liegen, markiert
- etwa 15 Klassenbezeichner je Sprache dienen als externer Kontext für Ortsnamen
- eine Liste mit wichtigen Persönlichkeiten dient zum Ausschluss unwahrscheinlicher Ortsnamen, z.B. *Annan*

Die Ergebnisse des Verfahrens liegen für deutsche Toponyme bei 72% (Precision, Recall, F_1 -Wert), wiederum inklusive der Geokodierung und somit nicht vergleichbar.

Michael Sonntag (2007) stellt ein System zur „Erkennung von Toponymen in deutschsprachigen Webseiten“ vor, ausgehend von einem engeren Toponymbegriff, der sich auf Oikonyme bis zur Ebene eines Dorfes oder Stadtteils, Hydro- und Oronyme und Namen für Länder, Regionen und geopolitische Einheiten beschränkt. Das System basiert auf UIMA, Eclipse und Java, die wichtigsten Verarbeitungsschritte sind (vgl. Abb. 5.1.5.2.1):

1. ein POS-Tagger markiert Eigennamen, die als Toponym-Kandidaten betrachtet werden. Durch den Tagger kann die Precision um 15% verbessert werden. Der Recall fällt mit Tagger allerdings um etwa 8%, verglichen mit einer auf großgeschriebene Wörter beschränkten Kandidatenmenge und bei Verwendung eines reduzierten Gazetteers (s.u.)
2. zusätzliche Toponym-Kandidaten werden durch Muster (räumliche Präpositionen, Klassenbezeichner, Postleitzahlen, für Ortsnamen charakteristische Prä- und Suffixe) extrahiert. Dieser Schritt ist notwendig, um den Recall zu verbessern, da die getesteten und verwendeten Tagger viele Ortsnamen als Substantive und nicht als Eigennamen annotieren, was zu einem maximalen Recall von ca. 80% führt (81)

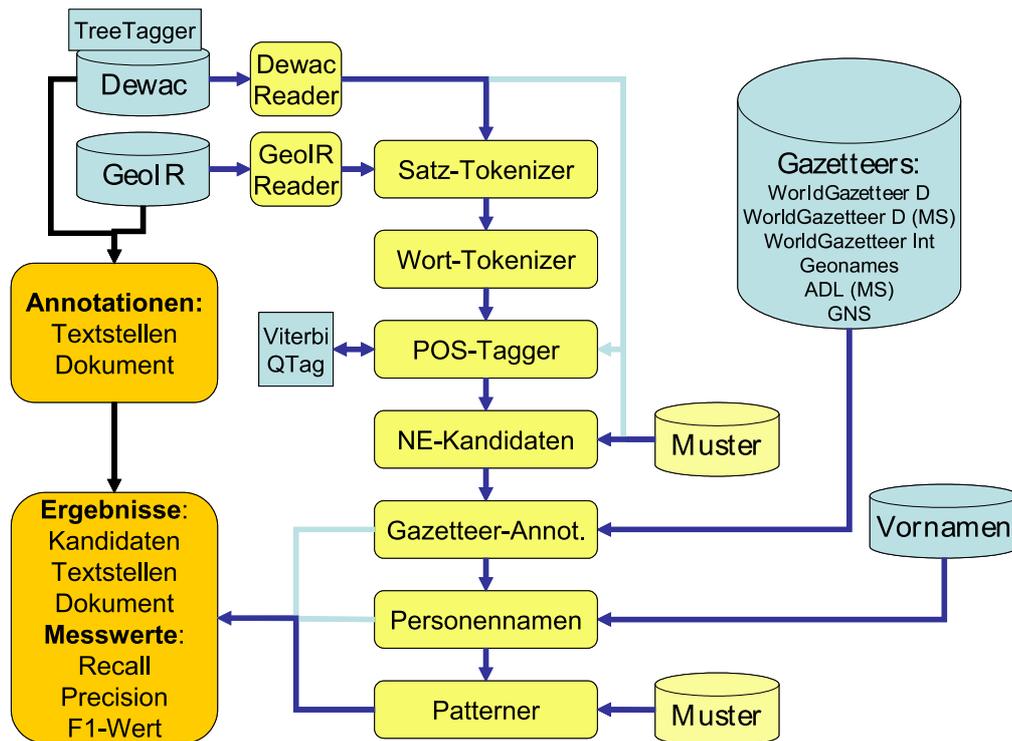


Abbildung 5.1: Schematischer Aufbau des Systems zur „Erkennung von Toponymen in deutschsprachigen Webseiten“ von Sonntag (2007: 72).

3. die Toponym-Kandidaten werden anhand eines Gazetteer verifiziert. Die besten Ergebnisse wurden mit dem deutschen Teil des WORLD-GAZETTEER erzielt, der außerdem um Länder und Regionen ergänzt wurde. Die anderen getesteten Gazetteers (NGA:GNS, geonames.org, ADL) ließen die Precision auf Werte zwischen 50–75% sinken.

Vor der Gazetteerabfrage wurden die Toponym-Kandidaten außerdem um Mehrwortgruppen, Genitivformen, präfigierte Formen (*Süd-*) erweitert, was den Recall um 2% ohne Verluste bei der Precision erhöht

4. eine weitere Verbesserung der Precision um 0,7% bei minimalem Einfluss auf den Recall bringen
 - „Killer-Muster“ (*Frau, Herr* vor einem Toponym, häufige Fehlerkennungen wie *Stein, Meine*) und
 - sichere „konstruktive Regeln“ für Toponyme (vorangestellte Klassenbezeichner oder Postleitzahlen) und andere NE-Klassen (Personen- und Organisationsnamen: Vornamen, *AG, GmbH*, usw.)

Der Wert der Arbeit von Sonntag liegt in den zahlreichen Experimenten, die den Einfluss der einzelnen Komponenten auf die Performance (Recall und Precision) des Systems untersuchen. Das System erreicht eine vergleichsweise gute Performance mit 81% Recall, 95% Precision und 87% F₁-Wert bezogen auf die beste Konfiguration und das beste der drei gemessenen Korpora. Die Ergebnisse für die anderen beiden zur Evaluation verwendeten Korpora sind deutlich schlechter, siehe Tab. 5.4. Die einzelnen Korpora bestehen aus je 100 Dokumenten und zwar:

Dewac 1: eine (mehr oder weniger zufällige) Auswahl (100 000 Token) aus dem Dewac-Korpus

des WACKY-Projektes (deutschsprachige Webseiten)

Dewac 2: ein Teilkorpus des Dewac-Korpus (90 000 Token): enthält Webseiten die mind. einen der fränkischen Ortsnamen *Nürnberg*, *Bamberg* und *Bayreuth* enthält und inhaltlich auf die Region Franken bezogen ist

GeoIR: inhaltlich auf Bayern bezogene deutschsprachige Webseiten (27 000 Token)

Ob es gerechtfertigt ist, die meisten Messungen und Experimente am inhaltlich sehr eingeschränkten Korpus Dewac 2 durchzuführen, sei dahingestellt. Eine grobe Schätzung der über alle drei gemittelten Werte liegt bei 73% Recall, 93% Precision und 82% F₁-Wert.

		Korpus und Konfiguration	p	r	f
Tabelle 5.4: Performance Sonntag (2007)		Dewac 2, Konfiguration mit zwei POS-Taggern	81	95	87
		Dewac 2, Standardkonfiguration	79	94	86
		GeoIR, Standardkonfiguration	75	93	83
		Dewac 1, Standardkonfiguration	63	90	74

Die folgenden Systeme lassen aufgrund der spärlichen Dokumentation kaum Aussagen zu ihrem Aufbau zu:

Schilder, Versley & Habel (2004) beschreiben ein System, das neben der Erkennung deutscher Ortsnamen, auch noch deren Disambiguierung (Geokodierung), Unterscheidung zwischen lokativer und nicht-lokativer Verwendung, und anaphorische Referenzen (z.B. durch Klassenbezeichner) leisten soll.

Martins, Silva & Chaves (2005) erreichen mit einem multilingualen System zur Erkennung und zur Geokodierung von Ortsnamen auf deutschen Webtexten nur für die Erkennung 72% Recall und 79% Precision.

SAILLABS-NER verwenden in ihrem System zur Erkennung von Entitäten in mehreren Sprachen, darunter Deutsch, ein Bigramm-Markov-Modell. Das System ist als Online-Demo verfügbar, und wurde anhand der zur Evaluierung erstellten Korpora (Kap. 5.1.5.2.2) getestet.

System / Autor	Prec.	Rec.	F₁	Typ / Strategie
Sonntag (2007)	95	81	87	Tagger, Rule, Gaz.
Volk & Clematide (2001)	81	91	86	Rule, Gaz., L-A-F
Rössler (2002)	77	86	81	Gaz., Tagger, Markov, L-A-F
Neumann & Piskorski (2002)	88	75	81	FSA, Gaz., L-A-F
Kühnlein (2003: 176)	73	84	78	Rule, Gaz.
Florian et al. (2003)	80	72	76	Gaz., MachLearn, NER
Schilder, Versley & Habel (2004)	94	74	83	Tagger
Martins, Silva & Chaves (2005)	79	72	75	Rule, Gaz./Ontol.
SAILLABS-NER	72	65	68	Markov (undok., da kommerziell)

Tabelle 5.5: Performance von Systemen zur Toponymerkennung in deutschen Texten. Die Werte sind nur sehr eingeschränkt vergleichbar, siehe die Anmerkungen im Text.

Einen Überblick über die Performance verschiedener NER-Systeme bei der Erkennung deutscher geographischer Entitäten gibt Tab. 5.5. Die Vergleichbarkeit ist jedoch gering, da meist

wenig über die zur Evaluierung verwendeten Korpora bekannt ist und sich die Standards, was als Named Entity bzw. Toponym annotiert wird, beträchtlich unterscheiden.

5.1.5.2 Deutsche Korpora mit NE-Annotationen Named-Entity-annotierte Korpora sind zur Evaluierung oder zum Training von NER-Systemen notwendig, ein deutsches Korpus ist hier unverzichtbar zur Evaluierung des eigenen Systems (Kap. 5.1.5.3.2).

Aufgrund der zur Annotation notwendigen Arbeit ist der naheliegende Gedanke, bereits vorhandene Korpora dafür zu nutzen. Allerdings eignen sich allgemeine wortarten- oder phrasenannotierte deutsche Korpora (z.B. TIGER-CORPUS oder NEGRA) nicht für diese Aufgabe, da sie (1.) nur eine Klasse ‘Eigennamen’ annotieren und nicht verschiedene Typen (Personen-, Ortsnamen, usw.) unterscheiden und (2.) viele komplexe Toponyme als Folge von Nicht-Eigennamen analysieren, z.B. *Dominikanische Republik*, *Mainzer Straße* (siehe S. 192).

Weiter bietet sich das deutsche Korpus des CONLL-2003-SHARED-TASK¹⁵² an: es ist mit dem Ziel der NER annotiert und wird vielfach für die NER verwendet (neben den Teilnehmern des Shared Task selbst, z.B. von Rössler & Morik 2005, Leveling & Hartrumpf 2006). Die deutschen Trainings- und Testkorpora sind ein Ausschnitt aus dem ECI Multilingual Text Corpus und entstammen der Frankfurter Rundschau aus dem Jahr 1992. Die Annotation der Named Entities erfolgte manuell an der Uni Antwerpen. Die Annotationen sind speziell was die Lokativa betrifft von fragwürdiger Qualität (vgl. auch Rössler 2007: 19–24). Wohnernamen und Adjektive werden mal als Entität der Klasse LOC, mal als MISC annotiert. Viele Nicht-Lokativa und damit klare Fehlannotationen wie *Burgfestspiele*, *Horkheimers* (hier der Soziologe, und nicht der Bewohner von Horkheim), *StadtteilRundschau* oder *BRD-Federführungen*, *Zwangstaufe* sind darunter, aber vor allem ist die Abgrenzung zwischen Orten im allgemeinen und lokativen Entitäten unglücklich: Neben Ländern, Städten, Straßen, Bauwerken etc. sind auch Schulen, Universitäten, Altenheime, Kirchengemeinden, Cafés und Restaurants als lokative Entitäten markiert. Zwar ist es gerechtfertigt und auch sinnvoll eindeutige Entitäten wie den *Bürgertreff Bockenheim* oder die *Deutsche Bibliothek* als Named Entities aufzunehmen, diese sind auf Landkarten oder Stadtplänen ja üblicherweise verzeichnet (vgl. ‚topographisches Objekt‘, S. 93). Wenn allerdings nur ein Klassenbezeichner (*Gartenschauland*, *Altortsmuseum*, *Bahnbetriebskantine*, *Bahnhof*, *Basilika*) auf diese Entität verweist, ist eine Markierung als Named Entity kaum zu rechtfertigen. Auch bei der Entscheidung, ob ein Klassenbezeichner mit zum Namen gehört, wäre eine konsequente Linie sinnvoll, und mit Absprache auch möglich gewesen (z.B. (*das Land*) *Hessen* vs. (*das Land*) *Hessen*, *Am Wiesenhof* vs. *Straße Am Kirchholz*; *Finanzplatz Berlin*, *Bankenplatz Berlin* vs. *Berlin*, (*Gemeindehaus*) *Christ-König* vs. *Bürgertreff Bockenheim*). Und schließlich bestehen Uneinigigkeiten, ob bei Kombinationen aus Straße und Hausnummer letztere als Teil der Ortsangabe zu annotieren sind, oder nicht. Gleiches gilt für Namensbestandteile wie *am Main*.

Klare Richtlinien zur Annotation fehlen, es existiert lediglich eine Übersicht,¹⁵³ welche Objekte zu welchen Klassen gehören. Eine Abschätzung der Fehler anhand der ersten 200 Lokationen aus dem Testkorpus A ergab eine Rate von 8–20% falschen und fehlenden Annotationen, je nachdem, was als falsch bewertet wird (vgl. die Aufstellung im Anhang auf S. 437). Dass

152 Siehe S. 200 für eine allgemeine Beschreibung des CONLL-2003-SHARED-TASK.

153 <http://www.cnts.ua.ac.be/conll2003/ner/annotation.txt>

mit solchen Trainings- und Testdaten selbst vom besten System (Florian et al. 2003) nur marginale 76% F_1 -Wert für Lokationen erreicht werden, verwundert nicht. Bezeichnend ist aber, dass von Seiten der Teilnehmer keine Kritik an der Qualität der Daten geäußert wurde. Allein die Algorithmen des maschinellen Lernens stehen im Vordergrund, die praktische wie linguistische Relevanz der Aufgabe wird nicht hinterfragt.

Wegen des Fehlens guter allgemein verfügbarer deutscher Named-Entity-annotierter Korpora, wurden in zwei Korpora manuell die lokativen Entitäten annotiert. Bei den Texten handelt es sich um einen Tag Süddeutsche Zeitung (vom 27. 05. 2003) und zehn zufällig¹⁵⁴ ausgewählte Texte aus der WIKIPEDIA.DE. Beide Texte wurden von mir annotiert. Ein drittes Korpus (Computer Zeitung) wurde an der Uni Duisburg annotiert (vgl. Rössler 2002: 31 und Kühnlein 2003: 138–41, bzw. siehe oben, S. 199) und mir von Marc Rössler freundlicherweise zur Verfügung gestellt. Zwei Korrekturen schienen nötig: zum einen wurden Derivationen von Toponymen separat markiert, zum anderen wurden Kombinationen aus Klassenbezeichner und Toponym (*Landgericht Stuttgart*) einheitlich annotiert. Insgesamt enthalten diese drei Korpora 210 000 Wortformen und knapp 3 500 geographische Entitäten (Details in Tab. 5.6).

5.1.5.3 Ein System basierend auf lokalen Grammatiken

In dem System zur Erkennung geographischer Entitäten, das hier vorgestellt wird, übernehmen lokale Grammatiken die Aufgabe, sowohl positive wie negative Kontexte für geographische Entitäten zu beschreiben. Die positiven Kontexte sind dabei sprachliche Einheiten wie (Teil-)Sätze, Präpositional- und Nominalphrasen. Die negativen Kontexte sind gleichzeitig positive Kontexte für Homonyme von geographischen Namen (z.B. *beim Essen*) oder für andere Typen von Entitäten (Sportvereine, Personen).

Bereits erste Versuche zeigten, dass die Beschränkung auf Prädikate mit geographischen Entitäten in Argumentposition einen zu geringen Recall liefern. Die Precision für (Teil-)Sätze wie *Nhum wohnt in X*, wobei *X* als geographische Entität (Stadt, Land etc.) im Lexikon verzeichnet sein muss, ist erwartungsgemäß optimal. Die Schwierigkeit besteht allerdings darin, genügend Prädikat-Argument-Strukturen zu beschreiben und diese in realen Texten mit zahlreichen Einschüben, Koordination auf verschiedenen Ebenen etc. zuverlässig zu erkennen. Auch positive Kontexte auf Phrasenebene sind im Recall zu schlecht. Deshalb werden zumindest eindeutige geographische Namen, d.h. Toponyme wie *Augsburg*, *Stuttgart*, nicht aber *Essen*, *Bergen* (siehe S. 108 für weitere Beispiele) auch ohne positiven Kontext als Entitäten annotiert.

Wichtiger Punkt beim Design des Systems war die Erweiterbarkeit: Lokale Grammatiken mit neuen positiven und negativen Kontexten können jederzeit hinzugefügt werden und die Performance des Systems verbessert sich – vorausgesetzt, die Precision der hinzugefügten Grammatik ist nicht schlechter als die von konkurrierenden Grammatiken oder einem Lexikon-Lookup allein. Die lokalen Grammatiken werden jede für sich angewendet. Anschließend wird anhand der numerischen Konkordanzen (*concord.ind* bei UNITEX) die endgültige Konkordanz berechnet. Die lokalen Grammatiken bzw. deren Konkordanzen werden dabei hierarchisch abgearbeitet, d.h. ein bereits bestehender „Treffer“ kann nur von einem längeren „Treffer“ überschrieben werden. Es gilt also das Prinzip des längsten Treffers, nicht das bei der Anwendung von Automaten, UNITEX eingeschlossen, übliche ‚leftmost-longest‘-Prinzip.

154 WIKIPEDIA bietet eine Funktion zur Auswahl eines zufälligen Artikels.

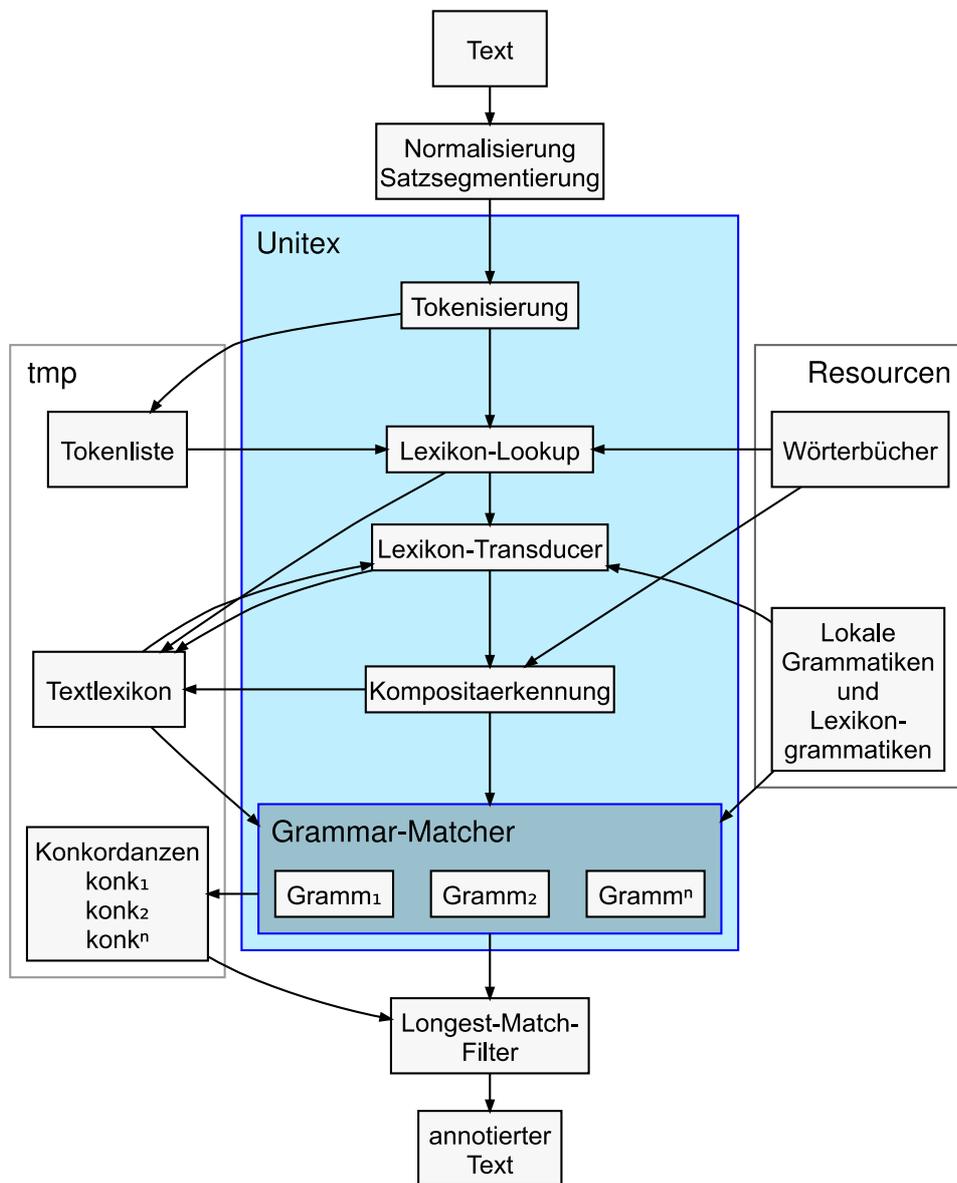


Abbildung 5.2: Verarbeitungspipeline, verwendete Ressourcen und Zwischenresultate bei der Toponymerkennung.

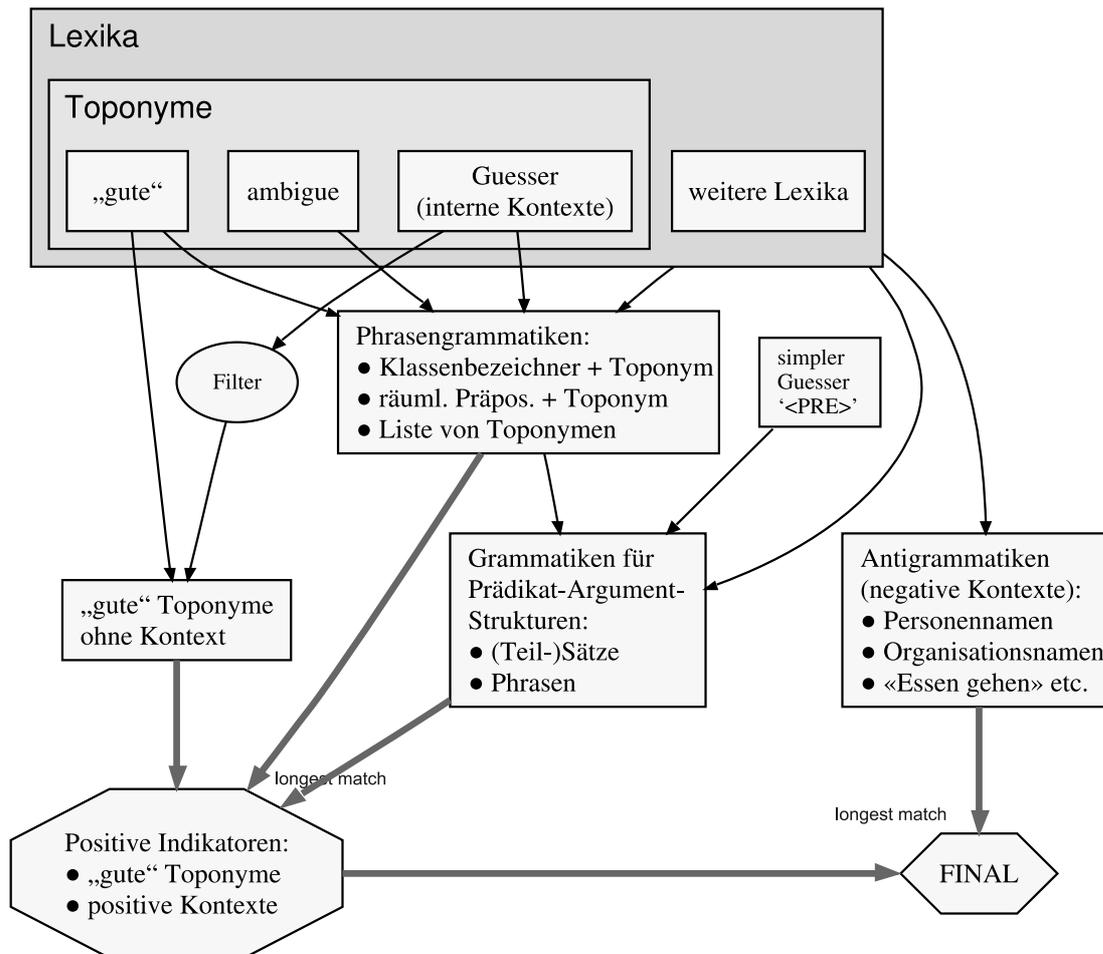


Abbildung 5.3: Abhängigkeiten zwischen den verschiedenen zur Toponymerkennung benutzten Lexika und Grammatiken.

Abb. 5.2 zeigt schematisch den Aufbau des Systems mit den einzelnen Verarbeitungsschritten. Die Abhängigkeiten zwischen den benutzten Lexika und lokalen Grammatiken bis hin zum Endergebnis ist in Abb. 5.3 dargestellt. Aus technischen Gründen (Tokenisierungsregeln von UNITEX) werden keine Toponyme unterhalb der Tokenebene erkannt, z.B. in *Irakkrieg*, *Deutschlandreise* (siehe auch Kap. 3.3.2.10 und 3.5.1.5).¹⁵⁵ Bindestrich-Komposita werden ebenso behandelt, da der Unterschied artifiziell und nicht linguistisch motiviert ist (*Irak-Krieg*).

5.1.5.3.1 Die lokalen Grammatiken des Systems

Im Folgenden werden die lokalen Grammatiken des Systems, ihre Intention und die erzielte Wirkung kurz beschrieben. Für Details der Evaluierung siehe Tab. 5.6.

5.1.5.3.1.1 Erkennung unbekannter Toponyme Im Lexikon nicht vorhandene Toponyme¹⁵⁶ werden erkannt (1.) anhand einer Suffixliste (siehe Fußnote 85 auf S. 97) und (2.) einer Reihe von Pattern, die (2a) aus mehreren Token bestehende Toponyme (z.B. Straßennamen, *Franz-Werfel-Straße*, *Place de la Concorde*), (2b) ausländische Ortsnamen mit generischen Bestandteilen (*Monte Brè*, *Lido di Jesolo*) oder spezifischem Aufbau (*Omonville-la-Rouge*), (2c) Buchstaben-Ziffern-Kombination für bestimmte „bürokratisierte“ Toponymtypen (Autobahnen *A9*, *N 13*) beschreiben, Weitere Pattern beschreiben (3.) Kombinationen aus Straße und Hausnummer (siehe Kap. 3.5.5).

Ein Guesser zur Erkennung von Ein-Wort-Toponymen, realisiert als lokale Grammatik (siehe Abb. B.9 und B.10), behandelt nur diejenigen Token, die nicht im Lexikon verzeichnet sind. Verhindert werden damit Fehlanalysen wie *Straßendorf*, *Gebirgsdorf*, *Marschhufendorf*, die aufgrund des Suffixes *-dorf* als Siedlungsnamen „geraten“ würden in Analogie zu *Unterpulendorf* oder *Ebermannsdorf*. Der Guess unbekannter Toponyme erfolgt aber noch vor der allgemeinen deutschen Kompositaerkennung. Diese würde z.B. die beiden letzten Ortsnamen in *unter-Pulle-* bzw. *Eber-Mann-Dorf* segmentieren, d.h. als „korrekt“ gebildete deutsche Nomina klassifizieren.

Dass die Beschreibung von Ortsnamen anhand der internen Struktur (vgl. Kap. 3.2.1.1) sehr schwierig ist, zeigt der geringe Recall (Tab. 5.6). Er ist aber auch deshalb so gering, da viele Ortsnamen bereits im Lexikon stehen und deshalb nicht mehr durch den Guesser erkannt werden. Die Precision liegt auch ohne externen Kontext im akzeptablen Bereich, im Zusammenspiel mit externen Kontexten ist sie gut.

5.1.5.3.1.2 Positive Kontexte nutzen die in Kap. 3.3.2 beschriebenen syntaktischen Eigenschaften von Toponymen. Als positive Kontexte wurden verwendet:

155 Von demselben Problem berichten Meyers et al. (2004b,a) im Zuge des Aufbaus der NOMBANK (Argumentstrukturen von Nomina). Hier sind es die Tokenisierungsregeln der Penn Treebank, die z.B. *Chicago-Paris flight* so tokenisieren, dass die Argumente nicht mehr erkennbar sind.

156 Zum Vergleich: Thielen (1995) benutzt eine Liste mit Suffixen (*-acker*, *-aich*, *-beuren*, *-hafen*, *-hausen*, *-stetten*, *-weiler*) und Präfixen (*Mittel-*, *Ost-*, *West-*, *Zentral-*) und versucht auch abgeleitete Adjektive anhand der Suffixe *-er* und *-aner* zu erkennen.

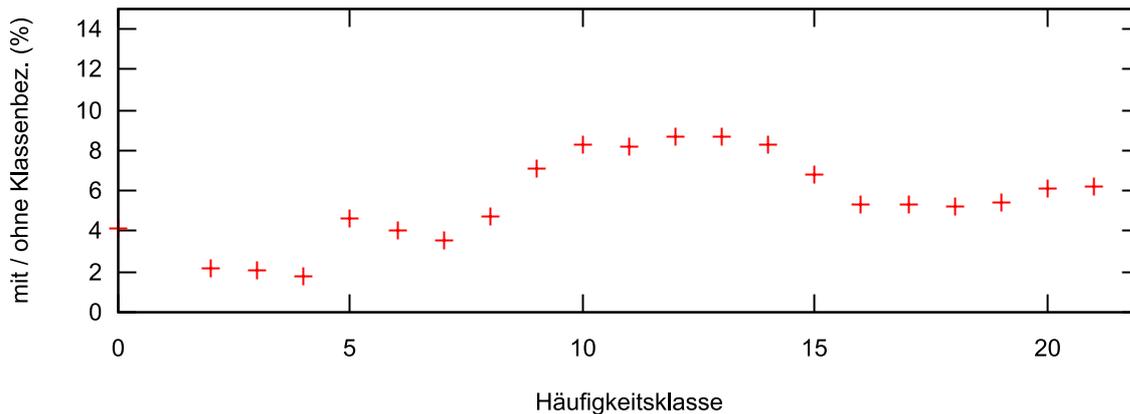


Abbildung 5.4: Anteil der durch Klassenbezeichner „gekennzeichneten“ Toponyme in Bezug zur Häufigkeitsklasse. Kontext ist eine lokative Präpositionalphrase, um das „Rauschen“ durch ambigüe Toponyme (*Essen*, usw.) möglichst zu minimieren. Der Klassenbezeichner selbst muss in Apposition zum Toponym stehen, d.h. in Relation stehen Häufigkeiten von Ausdrücken wie *in der Stadt X* und *in X*.

Die Häufigkeitsklasse gibt an, um welche Zweierpotenz das häufigste Wort (Toponym) häufiger als das gegebene ist. Um einen ungefähren Eindruck zu vermitteln, hier Beispiele für einige Klassen: 0: *Deutschland*, 2: *Berlin*, *USA*, 5: *Spanien*, *Moskau*, *Rhein*, 8: *Bukarest*, *Zwickau*, *Philadelphia*, *Mosel*, 10: *Waiblingen*, *Attersee*, *Röthenbach/Pegnitz*, *Katar*, *Gibraltar*, *Ärmelkanal*, 15: *Thüningersheim*, *Schönwies*, *Neuenburg am Rhein*, *Maikäfersiedlung*, *Campbell River* 21: *Steinach bei Mering*, *Leitzachtalstraße*. Demnach ist *Deutschland* 2^{10} mal häufiger als *Attersee*. Die Klasse wird durch $N = \left\lfloor \frac{1}{2} - \log_2 \left(\frac{freq \text{ aktuellesWort}}{freq \text{ häufigstesWort}} \right) \right\rfloor$ (als Wörter zählen hier nur Toponyme) berechnet.

- Kombinationen aus Klassenbezeichner und Toponym (siehe Kap. 3.5.1.3) werden in der Literatur meist als „der“ positive Kontext für Toponyme bezeichnet. Dass der damit erzielbare Recall gering ist, bemerkt z.B. Friburger (2006: 648): „Seuls 20% des noms de lieux ont un contexte gauche et parfois droit (ex.: *l'estuaire de la Seine*, *la mer Baltique*).“ Im Deutschen liegt dieser Wert mit ca. 6% deutlich niedriger – bedingt durch die Einbettung vieler „Klassenbezeichner“ in ein Nominalkompositum (*Ostsee*). Die immer wieder geäußerte These,¹⁵⁷ dass seltenere Toponyme häufiger mit Klassenbezeichner verwendet werden, konnte nur teilweise bestätigt werden. Nur sehr häufige Toponyme werden selten (< 5%) mit Klassenbezeichner verwendet. Für die ersten zehn Häufigkeitsklassen, die zusammen 2000 Toponyme enthalten, steigt der Wert an. Dann bleibt er mit Schwankungen bei deutlich unter 10% (siehe Abb. 5.4). Gerade für wirklich seltene Toponyme, die selbst in einem Toponymlexikon mit einigen zehntausend Einträgen fehlen, ist der Kontext des Klassenbezeichners keine wirksame Strategie.¹⁵⁸
- Präpositionalphrasen, bestehend aus räumlicher Präposition, (ev. Klassenbezeichner) und

157 Kimler (Z.B. 2004: 27), vgl. auch das Zitat von Mikheev, Moens & Grover (1999) auf S. 180.

158 Mögliche und höchst spekulative Erklärungen für dieses Phänomen und die nicht bestätigte, aber plausible Hypothese sind: (1.) Seltene Toponyme sind länger und verfügen damit über mehr inneren Kontext, vgl. die Beispiele bei Abb. 5.4. (2.) Die Verteilung über die Toponymklassen ist für verschiedene Häufigkeitsklassen unterschiedlich. In den obersten Häufigkeitsklassen sind durchweg Staaten, Großstädte und Kontinente zu finden, während die unteren „bunt“ gemischt sind. (3.) Die Beschränkung auf den Kontext einer räumlichen Präposition mag die Daten verfälscht haben. (4.) Klassenbezeichner können auch anaphorisch verwendet werden (*Glendalough ... Das kleine Dorf in den Wicklows ...*), was hier nicht erfasst wurde, und wovon vermutlich auch kein NER-System Gebrauch macht.

Toponym, geben mit deutlich höherem Recall einen positiven Kontext.

- In Aufzählungen und Listen von Toponymen können nicht nur ambigue Toponyme (*Weil, Bergen*) verifiziert, sondern auch unbekannte, sofern diese im Inneren der Liste eingebettet sind, sicher erkannt werden. Trennsymbole in den Listen können neben Komma und Konjunktionen auch typographische Symbole (z.B. ‘•’) sein.
- Prädikat-Argument-Strukturen, d.h. ‚linguistischen Situationen‘, an denen Toponyme als Aktanten beteiligt sind, stellen perfekte Kontexte dar, so perfekt, dass es möglich ist unbekannte Toponyme hier sicher allein an ihrer Großschreibung (<PRE>-Guesser) zu erkennen. Ihre Beschreibung in Form formaler Grammatiken erfordert jedoch einen hohen Aufwand, deshalb bleibt der Recall für diese Kontexte sehr klein.

5.1.5.3.1.3 Negative Kontexte sind notwendig, um Ambiguitäten von möglichen Toponymen aufzulösen und Ortsnamen als Bestandteil von anderen Namen zu erkennen und entsprechend zu annotieren.

Die Grammatiken zur Auflösung von Ambiguitäten bringen zwar nur Verbesserungen der Precision im Promille-Bereich, beseitigen aber „peinliche“ und trivial zu behebende Fehler, z.B. fälschlich erkanntes *Rom* in *CD-ROM*, *A4* (die Autobahn) in *DIN A4*, ‚Schweden‘ in *hinter schwedischen Gardinen*. Auch häufige Fehlerkennungen von ambiguen Toponymen, die durch einen positiven Kontext fälschlich validiert sind, werden so rückgängig gemacht. Z.B. wird der nicht seltene, aber ambigue Ortsname *Lachen* im Kontext einer räumlichen Präposition validiert: *in Lachen*. Bei anderen Präpositionen können aber unerwartete Seiteneffekte auftreten, so bei der Phrase *vor Lachen* (analog zu *vor München*). Diese Fälle aufzulösen, wird versucht, eine wirkliche Auflösung ist oft nur mit weiterem Kontext möglich, da die Beispiele ja oft tatsächlich ambig sind (so auch *an der Leine*).¹⁵⁹ Eine weitere Grammatik versucht homonyme Bewohnerbezeichnungen (*die Schweden, die Bayern, die Niedersachsen*) anhand des vorangestellten in Kasus und Numerus konkurrierenden Artikels zu erkennen.

Deutlich wirkungsvoller ist eine Grammatik, die allgemein Personennamen zu erkennen versucht, mit dem Ziel Ambiguitäten zu Toponymen aufzulösen. Für letztere hier einige Beispiele aus dem Evaluationskorpus (SZ) — die ambiguen Bestandteile sind hervorgehoben: *Heinrich Linz, Birgit Schönau, Cathrin Münster, Charles Darwin, Minister Peter Frankenberg (CDU), Gerd Landsberg, Markus Dornach, Cathleen Naundorf, Lars von Trier*. Die Erkennung der Personennamen erfolgt mittels einem Lexikon der Vor- und Nachnamen, von Funktionen und Berufsbezeichnern, und lokalen Grammatiken, die Anrede (*Mr.*), Titel (*Dr.rer.hort.*), Adels-Vorsatzworte (*von und zu*), nachgestellte Namenszusätze (*MdB, OSB*) und natürlich die Muster, wie alle Bestandteile kombiniert werden können, beschreiben.

159 Vgl. Satzpaare wie:

- (1) a. *Frank Müller (42) stammt aus Essen.*
b. *Der Rest des Abends besteht aus Essen und Fernsehen.*
- (2) a. *... Joint-venture ..., das vom Standort Hannover aus ein bundesweites Telekommunikationsnetz aufbauen will. Noch ist weder der Name noch der Starttermin des neuen Telecomunternehmens an der Leine klar.* (so im Evaluierungskorpus Computerzeitung)
b. *Herr Schmidt führt seinen Hund Waldi an der Leine.*

Ortsnamen sind häufig Bestandteil von Namen anderer Typen von Entitäten, z.B. Sportvereinen, Firmen, Institutionen. Die Problematik, ob in diesem Fall der Ortsname auch als Entität indiziert werden soll, wurde bereits in Kap. 5.1.5.1.2 diskutiert. Zwei Gebiete wurden dabei versucht durch „Antigrammatiken“ abzudecken: (1.) Namen von Sportvereinen inklusive der metonymischen Verwendung von Ortsnamen im Sport und (2.) in Namen von Firmen und Organisationen enthaltene Toponyme. Die Grammatiken aus dem Bereich Sport beschreiben einerseits den Aufbau von Vereinsnamen (Abb. B.11), andererseits aber auch weitere Kontexte, nämlich Phrasen und Sätze mit den Inhalten ‚X spielt gegen Y‘ (*Italiens Niederlage im EM-Finale gegen Frankreich, beim 5:1 gegen Kaiserslautern*) oder ‚Xhum spielt bei Y‘ (*der Linksaußen beim Zweitligisten TSG Oßweil*). Die Grammatik der Organisationsnamen fällt kleiner aus, sie konzentriert sich auf notorische Fälle wie *Bayern 3 Radio, Washington Post, New York Times* und in Namen von Universitäten, Gerichten usw. eingebettete Ortsnamen. Diese werden von der Grammatik auch als eingebettete annotiert.

Jede der großen „Antigrammatiken“ (Personen, Organisationen, Sport) verbessert für sich genommen die Precision des Systems um 2–5%, siehe Tab. 5.6. Alle drei zusammengenommen und mit den kleinen Antigrammatiken kombiniert führt das aber zu immerhin 15% höherer Precision, bei nur minimalen Verlusten beim Recall, d.h. auch die Antigrammatiken sind nicht fehlerfrei!

5.1.5.3.2 Performance und Evaluation des Systems Die Evaluierung (vgl. Kap. 5.1.4) erfolgte anhand der in Kap. 5.1.5.2.2 beschriebenen drei Korpora. Nur exakte Matches werden als korrekt gewertet, die Berechnung erfolgt auf Ebene der NE-Token, d.h. wenn ein Ortsname zehnmal im Korpus auftritt muss er vom System auch zehnmal erkannt werden. Details in Bezug auf die Werte für die einzelnen Korpora und die einzelnen Grammatiken und Module sind in Tab. 5.6 zu finden. Im Folgenden werden die (Teil-)Ergebnisse diskutiert.

Ein gutes, nicht zu kleines Lexikon ist notwendig, um eine solide Baseline beim Recall zu liefern. Da ein reiner Lookup ohne kontextsensitive Disambiguierung erfolgt, sollten möglichst keine „ambiguen“ Toponyme (*Essen, Laufen, Lippe*, vgl. S. 108) enthalten sein. Vergleicht man in Tab. 5.6 Zeile 2 mit 5 steigt die Precision allein durch den Ausschluss ambiguer Toponyme um 20%. Das ist auch der Grund, wieso ein großer „verunreinigter“ Gazetteer (USGS:GNIS und NGA:GNS, Zeile 1) zu einer nicht akzeptablen Precision führt. Der Recall ist passabel, zeigt aber im Vergleich mit dem Lexikon (Zeile 2), dass eine Anpassung an deutsche Besonderheiten, insbesondere die Aufnahme von Flexionsformen unumgänglich ist. Ein Guesser (Zeile 3) für Toponyme steigert zwar den Recall geringfügig (Zeile 6), aber auf Kosten der Precision. Er wird deshalb nur in Kombination mit einem Filter, der im Wesentlichen Mehr-Wort-Hodonyme akzeptiert oder mit zusätzlichem positivem Kontext verwendet. Hier bewirkt er eine kleine Erhöhung des Recalls, ohne die Precision negativ zu beeinflussen.

Die positiven Kontexte (Kap. 5.1.5.3.1.2) bringen eine geringe Verbesserung im Recall, indem sie auch ambigüe Toponyme akzeptieren. Für sich genommen wird der Anstieg im Recall zwar scheinbar durch eine schlechtere Precision aufgehoben, die aus Fehlerkennungen wie in *die Nordseeküste bei Nebel* oder *Lars von Trier ... stand von Trier im Wettbewerb* resultieren. Hier geben lokative Präpositionen einen falschen, da ambiguen Kontext. Die negativen Kontexte korrigieren jedoch einen Teil dieser Fehler, so dass der höhere Recall hier zu einem minimal höheren F-Wert führt, vgl. die Zeilen 17 und 18 in Tab. 5.6. Als Fazit lässt

	SZ (Development)			CompZ (Test)			Wikipedia (Test)			gesamt		
<i>Korpusgröße</i>	token	118493		token	152496		token	3346		token	274335	
	words	93912		words	119141		words	2360		words	215413	
	LOC	2326		LOC	995		LOC	140		LOC	3461	
Precision = p												
Recall = r												
f-score = f												
	p	r	f	p	r	f	p	r	f	p	r	f
<i>Gazetteers, Lexikon, Guesser</i>												
1. Gazetteers: GNIS+GNS	10.7	76.6	18.7	4.4	82.4	8.3	14.6	74.3	24.4	7.5	78.2	13.7
2. Lexikon	57.6	95.7	71.9	48.5	93.9	64.0	71.5	77.1	74.2	55.0	94.5	69.5
3. Guesser (interner Kontext)	62.0	3.2	6.1	7.9	0.6	1.1	60.9	10.0	17.2	43.2	2.7	5.2
4. 1, 2 und 3 kombiniert	12.4	96.4	21.9	4.6	94.1	8.8	16.4	87.1	27.6	8.4	95.4	15.5
5. Lex. ohne „ambigüe“ Topon.	75.9	92.1	83.2	84.3	92.6	88.2	89.1	70.0	78.4	78.5	91.4	84.4
6. 5 und 3 komb., gefiltert	76.0	93.6	83.9	82.7	92.6	87.3	88.6	72.1	79.5	78.2	92.5	84.7
<i>Positive (externe) Kontexte</i>												
<i>Phrasengrammatiken</i>												
7. lok. Präpositionalphrasen	95.3	33.8	49.9	93.8	42.9	58.9	85.1	28.6	42.8	94.4	36.2	52.3
8. Toponym mit Klassenbez.	94.0	6.1	11.5	96.6	5.7	10.8	100.0	12.1	21.7	95.2	6.2	11.7
9. Toponym-Aufzählungen	100.0	3.4	6.5	98.4	6.3	11.9	100.0	5.7	10.8	99.3	4.3	8.3
<i>Prädikat-Argument-Strukturen</i>												
10. alle	100.0	0.5	0.9	100.0	0.2	0.4	100.0	5.0	9.5	100.0	0.6	1.1
<i>Alle positiven Kontexte</i>												
11. Phrasengramm. (7, 8, 9) und PAS (10)	95.3	40.8	57.1	94.3	49.9	65.3	91.2	44.3	59.6	94.8	43.5	59.7
<i>Lexikon und positive Kontexte</i>												
12. Lexikon + Guesser (6) und pos. Kontexte (11)	75.5	95.1	84.2	81.0	93.6	86.8	87.0	71.4	78.4	77.3	93.7	84.7
<i>Negative Kontexte (komb. mit Lexikon und pos. Ktx.)</i>												
13. 12 und kleine Antigr.	75.9	95.0	84.4	83.8	93.5	88.4	88.5	71.4	79.1	78.3	93.6	85.3
14. 12 und Antigr. „Sport“	83.3	94.8	88.7	80.8	92.6	86.3	87.0	71.4	78.4	82.7	93.2	87.7
15. 12 und Antigr. „ORGs“	82.3	95.1	88.2	82.6	93.4	87.6	87.0	71.4	78.4	82.5	93.6	87.7
16. 12 und Personengramm.	77.1	95.1	85.1	84.5	93.6	88.8	87.7	71.4	78.7	79.4	93.7	85.9
17. 6 und Antigr. (13–16)	94.7	93.5	94.1	91.0	92.0	91.5	89.5	72.9	80.3	93.4	92.2	92.8
<i>NER-System (final)</i>												
18. 12 und Antigr. (13–16)	93.9	94.8	94.4	89.5	93.2	91.3	89.3	71.4	79.4	92.4	93.4	92.9
<i>Zum Vergleich</i>												
19. SAIL LABS Technology	73.6	60.2	66.2	68.0	77.5	72.4	83.8	45.9	59.3	71.9	64.4	67.9

Tabelle 5.6: Precision, Recall und F_1 -Wert des NER-Systems und seiner Teilmodule/-grammatiken für verschiedene Korpora (vgl. die Erläuterungen im Text).

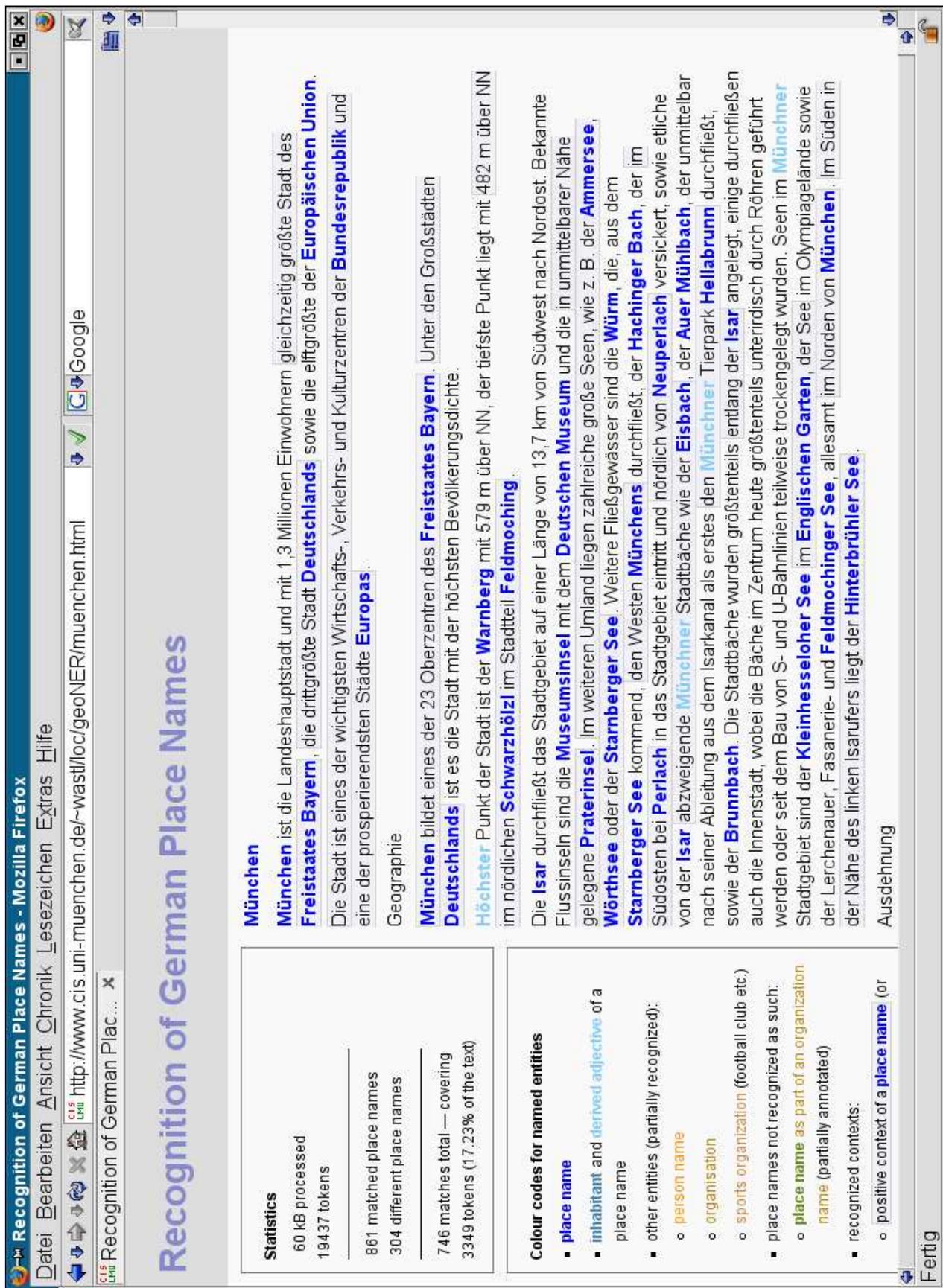


Abbildung 5.5: Screenshot des im Rahmen dieser Arbeit entwickelten Systems zur Erkennung von Toponymen in deutschen Texten.

sich zu den positiven Kontexten festhalten: ein hoher Aufwand, um viele sichere positive Kontexte zu beschreiben, bei gleichzeitig geringem Effekt.

Die negativen Kontexte sind dagegen unverzichtbar, steigern sie doch die Precision um etwa 15% auf gut 92%, der endgültige Wert des Systems. Der Recall ist mit gut 93% geringfügig besser, der F_1 -Wert liegt bei knapp 93%.

Fehleranalyse

Die Fehler, die das System macht, entfallen im Wesentlichen auf die folgenden Punkte:

1. unerkannte Toponyme: das Toponymlexikon ist mit 50 000 Toponymen noch deutlich zu klein, um z.B. zuverlässig Toponyme auch aus dem Reisetext einer Zeitung zu erkennen. Einige Klassen wie Straßennamen und Namen für Museen, Theater, Restaurants sind im Lexikon unterrepräsentiert.
2. die Bestimmung von metonymisch für Sportvereine verwendeten Ortsnamen ist schwierig und fehlerbehaftet (Evaluierungskorpus „Süddeutsche Zeitung“)
3. Organisations- und Produktnamen, die ein Toponym enthalten, werden nicht als solche erkannt, z.B. *U.S. Robotics*, *Texas Instruments*, *Timbuktu Pro for Networks*, *Cisco erwarb Grand Junction* (Evaluierungskorpus „Computerzeitung“)
4. Bewohnernamen homonym zu Toponymen (*Schweden*) werden nicht korrekt erkannt
5. unerkannte Personennamen werden fälschlich als Toponyme erkannt: *Hans von der Hagen*, *Hermann Graf von der Schulenburg*, *River Phoenix*
6. Ortsnamen in den Namen von Restaurants (*Harry's New York Bar*) führen zu doppelten Fehlern (korrekter Name nicht erkannt, fehlerhafte Erkennung eines Teilausdrucks)
7. Guesser und/oder positive Kontexte erkennen zu viel, z.B. *in Halle 12, Stand A 76* („Computerzeitung“)

Mögliche Erweiterungen

des Systems wären neben einer Vervollständigung des Lexikons die Verbesserung der Grammatiken zur Erkennung von Personen- und Organisationsnamen, was einerseits mehr und bessere negative Kontexte bedeuten würde, andererseits auch das System zu einem vollwertigen NER-System machen würde.

Bisher unberücksichtigt bleibt die Diskursebene bzw. die Strategie des Learn-Apply-Forget (Kap. 5.1.2.3). Mit lokalen Grammatiken in der gängigen Implementierung (INTEX/UNITEX) sind solche Strategien nur schwer zu realisieren. Auch werden mit Learn-Apply-Forget nicht alle ambigen Toponyme (*Weil*, *Essen*) aufzulösen sein. Für manche dieser Toponyme (vgl. Tab. 5.1) kann die Strategie des Learn-Apply-Forget sogar kontraproduktiv sein. Die Beschreibung positiver und negativer Kontexte macht sie ebensowenig überflüssig, wie ein gutes Lexikon. Zumal die „erlernten“ Toponyme im Deutschen auch in morphologische Varianten auftreten können.

Fazit

Das hier beschriebene System basiert auf drei Ressourcen:

- ein Toponymlexikon und ein Guesser für unbekannte Toponyme sind die notwendige Grundlage für einen hohen Recall
- positive Kontexte liefern Evidenz, um auch ambigue Toponyme (*Essen, Seine, Leiden*) annotieren zu können, d.h. sie verbessern den Recall, wenn auch nur geringfügig
- negative Kontexte verbessern Precision signifikant, indem sie fehlerhafte Annotationen allein aufgrund des Lexikons ausschließen

Das Gesamtergebnis ist mit gut 92% F₁-Wert durchweg besser als bei vergleichbaren Systemen (siehe Tab. 5.5).

5.2 Geokodierung

Geokodierung (auch Georeferenzierung, geo-coding, grounding, toponym resolution, place name disambiguation, geographic name disambiguation, GeoTagging) bezeichnet die Abbildung einer Ortsangabe in Form eines Toponyms (also einer textuellen Repräsentation) auf eine geographische Repräsentation (ein Koordinatenpaar oder ein Polygon). Die Abbildung erfolgt damit von einer nicht-formalen, ungenauen, ambiguen, unterspezifizierten, aber für den Menschen optimierten Repräsentation auf eine formale und exakte, dafür für den Menschen nur mittelbar nutzbare Repräsentation. Vor allem die Ambiguität von Toponymen, und damit die Verletzung der Monoreferentialität eines Eigennamens (siehe S. 94), macht die Geokodierung zu einer schwierigen Aufgabe. Laut Smith & Mann (2003) existieren in Nordamerika 41 Oxfords, 73 Springfields, 91 Washingtons und 97 Georgetown. Pouliquen et al. (2006) listen als „hochambigüe“ Ortsnamen Aleksandrovka (244), San Antonio (205), Santa Rosa (199), San Francisco (102), Buenos Aires (88). Nach Irie & Sundheim (2004) existieren 900 Orte weltweit mit dem Namen *San Francisco*. Und ein Blick in die Gazetteers NGA:GNS und USGS:GNIS mit zusammen knapp 8 Millionen Einträgen liefert als häufigste ambigue Einträge:

1727 San Antonio	1218 San Pedro
1510 Mill Creek	1176 San Juan
1488 San José	1065 Bear Creek
1468 Spring Creek	1033 San Francisco
1257 Dry Creek	1013 San Miguel
1231 Santa Rosa	912 Mud Lake

Auch wenn die Zahlen nicht einheitlich sind, viele Toponyme sind ambigue und die Zahl der Ambiguitäten ist unerwartet hoch.¹⁶⁰ Die Ambiguitäten sind weiter nicht gleichmäßig auf die Kontinente verteilt. Während in der alten Welt Allonymien (verschiedene Namen für ein und denselben Ort) häufiger sind, sind in der Neuen Welt mehr als die Hälfte der Toponyme ambigue (siehe Tab. 5.7).

Die Geokodierung wurde in den letzten Jahren u.a. von Smith & Crane (2001), Leidner, Sinclair & Webber (2003), Leidner (2004b), Kimler (2004), Garbin & Mani (2005), Smith & Mann (2003), Leidner (2005), Pouliquen et al. (2004, 2006) behandelt. Einen aktuellen, umfassenden Überblick über die verschiedenen Strategien inklusive einer Evaluation gibt Leidner (2006, 2007). Die Darstellung kann an dieser Stelle deshalb sehr knapp ausfallen.

¹⁶⁰ Vgl. auch die Liste ambiguer englischer Ortsnamen der WIKIPEDIA.

Continent	% places w/mult. names	% names w/mult. places
N. & Cent. Amerika	11.5	57.1
Oceania	6.9	29.2
South America	11.6	25.0
Asia	32.7	20.3
Africa	27.0	18.2
Europe	18.2	16.6

Tabelle 5.7: Homonymie und Allonymie von Toponymen im GETTY-Thesaurus (aus Smith & Mann 2003)

Voraussetzung für die Geokodierung ist selbstverständlich die Erkennung der Toponyme im Text (NER). Daran schließen sich einzelne oder mehrere der nachfolgenden Strategien zur Geokodierung an – die Angaben in eckigen Klammern beziehen sich auf Leidners Klassifikation (2006, 2007: 102–112):

0. [H₀: assign unambiguous referent]

der triviale Fall, dass im Gazetteer nur ein Eintrag für ein gg. Toponym vorhanden ist, also gar keine Ambiguität vorliegt

1. [H₁: “contained-in” qualifier following]

Angabe eines spezifizierenden Toponyms in unmittelbarer Folge auf das zu geokodierende Toponym, z.B. *Cambridge, MA, USA* (weitere Beispiele in Kap. 3.5.1.2). Das spezifizierende Toponym bzw. die geographische Entität, auf die es referiert, steht in einer Teil-von- oder administratives-Subjekt-von-Relation zum zu geokodierenden Toponym.

2. [H₂: superordinate mention]

wie Strategie 1, aber die Angabe erfolgt an beliebiger Stelle im Text, z.B. als im Kontext eines Dokuments auftretender Ländername (Pouliquen et al. 2004, 2006)

3. [H₃: largest population]

wähle den Ort mit der größten Bevölkerung, d.h. bevorzuge z.B. Paris (Frankreich) vor Paris (Texas)

4. [H₄: one referent per discourse]

basierend auf der Idee, dass auf der Diskursebene nur eine Referenz realisiert ist (One-Sense-per-Document, siehe Kap. 5.1.2.3), werden bereits aufgelöste Ambiguitäten innerhalb eines Textes gleichermaßen aufgelöst

5. [H₅: geometric minimality]

die Idee dieser Strategie lässt sich reduzieren auf: a) Orte stehen selten allein in einem Text, und b) die Orte innerhalb eines Textes stehen meist auch in räumlicher Nähe zueinander. Wenn also gemeinsam mit dem ambigen *Schwabhausen* (‘Schwabhausen bei Dachau’ vs. ‘Schwabhausen bei Landsberg’) auch die Toponyme *Erdweg*, *Markt Indersdorf* vorkommen, muss die erste Lesart die richtige sein, und zwar unabhängig von spezifizierenden Toponymen aus den Strategien 1 und 2. Die Schritte dieser Strategie sind:

- a) extrahiere alle Ortsnamen im Text

- b) trage die Ortsnamen mit all ihren möglichen (ambiguen) Referenten auf einer Landkarte ein
 - c) wähle von den ambiguen Referenten denjenigen aus, mit dem die Region, die alle Ortsnamen enthält, die kleinste Fläche aufweist. Dabei existieren verschiedene Verfahren, um die kleinste Region geometrisch zu bestimmen
6. [H₆: singleton capitals]
falls ein Toponym nur einmal im Text vorkommt, und einer der Referenten eine Hauptstadt eines selbstständigen Staates ist, wähle diese „Lesart“
7. [H₇: ignore small places]
die Größe des Gazetteers wird reduziert, indem kleine Orte ignoriert werden
8. [H₈: focus on geographic area]
hier wird der vollständige Gazetteer regionalisiert, d.h. in einzelne Länder/Regionen aufgeteilt. Ein Look-Up im vollständigen Gazetteer erfolgt nur, wenn der Text einem Land oder einer Region zugeordnet werden konnte, z.B. bei Kimler (2004), wenn von den Toponymen eines reduzierten Gazetteers (Strategie 7) mehr als drei Orte oder mind. 50% aller Orte des Textes in einem Land liegen
9. [H₉: distance to unambiguous text neighbors; H₁₄: textual-spatial correlation]
ähnlich Strategie 5 wird die räumliche Distanz zu im unmittelbaren textuellen Kontext stehenden Toponymen berechnet und der räumlich nächstgelegene Kandidat gewählt. Diese Strategie geht also von einer positiven Korrelation zwischen textueller und räumlicher Nähe aus.
10. [H₁₀: discard off-threshold]
eine Variante von Strategie 5 bei Smith & Crane (2001):
- a) Extraktion der Ortsnamen und erste Disambiguierung anhand einfacher Regeln
 - b) alle Orte werden mit der Anzahl ihrer Nennungen im Text als Gewicht auf einer Landkarte eingetragen
 - c) für die Wolke der gewichteten Punkte wird der Schwerpunkt und der durchschnittliche Abstand zum Schwerpunkt berechnet
 - d) alle Punkte, die mehr als das Zweifache des durchschnittlichen Abstands vom Schwerpunkt entfernt sind, werden gelöscht
11. [H₁₁: frequency weighting]
in Kombination mit anderen Strategien (z.B. 10): die Toponyme werden nach der Häufigkeit ihres Vorkommens im Text gewichtet
12. [H₁₂: prefer higher-level referents]
bevorzuge die Toponymen, deren Klasse in einer meronymisch orientierten Taxonomie (vgl. Abb. 3.4: Erde > Kontinente > Länder > ...) weiter oben angeordnet ist. *Amerika* wird also als KONTINENT interpretiert und die möglichen Interpretationen als Ortsteile der Gemeinde Garrel in Niedersachsen oder der Stadt Penig in Sachsen verworfen.
13. [H₁₃: feature type disambiguator]
Toponyme im Kontext eines Klassenbezeichners (Kap. 3.5.1.3) werden durch diesen disambiguiert: verwerfe die Interpretation als Region für *city of Scotland*

	p	r	f
Leidner (2007: 183): CoNLL-2003-SHARED-TASK			
H ₀ (unambiguous reference)	1.00	.16	.27
H ₀₊₃ (largest population)	.66	.28	.40
H ₀₊₄ (geometric minimality)	.50	.26	.34
H _{0+1,4+5,12} (Leidner, Sinclair & Webber 2003)	.38	.33	.36
Leidner (2007: 183): MUC-4			
H ₀ (unambiguous reference)	1.00	.16	.27
H ₀₊₃ (largest population)	.66	.28	.39
H ₀₊₄ (geometric minimality)	.47	.13	.21
H _{0+1,4+5,12} (Leidner, Sinclair & Webber 2003)	.47	.26	.34
Kimler (2004)			
Deutsch	.72	.72	.72
Englisch	.79	.78	.79
Steinberger & Pouliquen (2007: 148)			
Deutsch	.80	.68	.73
Englisch	.91	.78	.84

Tabelle 5.8: Ausgewählte Ergebnisse bei der Geokodierung

14. [H₁₄: textual-spatial correlation]

angenommen wird, dass zwei Orte oder auch ein Ort und eine andere Entität, die in Texten gemeinsam auftreten, auch in räumlicher Korrelation stehen. Als Beispiel nennt Leidner *Paris* und *Frankreich* oder *Kalkutta* und *Mutter Theresa*.

15. [H₁₅: default referent]

eine Generalisierung von Strategie 3, die einen Gazetteer, der verschiedene Klassen geographischer Entitäten nach ihrer Bedeutung, festgelegt durch spezifische Kriterien, hierarchisch rankt, z.B. der von Li et al. (2003) benutzte TIPSTER-Gazetteer.

16. [H₁₆: preference order]

werden verschiedene Gazetteers verwendet, können diese gewichtet werden, in Abhängigkeit von Größe und Qualität

Doch ist die Wahl der besten Strategien nur der Anfang. Entscheidend ist auch, wie, in welcher Abfolge und mit welchen Gewichtungen, diese zu einem Algorithmus zusammengefügt sind.¹⁶¹ Die Ergebnisse der Geokodierung bleiben deutlich hinter denen der NER zurück. Tab. 5.8 zeigt einige ausgewählte Ergebnisse. Ein Eindruck ist, dass simple Strategien wie Strategie 3 (wähle den Ort mit der größeren Bevölkerung) sich positiver auswirken als komplexe Strategien wie die der geometrisch kleinsten Region (Strategie 5). Das aber bedeutet den systematischen und linguistisch unbefriedigenden Ausschluss von kleineren Orten, selbst wenn aus dem Kontext eine Auflösung möglich ist wie in *Paris, Texas*. Die Optimierung der Strategien und der Gewichtungen bleibt damit auch weiter eine Herausforderung.

¹⁶¹ Vgl. Leidner (2007) für den Aufbau von verschiedenen Algorithmen und Kimler (2004: 33–34) zur Gewichtung der Einzelparameter.

5.2.1 Räumliche Definition von Toponymen

Eine Umformulierung der ‚Geokodierung‘ stellt die Aufgabe dar, für ein gegebenes eindeutiges oder bereits disambiguiertes Toponym eine adäquate räumliche Repräsentation zu finden. Interessant ist diese Aufgabe insbesondere für sog. ‚unscharfe Regionen‘, die nicht durch natürliche oder administrative Grenzen bestimmt sind: (*Schweizer*) *Mittelland*, *Nordbayern*, *Mid Wales*, *Südwesten* (der USA), *Cotswolds* (eine Hügellandschaft in England). Auch Menschen definieren die Grenzen solcher Regionen unterschiedlich. Purves et al. (2005), Purves, Clough & Joho (2005) lassen fünf Geographen das *Schweizer Mittelland* auf einer Landkarte verorten. Die Geographen erzielen dabei nur eine teilweise Übereinstimmung.

Zur annähernden Bestimmung unscharfer Regionen mittels Webdaten gehen Arampatzis et al. (2004), Purves et al. (2005), Purves, Clough & Joho (2005) wie folgt vor:

1. Über Anfragen an eine Internet-Suchmaschine müssen Texte (Webdokumente) gefunden werden, die die Namen von in dieser Region lokalisierten Orten enthalten. Als Anfragen kommen in Frage: (a) der Name der Region alleine (*the Rocky Mountains*), (b) in Kombination mit geographischen Objekten (*hotels in the Cotswolds*), oder (c) „Triggerphrasen“ (*X is located in Y, A is a town in B, MidWest [states|cities|towns] such as X, Y, and Z*). Die Autoren verwenden nur Triggerphrasen.
2. Erkennung und Geokodierung (Kap. 5.2) der Toponyme in den abgerufenen Webdokumenten bzw. in den aus der Ergebnisseite der Suchmaschine extrahierten Dokumenttiteln und Snippets.
3. Bestimmung eines Polygons, das die Region repräsentiert, anhand verschiedener geometrischer Verfahren.

Zur Bewertung des entwickelten Verfahrens vergleichen Purves et al. (2005), Purves, Clough & Joho (2005) zum einen die ermittelten Regionen mit den Angaben von Geographen, die als Versuchspersonen ebenfalls die Aufgabe hatten, die Grenzen unscharfer Regionen zu bestimmen. Zum anderen wenden Sie das Verfahren auf klar definierte administrative Regionen an. Die Ergebnisse erweisen sich grundsätzlich als brauchbar, allerdings sind die verwendeten Algorithmen sehr empfindlich bezüglich Fehlern bei NER (Geoparsing) und Geokodierung.

5.3 Automatische Extraktion lexikalischer Ressourcen

5.3.1 NER zur Lexikonerweiterung

Gazetteers sind die wichtigste Resource bei der Entitätenerkennung und für alle Systeme und Verfahren, regelbasierte wie statistische, unverzichtbar. Eine naheliegende Idee ist deshalb, die Aquisierung von Gazetteers und Lexika unter dem Stichwort „Gazetteer-Learning“ zu automatisieren. Hier werden einige Systeme besprochen, die versuchen speziell Toponyme zu „lernen“, d.h. aus großen, unannotierten Korpora zu extrahieren.

Das von Jacquemin & Bush (2000) entwickelte System sammelt Named Entities aus dem Web, die in Aufzählungen vorkommen. Dazu werden (1.) zu Mustern wie „*following* <CL>“, „<CL> *such as*“, *list of* <CL>“ oder „*such* <CL> *as*“ über zwei Internetsuchmaschinen 3 GB

Webdokumente gecrawlt. ⟨CL⟩ steht dabei für einen von 34 Klassenbezeichnern wie *universities*, *politicians* oder *car makers*. (2.) Üblicherweise leiten solche Muster Aufzählungen ein in Form von (a) rein textuellen, durch Kommata und Konjunktionen separierten oder (b) durch das HTML-Format ausgezeichneten Listen. Aus diesen Listen werden dann die Named Entities extrahiert. Die Precision für ein Sample von 50 000 so aquirierten named entities betrug 63% für 34 angenommen Klassen bzw. 74% für die drei Klassen Organisation, Person, Lokation.

Erp (2006) versucht mittels memory-based learning Toponyme aus Texten zu lernen und in Texten zu erkennen (NER). Sie arbeitet mit einem Kontext von fünf Wörter (jeweils zwei links und rechts des Toponyms). Als Seed-Liste dienen 25 englische¹⁶² Toponyme für alle drei bearbeiteten Sprachen (Englisch, Holländisch, Deutsch). Die Ergebnisse sind dünn. Beim Gazetteer-learning wurden bezogen auf die verwendeten Korpora mit einigen zehn bis hundert Millionen Wortformen sprachabhängig nur einige Dutzend bis 200 Toponyme gelernt, bei Fehlerraten bis zu 25%. Angewendet auf NER sind die Ergebnisse ebenfalls ernüchternd: im holländischen Evaluationskorpus wurden zusätzlich zu den Toponymen der Seed-Liste zwar fünfmal mehr Toponyme mit einer Precision von 100% erkannt, der Recall liegt aber nur bei 13%. Bei einem Lauf mit dem deutschen CoNLL-Korpus (siehe Kap. 5.1.5.2.2) wurden zusätzlich zur Seed-Liste keine neuen Toponyme erkannt.

Uryupina (2002, 2003) will unbekannte Toponyme aus sechs Klassen (CITY, REGION, COUNTRY, ISLAND, RIVER, MOUNTAIN) automatisch aus Webseiten extrahieren, um damit einen Gazetteer zu vervollständigen. Dazu werden zunächst Pattern zur Erkennung der Toponyme (z.B. "*island of X*") und Seed-Wortlisten manuell erarbeitet. Für jedes Element der Patternliste instantiiert mit den Elementen der Seed-Wortliste werden Webseiten mittels einer Suchmaschine ermittelt. Diese fließen in einen Bootstrap-Prozess maschinellen Lernens ein, dessen Ergebnis neue, unbekannte Toponyme und weitere Pattern sind, anhand derer sicher unbekannte Toponyme erkannt und einer Klasse zugeordnet werden können. Die Extraktion neuer Ortsnamen erfolgt mit 86,5% Treffergenauigkeit nach zwei Bootstrapping-Läufen. Besonders schlecht war das Verfahren für die Klasse CITY (62% Genauigkeit), für die nach Uryupina einerseits die vielen ambiguen Ortsnamen, andererseits die nur wenigen sicheren Pattern verantwortlich sind.

Toral & Muñoz (2006) wollen Artikel aus der WIKIPEDIA (simple.wikipedia.org, vereinfachtes Englisch) einer der drei Eigennamenkategorie Person, Ort, Organisation zuordnen. Dazu wird der Text des Artikels getaggt und für alle Nomina wird mittels WORDNET bestimmt, ob für ihre erste Lesart/Bedeutung ein Hyperonym einer der gesuchten Kategorien existiert. Anschließend wird der gesamte Artikel der Klasse zugeordnet, der die meisten Nomina zugeordnet wurden. Diese Zuordnung kann durch einen Kappa-Wert modifiziert werden, der einen Schwellwert angibt, um das wievielfache mehr Nomina der häufigsten Klasse zugewiesen sein müssen als der zweithäufigsten Klasse. Kombiniert mit einer „is-instance“-Strategie, die nur die Stichwörter akzeptiert, die in WORDNET als Instanzen (Eigennamen) annotiert sind, erreichen Toral & Muñoz für ein Testset aus der WIKIPEDIA F₁-Werte von 78% (Orte) und 68% (Personen) bei niedriger Precision (67 bzw. 61%). Ein höherer Kappa-Wert kann die Precision jedoch auf Kosten des Recalls steigern. Die Ergebnisse können damit nicht überzeugen, vor allem im Hinblick darauf, dass die Stichwörter ja bereits Einträge in WORDNET und in einer der kleineren der WIKIPEDIAS sind, und die nötige Kategorisierung

162 Sic! Und die Seed-Liste enthält keineswegs „neutrale“, in allen drei Sprachen orthographisch gleich geschriebene Toponyme wie *London*, sondern auch z.B. *The Hague* und *Great Britain*.

ohne viel Aufwand dort zu entnehmen ist. Stichwörter, die nicht in WORDNET enthalten sind – und das gilt für viele Eigennamen, sind durch die „is-instance“-Strategie von der Erkennung ausgeschlossen.

Nadeau (, vgl. Nadeau 2007: 49) extrahiert aus einem Web-Korpus 15 000 Städtenamen (und andere NE-Typen) mittels HTML-Struktur und sprachbasierter Pattern. Er erreicht 97% Precision für ein Sample von 100 zufällig ausgewählten Städtenamen.

Ob der manuelle Aufbau von Ressourcen jemals durch Verfahren zum automatischen Lexikonaufbau überflüssig wird, ist unwahrscheinlich. Von den hier vorgestellten Verfahren und Experimenten, erreicht allein Nadeau in der Precision brauchbare Ergebnisse bei einem einträglichen – im Sinne der *aquirierten Entities* – Recall. Der Aufwand für die Nachbearbeitung, um zu 100% korrekte Einträge zu erhalten, dürfte aber auch deutlich über dem liegen, was für die Anpassung und manuelle Ergänzung bereits vorhandener Gazetteers – für Toponyme sind mehrere frei verfügbar – benötigt wird.

5.3.2 Automatische Extraktion von Relationen zwischen Toponymen

War die Extraktion von Toponymen zum Lexikonaufbau bereits Thema für eine Automatisierung, gilt das noch mehr für die Extraktion von Relationen zwischen Toponymen. Die Erkennung von Relationen zwischen Entitäten in Texten gehört zum Themengebiet „Ontology Learning and Population“¹⁶³ Keine der zahlreichen Arbeiten konzentriert sich dabei Relationen zwischen Toponymen. Suchanek, Ifrim & Weikum (2006a,b) extrahieren für Toponyme nur die Synonymie-Relation. Ihr System lernt dabei Pattern wie *X was known as Y* oder *X (formerly Y)* automatisch.

In Texten (hier WIKIPEDIA.DE) sind wunderbare Sätze wie

(5.17) *Apples ist eine politische Gemeinde im Distrikt Aubonne des Kantons Waadt in der Schweiz.*

(5.18) *Die Elsoff ist ein linker Zufluss der Eder.*

zu finden, was die Hoffnung rechtfertigt, daraus Relationen zu extrahieren und in eine Ontologie einzuspeisen. Einige der Relationen wurden in Kap. 4 beschrieben, ohne sie für den Ontologieaufbau zu nutzen.

5.4 Usernavigation und geographische Queries

Named Entity Recognition und Geokodierung sind die Voraussetzung für eine Reihe von Anwendungen, die es einerseits dem User erleichtern sollen durch die Suchergebnisse zu navigieren oder die Suche weiter einzuschränken, andererseits eine Anfrage des Users als geographische Query zu erkennen und dann anders und besser zu verarbeiten.

163 Siehe Buitelaar, Cimiano & Magnini eds. (2005), Biemann (2005) für einen Überblick.

5.4.1 Usernavigation

Techniken wie NER und Phrasenextraktion ermöglichen es, den Benutzer beim Browsen der Ergebnisse und bei der Verfeinerung und Umformulierung der Query zu führen. Die Benutzerführung erfolgt über dynamisch erzeugte Links auf in der Resultatmenge gefundene Keyword-Phrasen und Entitäten, die dann zu einer Einschränkung bzw. Umformulierung der Query führen, so dass die Treffermenge kleiner und im Sinne des Users gefiltert wird. Crestan & Loupy (2004) haben in Experimenten gezeigt, dass solche Techniken dem User ermöglichen, schneller das gewünschte Dokument oder die gewünschte Information zu finden.

Diese ‚faceted search‘ oder ‚faceted browsing‘ genannte Technik (vgl. auch Hearst et al. 2002) sind aber nicht spezifisch für geographische Entitäten. Das Browsen über Relationen zwischen den geographischen Entitäten und die Visualisierung auf und mittels Landkarten dagegen schon.

5.4.1.1 Visualisierung auf Landkarten

Die Visualisierung auf Landkarten ist die naheliegendste Methode um dem User einen schnellen Überblick über die gefundenen geographischen Entitäten in einem Dokument oder einer Dokumentenkollektion zu geben. Der simple Plot der Koordinaten der gefundenen Entitäten auf einer Karte oder einem Satellitenbild kann mit dem Highlighting von Regionen (Polygonen oder geopolitischen Einheiten), oder auch Vektoren, um den räumlich-zeitlichen Verlauf von Ereignissen darzustellen, kombiniert werden. Der User erhält so schnell erfassbare visuelle Zusammenfassungen von Ereignissen und Abläufen. Unabhängig vom Nachrichtenressort lassen sich Nachrichten nach Regionen browsen. Details und beispielhafte Anwendungen sind z.B. bei Hauptmann & Olligschlaeger (1999), Pouliquen et al. (2004, 2006) und Leidner (2007: 187–98) zu finden.

5.4.1.2 Benutzerführung über Ontologien

Neben dem Highlighting gefundener Entitäten im Text und ihrer Anzeige als Liste zum Refinement, kann auch das Einblenden eines Teilbaums einer Taxonomie, die gefundene oder fokussierte Entitäten im Kontext verlinkter Entitäten zeigt, den User bei der Navigation, Verfeinerung oder Umformulierung der Query unterstützen. Systeme, die die Anbindung einer Ontologie nicht nur, aber auch geographischer Entitäten umsetzen sind KIM (Manov et al. 2003) und EFGT Net (Weigel et al. 2006). Beispiele sind in der angegebenen Literatur zu finden.

5.4.2 Lokalisierung von Texten

Die Lokalisierung von Texten, also die Zuordnung eines Textes zu einem Ort oder einer Region, ist als Aufgabe durchaus der Geokodierung ähnlich (vgl. Leidner 2007: 198–211), mit dem großen Unterschied, dass Texte ja keine geographischen Entitäten sind, und deshalb zunächst definiert werden muss, was der ‚Ort‘ oder der ‚geographic scope‘ eines Dokuments ist. Silva et al. (2004) definieren in als: „We define the geographic scope of a page as the region, if it exists, where more people than average would find that page relevant.“ Eine

ähnliche Definition ist bei Ding, Gravano & Shivakumar (2000) zu finden: „The geographical scope of a web resource w is the geographical area that the creator of w intends to reach.“ Diese Definitionen sind natürlich dehnbar und interpretationsabhängig. Aber das eigentliche Problem besteht in der Bestimmung des „geographic scope“.

Buyukkokten et al. (1999) verknüpfen die IP-Adressen mit den Telefonvorwahlen der zuständigen Netzwerkadministratoren und ermitteln daraus die Koordinaten für die unter der IP angebotenen Webseiten. Der Nachteil dieser Methode liegt auf der Hand: der Inhalt eines Dokuments ist unabhängig von dem Ort der „Aufbewahrung“ des Dokuments. Gut ist aber die Idee, die Buyukkokten et al. als Anwendung vorschlagen. Dabei werden für eine gegebene URL mittels der Linksuche von Google hierauf verweisende Seiten bzw. URLs ermittelt und deren Koordinaten auf eine Landkarte geplottet.

Diese Idee greifen Ding, Gravano & Shivakumar (2000) auf, um den ‚geographic scope‘ einer Webseite zu bestimmen: „Our assumption in this section is that this interest will translate in web pages across the country containing HTML links to this web resource. Conversely, a resource with a much more limited geographical scope will exhibit a significantly different link distribution pattern across the country.“ Neben der Verlinkung schlagen Ding, Gravano & Shivakumar eine zweite Methode vor. Dazu werden mittels eines NE-Taggers Toponyme aus dem Text extrahiert und anschließend geokodiert. Dazu werden die nicht unambigen Toponyme den US-Bundesstaaten zugeordnet. Der Bundesstaat mit den meisten zugeordneten Namen wird als Region zur Disambiguierung der verbleibenden. Die Koordinaten der Städte werden dann gleich wie die Ausgangsorte von Links behandelt. Das Ziel ist schließlich, Seiten entweder „global“ den gesamten Vereinigten Staaten, einem der Bundesstaaten oder einer Stadt zuzuordnen.

Neben diesen empirischen Ansätzen, um einen Text (Webseite) einem Ort zuzuordnen, wird auch versucht, diese Aufgabe den Autoren der Webseite zu übertragen. Z.B. als Empfehlung, wie Adressen zu schreiben sind, damit sie von einem Dienst wie GOOGLE:LOCAL erkannt werden. Das Stanford Research Institute schlägt vor, eine Top-Level-Domain `.geo` einzuführen, die in Quadranten/Parzellen aufgeteilt ist, in die andere Domains hineingemappt werden (Jones et al. 2002). Und schließlich existieren eine Reihe von Quasi-Standards, um die Koordinaten eines Dokuments oder einer Seite als Meta-Angaben (in sog. ‚Microformaten‘) direkt im HTML-Kode anzugeben:

GEOURL erlaubt die Angabe von Koordinaten im HTML-Header:

```
<meta name="ICBM" content="48.15035, 11.59470">
<meta name="DC.title" content="Centrum für Informations-
                        und Sprachverarbeitung (CIS)">
```

GEO-tags darüberhinaus auch noch Ortsnamen und Zuordnung zu einer Region gemäß der ISO 3166:

```
<meta name="geo.position" content="48.15035;11.59470">
<meta name="geo.placename"
        content="Centrum für Informations- und
                Sprachverarbeitung (CIS), Munich, Germany">
<meta name="geo.region" content="de-by">
```

GEOMICROFORMAT ermöglicht es, Koordinaten an beliebiger Stelle im Text anzugeben:

```
<div class="geo">
```

```
<abbr class="latitude" title="48.15035">N 48° 9' 1"</abbr>  
<abbr class="longitude" title="11.59470">E 11° 35' 41"</abbr>  
</div>
```

5.4.3 Geographische Queries

Definition einer geographischen Query

Sanderson & Kohler (2004) definieren eine geographische Query wie folgt: sie enthält mindestens ein Toponym, eine Postleitzahl, ein Ortsadjektiv, einen geographischen Term oder Klassenbezeichner. In dem von ihnen bearbeiteten Querylog (1 Mio. Anfragen) der Suchmaschine Excite sind nach der Definition 16% aller Queries ‚geographisch‘.¹⁶⁴ Sanderson & Kohler (2004) stellen weiter fest, dass ‚geographische Queries‘ länger sind als der Durchschnitt aller Queries (Ø 3,3 vs. 2,6 Terme). Nur 0,5% aller Queries haben die Form „(Objekt) in (Ort)“, und drücken damit die räumliche Relation explizit aus. Den geringen Anteil an Queries, die eine räumliche Relation direkt mit *in*, *near* etc. ausdrücken, erklären Sanderson & Kohler (2004) auch damit, dass der Benutzer wisse, dass diese von der Suchmaschine als Stopwörter ignoriert würden.

Die Definition einer geographischen Query als Tripel $\langle \textit{what}, \textit{relator}, \textit{where} \rangle$ ist die gängigste und am meisten plausible Definition (z.B. Fu, Jones & Abdelmoty 2005a,b, GEO-CLEF07QUERY).

Gravano, Hatzivassiloglou & Lichtenstein (2003) nehmen an, dass manche Web Queries von Natur aus lokal (*houses for sale*), andere dagegen global (*wildflowers*) sind. D.h. entweder sind die Queries aus der Perspektive des Users bzgl. der räumlichen Dimension auf ein bestimmtes Gebiet beschränkt – auch wenn dieses nicht genannt ist, und die Query damit unterspezifiziert ist. Oder die Query hat einen globalen räumlichen Bezug (hier die gesamten USA). Gravano, Hatzivassiloglou & Lichtenstein versuchen dann 1200 Queries einer der beiden Kategorien zuzuordnen. Sie holen dafür für jede Query 50 Dokumente via Google API und klassifizieren diese anhand von 1600 Ortsnamen aus den drei Verwaltungsebenen Country, State und City mittels verschiedenen maschinellen Lernalgorithmen. Die Ergebnisse – maximal 50% F₁-Wert bei der Klassifikation – sind mager, aber eher ein Beleg dafür, dass die binäre Kategorisierung zumal im Hinblick auf die gewählten Beispiele nicht sinnvoll ist. Als Antwort auf die Query *wild flowers* kann ein Link zu einem lokalen oder regionalen spezialisierten Blumensamenhändler oder botanischen Garten ebenso nützlich sein, wie auf ein überregionales Immobilienportal auf die Anfrage *houses for sale*.

Verarbeitung einer geographischen Query

Zur Verarbeitung einer geographischen Query muss zunächst einmal die nicht ganz triviale Entscheidung getroffen werden, ob die Query überhaupt eine geographische ist. Im nächsten Schritt müssen die Bestandteile *what*, *relator* und *where* erkannt werden. Das weitere Vorgehen hängt von der gewählten Strategie ab. Fu, Jones & Abdelmoty (2005b) unterscheiden zwei grundsätzlich verschiedene Strategien:

164 Zum Vergleich: Nach Silva et al. (2004) enthalten 4% aller Queries der portugiesischen Suchmaschine tumba einen (von nur 308) Ortsnamen.

- termbasiert, dabei wird die Query um „verwandte“ Terme, insbesondere Meronyme des *where*-Toponyms, expandiert (,Query Expansion‘)
- footprint-based, dabei werden Query und Dokumente räumlich lokalisiert und dann räumlich verarbeitet, d.h. nach räumlicher Nähe zueinander gefiltert oder gerankt

Query Expansion

Petras & Gey (2005) expandieren die Query *Hochwasser in Holland und Deutschland* um wichtige Konzepte/Termini (*Hochwassergebiet, Deich, Flut*) und mit *Waal* und *Maas* um wichtige (holländische) Flüsse. Aber nicht jede Erweiterung der Query bringt Verbesserungen. Als Negativbeispiel nennen Petras & Gey die Query *Aktionen gegen die Pelzindustrie in Europa und den U.S.A.* erweitert um die Namen aller Länder Europas.

Die gleiche Beobachtung machen Larson (2005): „Another conclusion is that, in some cases, query expansion of region names to a list of names of particular countries in that region is modestly effective (although we haven’t yet been able to test for statistical significance). In other cases, however it can be quite detrimental.“

Leidner (2005) experimentiert mit der Queryexpansion über WORDNET-Relationen. Die Expansion bringt aber keine Verbesserung: „we discovered that geographic query expansion with WordNet meronyms appears not to improve retrieval performance.“

Buscaldi, Rosso & Arnal Sanchis (2005) versuchen die Antworten auf Queries, die mindestens ein Toponym enthalten, zu verbessern, in dem sie die Query um Synonyme und Meronyme des Toponyms erweitern. Synonyme und Meronyme erhalten sie aus dem englischen WordNet, das z.B. für *Australia* als Synonym *Commonwealth of Australia* und als Meronym *Canberra* enthält. Für die Expansion auf der Queryseite erzielen Buscaldi, Rosso & Arnal Sanchis (2005) nur negative Effekte in Precision und Recall; lediglich bei der Expansion der Terme zum Indizierungszeitpunkt zeigt sich ein sehr kleiner positiver Effekt.

‘Footprint-based spatial query expansion’ und Ranking nach räumlicher Nähe

Die Gefahr, die Query um zu viele Terme zu erweitern, versuchen Techniken die mit Footprints arbeiten zu entgehen. Dazu werden sowohl für Query wie Dokumente Footprints ermittelt. Der Footprint der Query ist identisch mit den Koordinaten des *where*-Toponyms. Der Footprint des Dokuments wird aus den dort enthaltenen Toponymen bestimmt (siehe auch Kap. 5.4.2). Nach dem die Query gestellt wurde, wird versucht, Dokumente mit einem Footprint zu finden, die in der durch Queryfootprint und Relator spezifizierten Region liegen. Unscharfe Relatoren wie *near* oder *north* müssen dabei speziell verarbeitet werden. Ist die Ergebnismenge nicht ausreichend, wird die Query expandiert und die Expansion solange iterativ in Teilschritten wiederholt, bis die Ergebnismenge ausreichend ist. Dies garantiert aber auch, dass die räumlich nächsten und damit räumlich relevantesten Dokumente an oberster Stelle in der Ergebnismenge erscheinen. Die Expansion der Query kann dabei auf zweierlei Weise erfolgen – über die Anreicherung der Query um räumlich verwandte Ortsnamen aus einer Ontologie, oder üblicherweise über die aus den Koordinaten berechneten Abstände.

Footprint-basierte Expansion wird von diversen lokalen Suchdiensten (z.B. GOOGLE:LOCAL und MSN:LOCAL) eingesetzt. Eine detaillierte Beschreibung ist in Fu, Jones & Abdelmoty (2005b) zu finden, mit dem Ranking nach räumlich geographischer Relevanz beschäftigen

sich Larson & Frontiera (2004), ontologiebasierte Systeme werden von Jones et al. (2002) und Larson (2005) vorgestellt.

Für die Abbildung einer Query auf Koordinaten steht mit Spatial SQL (vgl. Egenhofer 1994, Anderegg 2002, OPENGIS-SQL) ein Werkzeug zur Verfügung, um räumliche Anfragen von Datenbankservern (z.B. ESRI, Oracle, PostgreSQL/POSTGIS) verarbeiten zu lassen.

5.5 Räumliche Repräsentation \Leftrightarrow sprachliche Beschreibung

Der Wunsch, räumliche Repräsentation und sprachliche Beschreibung aufeinander abbilden zu können, war bereits auf dem Höhepunkt der KI-Forschung präsent (vgl. das Projekt LI-LOG, Herzog & Rollinger eds. 1991). Aber auch aus der heutigen nüchternen Perspektive rückt diese Aufgabe immer wieder in den Fokus.

Naaman et al. (2004, 2006) versuchen für ein gegebenes Koordinatenpaar eine möglichst treffende natürlichsprachige Beschreibung zu finden, z.B. ein Toponym oder auch eine Ortsbeschreibung wie *40 kms south of San Francisco*. Nun reicht es nicht aus einem Gazetteer oder Geo-Datenbanken den nächstbesten Eintrag zu extrahieren. Faktoren wie der Bekanntheitsgrad eines Toponyms spielen eine wichtige Rolle, soll die Ortsbeschreibung auch verständlich sein. Die Autoren versuchen dazu über die Trefferanzahl bei Google den Bekanntheitsgrad eines Toponyms zu ermitteln. Mögliche Anwendung des Verfahrens ist das Taggen von Bildern in einer Kamera mit integriertem GPS.

Die von Naaman et al. formulierte Aufgabe nennt Leidner (2007: 32) „toponym generation“. Von einer wirklichen Versprachlichung räumlicher Konstellationen ist sie in der Komplexität noch weit entfernt. Ein interessantes Projekt in dieser Richtung stellen Bähr & Müller (2004), Müller & Bähr (2006), Müller (2006, 2008) vor. Das brasilianische Kataster repräsentiert Grundstücke nicht mittels Karten, sondern einer natürlichsprachlichen Beschreibung. Existieren Karten sind diese rechtlich nicht bindend. Ein kurzer Ausschnitt einer solchen Grundstücksbeschreibung:

- (5.19) *Es beginnt am Punkt des Zusammentreffens von dem Grundstück von A und dem Grundstück von A (vormals Z). Von diesem Punkt aus folgt es in einer geraden Linie einer Strecke von 301,00 m; von diesem Punkt aus biegt es nach links und folgt einem kleinen Bachlauf und begleitet auf dieser Strecke seine Windungen auf einer Länge von 68,70 m, während es dabei an das Grundstück von A (vormals Z) grenzt. Von diesem Punkt aus folgt es weiterhin ...* (Müller & Bähr 2006)

Ziel des noch laufenden Projekts ist es, eine Abbildung zwischen der sprachlichen und der Kartendarstellung herzustellen.

Zusammenfassung

Aufgabe der Arbeit war, lokative Sätze mittels lokaler Grammatiken im Hinblick auf eine Anwendung im Information Retrieval zu beschreiben. Somit sind es drei Themengebiete, die sich als rote Fäden durch die verschiedenen Kapitel der Arbeit ziehen: (1.) Lokativa und der Komplex ‚Sprache und Raum‘, (2.) die Beschreibung von elementaren Sätzen und ihrer Bestandteile, der Prädikate und ihrer Argumente inklusive Transformation und Paraphrase anhand von lokalen Grammatiken und Lexikogrammatiken, und (3.) die Anwendung der erarbeiteten Ressourcen, der Grammatiken und Lexika, im Information Retrieval (IR) bei der Erkennung von lokativen Named Entities.

Kapitel 1 gibt einen Überblick über die beiden als Beschreibungsrahmen und Plattform zur Grammatikimplementierung gewählten Modelle der lokalen Grammatik und Lexikogrammatik. Detailliert wird auf ihre Eigenschaften, Anwendungsgebiete, Repräsentation und Implementierung eingegangen. Gewählt wurden die beiden durch Maurice Gross begründeten grammatischen Beschreibungsmodelle wegen ihrer empirischen Ausrichtung, der intuitiven Repräsentation als Graphen und der effizienten Anwendung in Form von endlichen Automaten gerade auch im IR.

Das folgende Kapitel 2 ist lokativen Prädikaten gewidmet und fasst zunächst Forschungsstand und deskriptive Probleme aus dem Themengebiet ‚Sprache und Raum‘ überblickshaft zusammen. Für die semantischen Grundtypen räumlicher Prädikationen, räumliche Relationen und Bewegung im Raum, werden verschiedene Beschreibungsansätze in knapper Form vorgestellt und diskutiert. Im Anschluss wird versucht, eine Definition zu finden, welche Sätze als ‚lokativ‘ zu gelten haben, nämlich einfache Sätze, die eine räumliche Relationen oder eine Bewegung im Raum ausdrücken, und dabei üblicherweise mind. ein lokatives Argument selegieren (vgl. Kap. 2.3.2 für Details der Definition). Die Prädikate (elementarer) lokativer Sätze können im Deutschen durch die vier Wortarten Verb, Nomen, Adverb/Adjektiv und Adposition realisiert sein. Diese vier Gruppen lokativer Prädikate werden jeweils auf dem Hintergrund bestehender linguistischer Arbeiten und Beschreibungen, insbesondere aus dem Bereich der lokalen Grammatik, Lexikogrammatik und des Inhalt \Leftrightarrow Text-Modells intensiv behandelt.

Ein ähnliches Bild hinsichtlich ihrer Realisierung ergab auch die Untersuchung von lokativen Argumenten (Kap. 3): lokative Argumente in (lokativen) Sätzen können ebenfalls in verschiedener Form auftreten, als Nomina/Eigennamen, Adjektive, Adverbien und unterhalb der Wortebene als Bindestrich- oder Fugenformen (Kompositaerstglieder). Die Variationsmöglichkeiten bei der Realisierung von Prädikaten wie Argumenten ermöglichen die Bildung synonymmer elementarer Sätze in vielfacher syntaktischer Form. Diese Paraphrasen elementarer Sätze sind ein zentrales Ziel der Beschreibung.

Traditionell werden unter dem Begriff ‚Lokativa‘ vor allem Nomina, die Orte bezeichnen, zusammengefasst. Sie stellen die mit Abstand größte Gruppe lokativer Argumente dar. Hier wurden Lokativa pragmatisch und anthropozentrisch definiert als Bezeichnungen für (unbewegliche) Orte, die als Aufenthaltsorte für Menschen geeignet sind. Mit der Definition

eng verbunden ist die Frage, wie sich Lokativa semantisch untergliedern und als Taxonomie organisieren lassen. Neben ausführlichen Kommentaren zu bestehenden Taxonomien wurde nach einem eigenen, pragmatischen orientierten Entwurf ein größeres Toponymlexikon als Taxonomie und Ontologie strukturiert.

Den Toponymen ist insgesamt viel Raum gewidmet. Für die Anwendung im IR sind sie sicher die wichtigste Klasse lokativer Argumente, da sie in der Referenz (meist) eindeutig und hochinformativ sind. Toponyme können in vielfacher morphologischer Gestalt realisiert sein. Sie flektieren, sie treten als Adjektive oder als Erstglieder in Nominalkomposita auf, und sie unterliegen stark konventionalisierten Regeln zur Abkürzung. Diese morphologischen Eigenschaften wurden zusammen mit den syntaktisch-kombinatorischen Eigenschaften ausführlich beschrieben und die Relevanz der Beschreibung für Aufgaben wie die Paraphrasierung untersucht. So kann ein Toponym in dreifach verschiedener Form mit dem gleichen Prädikat auftreten: *Frankreichs Präsident* vs. *der Präsident von Frankreich* vs. *der französische Präsident* – die drei Ausdrücke sind Paraphrasen voneinander.

Auf Basis der beschriebenen Eigenschaften wurde ein 50 000 Toponyme und 13 000 Derivationen umfassendes Lexikon erarbeitet. Mit Hilfe dieses Lexikons und weiterer kleinerer Lexika für lokative Nomina, Adjektive und Adverbien, wurden Phrasen, die die Argumentstellen der lokativen Prädikate auffüllen können, bis zur Ebene der Präpositionalphrase und des Adverbials untersucht, beschrieben und als lokale Grammatiken implementiert. Dazu wurden zahlreiche Phrasengrammatiken erstellt, die feste adverbiale Wendungen, phrasale Konstruktionen, Nominal- und Präpositionalphrasen, und Kombinationen aus Adverbien und Adverbialen (z.B. *hier in München*) beschreiben, und dabei Besonderheiten im Kontext von räumlichen Ausdrücken und Toponymen berücksichtigen, die über eine allgemeine deutsche Syntax hinausgehen.

Aufbauend auf den im zweiten Kapitel dargelegten theoretischen Grundlagen, wird in Kap. 4 eine syntaktische Beschreibung lokativer Prädikate versucht. Etwa 400 Verben, 250 Nomina, 200 Präpositionen und 40 Adverbien wurden vierzig elementaren Satz schemata zugeordnet und für etwa zwanzig Satz schemata die syntaktisch kombinatorischen Eigenschaften der Lemmata in Lexikongrammatiken erfasst. Die beschriebenen Eigenschaften reichen dabei von generell grammatischen wie Anzahl und syntaktische Realisierung der Aktanten (Valenz, Kasusreaktion, Selektionseigenschaften, usw.) bis hin zu klar lexikalischen wie spezifischen Modifikatoren oder morphologischen Derivationen. Die Lexikongrammatiken wurden wiederum als lokale Grammatiken implementiert, unter Verwendung der in Kap. 3 erarbeiteten Wörterbücher und Phrasengrammatiken.

Wenn immer möglich, wurde in dieser Arbeit versucht, sprachliche Ressourcen ausgehend von einem Korpus zu erstellen. Die empirisch deskriptive Ausrichtung ist bei der Erstellung lokaler und Lexikongrammatiken selbstverständlich. Teilweise automatisiert werden konnte die Nutzung von Korpora bei der syntaktischen Klassifikation der Toponyme und bei der Ermittlung der Toponymderivationen (Beziehungsadjektive und Wohnernamen).

Über Anwendungen aus dem Gebiet „Sprache und Raum“ im Bereich des IR vermittelt Kap. 5 einen weiten, wenn auch nicht vollständigen, so doch aktuellen Überblick, angefangen von der Erkennung und Geokodierung (räumliche Disambiguierung) lokativer Named Entities, über die Extraktion lexikalischer und ontologischer Ressourcen, die Zuordnung von Dokumenten zu Regionen, bis hin zur Verarbeitung geographischer Queries. Der Schwerpunkt

liegt auf der Named Entity Erkennung (NER), wo einerseits bestehende Ansätze, Annotierungsrichtlinien und -standards kritisch gewürdigt werden, andererseits ein eigenes System zur Erkennung von Toponymen in deutschsprachigen Texten implementiert wurde. Die bei der Beschreibung lokativer Sätze erarbeiteten Ressourcen – Wörterbücher und Grammatiken – werden dabei eingesetzt. Integration und Erarbeitung großer und deskriptiv tiefer linguistischer Ressourcen für Aufgaben wie die Named Entity Recognition stellen zwar kein Novum dar, werden aber wegen des Aufwands meist gemieden. Resultat ist ein System, das mit einem F_1 -Wert von gut 92% (93,4% Precision, 92,2% Recall), deutlich vor existierenden, weniger ressourcenintensiven Systemen liegt, die um etwa zehn Prozentpunkte übertroffen werden. Zu verdanken ist dies einerseits dem Toponymlexikon, das – wichtig für Sprachen wie das Deutsche – Flexionsformen und Varianten enthält, andererseits einer Reihe von „Antigrammatiken“, die versuchen Toponyme als Bestandteile anderer Entitäten (z.B. als Bestandteil im Name eines Sportvereins) herauszufiltern, und auftretende Ambiguitäten (*Essen*, *Bergen*) zu disambiguieren. Der Einfluss der intensiven Beschreibung positiver Kontexte, auf Phraseebene und als elementare (Teil-)Sätze, ist zwar spürbar, aber im Verhältnis zu den anderen beiden Faktoren doch klein.

Eine Reihe von Aufgaben konnte in dieser Arbeit nicht in Angriff genommen werden und bleibt weiteren Studien vorbehalten. Die Beschreibung der lokativen Verben, Nomina und Adverbien ist noch nicht abgeschlossen und die erstellten Tabellen sollten anhand eines Korpus verifiziert werden. Weiter sind die Lexikontabellen grundsätzlich geeignet, um für eine Ontologie relevante Relationen aus Texten zu extrahieren. Da die Relationen durch verschiedene Prädikate ausgedrückt werden können, lassen sich die Tabellen auch dazu nutzen, Paraphrasen in einem Korpus aufzuspüren. Beide Aufgaben konnten leider nicht mehr berücksichtigt werden.

Summary

The way in which language structures space and expresses spatial relations between entities has been the object of intensive linguistic research during the last three decades. In computational linguistics the analysis of spatial expressions is done in the fields of machine translation and artificial intelligence. More recent, spatial expressions have become an object of interest in Geographic Information Systems (GIS) and information retrieval on the world wide web and various other domains.

The present work focuses on the linguistic dimension of spatial expressions, and the description of ‘locative sentences’, in particular. The description includes elementary sentences and their components, predicates and arguments, and also transformations and paraphrases. The emphasis is also on syntactically reduced predications within a single noun phrase. In practice, lexicon grammars and local grammars are used as technique and description method, and are applied to spatial expressions. The dictionaries and grammars compiled in the course of the descriptive work are further used in the field of information retrieval to detect locative named entities in German texts.

Chapter 1 introduces the concepts of local grammars and lexicon grammars and gives a short overview of their properties, possible applications, their representations and implementations. The two grammar models have been chosen because of their empirical orientation, their intuitive representation by graphs and their efficient application in terms of finite state automata.

The second chapter focuses on locative predicates and the expression of space in language. First, an overview is given about various approaches in the linguistic description of spatial relations and movement. Subsequently, an attempt was made to develop a definition of ‘locative’ sentences. Locative sentences are hence defined as simple sentences expressing a spatial relation or the motion of an object, and governing at least one ‘locative’ argument. The kernel of a locative sentence, the predicate can be realized syntactically in German by the following parts-of-speech:

- (full) verbs

- (1) *Die Isar **mündet** in die Donau.* ‘The Isar River **flows** into the Danube.’
- (2) *Die Isar **speist** die Donau.* ‘The Isar River **feeds** [into] the Danube.’
- (3) *Max **reist** durch Europa.* ‘Max **travels** through Europe.’

- nouns

- (4) *Die Isar ist ein **Nebenfluss** der Donau.* ‘The Isar River is a **tributary** of the Danube.’
- (5) *Max unternimmt eine **Reise** nach Europa.* ‘Max takes a trip to Europe.’
- (6) *Max macht eine **Europareise**.* “Max takes an *Europe-trip*.”

- (7) Die **Entfernung** zwischen München und Windach beträgt nur $\frac{1}{2}$ Autostunde.
 “The **distance** between Munich and Windach is only half an hour by car.”

- adverbs/adjectives

- (8) *Max ist in Europa unterwegs.* ‘Max is **on tour** in Europe.’
 (9) *Windach liegt $\frac{1}{2}$ Autostunde von München entfernt.* ‘Windach is a half hour’s drive **away** from Munich.’
 (10) *das eine halbe Autostunde von München entfernte Windach* ‘Windach which is a half hour’s drive **away** from Munich’ (adjective in German)

- adpositions

- (11) *München liegt in Bayern.* ‘Munich is located **in** Bavaria.’ (*liegen* ‘to be located’ is considered to be a support verb)
 (12) *Ingolstadt liegt im Herzen Bayerns.* ‘Ingolstadt is located **in the heart** of Bavaria.’ (a complex, idiomatic preposition)
 (13) *Windach liegt $\frac{1}{2}$ Autostunde außerhalb Münchens.* ‘Windach is located a half hour’s drive outside of Munich.’

Similar to predicates, the realizations of the locative arguments also allow for different parts-of-speech: nouns resp. proper nouns, and adjectives, see (15). In addition, arguments can also be expressed below the word level by components of morphological compounds, cf. *Europareise* ‘trip to Europe’ in (6) or *Donauzufluss*, ‘tributary of Danube’. Consequently, this allows for the realization of locative predications not only as sentences but also as noun phrases:

- (14) *die Isarmündung in die Donau* ‘the mouth of the Isar River into the Danube’
 (15) a. *die französische Hauptstadt Paris* ‘the French capital Paris’
 b. = *Frankreichs Hauptstadt Paris* ‘France’s capital, Paris’
 c. = *Paris, die Hauptstadt von Frankreich* ‘the capital of France, Paris’

A broad discussion of existing linguistic studies referring to locative predicates, with an emphasis on local, lexicon, and dependency grammars, completes the second chapter.

The third chapter deals with locative arguments, that is locative nouns and place names (toponyms), as well as noun or adverbial phrases filling the syntactic argument slots. (Proper) nouns denoting places constitute an important semantic class. The subdivision of places and their hierarchical organization in terms of a taxonomy has been studied by many linguists and geographers. In addition to detailed comments on existing taxonomies, a toponym dictionary has been developed and (re)structured into an ontology.

Place names (toponyms) are unmistakably the most important class of locative arguments in information retrieval: they are informative to a large extent, and usually have a clear-cut reference. Toponyms can be realized in a variety of morphological forms: toponyms do inflect, toponyms may appear as adjectives or as first part of morphological compounds. Finally, toponyms are subject to very specific rules of abbreviation. These morphological characteristics together with the syntactic-combinatorial properties of toponyms have been described

in detail. The relevancy of the description becomes obvious in tasks such as paraphrasing: in (15) the toponym *Frankreich* ‘France’ appears in three different forms as argument of the same predicate. Based on the description and the evolved morphological and syntactic properties a comprehensive dictionary of 50.000 toponyms and 13.000 derivative forms (adjectives and names of inhabitants) was also developed. With the help of this dictionary and other smaller dictionaries for locative nouns, adjectives, and adverbs, phrases that can fill the argument positions of locative predicates are described, while taking into account specific features in the context of geographic terms and place names which go beyond the general German syntax. Phrase grammars are implemented as local grammars for noun and prepositional phrases, combinations with adverbs (e.g., *here in Munich*), and finally, fixed adverbial phrases and idiomatic constructions (e.g., *einen Steinwurf entfernt* ‘a stone’s throw away’).

The syntactic properties of locative predicates are described in chapter 4 based on the previously outlined theoretical principles. About 400 verbs, 250 nouns, 200 prepositions, and 40 adverbs were assigned to forty basic sentence patterns. For about twenty sentence patterns the syntactic and combinatorial properties of the lemmas are described in terms of a lexicon grammar. The grammar tables describe properties starting from the number and syntactic realization of the actants (valency, governed case, selectional restrictions, etc.) up to specific modifiers and lexical or morphological derivations. These lexicon grammars were then compiled into local grammars, making use of the dictionaries and phrase grammars of locative nouns and toponyms.

Whenever possible, an attempt was made in this work to create linguistic resources starting from a corpus. The empirical approach could be partially automated during the tasks of the syntactic classification of toponyms and in finding morphological derivations of toponyms. Here on the basis of considerably large corpora suggestions were prepared which were then checked manually to correct errors and overgenerations.

Chapter 5 gives a general overview of applications in geo-spatial Information Retrieval: geo-tagging (recognition of geographic named entities), geo-coding (spatial disambiguation), the extraction of lexical and ontological resources, the localization of documents, and the processing of geographical queries. The focus is set on Named Entity Recognition (NER). On the one hand, existing approaches and annotation standards are discussed, on the other hand, a system for the detection of toponyms in German texts is presented. This system integrates the developed resources, dictionaries as well as grammars. A F-score of 92 percent (93.4 precision, 92.2 recall) was achieved by the system, and as such exceeds less resource intensive systems by about ten percentage points. A detailed analysis shows that two resources were mainly responsible for the quality of the system. The dictionary of place names includes inflectional forms and variants which are important for languages such as German—hence it has a broader coverage on natural language texts than a mere gazetteer. Second, disambiguation of ambiguous place names (*Essen* ‘food’ resp. ‘the town of Essen’) and detection of toponyms within other entities, e.g., as components of names of soccer teams, is performed by a set of "anti-grammars." In contrast, the intense description of positive contexts (phrases and predications) provides only small but still remarkable improvements in quality.

A number of areas are not examined and remain subject to future works. The description of locative verbs, nouns and adverbs is not yet completed and the resulting tables should be carefully verified on a corpus. Further, the lexicon grammar tables are suitable to extract ontologically relevant relations from texts. This would allow for the automatic enrichment of

lexical resources by inter-linking the lexical items. Finally, the tables could also be used to detect paraphrases since they contain many “synonymous” predicates referring to identical relations.

Literaturverzeichnis

Abney, Steven

- 1991 Parsing by chunks. In: Abney, Steven; Berwick, Robert; Tenny, Carol (eds.): *Principle-based parsing*. Dordrecht: Kluwer.
- 1996a Partial parsing via finite-state cascades. *Proceedings of the Workshop on Robust Parsing, 8th European Summer School in Logic, Language and Information, Prague, Czech Republic* 8–15. <http://www.vinartus.net/spa/96h.pdf>, <http://stp.ling.uu.se/~torbjorn/kurser/FST/papers/96h.ps>
- 1996b Partial parsing via finite-state cascades. *Journal of Natural Language Engineering* 2/4, 337–344. New York: Cambridge University Press. <http://www.vinartus.net/spa/97a.pdf>

Abraham, Werner

- 2003 The myth of doubly governing prepositions in German. In: Shay & Seibert (eds.): *Motion, Direction and Location in Language* 19–38.

ACE

Automatic Content Extraction (ACE). <http://www.nist.gov/speech/tests/ace/>, <http://projects.ldc.upenn.edu/ace/>

Ackeren, Isabell van

- 1999 Präpositionen als 'Raumausstatter'? Sprachliche Verarbeitung räumlicher Wahrnehmung durch Präpositionen mit lokaler Lesart. http://www.linse.uni-essen.de/esel/pdf/pr_raum.pdf

Ackermann, Chantal

- 1999 Codierung von semantischen Informationen in einem Lexikon des Deutschen. Studie: Lokativa. Hauptseminararbeit, Ludwig-Maximilians-Universität München.

ADL

Alexandria Digital Library Gazetteer. <http://www.alexandria.ucsb.edu/gazetteer/>

Ágel, Vilmos

- 2000 *Valenztheorie*. Ein Arbeitsbuch. Tübingen: Gunter Narr.

Ágel et al. eds. [Ágel, Vilmos; Eichinger, Ludwig M.; Eroms, Hans Werner; Hellwig, Peter; Heringer, Hans Jürgen; Lobin, Henning (eds.)]

- 2003–2006 *Dependenz und Valenz*. Handbücher zur Sprach- und Kommunikationswissenschaft 25/1-2. Berlin, New York: Mouton de Gruyter.

Altfeld, Jürgen

- 2000 Maschinelles Adressen-Abgleich – Extraktion und Struktur-Analyse von Adressen aus Web-Seiten. Magisterarbeit im Studiengang Computerlinguistik, Centrum für Informations- und Sprachverarbeitung, Ludwig-Maximilians-Universität München.

Amores, J. Gabriel; Mora, Juan P.

- 1998 Machine translation of motion verbs. In: Martín Vide, Carlos (ed.): *Mathematical and computational analysis of natural languages*. Studies in Functional and Structural Linguistics. Selected papers from the 2nd International Conference on Mathematical Linguistics, Tarragona (ICML '96), 2 - 4 May 1996 45, 191–205. Amsterdam: John Benjamins.

Anderegg, Frank

- 2002 Raumbezogene Datentypen in SQL/MM Spatial und verwandten Standards.
http://www.integis.ch/documents/ISem_Anderegg_SQLMM_2002-07-04.pdf

André, Elisabeth; Herzog, Gerd; Rist, Thomas

- 1989 Natural language access to visual data: Dealing with space and movement.
<http://www.ai.mit.edu/people/jimmylin/papers/Andre89.pdf> [This paper appeared as: Report 63, Universität des Saarlandes, SFB 314 (VITRA), November 1989. It has been presented at the 1st Workshop on Logical Semantics of Time, Space and Movement in Natural Language, Toulouse, France, 1989.]

Antti Arppe, Lauri Carlson, Krister Lindèn, Jussi Piitulainen, Mickael Suominen, Martti Vainio, Hanna Westerlund, and Anssi Yli-Jyrä (eds.)

- 2005 *Inquiries into words, constraints, and contexts*. Festschrift in Honour of Kimmo Koskenniemi on his 60th Birthday. Stanford: CSLI Publications.
<http://csli-publications.stanford.edu/koskenniemi-festschrift/>

Apresjan, Jurij Derenikovič

- 1964 O sil'nom i slabom upravlenii. *Voprosy jazykoznanija* 3, 32–49.
1966 *Idei i metody sovremennoj strukturnoj lingvistiki*. Moskva: Prosveščenie.
1973 Regular polysemy. *Linguistics* 142, 5–32. [Übersetzung von 1971, *Izvestija akademii nauk SSSR: Serija literatury i jazyka* 30, 509–23]
1974 *Leksičeskaja semantika*. Sinonimičeskie sredstva jazyka. Moskva: Nauka.

Apresjan et al. [Apresjan, Jury D.; Boguslavsky, Igor M.; Iomdin, Leonid L.; Tsinman, Leonid L.]

- 2002 Lexical functions in NLP: possible uses. *Computational linguistics for the new millennium: divergence or synergy? Festschrift in honour of Peter Hellwig on the occasion of his 60th birthday* 55–72. Peter Lang. http://proling.iitp.ru/bibitems/LF_uses.pdf

Apresjan et al. [Apresjan, Jurij D.; Boguslavskij, Igor M.; Iomdin, Leonid L.; Lazurskij, Alexandre V.; Sannikov, Vladimir Z.; Tsinman, Leonid L.]

- 1992 ETAP-2: The Linguistics of a Machine Translation System. *Meta: Journal des Traducteurs – Translators' Journal* 37/1, 97–112.

Apresjan et al. [Apresjan, Jurij Derenikovič; Boguslavskij, Igor' Michajlovič; Iomdin, Leonid Lejbovič; Lazurskij, Aleksandr Vadimovič; Percov, Nikolaj Viktorovič; Sannikov, Vladimir Zinov'evič; Cinman, Leonid L'vovič]

- 1989 *Lingvističeskoe obespečenie sistemy ÉTAP-2*. Moskva: Nauka.

Arampatzis et al. [Arampatzis, Avi; Kreveld, Marc van; Reinbacher, Iris; Clough, Paul; Joho, Hideo; Sanderson, Mark]

- 2004 Web-based delineation of imprecise regions. *SIGIR 2004*.
<http://www.geo.unizh.ch/~rsp/gir/abstracts/arampatzis.pdf>

Arapov, M. V.

- 1972 Sootnošenie suščestvitel'nych so značeniem „nazvanie mestnosti“ i „nazvanie žitelej dannoj mestnosti“ v sovremennom russkom jazyke. *Problemy strukturnoj lingvistiki* 1971, 469–503. Moskva.

ARCHINFORM

Hendel, Sascha: International Architecture Database. <http://www.archinform.net/>

Arppe, Antti

- 2005 The very long way from basic linguistic research to commercially successful language business: the case of two-level morphology. In: Antti Arppe (eds.): *Inquiries into words, constraints, and contexts* 2–17.
<http://csli-publications.stanford.edu/koskenniemi-festschrift/1-arppe.pdf>

- Asher, Nicholas; Sablayrolles, Pierre
 1994a A typology and discourse semantics for motion verbs and spatial PPs in French. *Journal of Semantics* 12/1, 163–209.
 1994b A compositional spatio-temporal semantics for French motion verbs and spatial PPs. *Proceedings of SALT4, Semantics and Linguistic Theory, Rochester, NY, May 6-8, 1994*.
 1996 A typology and discourse semantics for motion verbs and spatial PPs in French. In: Pustejovsky; Boguraev (eds.): *Lexical semantics* 163–209.
- Asher, Nicholas; Vieu, Laure
 1995 Toward a geometry of common sense: A semantics and a complete axiomatization of mereotopology. In: Mellish, C. S. (ed.): *Proceedings of IJCAI-95* 1, 846–852. Montréal.
- Aurnague et al. [Aurnague, Michel; Bras, Myriam; Vieu, Laure; Asher, Nicholas]
 2001 The syntax and semantics of locating adverbials. *Cahiers de Grammaire* 11–35.
<http://www.irit.fr/~Laure.Vieu/Publis/CG01.pdf>
- Aurnague, Michel; Vieu, Laure
 1996 A three-level approach to the semantics of space. In: Zelinsky-Wibbelt (ed.): *The semantics of prepositions from mental processing to natural language processing* 393–439.
- Auroux et al. eds. [Auroux, Sylvain; Koerner, E. F. K.; Niederehe, Hans-Josef; Versteegh, Kees (eds.)]
 2000–2006 *History of the Language Sciences. Geschichte der Sprachwissenschaften. Histoire des. Ein internationales Handbuch zur Entwicklung der Sprachforschung von den Anfängen bis zur Gegenwart. Handbücher zur Sprach- und Kommunikationswissenschaft* 18. Berlin, New York: Mouton de Gruyter.
- Axelrod, Amittai E.
 2003 On building a high performance gazetteer database. *Workshop on the Analysis of Geographic References, NAACL'03*.
<http://www.metacarta.com/kornai/NAACL/WS9/Conf/ws910.pdf>
- Babkin, Aleksandr Michajlovič; Levašov, Evgenij Aleksandrovič (eds.)
 1975 *Slovar' nazvanij žitelej SSSR. Okolo 10 000 nazvanij*. Moskva: Russkij jazyk.
- Back, Otto
 1995 Typologie der Ländernamen: Staaten-, Länder-, Landschaftsnamen. In: Eichler et al. (eds.): *Namenforschung: ein internationales Handbuch zur Onomastik* 1348–1356.
- Bähr, Hans-Peter; Müller, Marina
 2004 Graphics and language as complementary formal representations for geospatial descriptions. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences (IAPRSIS)* 35/B4.
http://www.ipf.uni-karlsruhe.de/Personen/mueller/literatur/Baehr_Mueller.pdf
- Bauer, Gerhard
 1998 [1985] *Namenkunde des Deutschen*. Germanistische Lehrbuchsammlung 21. Berlin.
- Bauer, Reinhard
 1995 Amtliche Geltung und Schreibung von Orts- und Flurnamen. In: Eichler et al. (eds.): *Namenforschung: ein internationales Handbuch zur Onomastik* 1790–1795.
- BBN
 BBN Technologies. <http://bbn.com/>
- BBNTAXON
 Brunstein, Ada [= BBN Technologies] 2002: Annotation guidelines for answer types.
<http://www ldc.upenn.edu/Catalog/docs/LDC2005T33/BBN-Types-Subtypes.html>

- Becker, A. L.; Arms, D. G.
 1969 Prepositions as predicates. In: Binnick, Robert I.; Davison, Alice; Green, Georgia M.; Morgan, Jerry L. (eds.): *Papers from the fifth regional meeting of the Chicago Linguistic Society, April 18-19, 1969* 1–11. Chicago (Illinois): Department of Linguistics, University of Chicago.
- Becker, Angelika
 1994 *Lokalisierungsausdrücke im Sprachvergleich*. Eine lexikalisch-semantische Analyse von Lokalisierungsausdrücken im Deutschen, Englischen, Französischen und Türkischen. Tübingen: Niemeyer.
- Bekavac, Božo; Tadić, Marko
 2007 Implementation of Croatian NERC system. *Balto-Slavonic Natural Language Processing 2007, Workshop at ACL 2007, June 29th, 2007, Prague, Czech Republic*.
<http://hmk.ffzg.hr/txts/BBMT4ACL2007.pdf>,
<http://www.aclweb.org/anthology-new/W/W07/W07-1702.pdf>
- Belleil, Claude
 1997 Reconnaissance, typage et traitement des coréférences des toponymes français et de leurs gentilés par dictionnaire électronique relationnel. Thèse de doctorat d'Informatique, Université de Nantes.
- Belleil, Claude; Maurel, Denis
 1997 Un dictionnaire relationnel des noms propres liés à la géographie, consultés par transducteurs. *Meta: Journal des traducteurs* 42/2, 273–282.
<http://www.erudit.org/revue/meta/1997/v42/n2/002053ar.pdf>
- Beneš, Eduard
 1974 Präpositionswertige Präpositionalwendungen. In: Engel, Ulrich; Grebe, Paul (ed.): *Sprachsystem und Sprachgebrauch* 1, 33–52. Düsseldorf.
- Bennett, David C.
 1976 *Spatial and temporal uses of English prepositions*. London: Longman.
- Berger, Dieter; Drosdowski, Günter et al. [= Dudenredaktion]
³1985 *Duden „Richtiges und gutes Deutsch“*. Mannheim.
- Berman, Judith
 2003 *Clausal Syntax of German*. Studies in constraint-based lexicalism. Stanford: CSLI.
- Biemann, Christian
 2005 Ontology learning from text: A survey of methods. *LDV-Forum 2005* 20/2, 75–93.
- Bierwisch, Manfred
 1988 On the grammar of local prepositions. In: Bierwisch, Manfred; Motsch, Wolfgang; Zimmermann, Ilse (eds.): *Syntax, Semantik und Lexikon*. Rudolf Růžička zum 65. Geburtstag. *Studia Grammatica* 29, 1–65. Berlin: Akademie.
- Bierwisch, Manfred; Lang, Ewald
 1989 Somewhat longer – much deeper – further and further. Epilogue to the dimension adjective project. In: Bierwisch; Lang (eds.): *Dimensional adjectives* 471–514.
- Bierwisch, Manfred; Lang, Ewald (eds.)
 1987 *Grammatische und konzeptionelle Aspekte von Dimensionsadjektiven*. *Studia grammatica* 26-27. Berlin: Akademie.

Bierwisch, Manfred; Peter Bosch (eds.)

- 1995 Semantic and conceptual knowledge. *Papers for a joint workshop of the Arbeitsgruppe Strukturelle Grammatik and the Institut für Logik und Linguistik, 21-23 April, 1994, in Berlin. Arbeitsberichte des Sonderforschungsbereichs 340. Sprachtheoretische Grundlagen für die Computerlinguistik 71.* Stuttgart, Tübingen.

Blanc, Olivier

- 2006 Algorithmes d'analyse syntaxique par grammaires lexicalisées : optimisation et traitement de l'ambiguïté. Thèse de doctorat en linguistique-informatique, Université de Marne-la-Vallée. <http://igm.univ-mlv.fr/LabInfo/theses/2006/blanc.ps>

Blanc, Olivier; Constant, Mathieu

- 2005 Lexicalisation of grammars with parameterized graphs. *Proceedings of the international conference "Recent Advances in Natural Language Processing", RANLP 2005, 21-23 September 2005, Borovets, Bulgaria* 117–121.
http://igm.univ-mlv.fr/~mconstan/papers/blanc_constant_ranlp05.pdf
- 2006 Outilex, a Linguistic Platform for Text Processing. *ACL 2006, 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference, Sydney, Australia, 17-21 July 2006.* <http://acl.ldc.upenn.edu/P/P06/P06-4019.pdf>

Blanc, Olivier; Constant, Mathieu; Laporte, Éric

- 2006 Outilex, plate-forme logicielle de traitement de textes écrits. *Actes de TALN 2006, Louvain (Belgique), 10-12 avril 2006.* <http://arxiv.org/pdf/0711.3691v2>

Blanc, Olivier; Constant, Mathieu; Watrin, Patrick

- 2007a A finite-state super-chunker. In: Holub, Jan; Zdárek, Jan (eds.): *Implementation and Application of Automata, 12th International Conference, CIAA 2007, Prague, Czech Republic, July 16-18, 2007, Revised Selected Papers.* Lecture Notes in Computer Science 4783, 306–308. Springer.
- 2007b Segmentation in super-chunks with a finite-state approach. *Finite-State Methods and Natural Language Processing (FSMNLP) 2007, Potsdam, 14 – 16 September.*

Bloomfield, Leonhard

- 1933 *Language.* New York: Holt, Rinehart & Winston.

Bloom et al. eds. [Bloom, Paul; Peterson, Mary A.; Nadel, Lynn; Gerrett, Merrill F. (eds.)]

- 1996 *Language and space.* Cambridge (Massachusetts): MIT Press.

Blutner, Reinhard

- 1995 Systematische Polysemie: Ansätze zur Erzeugung und Beschränkung von Interpretationsvarianten. In: Bierwisch & Bosch (eds.): *Semantic and conceptual knowledge* 33–67.

Boas, Hans Ulrich

- 1998 Locational verbs in a lexicon grammar of English. In: Caron, Bernar (ed.): *ICL'97: Proceedings of the XVIth International Congress of Linguists, Paris, 20 - 25 July 1997.* Amsterdam: Pergamon. [CD]

Boguslavsky, Igor M.; Barrios Rodríguez, María Auxiliadora; Diachenko, P.

- 2006 CALLEX-ESP: a software system for learning Spanish lexicon and collocations. In: Méndez-Vilas, A.; Solano Martín, A.; Mesa González, J. A.; Mesa González, J. (eds.): *Current Developments in Technology-Assisted Education* 1, 22–26.
<http://www.formatex.org/micte2006/pdf/22-26.pdf>

Bohnenmeyer et al. [Bohnenmeyer, J.; Enfield, N. J.; Essegbey, J.; Ibarretxe-Antuñano, I.; Kita, S.; Lüpke, F.; Ameka, F. K.]

- 2007 Principles of event representation in language: The case of motion events. *Language* 83/3. <http://linguistics.buffalo.edu/people/faculty/bohnenmeyer/principles7.pdf>

- Boons, Jean Paul
- 1985 Préliminaires à la classification des verbes locatifs: les compléments de lieu, leurs critères, leurs valeurs aspectuelles. *Lingvisticae investigationes* 9/2, 195–267.
- 1987 La notion sémantique de déplacement dans une classification syntaxique des verbes locatifs. *Langue Française* 76, 5–40. Paris: Larousse.
- Boons, Jean-Paul; Guillet, Alain; Leclère, Christian
- 1976 *La structure des phrases simples en français*. Constructions intransitives. Langue et cultures 8. Genève: Droz.
- 1978 Remarks on the principles and the goals of the syntactic tables of lexical elements. *SMIL, Journal of Linguistic Calculus* 1978/3-4, 10–24. Stockholm: Skriptor.
- Breindl, Eva
- 2006 Präpositionalphrasen. In: Ágel et al. (eds.): *Dependenz und Valenz* 936–951.
- o.J. [¹1999] Präpositionen-Bibliografie. Projekt 'Handbuch der deutschen Konnektoren', Institut für Deutsche Sprache. <http://www.ids-mannheim.de/gra/konnektoren/prp-zitat.html>
- Brekle, Herbert E.
- 1976 An early plea for a relational treatment of verbs and prepositions: John Fearn's Anti-Tookey (1824/27). In: Parret, Herman (ed.): *History of linguistic thought and contemporary linguistics* 503–517. Berlin: de Gruyter.
- Brendler, Andrea; Brendler, Silvio (eds.)
- 2004 *Namenarten und ihre Erforschung*. Ein Lehrbuch für das Studium der Onomastik. Lehr- und Handbücher zur Onomastik 1. Hamburg: Baar.
- Brendler, Silvio
- 2004 Klassifikation der Namen. In: Brendler & Brendler (eds.): *Namenarten und ihre Erforschung* 69–92.
- Brill, Eric
- 1993 A corpus-based approach to language learning. Dissertation, Department of Computer and Information Science, University of Pennsylvania. <http://www.cs.jhu.edu/~brill/dissertation.ps>, <http://citeseer.ist.psu.edu/223500.html>
- Brill, Eric; Resnik, Phillip
- 1994 A rule-based approach to prepositional phrase attachment disambiguation. *Proceedings of 15th International Conference on Computational Linguistics (COLING94)*, Kyoto, Japan, 1994. <http://citeseer.ist.psu.edu/brill94rulebased.html>, <http://acl.lldc.upenn.edu/C/C94/C94-2195.pdf>
- Brown, Edward Keith; Jim E. Miller (eds.)
- 1996 *Concise encyclopedia of syntactic theories*. Cambridge: Elsevier.
- Bsiri, Sandra; Geierhos, Michaela; Ringlsetter, Christoph
- 2008 Structuring job search via local grammars. In: Gelbukh, Alexander (ed.): *Advances in natural language processing and applications*. Research in Computing Science 33, 201–212. <http://www.gelbukh.com/cicling/2008/RCS-vol-33/16-Bsiri.pdf>
- BUFFALOONTOLOGY
- Mark, David M.; Smith, Barry: Geographic categories: an ontological investigation. <http://www.geog.buffalo.edu/ncgia/ontology/>
- Buitelaar, Paul; Cimiano, Philipp; Magnini, Bernardo (eds.)
- 2005 *Ontology learning from text*. Methods, evaluation and applications. Frontiers in Artificial Intelligence and Applications 123. Amsterdam: IOS Press.

- Buscaldi, Davide; Rosso, Paolo; Arnal Sanchis, Emilio
 2005 A WordNet-based query expansion method for geographical information retrieval. *Working Notes for the CLEF 2005 Workshop, 21-23 September, Vienna, Austria*.
http://www.clef-campaign.org/2005/working_notes/workingnotes2005/buscaldi05.pdf
- Bußmann, Hadumod (ed.)
³2002 *Lexikon der Sprachwissenschaft*. Stuttgart: Kröner.
- Buyukkokten et al. [Buyukkokten, Orkut; Cho, Junghoo; Garcia-Molina, Hector; Gravano, Luis; Shivakumar, Narayanan]
 1999 Exploiting geographical location information of web pages. *Proceedings of the ACM SIGMOD Workshop on the Web and Databases (WebDB'99)*.
<http://www1.cs.columbia.edu/~gravano/Papers/1999/webdb.pdf>
- Byrne, Kate
 2007 Nested Named Entity recognition in historical archive text.
<http://homepages.inf.ed.ac.uk/s0233752/blog/refs/icsc07kb.pdf>
- CANOO.NET
 Canoo.net – free online german language resources. <http://www.canoo.net/>
- Caroli, Folker
 1984a La notion de directionnalité dans les verbes de déplacement en allemand. In: Guillet, Alain; Fauci, Nuncio La (eds.): *Actes du 1er colloque Européen sur la grammaire et le lexique comparés des langues romanes, 21 au 24 sept. 1981 à Palerme* 1–35. Amsterdam, Philadelphia: John Benjamins.
 1984b Les verbes transitifs à complément de lieu en allemand. *Lingvisticae investigationes* 8, 225–267. Amsterdam, Philadelphia: John Benjamins.
- Caroli, Folker; Figge, Udo L.
 1989 Vergleichende Klassifikation deutscher und französischer Verben mit Lokalgängung. *Archiv für das Studium der neueren Sprachen und Literaturen* 226, 241–265.
- Caroli et al. [Caroli, Folker; Haller, Johann; Schütz, Jörg; Truar, Mathias]
 1990 Lexicon Importation/Exportation. In: Rieger, Burghard; Schaefer, Burkhard (eds.): *Lexikon und Lexikographie. Maschinell - maschinell gestützt. Grundlagen, Entwicklungen, Produkte. Vorträge im Rahmen der Jahrestagung 1990 der Gesellschaft für Linguistische Datenverarbeitung (GLDV) e.V., Siegen, 26. - 28. März 1990*. Sprache und Computer 11, 62–78. Hildesheim: Olms. [76–78 Tabelle transitiver Verben mit Lokalgängung]
- Carstensen, Kai-Uwe
 2000 Räumliche Mikroperspektiven und die Semantik lokaler Präpositionen. In: Habel; Stutterheim (eds.): *Räumliche Konzepte und sprachliche Strukturen* 237–260.
- Carstensen, Kai-Uwe; Simmons, Geoff
 1991 Why a hill can't be a valley: Representing gestalt and position properties of objects with object schemata. In: Herzog & Rollinger (eds.): *Text understanding in LILOG* 632–644.
- Chaves, Marcirio Silveira
 2005a Em direção a uma Base de Conhecimento de Senso Comumem Língua Portuguesa.
<http://xldb.fc.ul.pt/~mchaves/doctor/articles/iiia.pdf>
 2005b Extrairdo Informação através de Gramáticas Locais.
http://xldb.fc.ul.pt/~mchaves/doctor/articles/lc_2.pdf

- Chaves, Marcirio Silveira; Rodrigues, Catarina; Silva, Mário J.
 2007 Data model for geographic ontologies generation. *XATA2007 - XML: Aplicações e Tecnologias Associadas* 47–58.
http://xldb.fc.ul.pt/data/Publications_attach/xata2007-camera-ready.pdf
- Chaves, Marcirio Silveira; Silva, Mário J.; Martins, Bruno
 2005 GKB - Geographic Knowledge Base. DI-FCUL TR-5-12.
<http://www.di.fc.ul.pt/tech-reports/05-12.pdf>
- Chinchor, Nancy
 1997 MUC-7 named entity task definition, version 3.5.
http://www.itl.nist.gov/iaui/894.02/related_projects/muc/proceedings/ne_task.html
- Chinchor et al. [Chinchor, Nancy; Brown, Erica; Ferro, Lisa; Robinson, Patty]
 1999 1999 Named Entity Recognition Task Definition.
ftp://jaguar.ncsl.nist.gov/ace/phase1/ne99_taskdef_v1_4.pdf
- Chinchor, Nancy; Robinson, Patty; Brown, Erica
 1998 The HUB-4 IE-NE Task Definition Version 4.8.
http://www.nist.gov/speech/tests/bnr/hub4_98/hub4_98.htm,
http://www.nist.gov/speech/tests/bnr/hub4_98/h4_iene_task_def.4.8.ps
- Chomsky, Noam
 1956 Three models for the description of language. *IRE Transactions on Information Theory* 2, 113–124.
 1957 *Syntactic structures*.
- Chomsky, Noam; Miller, George A.
 1963 Introduction to the formal analysis of natural languages. In: Luce, R. D.; Bush, R. R.; Galanter, E. (eds.): *Handbook of mathematical psychology* 2, 269–321. New York: John Wiley and Sons.
- Chrobot, Agata [= Savary, Agata]
 1998 Flexion automatique des mots composés. In: Klein, J.; Lamiroy, Béatrice; Pierret, J.-M. (eds.): *Théorie linguistique et applications informatiques*. Actes du XVI colloque européen sur la grammaire et le lexique comparés, Louvain-la-Neuve, septembre 1997. Cahiers de l'Institut de Linguistique de Louvain 145–159. Louvain-la-Neuve.
- Clemenceau, David
 1993 Structuration du lexique et reconnaissance des mots dérivés. PhD thesis, University of Paris VII, Paris, France.
- Clément, Lionel; Gerdes, Kim; Kahane, Sylvain
 2002 An LFG-type grammar for German based on the topological model. *Proceedings of the LFG02 Conference*. <http://csli-publications.stanford.edu/LFG/7/lfg02cgk-num.pdf>
- Coates-Stephens, Sam
 1991 Automatic acquisition of proper noun meanings. In: Ras, Zbigniew W.; Zemankova, Maria (eds.): *Methodologies for Intelligent Systems*. 6th International Symposium, ISMIS '91, Charlotte, N.C., USA, October 16-19, 1991, Proceedings. Lecture Notes in Computer Science 542, 306–315. Springer.
 1992 The analysis and acquisition of proper names for the understanding of free text. *Computers and the Humanities* 26/5-6, 441–456. Springer.
- Collier et al. [Collier, Nigel; Takeuchi, Koichi; Nobata, Chikashi; Fukumoto, Junichi; Ogata, Norihiro]
 2002 Progress on multi-lingual named entity annotation guidelines using RDF(S). *Third International Conference on Language Resources and Evaluation, Las Palmas de Gran*

Canaria, Spain. 29th - 31st May 2002 2074–2081.
<http://research.nii.ac.jp/~collier/papers/LREC2002.ps>

CoNLL-2003-SHARED-TASK

CoNLL-2003 shared task: Language-Independent Named Entity Recognition.
<http://cnts.ua.ac.be/conll2003/ner/>

Constant, Matthieu

- 2002 On the analysis of locative phrases with graphs and lexicon-grammar: the classifier/proper noun pairing. In: Ranchhod, Elisabete; Mamede, Nuno J. (eds.): *Advances in natural language processing*. Proceedings of Third International Conference, PorTAL 2002, Faro, Portugal, June 23-26, 2002. Lecture notes in artificial intelligence 2389, 33–42. Berlin: Springer.
- 2003a Converting linguistic systems of relational matrices into finite state transducers. *Proceedings of the EACL Workshop on Finite State Methods in Natural Language Processing, Budapest*.
- 2003b Grammaires locales pour l'analyse automatique de textes: méthodes de construction et outils de gestion. Thèse de doctorat en informatique, Université de Marne-la-Vallée.
<http://igm.univ-mlv.fr/LabInfo/theses/2003/constant.pdf>
- 2004 GRAAL, une bibliothèque de graphes : mode d'emploi. In: Muller, C.; Royauté, J.; Silberstein, M. (eds.): *Cahiers de la MSH Ledoux 1, INTEX pour la linguistique et le traitement automatique des langues* 321–330. Besançon: Presse Universitaire de Franche-Comté.
- 2007 GraalWeb ou accéder à une bibliothèque décentralisée de grammaires locales. *Colloquium Bases de données lexicales: construction et applications (BDL-CA)*, Montréal, 23 avril 2007. <http://www.olst.umontreal.ca/pdf/BDL-CA07.pdf>

CORINE

CORINE Land Cover 2000. <http://www.corine.dfd.dlr.de/>

Costa, Lucio; Steinhauer, Zinnia Millincent

- 1995 Parsing with the lexicon grammar, Harrisian transformations and local subgrammars. In: D'Agostino, Emilio (ed.): *Tra sintassi e semantica: descrizioni e metodi di elaborazione automatica della lingua d'uso*. Pubblicazioni dell'Università degli Studi di Salerno, sezione di studi filologici, letterari e artistici 28, 449–463. Napoli: Edizioni Scientifiche Italiane.

Courtois, Blandine

- 1990 Un système de dictionnaires électroniques pur les mots simples du français. In: Courtois, Blandine, Silberstein, Max D. (eds.): *Dictionnaires électroniques du français*. Langue Française 87, 11–22. Paris: Larousse.
- 2004 Dictionnaires électroniques *DELA*F anglais et français. In: Leclère et al. (eds.): *Lexique, syntaxe et lexique-grammaire* 113–123.

Coventry, Kenny

- 2003 Spatial prepositions, spatial templates, and 'semantic' versus 'pragmatic' visual representations. In: van der Zee & Slack (eds.): *Representing direction in language and space* 255–267.

Creary, Lewis J.; Gawron, Mark; Nerbonne, John

- 1987 Towards a theory of locative reference. *Proceedings of the ASL 1987*.
- 1989 Reference to locations. *Proceedings of the ACL 1989*. <http://citeseer.ist.psu.edu/creary89reference.html>, <http://acl.ldc.upenn.edu/P/P89/P89-1006.pdf>

- Crestan, Eric; Loupy, Claude de
 2004 Browsing help for a faster retrieval. *Proceedings of the 20th international Conference on Computational Linguistics (Geneva, Switzerland, August 23 - 27, 2004)* 576–582.
<http://acl.ldc.upenn.edu/C/C04/C04-1083.pdf>,
<http://dx.doi.org/10.3115/1220355.1220438>
- Crysmann, Berthold
 2005 Syncretism in German: a unified approach to underspecification, indeterminacy, and likeness of case. In: Müller, Stefan (ed.): *The Proceedings of the 12th International Conference on Head-Driven Phrase Structure Grammar Department of Informatics, University of Lisbon, August 23-24, 2005* 91–107.
<http://csli-publications.stanford.edu/HPSG/6/crysmann.pdf>
- Cucchiarelli, Alessandro; Luzi, Danilo; Velardi, Paola
 1998 Automatic semantic tagging of unknown proper names.
<http://www.cs.mu.oz.au/acl/P/P98/P98-1045.pdf>
 1999 Semantic tagging of unknown proper nouns. *Natural Language Engineering* 5/2, 171–185. Cambridge: Cambridge University Press.
- Cunqueiro, Cibrán Beiras
 2005 Lösung von lexikalischen Ambiguitäten in der spanischen Sprache mittels des Formalismus Elag. Abschlussarbeit für das Aufbaustudium Computerlinguistik an der Ludwig-Maximilians-Universität München. <http://igm.univ-mlv.fr/~unitex/ELAG.pdf>
- Curran, James R.; Clark, Stephen
 2003 Language Independent NER using a Maximum Entropy Tagger. *Proceedings of the seventh conference on Natural language learning at HLT-NAACL 2003, Edmonton, Canada* 4, 164–167. <http://citeseer.ist.psu.edu/655980.html>
- DAML:GEOFILE
 DARPA Agent Markup Language (DAML) 2001: DAML Geofile.
<http://www.daml.org/2001/02/geofile/>
- DEC
 Mel'čuk et al. [Mel'čuk, Igor; N. Arbatchewsky-Jumarie; Elnitsky L.; L. Iordanskaja; A. Lessard] 1984: *Dictionnaire explicatif et combinatoire du français contemporain. Recherches lexico-sémantiques I*. Montréal: Les Presses de l'Université de Montréal.
 Mel'čuk et al. [Mel'čuk, Igor; N. Arbatchewsky-Jumarie; L. Dagenais; L. Elnitsky; L. Iordanskaja; M.-N. Lefebvre; S. Mantha] 1988: *Dictionnaire explicatif et combinatoire du français contemporain. Recherches lexico-sémantiques II*. Montréal: Les Presses de l'Université de Montréal.
 Mel'čuk et al. [Mel'čuk, Igor; N. Arbatchewsky-Jumarie; L. Iordanskaja; S. Mantha] 1992: *Dictionnaire explicatif et combinatoire du français contemporain. Recherches lexico-sémantiques III*. Montréal: Les Presses de l'Université de Montréal.
 Mel'čuk et al. [Mel'čuk, Igor; N. Arbatchewsky-Jumarie; L. Iordanskaja; S. Mantha; A. Polguère] 1999: *Dictionnaire explicatif et combinatoire du français contemporain. Recherches lexico-sémantiques IV*. Montréal: Les Presses de l'Université de Montréal.
- Diersch, Helga
 1972 *Verben der Fortbewegung in der deutschen Sprache der Gegenwart*. Eine Untersuchung der syntagmatischen und paradigmatischen Beziehungen des Wortinhalts. Berlin: Akademie.
- Di Meola, Claudio
 1994 *Kommen und gehen*. Eine kognitiv-linguistische Untersuchung der Polysemie deiktischer Bewegungsverben. Linguistische Arbeiten 325. Tübingen: Niemeyer.

- 2000 *Die Grammatikalisierung deutscher Präpositionen*. Studien zur deutschen Grammatik 62. Tübingen: Stauffenburg.
- Ding, Junyan; Gravano, Luis; Shivakumar, Narayanan
 2000 Computing geographical scopes of web resources. *Proceedings of the 26th International Conference on Very Large Data Bases (VLDB'00)*.
<http://www1.cs.columbia.edu/~gravano/Papers/2000/vldb00.pdf>
- DUDEN2006
 Eisenberg, Peter et al. [= Dudenredaktion] ⁷2006: *Duden: Grammatik der deutschen Gegenwartssprache*. Der Duden in 12 Bänden 4. Mannheim: Dudenverlag. [völlig neu erarbeitete und erweiterte Auflage]
- DUDEN-GRAMMATIK
 [Dudenredaktion]: Grebe et al. [Grebe, Paul; Gipper, Helmut; Mangold, Max; Mentrup, Wolfgang; Winkler, Christian] ³1973: *Duden: Grammatik der deutschen Gegenwartssprache*. Mannheim: Dudenverlag.
 [Dudenredaktion]: Drosdowski, Günter et al. ⁴1984: *Duden: Grammatik der deutschen Gegenwartssprache*. Mannheim: Dudenverlag.
 [Dudenredaktion]: Eisenberg, Peter et al. ⁶1998: *Duden: Grammatik der deutschen Gegenwartssprache*. Mannheim: Dudenverlag. [neu bearb. Auflage]
- Egenhofer, Max J.
 1994 Spatial SQL: A query and presentation language. *IEEE Transactions on Knowledge and Data Engineering* 6, 86–95. <http://www.spatial.maine.edu/~max/SSQL.pdf>
- Eggert, Elmar
 2005 *Bisontins ou Besançonnais? À la recherche des règles pour la formation des gentils pour une application au traitement automatique*. Tübinger Beiträge zur Linguistik 480. Tübingen: Gunter Narr. [Dissertation Universität Münster und Universität Tours, 2002]
- Eggert, Elmar; Maurel, Denis; Belleil, Claude
 1998 Allomorphies et suppléments dans la formation des gentils : Application au traitement informatique. *Cahiers de Lexicologie* 73/2, 167–179.
- Ehrich, Veronika
 1996 Verbbedeutung und Verbgrammatik: Transportverben im Deutschen. In: Lang; Zifonun (eds.): *Deutsch – typologisch* 229–260.
- Eichler, Ernst (ed.)
 2004 *Völkernamen, Ländernamen, Landschaftsnamen*. Protokoll der gleichnamigen Tagung im Herbst 2003 in Leipzig. *Onomastica Lipsiensia* 2. Leipzig: Leipziger Universitäts-Verlag.
- Eichler et al. eds. [Eichler, Ernst; Hilty, Gerold; Löffler, Heinrich; Steger, Hugo; Zgusta, Ladislav (eds.)]
 1995 *Namenforschung: ein internationales Handbuch zur Onomastik*. Handbücher zur Sprach- und Kommunikationswissenschaft 11.
- Eichler, Ernst et al. [= Autorenkollektiv]
 1971 *Namenforschung heute*. Ihre Ergebnisse und Aufgaben in der Deutschen Demokratischen Republik. Berlin: Akademie.
- Engelke, Sabine
 2003 *Freie und feste Adverbiale im Deutschen*. Studien zur Informations- und Sprachverarbeitung 8. München: Centrum für Informations- und Sprachverarbeitung. [Dissertation Ludwig-Maximilians-Universität München]

Engel, Ulrich

³1996 [1988] *Deutsche Grammatik*. Heidelberg: Julius Groos.

EN.WIKIPEDIA.ORG

Wikipedia. <http://en.wikipedia.org/>

Eroms, Hans-Werner

1981 *Valenz, Kasus und Präposition*. Untersuchungen zur Syntax und Semantik präpositionaler Konstruktionen in der deutschen Gegenwartssprache. Monographien zur Sprachwissenschaft 11. Heidelberg: Winter.

2000 *Syntax der deutschen Sprache*. Berlin, New York: Walder de Gruyter.

Erp, Marieke van

2006 Bootstrapping multilingual geographical gazetteers from corpora. *Proceedings of the 11th ESSLLI Student Session, Malaga, Spain, 31 July - 11 August 2006*.

<http://ilk.uvt.nl/mitch/papers/esslli06.pdf>,

<http://staff.science.uva.nl/~katrenko/stus06/images/erp.pdf>

Eschenbach et al. [Eschenbach, Carola; Tschander, Ladina; Habel, Christopher; Kulik, Lars]

2000 Lexical specifications of paths. In: Freksa, Christian; Habel, Christopher; Wender, Karl Friedrich (eds): *Spatial Cognition II - Integrating Abstract Theories, Empirical Studies, Formal Methods, and Practical Applications* 127–144. Berlin: Springer.

http://www.cs.mu.oz.au/~lars/Publications/SC2_ETHK.pdf

EURCOM:LAU

European Commission: Local Administrative Units (LAU).

http://ec.europa.eu/comm/eurostat/ramon/nuts/lau_en.html

Fabricius-Hansen, Cathrine

1975 *Transformative, intransformative und kursive Verben*. Linguistische Arbeiten 26. Heidelberg: Niemeyer.

Fairon, Cédric; Paumier, Sébastien; Watrin, Patrick

2005 Can we parse without tagging? In: Vetulani, Zygmunt (ed.): *Human language technologies as a challenge for computer science and linguistics. Proceedings of the 2nd Language & Technology Conference: Poznań, Poland, April 21-23, 2005*.

<http://cedux.fltr.ucl.ac.be/~pat/cv/download/LTC05.pdf>

Fairon, Cédric; Senellart, Jean

1999 Réflexions sur la localisation, l'étiquetage, la reconnaissance et la traduction d'expressions linguistiques complexes. *TALN'99. Cargèse, Juillet 1999*.

<http://talana.linguist.jussieu.fr/taln99/ps/A26/A26.pdf>,

http://www.atala.org/doc/actes_taln/AC_0007.pdf

Fearn, John

1824–1827 *Anti-Tooke, or, an analysis of the principles and structure of language, exemplified in the English tongue*. London: Valby.

Fellbaum, Christiane (ed.)

1998 *WordNet: an electronic lexical database*. Language, speech, and communication. Cambridge (Massachusetts): MIT Press.

Ferrés et al. [Ferrés, Daniel; Massot, Marc; Padró, Muntsa; Rodríguez, Horacio; Turmo, Jordi]

2004 Automatic Classification of Geographical Named Entities. *LREC 2004, Fourth International Conference on Language Resources and Evaluation, Lisbon - Portugal, 26 May - 28 May 2004*. <http://gandalf.aksis.uib.no/non/lrec2004/pdf/237.pdf>

- Ferrés Domènech, Daniel
- 2007 Geographical information resolution and its application to the question answering systems. Memòria del DEA i Projecte de Tesi, Programa de Doctorat en Intel·ligència Artificial, Departament de Llenguatges i Sistemes Informàtics, Universitat Politècnica de Catalunya. <http://www.lsi.upc.edu/~dferres/dea-dferres.pdf>
- Fisher, P. F.; Wood, J.
- 1998 What is a mountain? or the Englishman who went up a Boolean geographical concept and realised it was fuzzy. *Geography* 83/3, 247–256.
- Fitschen, Arne
- 2004 Ein Computerlinguistisches Lexikon als komplexes System. Dissertation, Institut für maschinelle Sprachverarbeitung, Universität Stuttgart. <http://www.ims.uni-stuttgart.de/projekte/IMSLex/fitschendiss.pdf>
- Fleischer, Wolfgang
- 1964 Zum Verhältnis von Name und Appellativum im Deutschen. *Wissenschaftliche Zeitschrift der Karl-Marx-Universität Leipzig, Gesellschafts- und sprachwissenschaftliche Reihe* 13, 369–378.
- 1967 Zur Funktion des Artikels in der deutschen Sprache der Gegenwart. *Acta Universitatis Wratislaviensis* 60 (Germanica Wratislaviensia 11), 131–160. Wrocław.
- Fleischer, Wolfgang; Barz, Irmhild
- 1992 *Wortbildung der deutschen Gegenwartssprache*. Tübingen: Niemeyer.
- Florian et al. [Florian, Radu; Ittycheriah, Abe; Jing, Hongyan; Zhang, Tong]
- 2003 Named entity recognition through classifier combination. *Proceedings of CoNLL-2003, Edmonton, Canada, 2003* 168–171. <http://www.cnts.ua.ac.be/conll2003/pdf/16871flo.pdf>
- Fonseca, Frederico Torres
- 2001 Ontology-driven geographic information systems. PhD thesis. http://www.spatial.maine.edu/Publications/phd_thesis/Fonseca2001.pdf
- Fonseca, Frederico Torres; Davis, Clodeveu; Câmara, Gilberto
- 2003 Bridging ontologies and conceptual schemas in geographic information integration. *GeoInformatica* 7/4, 355–378. http://www.dpi.inpe.br/gilberto/papers/ontologies_models.pdf
- Fonseca et al. [Fonseca, Frederico Torres; Egenhofer, Max J.; Agouris, Peggy; Câmara, Gilberto]
- 2002 Using ontologies for integrated geographic information systems. *Transactions in GIS* 6/3, 231–257. http://lbdwww.epfl.ch/f/teaching/courses/SlidesST/UsingOntologies_GIS.pdf, <http://citeseer.ist.psu.edu/462719.html>
- Forsgren, Mats; Jonasson, Kerstin; Kronning, Hans (eds.)
- 1998 *Prédication, assertion, information*. Actes du colloque d’Uppsala en linguistique française, 6 - 9 juin 1996. *Studia Romanica Upsaliensis* 56. Uppsala: Acta Universitatis Upsaliensis.
- Francis, Wintrop Nelson; Kučera, Henry
- ²1970 *Computational analysis of present-day American English*. Providence (Rhode Island): Brown University Press.
- 1982 *Frequency analysis of English usage. lexicon and grammar*. Boston (Massachusetts): Houghton Mifflin.
- Freksa, Christian; Mark, David M. (eds.)
- 1999 *Spatial information theory – cognitive and computational foundations of Geographic*. Berlin: Springer.

- Frense, Jutta; Bennet, Paul
 1996 Verb alternations and semantic classes in English and German. *Language Sciences* 18, 305–317.
- Friburger, Nathalie
 2002 Reconnaissance automatique des noms propres: Application à la classification automatique de textes journalistiques. These pour obtenir le grade de docteur de l'Université de Tours. Tours. http://tln.li.univ-tours.fr/Tln_Biblio/2002theseFriburger.zip
 2006 Linguistique et reconnaissance automatique des noms propres. *Meta: Journal des traducteurs* 51/4, 637–650.
- Friburger, Nathalie; Maurel, Denis
 2001a Elaboration d'une cascade de transducteurs pour l'extraction de motifs: l'exemple des noms de personnes. *TALN 2001, Tours, 2-5 juillet 2001*. http://www.li.univ-tours.fr/taln-recital-2001/Actes/tome1_PDF/partie2_p30_322/art16_p183_192.pdf
 2001b Finite state transducer cascade to extract proper nouns in French text. *Lecture notes in computer science*. Second Conference on Implementing and Application of Automata (CIAA'2001), Pretoria, South Africa, 23-25 July 2008, 115–124. Berlin: Berlin. http://tln.li.univ-tours.fr/Tln_Biblio/2001ciaa.zip
 2002 Textual similarity based on proper names. *Mathematical Formal Information Retrieval (MFIR'2002) at the 25th ACM SIGIR Conference (Tampere, Finland, 2002)* 155–167. http://cir.dcs.vein.hu/cikkek/friburger_MFIR02.pdf, <http://citeseer.ist.psu.edu/536548.html>
 2004 Finite-state transducer cascades to extract named entities in texts. *Theoretical Computer Science* 313/1, 94–104. Elsevier.
- Fuchshuber-Weiß, Elisabeth
 1995 Straßennamen: deutsch. In: Eichler et al. (eds.): *Namenforschung: ein internationales Handbuch zur Onomastik* 1468–1475.
- Fu, Gaihua; Abdelmoty, Alia I.; Jones, Christopher B.
 2003 Design of a geographical ontology. SPIRIT Public Deliverables (Reports) D5 3101. http://www.geo-spirit.org/publications/SPIRIT_WP3_D5.pdf
- Fu, Gaihua; Jones, Christopher B.; Abdelmoty, Alia I.
 2005a Building a geographical ontology for intelligent spatial search on the web. *Proceedings of IASTED International Conference on Databases and Applications (DBA2005)*. http://www.geo-spirit.org/publications/geoontology_DBA.pdf
 2005b Ontology-based spatial query expansion in information retrieval. *International Conference on Ontologies, Databases and Applications of SEMantics (ODBASE 2005), 31 Oct - 4 Nov 2005, Agia Napa, Cyprus*. <http://www.geo-spirit.org/publications/gaihua-odbase2005.pdf>, <http://citeseer.ist.psu.edu/fu05ontologybased.html>
- Fuhrhop, Nanna
 1998 *Grenzfälle morphologischer Einheiten*. Studien zur deutschen Grammatik 57. Tübingen: Stauffenburg. [Berlin, Freie Universität, Dissertation, 1997]
 2001 ‚Berliner‘ Ballen und ‚Potsdamer‘ Bürgermeister – Substantive oder Adjektive? In: Lang, Ewald; Rochon, Marzena; Schwabe, Kerstin; Teuber, Oliver (eds.): *Papers for Ewald Lang*. *ZAS Papers in Linguistics* 21, 45–58.
 2003 »Berliner« Luft und »Potsdamer« Bürgermeister: Zur Grammatik der Stadtadjektive. *Linguistische Berichte* 193, 91–108.
- Furlan, Aidan; Baldwin, Timothy; Klippel, Alex
 2007 Landmark classification for route directions. *Proceedings of the Fourth ACL-SIGSEM Workshop on Prepositions, Prague, Czech Republic, pp. 9–16*. <http://acl.ldc.upenn.edu/W/W07/W07-1602.pdf>

- Gale, William A; Church, Kenneth W.; Yarowsky, David
 1992 One sense per discourse. *Proceedings DARPA speech and Natural Language Workshop, Harriman, NY*. <http://acl.ldc.upenn.edu/H/H92/H92-1045.pdf>
- Gallmann, Peter
 1997 Zur Morphosyntax der Eigennamen im Deutschen. In: Löbel, Elisabeth; Rauh, Gisa (eds.): *Lexikalische Kategorien und Merkmale*. Linguistische Arbeiten 366, 72–84. Tübingen: Niemeyer.
http://www.personal.uni-jena.de/~x1gape/Pub/Morphosyntax_EN_1997.pdf
- Gambarotto, Pierre; Muller, Philippe
 2003 Vectors across spatial domains: From place to size, orientation, shape, and parts. In: van der Zee & Slack (eds.): *Representing direction in language and space* 144–165.
- Garbin, Eric; Mani, Inderjeet
 2005 Disambiguating toponyms in news. *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP), Vancouver, October 2005* 363–370. <http://complingone.georgetown.edu/~linguist/papers/hlt05.pdf>, <http://acl.ldc.upenn.edu/H/H05/H05-1046.pdf>
- Garrigues, Mylène
 1993 Prépositions et noms de pays et d'îles: une grammaire locale pour l'analyse automatique des textes. *Linguisticæ Investigationes* 17/2, 281–306. Amsterdam, Philadelphia: John Benjamins.
 1995 Prepositions and the names of countries and islands: a local grammar for the automatic analysis of texts. *Language Research* 31/2, 309–334. Seoul: Language Research Institute, Seoul National University.
- GATE
 GATE Information Extraction. <http://gate.ac.uk/ie/>
- GENTILÉS DU QUÉBÉC
 Commission de toponymie Québec: Liste des gentils (noms des habitants) du Québec.
<http://www.toponymie.gouv.qc.ca/lesgentilesliste.asp>
- GEOBASE.CA
 GeoBase. <http://www.geobase.ca/>
- GEOCLEF
 Evaluation of cross-language Geographic Information Retrieval (GIR) systems.
<http://ir.shef.ac.uk/geoclef/>
- GEOCLEF07QUERY
 Query Parsing Task Proposal for GeoCLEF 2007.
<http://ir.shef.ac.uk/geoclef/2007/Query-Parsing.htm>
- GEOCODER.US
 geocoder.us – find the latitude & longitude of any US address – for free.
<http://geocoder.us/>
- GEODATENZENTRUM
 Bundesamt für Kartographie und Geodäsie: Geodatenzentrum.
<http://www.geodatenzentrum.de/>
- GEOFEATURES.OWL
 geoFeatures.owl. <http://www.mindswap.org/2003/owl/geo/geoFeatures.owl>
- GEOMICROFORMAT
 geo – Microformats. <http://microformats.org/wiki/geo>

GEONAMES.ORG

GeoNames. <http://www.geonames.org/>

GEO-tags

GeoTags Location-Based Search Engine. <http://geotags.com/>

GEOURL

The GeoURL ICBM Address Server. <http://geourl.org/>

Gerdes, Kim; Kahane, Sylvain

- 2001 Word order in German: a formal dependency grammar using a topological hierarchy. *Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics*. <http://acl1.ldc.upenn.edu/P/P01/P01-1029.pdf>
- 2006 A polynomial parsing algorithm for the topological model. *Actes de la conférence ACL 2006, Sydney*. Synchronizing constituent and dependency grammars, illustrated by German word order phenomena. <http://gerdes.fr/papiers/2006/ac12006.pdf>
- 2007 Phrasing it differently. In: Wanner, Leo (ed.): *Selected lexical and grammatical issues in the Meaning-Text Theory*. WannerEd2007Selected 297–335.

Gerling, Martin; Orthen, Norbert

- 1979 *Deutsche Zustands- und Bewegungsverben*. Eine Untersuchung zu ihrer semantischen Struktur und Valenz. Studien zur deutschen Grammatik 11. Tübingen: Gunter Narr.

GERMANET

GermaNet. <http://www.sfs.uni-tuebingen.de/lsd/>

GERTWOL

Lingsoft Language Solutions: GERTWOL: Morphologisches Analysesystem für das Deutsche. <http://www2.lingsoft.fi/cgi-bin/gertwol>

GETTY

Getty Thesaurus of Geographic Names Online.
http://www.getty.edu/research/conducting_research/vocabularies/tgn/

Gey, Frederic C.

- 2001 Research to improve cross-language retrieval – position paper for CLEF. In: Peters, C. (ed.): *Cross-language information retrieval and evaluation. Workshop of Cross-Language Evaluation Forum, CLEF 2000, Lisbon, Portugal, September 21-22, 2000, Revised Papers*. Lecture notes in computer science 2069, 83–88. Berlin: Springer.
<http://ucdata.berkeley.edu:7101/personal/fred/my-papers/position.pdf>

Gey, Fredric C.; Carl, Kim

- 2004 Geotemporal access to multilingual documents. *SIGIR 2004*.
<http://www.geo.unizh.ch/~rsp/gir/abstracts/gey.pdf>

Gey et al. [Gey, Fredric; Larson, Ray; Sanderson, Mark; Bischoff, Kerstin; Mandl, Thomas; Womser-Hacker, Christa; Santos, Diana; Rocha, Paulo]

- 2006 GeoCLEF 2006: the CLEF 2006 cross-language geographic information retrieval track overview. *GeoCLEF 2006*.
http://dis.shef.ac.uk/mark/cv/publications/papers/my_papers/geoCLEF2006.pdf

Gey et al. [Gey, Fredric; Larson, Ray; Sanderson, Mark; Bischoff, Kerstin; Mandl, Thomas; Womser-Hacker, Christa; Santos, Diana; Rocha, Paulo; Montoyo, Andres]

- 2007 Challenges to evaluation of multilingual geographic information retrieval in GeoCLEF. *The First International Workshop on Evaluating Information Access (EVIA), May 15, 2007, Tokyo, Japan*.
<http://research.nii.ac.jp/ntcir/workshop/OnlineProceedings6/EVIA/16.pdf>

- Gey et al. [Gey, Fredric; Larson, Ray; Sanderson, Mark; Joho, Hideo; Clough, Paul]
 2005 GeoCLEF: the CLEF 2005 Cross-Language Geographic Information Retrieval Track. *Working Notes for the CLEF 2005 Workshop, 21-23 September, Vienna, Austria.*
http://www.clef-campaign.org/2005/working_notes/workingnotes2005/sanderson05.pdf
- Geyken, Alexander
 2001 *Aufbau einer kontrastiven französisch-deutschen Lexikongrammatik am Beispiel der Kommunikationsverben.* Studien zur Informations- und Sprachverarbeitung 2. München: Centrum für Informations- und Sprachverarbeitung.
- Glättli, Hugo
 1966 Observations sur l'emploi des prépositions devant les noms de pays. *Mélanges de grammaire française* 131–141. Gembloux: Duculot.
- GML
 Open GIS Consortium 2004: GML – the Geography Markup Language.
<http://www.opengis.net/gml/>
- GOOGLE:LOCAL
 Google Local. <http://local.google.com/>
- Gorodeckaja, Inna Leonidovna; Levašov, Evgenij Aleksandrovič
 2003 *Russkie nazvanija žitelej.* Slovar'-spravočnik; bolee 14000 nazvanij žitelej Rossii, respublik byvšego SSSR i zarubežnyh stran. Moskva: Russkie Slovare.
- Görz, Günther; Rollinger, Claus-Rainer; Schneeberger, Josef (eds.)
 42003 [12000] *Handbuch der künstlichen Intelligenz.* München: Oldenbourg.
- GRAALWEB
 GraalWeb: an on-line library of local grammars.
http://igm.univ-mlv.fr/~mconstan/library/index_graalweb.html
- GRAMMIS
 Grammis – das grammatische Informationssystem des IDS.
<http://hypermedia.ids-mannheim.de/index.html>
- Grass, Thierry
 2000 Typologie et traductibilité des noms propres de l'allemand vers le français. In: Maurel, Denis; Guenther, Franz (ed.): *Traitement automatique des noms propres.* Traitement automatique des langues 41/3, 643–669. Paris: ATALA/Hermes Sciences.
 2006 La traduction comme appropriation : le cas des toponymes étrangers. *Meta: Journal des traducteurs* 51/4, 660–670.
- Gravano, Luis; Hatzivassiloglou, Vasileios; Lichtenstein, Richard
 2003 Categorizing web queries according to geographical locality. *Proceedings of the 12th ACM Conference on Information and Knowledge Management (CIKM 2003).*
<http://www1.cs.columbia.edu/~gravano/Papers/2003/cikm03.pdf>
- Grefenstette, Gregory
 1996 Light parsing as finite state filtering. In: Kornai (ed.): *Extended finite state models of language* 86–94. <http://www.ling.gu.se/~lager/kurser/FST/papers/mltt-96-12.ps>
- Greule, Albrecht
 1990 Ortsnamenwörterbücher. In: Hausmann et al. (eds.): *Wörterbücher* 1276–1284.
- Grimaud, Michel
 1988 Toponyms, prepositions, and cognitive maps in English and French. *Geolinguistics.* Journal of the American Society of Geolinguistics 14, 54–76.

- Grimm, Hannelore; Schöler, Hermann
 1976 Semantic organization of German spatial prepositions. *International Review of Applied Linguistics in Language Teaching* 14/2, 165–83.
- Grishman, Ralph
 2000 Entity Annotation Guidelines. Entity Detection and Tracking - Phase 1. ACE Pilot Study Task Definition. 3–7. ftp://jaguar.ncsl.nist.gov/ace/phase1/edt_phase1_v2.2.pdf
- Grishman, Ralph; Sundheim, Beth M.
 1996 Message Understanding Conference - 6: A Brief History. *Proceedings of the 16th International Conference on Computational Linguistics (COLING), I, Copenhagen, 1996* 466–471. <http://acl.ldc.upenn.edu/C/C96/C96-1079.pdf>
- Gross, Gaston
 1981 Les prépositions composées. In: Schwarze, Christoph (ed.): *Analyse des prépositions*. Linguistische Arbeiten. III^e Colloque franco-allemand de linguistique théorique du 2 au 4 février 1981 à Constance 110, 29–39. Tübingen: Niemeyer.
 1989 *Les constructions converses du français*. Langue et cultures 22. Genève: Droz.
 1994 Classes d'objets et description des verbes. In: Giry-Schneider, Jacqueline (ed.): *Sélection et sémantique: classes d'objets, compléments appropriés, compléments analysables*. Langages 115, 15–30. Paris: Larousse.
 1996 Une typologie sémantique des connecteurs : l'exemple de la cause. *Studi italiani di linguistica teoretica e applicata* 25/1, 153–179.
 1998a Eliminating semantic ambiguity by means of a lexicon-grammar. In: Bresson, Daniel; Kubczak, Jacqueline (eds.): *Abstrakte Nomina. Vorarbeiten zu ihrer Erfassung in einem zweisprachigen syntagmatischen Wörterbuch*. Studien zur deutschen Sprache 10, 253–272. Tübingen: Gunter Narr.
 1998b Pour une typologie des prédicats nominaux. In: Forsgren, Jonasson & Kronning (eds.): *Prédication, assertion, information* 221–230.
- Gross, Gaston; Clas, André
 1997 Synonymie, polysémie et classes d'objets. *Meta: Journal des traducteurs* 42/1, 147–155.
- Gross, Maurice
 1975 *Méthodes en syntaxe*. Régime des constructions complétives. Actualités scientifiques et industrielles 1365. Paris: Hermann.
 1979 On the failure of generative grammar. *Language* 55/4, 859–885.
 1984 Lexicon-grammar and the syntactic analysis of French. *10th International Conference on Computational Linguistics and 22nd Annual Meeting of the Association for Computational Linguistics* 275–282. <http://acl.ldc.upenn.edu/P/P84/P84-1058.pdf>
 1986 Lexicon-grammar: The representation of compound words. *COLING-1986, Proceedings, Bonn* 1–6. <http://acl.ldc.upenn.edu/C/C86/C86-1001.pdf>
 1989 The use of finite automata in the lexical representation of natural language. In: Gross, Maurice; Perrin, Dominique (eds.): *Electronic dictionaries and automata in computational linguistics*. Proceedings of the LITP Spring School on Theoretical Computer Science, Saint-Pierre d'Oléron, France, May 25–29, 1987. Lecture notes in computer science 377, 34–50. Berlin: Springer.
 1993 Local grammars and their representation by finite automata. In: Hoey, Michael (ed.): *Data, description, discourse: Papers on the English language in honour of John McH. Sinclair* 26–38. London: Harper Collins.
 1994 Constructing lexicon-grammars. In: Atkins, B. T. S.; Zampolli, A. (eds.): *Computational Approaches to the Lexicon* 213–263. Oxford: Oxford University Press.
 1995a La notion de lieu argument du verbe. In: Shyldkrot, Hava Bat-Zeev; Kupferman, Lucien (eds.): *Tendances récentes en linguistique française et générale, volume dédié à David*

- Gaalone*. *Lingvisticæ Investigationes Supplementa* 20, 173–200. Amsterdam, Philadelphia: John Benjamins.
- 1995b Representation of finite utterances and the automatic parsing of texts. *Language Research* 31/2, 291–307. Seoul: Language Research Institute, Seoul National University.
- 1996 Lexicon grammar. In: Brown & Miller (eds.): *Concise encyclopedia of syntactic theories* 244–258. [= 1994a]
- 1997 The construction of local grammars. In: Roche; Schabes (eds.): *Finite-state language processing* 329–354.
- 1999 A bootstrap method for constructing local grammars. *Contemporary Mathematics: Proceedings of the Symposium, University of Belgrad* 229–250. Belgrad.
- 2002a Les ambiguïtés. *Lingvisticæ investigationes* 24/1, 3–41. Amsterdam, Philadelphia: John Benjamins.
- 2002b Consequences of the metalanguage being included in the language. In: Nevin (ed.): *The legacy of Zellig Harris* 57–67.
- 2006 État du lexique-grammaire du français et perspectives d’extension. In: Auroux et al. (eds.): *History of the Language Sciences* 2122–2129.
- Gross, Maurice; Schützenberger, Marcel-Paul
- 1971 Review of Z. Harris’ *Mathematical Structures of Language*. *The American Scientist* 58.
- 1972 Review of Z. Harris’ *Mathematical Structures of Language*. In: Plötz (ed.): *Transformationnelle Analyse* 307–312. [= Gross & Schützenberger (1971)]
- Grote, Brigitte
- 2001 A German prepositional phrase grammar: paradigmatic choices and syntagmatic structure. *Functions of Language* 7/2, 231–72. Amsterdam: John Benjamins.
- Guillet, Alain
- 1984 Prépositions de lieu et verbes supports. *Revue québécoise de linguistique* 13/2, 59–93. Montréal: Université du Québec à Montréal.
- Guillet, Alain; Leclère, Christian
- 1981 Restructuration du groupe nominal. In: Guillet & Leclère (eds.): *Formes syntaxiques et prédicats sémantiques* 99–125.
- 1992 *La structure des phrases simples en français*. Les constructions transitives locatives. *Langue et cultures* 26. Genève, Paris: Droz.
- Guillet, Alain; Leclère, Christian (eds.)
- 1981 *Formes syntaxiques et prédicats sémantiques*. *Langages* 63. Paris: Larousse.
- Gutermann, Barbara
- 1996 *Die Raumdarstellung in deutschen und französischen Texten*. Sprachvergleichende Untersuchungen zur Frequenz der Darstellung räumlicher Relationen, zu ihrer sprachlichen Form und zu Inhalt-Form-Beziehungen. Europäische Hochschulschriften 160. Frankfurt am Main: Peter Lang. [Dissertation Universität Stuttgart 1995]
- Habel, Christopher
- 1989a *Zwischen*-Bericht. In: Habel; Herweg; Rehkamper (eds.): *Raumkonzepte in Verstehensprozessen* 37–69.
- 1989b Propositional and depictorial representations of spatial knowledge: The case of path-concepts. Report FBI-HH-M-171/89, University of Hamburg.
- 1990 Propositional and depictorial representations of spatial knowledge: The case of path-concepts. In: Studer, Rudi (ed.): *Proceedings of the International Symposium on Natural Language and Logic, Hamburg, Germany, May 9-11, 1989*. Lecture Notes in Artificial Intelligence 459, 94–117. Springer.
- 1991 Processing of spatial expressions in LILOG. In: Herzog & Rollinger (eds.): *Text understanding in LILOG* 598–608.

- Habel et al. [Habel, Christopher; Herweg, Michael; Pribbenow, Simone; Schlieder, Christoph]
 2003 Wissen über Raum und Zeit. In: Görz, Rollinger & Schneeberger (eds.): *Handbuch der Künstlichen Intelligenz* 349–405.
- Haider, Hubert
 2002 Adverbials at the syntax-semantics interface. In: Kamp, Hans; Reyle, Uwe (eds.): *How we say when it happens* 53–70.
- Harris, Zellig Sabbettai
 1962 *String analysis of sentence structure*. Papers on Formal Linguistics 1. The Hague: Mouton.
 1964 *The elementary transformations*. Transformations and Discourse Analysis Papers 54. Philadelphia: University of Pennsylvania. [Excerpted in Harris 1970: 482-532, Plötz ed. (1972): 57-75, and, in abbreviated form, in Harris 1981: 211-235]
 1965 Transformational Theory. *Language* 41/3, 363–401.
 1968 *Mathematical structures of language*. Interscience tracts in pure and applied mathematics 21. New York: Interscience Publishers.
 1972 Transformational theory. In: Plötz (ed.): *Transformationnelle Analyse* 103–154. [= Harris (1965)]
 1976 *Notes du cours de syntaxe*. Paris: Le Seuil.
 1988 *Language and information*. Bampton lectures in America 28. New York: Columbia University Press.
 1991 *A theory of language and information. A mathematical approach*. Oxford: Oxford University Press.
 2002 The background of transformational and metalanguage analysis. In: Nevin (ed.): *The legacy of Zellig Harris* 1–15.
- Hauff, Claudia; Trieschnigg, Dolf; Rode, Henning
 2006 University of Twente at GeoCLEF 2006: geofiltered document retrieval. *GeoCLEF 2006*. <http://wwwhome.cs.utwente.nl/~hauffc/CLEF2006.pdf>
- Hauptmann, Alexander G.; Olligschlaeger, Andreas M.
 1999 Using location information from speech recognition of television news broadcasts. In: Robinson, Tony; Renals, Steve (eds.): *Proceedings of the ESCA ETRW Workshop on Accessing Information in Spoken Audio* 102–106. Cambridge (England): University of Cambridge. <http://citeseer.ist.psu.edu/61323.html>, <http://svr-www.eng.cam.ac.uk/~ajr/esca99/Hauptmann.ps>
- Hausmann et al. eds. [Hausmann, Franz Josef; Reichmann, Oskar; Wiegand, Herbert E.; Zgusta, Ladislav (eds.)]
 1989–1991 *Wörterbücher – Dictionaries*. Ein internationales Handbuch zur Lexikographie. Handbücher zur Sprach- und Kommunikationswissenschaft 5. Berlin: de Gruyter.
- Hays, Ellen M.
 1987 A computational treatment of locative relations in natural language. Report MS-CIS-87-31, LINC LAB 58, University of Pennsylvania. http://repository.upenn.edu/cis_reports/733/
- Hearst et al. [Hearst, Marti; Elliott, Ame; English, Jennifer; Sinha, Rashmi; Swearingen, Kirsten; Yee, Ka-Ping]
 2002 Finding the flow in web site search. *Communications of the ACM* 45/9, 42–49. <http://flamenco.berkeley.edu/papers/cacm02.pdf>

- Heintzeler, Mirjam
 1992 *Raumausdrücke im Konzeptlexikon*. Die Darstellung der Komposition lokaler Verben und Präpositionen in einem konzeptuellen Lexikon. Konstanzer Dissertationen 362. Konstanz: Hartung-Gorre.
- Heinzle, Frauke; Kopczynski, Matthias; Sester, Monika
 2003 Spatial data interpretation for the intelligent access to spatial information in the internet. *Proceedings of 21st International Cartographic Conference, August 10-16, 2003, Durban, South Africa*.
http://www.geo-spirit.org/publications/Spatial_Data_Interpretion_Hannover.pdf
- Helbig, Gerhard; Buscha, Joachim
¹⁴1991 *Deutsche Grammatik*. Ein Handbuch für den Ausländerunterricht. Berlin, München: Langenscheidt / Enzyklopädie.
- Helbig, Gerhard; Schenkel, Wolfgang
³1975 [1969] *Wörterbuch zur Valenz und Distribution deutscher Verben*. Leipzig: Bibliographisches Institut.
- Hellfritzsich, Volkmar
 1995 Satznamen. In: Eichler et al. (eds.): *Namenforschung: ein internationales Handbuch zur Onomastik* 435–439.
- Herskovits, Annette
 1985 Semantics and pragmatics of locative expressions. *Cognitive science* 9, 341–378.
- Herweg, Michael
 1988 *Zur Semantik einiger lokaler Präpositionen des Deutschen*. Überlegungen zur Theorie der lexikalischen Semantik am Beispiel von *in*, *an*, *bei* und *auf*. LILOG-Report 21. Stuttgart: IBM.
 1989 Ansätze zu einer semantischen Beschreibung topologischer Präpositionen. In: Habel; Herweg; Rehkamper (eds.): *Raumkonzepte in Verstehensprozessen* 99–127.
- Herzog, Otthein; Rollinger, Claus-Reiner (eds.)
 1991 *Text understanding in LILOG*. Integrating computational linguistics and artificial intelligence. Lecture notes in computer science 546. Berlin: Springer.
- Hill, Linda L.
 2000 Core elements of digital gazetteers: placenames, categories, and footprints. In: Borbinha, J.; Baker, T. (eds.): *Research and Advanced Technology for Digital Libraries. Proceedings of the 4th European Conference, ECDL 2000 Lisbon, Portugal, September 18-20, 2000* 280–290. Berlin: Springer.
http://www.alexandria.ucsb.edu/%7Elhill/paper_drafts/ECDL2000_paperdraft7.pdf
- Hill, Linda L.; Goodchild, Michael F.; Janée, Greg
 2004 Research directions in georeferenced IR based on the Alexandria Digital Library Project. *SIGIR 2004*. <http://www.geo.unizh.ch/~rsp/gir/abstracts/hill.pdf>
- Hinrichs, Erhard W.
 2005 Finite-state parsing of German. In: Antti Arppe (eds.): *Inquiries into words, constraints, and contexts* 35–44.
<http://csli-publications.stanford.edu/koskenniemi-festschrift/4-hinrichs.pdf>
- Hinrichs, Erhard W.; Trushkina, Julia S.
 2002 Getting a grip on morphological disambiguation. *Tagungsband der 6. Konferenz zur Verarbeitung natürlicher Sprache (KONVENS 2002), Saarbrücken* 59–66.
<http://www.sfb441.uni-tuebingen.de/a1/Publikationen/grip.pdf>

- 2004 Forging Agreement: morphological disambiguation of noun phrases. *Research on Language & Computation* 2/4, 621–648. Springer. <http://www.bultreebank.org/proceedings/paper06.pdf>, <http://www.sfb441.uni-tuebingen.de/a1/Publikationen/forge.pdf>

HISTORTSBUCH

Rademacher, Michael: Deutsch-österreichisches Ortsbuch 1871-1945.
<http://www.literad.de/geschichte/ortsbuch39.html>

Hobbs, Jerry R.

- 2006 Interoperability among geospatial ontologies.
<http://colab.cim3.net/file/work/SICoP/2006-06-20/JHobbs06202006.ppt>

Hockett, Charles F.

- 1963 The problem of universals in language. In: Greenberg, Joseph H. (ed.): *Universals of language* 1–29. Cambridge (Massachusetts), London.

Höhle, Tilman

- 1986 Der Begriff ‚Mittelfeld‘: Anmerkungen über die Theorie der topologischen Felder. In: Weiss, Walter; Wiegand, Herbert Ernst; Reis, Marga (eds.): *Textlinguistik contra Stilistik? Kontroversen, alte und neue: Akten des VII. Internationalen Germanisten-Kongresses, Göttingen 1985* 3, 329–340. Tübingen: Niemeyer.

Hottenroth, Priska-Monika

- 1986 Die Semantik lokaler Präpositionen. Ein prototypen-semantisches Modell für die französische Präposition *dans* mit einer Analyse der Beziehungen zwischen der Präposition und den Objektbezeichnungen in den Präpositionalsyntagmen. Habilitationsschrift Konstanz.

Hunston, Susan; Francis, Gill

- 2000 *Pattern grammar*. A corpus-driven approach to the lexical grammar of English. Studies in Corpus Linguistics 4. Amsterdam, Philadelphia: John Benjamins.

INTEX

INTEX: an Linguistic Development Environment. <http://intex.univ-fcomte.fr/>

Irie, Robert; Sundheim, Beth M.

- 2004 Resources for place name analysis. *LREC 2004, Fourth International Conference on Language Resources and Evaluation, Lisbon - Portugal, 26 May - 28 May 2004*.
<http://gandalf.aksis.uib.no/non/lrec2004/pdf/785.pdf>

Jackendoff, Ray

- 1990 *Semantic structures*. Cambridge (Massachusetts): MIT Press.

Jacquemin, Christian; Bush, Caroline

- 2000 Lexical and formatting clues for named entity acquisition from the web. *Proceedings of Joint SIGDAT Conference on Empirical Methods in NLP and Very Large Corpora. Hongkong* 181–189. <http://acl.ldc.upenn.edu/W/W00/W00-1323.pdf>

Janssen, Maarten

- 2002 SIMuLLDA: a Multilingual Lexical Database Application using a Structured Interlingua. Proefschrift Universiteit Utrecht.
<http://igitur-archive.library.uu.nl/dissertations/2002-0905-111545/inhoud.htm>

Johnson, Stephen

- 2002 Foreword. In: Nevin & Johnson (eds.): *The legacy of Zellig Harris* ix–xxviii.
<http://www.dmi.columbia.edu/zellig/Foreword2.html>

- Jones, Christopher B.; Abdelmoty, Alia I.; Fu, G.
 2003 Maintaining ontologies for geographical information retrieval on the web. *Proceedings of OTM Confederated International Conferences CoopIS, DOA, and OOBASE* 934–951.
http://www.geo-spirit.org/publications/SPIRIT_maintaining_ontologies.pdf
- Jones et al. [Jones, Christopher B.; Purves, R.; Ruas, A.; Sanderson, Mark; Sester, Monika; Kreveld, Mark van; Weibel, R.]
 2002 Spatial information retrieval and geographical ontologies: an overview of the SPIRIT project. *SIGIR 2002: Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval August 11-15, 2002, Tampere, Finland* 387–388. ACM Press.
http://www.geo-spirit.org/publications/SPIRIT_project_overview.pdf
- Jørgensen, Fredrik
 2004 *The semantic representation of locatives in machine translation*. Master thesis, University of Oslo. http://folk.uio.no/fredrijo/publications/pdf/jorgensen_2004.pdf
- Jørgensen, Fredrik; Lønning, Jan Tore
 2005 A minimal recursion semantic analysis of locatives. *Proceedings of the Second ACL-SIGSEM Workshop on The Linguistic Dimensions of Prepositions and their Use in Computational Linguistics Formalisms and Applications. Colchester, UK, April 2005. University of Essex* 64–73.
http://folk.uio.no/fredrijo/publications/pdf/jorgensen_and_lonning_2005.pdf
- Joshi, Aravind K.
 1996 A parser from antiquity: An early application of finite state transducers to natural language parsing. *Workshop on Extended finite state models of language, August 11-12, 1996 ECAI '96, Budapest, Hungary*.
<http://www.metacarta.com/kornai/ECAI/joshi.ps.gz>
- Joshi, Aravind K.; Hopely, Philip
 1999 A parser from antiquity: an early application of finite state transducers to natural language parsing. In: Kornai (ed.): *Extended finite state models of language* 6–15.
- Kadmon, Naftali
 2001 *Toponymy*. The lore, laws and language of geographical names.
 2006a Language and the conversion of geographical names. In: UNGEGN: *Manual for the national standardization of geographical names* 99–106. [aus Kadmon 2001: Toponymy]
 2006b Exonyms, also called conventional names. In: UNGEGN: *Manual for the national standardization of geographical names* 129–131. [aus Kadmon 2001: Toponymy]
- Kahane, Sylvain
 2003 The meaning-text theory. In: Ágel et al. (eds.): *Dependenz und Valenz* 546–570.
- Kalverkämper, Hartwig (ed.)
 1978 *Textlinguistik der Eigennamen*. Stuttgart: Klett-Cotta.
- Kamianets, Wolodymyr
 2000 Zur Einteilung der deutschen Eigennamen. *Grazer linguistische Studien* 54, 41–58.
<http://www-gewi.kfunigraz.ac.at/ling/cgi-bin/data/gls/artikel/54/05-kamianets.pdf>
- Karlsson, Fred
 2005 Constraints on clausal embedding complexity in Standard Average European (SAE).
<http://humanities.uchicago.edu/depts/linguistics/abstract/karlsson.html>
 2006 Recursion in Natural Languages. In: Salakoski, Tapio; Ginter, Filip; Pyysalo, Sampo; Pihkala, Tapio (eds.): *Advances in Natural Language Processing. 5th international conference on NLP, FinTAL 2006 Turku, Finland, August 23-25, 2006*. Lecture Notes in Artificial Intelligence 4139. Berlin, Heidelberg: Springer.

- 2007 Constraints on multiple initial embedding of clauses. *International Journal of Corpus Linguistics* 12/1, 107–118.
http://www.ling.helsinki.fi/~fkarlsson/Initial_Embedding.pdf
- 2008 Early generative linguistics and empirical methodology. In: Lüdeling & Kytö (eds.): *Corpus linguistics*. <http://www.ling.helsinki.fi/~fkarlsson/earlygen.pdf>
- Karttunen, Lauri; Beesley, Kenneth R.
 2001 A Short History of Two-Level Morphology.
<http://www.ling.helsinki.fi/~koskenni/esslli-2001-karttunen/>
- Kaufmann, Henning
 1977 [¹1959] *Bildungsweise und Betonung der deutschen Ortsnamen*. Grundfragen der Namenskunde 1. München: Fink. [Zweite verbesserte und erweiterte Auflage]
- Kavouras, Marinos; Margarita, Kokla; Tomai, Eleni
 2005 Comparing categories among geographic ontologies. In: Gould, Michael (ed.): *Geospatial research in Europe: AGILE 2003*. Computers & Geosciences 31/2, 145–154, Special Issue.
<http://www.iacm.forth.gr/regional/papers/computersandgeosciences.pdf>
- Kawaguchi et al. eds. [Kawaguchi, Yuji; Zaima, Susumu; Takagaki, Toshihiro; Shibano, Kohji; Usami, Mayumi (eds.)] [= Tokyo University of Foreign Studies]
 2005 Linguistic informatics – State of the art and the future. *Proceedings of the First International Conference on Linguistic Informatics*. Usage-based linguistic informatics 1. Amsterdam: John Benjamins.
- KIM
 Ontotext Lab: The KIM Platform: Knowledge & Information Management.
<http://www.ontotext.com/kim/>
- Kimler, Marco
 2004 Geo-Coding: Recognition of geographical references in unstructured text, and their visualisation. Diploma thesis submitted to the University of Applied Sciences in Hof, Germany, in August 2004.
http://langtech.jrc.it/Documents/0408_Kimler_Thesis-GeoCoding.pdf
- King, Robert Thomas
 1974 Modern semantic theory: a critical analysis and an application to the study of German prepositions. PhD thesis, Rice University, Houston, Texas. [! Anglisten S AO 1460/1+2]
- Kittredge, Richard
 1982 Sublanguages. *Computational Linguistics* 8/2, 79–84.
- Kittredge, Richard; Lehrberger, J. (eds.)
 1981 *Sublanguage*. Studies of language in restricted semantic domains. Berlin: deGruyter.
- Klatt, Stefan
 2004 A high quality partial parser for annotating German text corpora. *LREC 2004, Fourth International Conference on Language Resources and Evaluation, Lisbon - Portugal, 26 May - 28 May 2004*. <http://gandalf.aksis.uib.no/non/lrec2004/pdf/575.pdf>, <http://www.informatik.uni-stuttgart.de/ifi/is/Forschung/Papiere/klatt/chunker04.pdf>
- KML
 KML – Google Code. <http://code.google.com/apis/kml/>
- Kolde, Gottfried
 1995 Grammatik der Eigennamen (Überblick). In: Eichler et al. (eds.): *Namenforschung: ein internationales Handbuch zur Onomastik* 400–408.

- Kornai, András (ed.)
 1999 *Extended finite state models of language*. Studies in natural language processing. Cambridge: Cambridge University Press.
- Kornai, András; Sundheim, Beth M. (eds.)
 2003 *Analysis of geographic references*. Workshop held at the HLT/NAACL Conference 2003, Edmonton, Alberta, Canada. Association for Computational linguistics.
<http://gunsight.metacarta.com/kornai/NAACL/WS9/>
- Koskenniemi, Kimmo
 1983 *Two-level morphology: a general computational model for word-form recognition and production*. Dissertation, Department of General Linguistics, University of Helsinki. University of Helsinki, Department of General Linguistics, Publications 11. Helsinki.
- Koß, Gerhard
 1995 Die Bedeutung der Eigennamen: Wortbedeutung/Namenbedeutung. In: Eichler et al. (eds.): *Namenforschung: ein internationales Handbuch zur Onomastik* 458–463.
- Kracht, Marcus
 2002 On the semantics of locatives. *Linguistics and Philosophy* 25/2, 157–232.
<http://www.linguistics.ucla.edu/people/Kracht/html/lokal.pdf>,
<ftp://ftp.math.fu-berlin.de/pub/usr/kracht/sprache/lokal.ps>
 2006 Directionality selection. In: Saint-Dizier (ed.): *Syntax and semantics of prepositions* 101–114. <http://kracht.humnet.ucla.edu/marcus/html/direction.pdf>
 2008a The fine structure of spatial expressions. In: Asbury, Anna; Dotlacil, Jakub; Gehrke, Berit; Nouwen, Rick (eds.): *The structure of local P* 35–62. Amsterdam: John Benjamins. <http://kracht.humnet.ucla.edu/marcus/html/finelokal.pdf>
 2008b Language and space. Vorlesungsmanuskript.
<http://wwwhomes.uni-bielefeld.de/mkracht/html/lokative.pdf>
- Kroes-Hecht, Birgit M.
 1967 Präpositionen in Übersetzungsmaschinen I. *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung* 20, 243–246. Berlin.
 1969 Präpositionen in Übersetzungsmaschinen II. *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung* 22, 11–15. Berlin.
 1970 Funktion der Präpositionen. Versuch einer Systematisierung. *Actes du Xe Congrès International des Linguistes IV*, 375–396.
- Krovetz, Robert
 1998 More than one sense per discourse. *Proceedings of the ACL-SIGLEX Workshop, SENSEVAL, Herstmonceux Castle September 1998, Web-Based Proceedings*.
<http://www.itri.brighton.ac.uk/events/senseval/ARCHIVE/PROCEEDINGS/krov2.ps>,
<http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.27.2864>
- Krstev, Cvetana; Vitas, Duško
 2007 How to find the right path? (On the morphological disambiguation of sentence in Serbian). *Formal Description of Slavic Languages FDSL-7, Leipzig 30 November - 2 December, 2007*. http://www.uni-leipzig.de/~jungslav/fdsl/fdsl7/abstracts/Krstev_Vitas_FA.pdf, <http://www.matf.bg.ac.yu/~cvetana/biblio/FDSL7-Krstev.pdf>
- Kühnlein, Carola
 2003 Eigennamenerkennung über Konzeptsensoren. Lizentiatsarbeit der Philosophischen Fakultät der Universität Zürich im Bereich Computerlinguistik.
<http://www.ifi.unizh.ch/cl/study/lizarbeiten/lizcarolakuenlein.pdf>

- Langer, Stefan
- 1996 *Selektionsklassen und Hyponymie im Lexikon*. Dissertation. CIS-Bericht-96-94. München. http://www.cis.uni-muenchen.de/pub/cis-berichte/diss_langer.ps.gz [Dissertation Ludwig-Maximilians-Universität München]
- 2003 Kohyponymie in elektronischen Wörterbüchern. *LDV Forum* 18/1-2, 133–144. http://ariadne.coli.uni-bielefeld.de/gldv/site/2003_Doppelheft/133-144_Langer.pdf, http://www.cis.uni-muenchen.de/people/langer/veroeffentlichungen/gldv_kohyponymie.pdf
- Lang, Ewald
- 1989 The semantics of dimensional designation of spatial objects. In: Bierwisch; Lang (eds.): *Dimensional adjectives* 263–417.
- Lang, Ewald; Zifonun, Gisela (eds.)
- 1996 *Deutsch – typologisch*. Jahrbuch 1995 des Instituts für deutsche Sprache. Berlin, New York: de Gruyter.
- Laporte, Eric
- 2001 Reduction of lexical ambiguity. *Linguisticæ investigationes* 24/1, 67–103. Amsterdam, Philadelphia: John Benjamins.
- Laporte, Éric
- 1995a Appropriate nouns with obligatory modifiers. *Language Research* 31/2, 251–289. Seoul: Language Research Institute, Seoul National University.
- 1995b Noms appropriés à modifieur obligatoire. In: Labelle, Jacques: *Lexiques-grammaires comparés et traitements automatiques, Actes du deuxième colloque international LGC, Service des Publications, Québec, UQAM* 31–64.
- 2004a Foreword. In: Leclère et al. (eds.): *Lexique, syntaxe et lexique-grammaire* xi–xxi.
- 2004b Restructuration and the subject of adjectives. In: Leclère et al. (eds.): *Lexique, syntaxe et lexique-grammaire* 373–388. <http://hal.archives-ouvertes.fr/hal-00189718/en/>
- 2005 Une classe d’adjectifs de localisation. *Actes du Colloque, Maison des Sciences de l’Homme, novembre 2002*. Cahiers de lexicologie 86, Adjectifs non prédicatifs, 145–161. Paris: Garnier. <http://igm.univ-mlv.fr/~laporte/publi/adj-loc.htm>, <http://halshs.archives-ouvertes.fr/halshs-00189736/en/>
- Laporte, Éric; Monceaux, Anne
- 1998–1999 Elimination of lexical ambiguities by grammars: the ELAG system. In: Fairon, Cédric (ed.): *Analyse lexicale et syntaxique: le système INTEX* 341–367. <http://citeseer.nj.nec.com/laporte98elimination.html>, <http://www-igm.univ-mlv.fr/~laporte/publi/aji.ps>
- La Robertie, Pierre de
- 2006 Le nom propre en chinois. *Corela – Cognition, Représentation, Langage. Colloque Traitement lexicographique des noms propres, 24 mars 2005, Laboratoire d’Informatique de l’Université François-Rabelais de Tours* Numéros spéciaux, Le traitement lexicographique des noms propres. <http://edel.univ-poitiers.fr/corela/document.php?id=394>
- Larson, Ray R.
- 2005 Cheshire II at GeoCLEF: Fusion and Query Expansion for GIR. *Working Notes for the CLEF 2005 Workshop, 21-23 September, Vienna, Austria*. http://www.clef-campaign.org/2005/working_notes/workingnotes2005/larson205.pdf
- Larson, Ray R.; Frontiera, Patricia
- 2004 Ranking and representation for geographic information retrieval. *SIGIR 2004*. <http://www.geo.unizh.ch/~rsp/gir/abstracts/larson.pdf>

Laur, Wolfgang

- 1995 Morphologie und Wortbildung der Ortsnamen. In: Eichler et al. (eds.): *Namenforschung: ein internationales Handbuch zur Onomastik 1370–1375*.

Law, Gwillim

- 1999 *Administrative subdivisions of countries*. A comprehensive world reference, 1900 through 1998. <http://www.mindspring.com/~gwil/statoids.html>

LDC

- Linguistic Data Consortium. <http://www ldc.upenn.edu>
- 2008a ACE (Automatic Content Extraction) English annotation guidelines for entities, version 6.1 – 2008.01.15.
http://projects ldc.upenn.edu/ace/docs/English-Entities-Guidelines_v6.1.pdf
- 2008b ACE (Automatic Content Extraction) English annotation guidelines for relations, version 6.0 – 2008.01.07.
http://projects ldc.upenn.edu/ace/docs/English-Relations-Guidelines_v6.0.pdf

LDC:LCTL

- LDC - Projects - Less Commonly Taught Languages.
<http://projects ldc.upenn.edu/LCTL/Specifications/>

LDC:LCTL:NE-GUIDELINES

- Linguistic Data Consortium – LCTL Team 2006: Simple Named Entity Guidelines. For Less Commonly Taught Languages, Version 6.5 – March 28, 2006.
<http://crl.nmsu.edu/say/SimpleNamedEntityGuidelinesV6.5.pdf>
- Strassel, Stephanie [Linguistic Data Consortium] 2006: Simple Named Entity Guidelines V6.4. For the Low Density Languages Project. <http://projects ldc.upenn.edu/LCTL/Specifications/SimpleNamedEntityGuidelinesV6.4.pdf>

Leclère, Christian

- 1990 Organisation du lexique-grammaire des verbes français. In: Courtois, Blandine, Silberztein, Max D. (eds.): *Dictionnaires électroniques du français*. Langue Française 87, 112–122. Paris: Larousse.
- 2002 Organization of the lexicon-grammar of French Verbs. *Lingvisticæ Investigationes* 25/1, 29–48. Amsterdam, Philadelphia: John Benjamins.
- 2003 The lexicon-grammar of French verbs: a syntactic database. In: Kawaguchi, Y. et al. (eds.): *Proceedings of the First International Conference on Linguistic Informatics*. UBLI, Tokyo University of Foreign Studies.
- 2005 The lexicon-grammar of French verbs: a syntactic database. In: Kawaguchi et al. (eds.): *Linguistic informatics – State of the art and the future* 29–45. [= Leclère (2003)]

Leclère, Christian; Brisbois-Lenehardt, Jacqueline

- 2004 Synonymie de mots et synonymie de phrases: une approche formelle. In: Leclère et al. (eds.): *Lexique, syntaxe et lexique-grammaire* 389–404.

Leclère et al. eds. [Leclère, Christian; Laporte, Éric; Piot, Mireille; Silberztein, Max D. (eds.)]

- 2004 *Lexique, syntaxe et lexique-grammaire; syntax, lexis & lexicon-grammar*. Papers in honour of Maurice Gross. *Lingvisticæ Investigationes Supplementa* 24. Amsterdam, Philadelphia: John Benjamins.

Lee, Seungwoo; Lee, Gary Geunbae

- 2005 Heuristic methods for reducing errors of geographic named entities learned by bootstrapping. http://isoft.postech.ac.kr/publication/lncs/ijcnlp05_lee.pdf

- Lee, Seungwoo; Lee, Gary Geunbae; Myaeng, Sung Hyon (ed.)
 2005 A bootstrapping approach for geographic named entity annotation. *Information Retrieval Technology. Asia Information Retrieval Symposium 1, 2004, Beijing*. Lecture Notes in Computer Science 3411, 178–189. Berlin, Heidelberg: Springer.
http://isoft.postech.ac.kr/publication/lncs/airs04_lee.pdf,
<http://hri.kaist.ac.kr/publications/22gblee3e.pdf>
- Lee, Sun-Muk
 1994 *Untersuchungen zur Valenz des Adjektivs in der deutschen Gegenwartssprache*. Die morphosyntaktische und logisch-semantische Bestimmung der Ergänzungen zum Adjektiv. Europäische Hochschulschriften, Series I: Deutsche Sprache und Literatur 1429. Frankfurt am Main: Peter Lang.
- Leidner, Jochen L.
 2004a Towards a reference corpus for automatic toponym resolution evaluation. *SIGIR 2004*.
<http://www.geo.unizh.ch/~rsp/gir/abstracts/leidner.pdf>
 2004b Toponym Resolution in Text: Which Sheffield is it? *SIGIR 2004*.
<http://www.iccs.inf.ed.ac.uk/~s0239229/documents/leidner-2004-fyrsigir.pdf>
 2005 Preliminary Experiments with Geo-Filtering Predicates for Geographic IR. *Working Notes for the CLEF 2005 Workshop, 21-23 September, Vienna, Austria*.
http://www.clef-campaign.org/2005/working_notes/workingnotes2005/leidner05.pdf
 2006 Toponym resolution: a first large-scale comparative evaluation. Research Report, School of Informatics, University of Edinburgh. EDI-INF-RR-0839.
<http://www.inf.ed.ac.uk/publications/online/0839.pdf>
 2007 Toponym resolution in text: annotation, evaluation and applications of spatial grounding of place names. PhD thesis, School of Informatics, University of Edinburgh, Edinburgh, Scotland, UK.
- Leidner, Jochen L.; Sinclair, Gail; Webber, Bonnie
 2003 Grounding spatial named entities for information extraction and question answering. *Proceedings of the HLT-NAACL 2003 Workshop on Analysis of Geographic References* 31–38. Edmonton (Alberta, Canada). <http://www.iccs.informatics.ed.ac.uk/~s0239229/documents/pub/pdf/Leidner-Sinclair-Webber-2003-HLTNAACL.pdf>,
<http://people.mokk.bme.hu/~kornai/NAACL/WS9/ws908.pdf>
- Le Pesant, Denis
 2000 Six études de sémantique lexicale sur les noms communs de lieux. Ouvrage rédigé en vue d'obtenir l'Habilitation à Diriger des Recherches. Villetaneuse: Université Paris 13.
<http://perso.orange.fr/denis.lepesant/>
- Le Pesant, Denis; Mathieu-Colas, Michel
 1998 Introduction aux classes d'objets. In: Le Pesant & Mathieu-Colas (eds.): *Les classes d'objets* 6–33.
- Le Pesant, Denis; Mathieu-Colas, Michel (eds.)
 1998 *Les classes d'objets*. Langages 131. Paris: Larousse.
- Le Pesant, Denis; Trégouët, Marta
 1997–2000 Dictionnaire des noms locatifs.
http://www-lli.univ-paris13.fr/rapports/1997_2000/3.1locatifs.htm
- Levašov, Evgenij Aleksandrovič
 2000 *Geografičeskie nazvanija*. Prilagatel'nye, obrazovannye ot nich, nazvanija žitelej. Sankt-Peterburg.

- Leveling, Johannes; Hartrumpf, Sven
- 2006 On metonymy recognition for GIR. *Proceedings of GIR-2006, the 3rd Workshop on Geographical Information Retrieval (held at SIGIR 2006)*. Seattle, Washington. <http://www.geo.unizh.ch/~rsp/gir06/papers/individual/leveling.pdf>
- 2007 University of Hagen at GeoCLEF 2007: Exploring location indicators for geographic information retrieval. *Results of the CLEF 2007 Cross-Language System Evaluation Campaign, Working Notes for the CLEF 2007 Workshop, Budapest, Hungary*. http://www.clef-campaign.org/2007/working_notes/levelingCLEF2007.pdf
- Leveling, Johannes; Hartrumpf, Sven; Veiel, Dirk
- 2005 University of Hagen at GeoCLEF 2005: Using Semantic Networks for Interpreting Geographical Queries. *Working Notes for the CLEF 2005 Workshop, 21-23 September, Vienna, Austria*. http://www.clef-campaign.org/2005/working_notes/workingnotes2005/veiel05.pdf
- Leveling, Johannes; Veiel, Dirk
- 2006 University of Hagen at GeoCLEF 2006: Experiments with metonymy recognition in documents. In: Nardi, Alessandro; Peters, Carol; Vicedo, José Luis: *Results of the CLEF 2006 Cross-Language System Evaluation Campaign, Working Notes for the CLEF 2006 Workshop, Alicante, Spain*. http://www.clef-campaign.org/2006/working_notes/workingnotes2006/veielCLEF2006.pdf
- Levin, Beth
- 1993 *English verb classes and alternations: a preliminary investigation*. Chicago: The University of Chicago Press.
- Levinson, Stephen C.
- 2003 *Space in language and cognition*. Explorations in cognitive diversity. Cambridge, New York: Cambridge University Press.
- Levinson, Stephen C.; Wilkins, David P.
- 2006 *Grammars of space*. Language, culture, and cognition 6. Cambridge, New York: Cambridge University Press.
- Leys, Odo
- 1967 Zur Funktion des Artikels beim Eigennamen. In: Fischer, Rudolf (ed.): *Onomastica Slavogermanica 3*. Abhandlungen der sächsischen Akademie der Wissenschaften zu Leipzig, Philologisch-historische Klasse 58/4, 19–24. Berlin: Akademie.
- Liedtke, Herbert
- ³2002 [1984] *Namen und Abgrenzungen von Landschaften in der Bundesrepublik Deutschland*. Forschungen zur deutschen Landeskunde 239. Deutsche Akademie für Landeskunde.
- Li et al. [Li, Huifeng; Srihari, Rohini K.; Niu, Cheng; Li, Wei]
- 2003 Location Normalization for Information Extraction. In: Kornai & Sundheim (eds.): *Analysis of geographic references*. <http://edward.comp.lancs.ac.uk/acl/C/C02/C02-1127.pdf>, <http://www.aclweb.org/anthology-new/C/C02/C02-1127.pdf>
- Li, Jie
- 1994 *Räumliche Relationen und Objektwissen am Beispiel von an und bei*. Studien zur deutschen Grammatik 49. Tübingen: Gunter Narr.
- LINGPIPE
- LingPipe Home. <http://www.alias-i.com/lingpipe/>
- Liu, Haitao
- 1999–2006 Basic Bibliography for Dependency grammar and Valency Theory. <http://ling.cuc.edu.cn/htliu/dgb.htm>

- Lötscher, Andreas
 1995 Der Name als lexikalische Einheit: Denotation und Konnotation. In: Eichler et al. (eds.): *Namenforschung: ein internationales Handbuch zur Onomastik* 448–457.
- Lüdeling, Anke; Kytö, Merja (eds.)
 2008–2009 *Corpus linguistics*. An international handbook. Handbücher zur Sprach- und Kommunikationswissenschaft 29. Berlin: de Gruyter.
- Maienborn, Claudia
 1990a *Lokale Verben und Präpositionen*. Semantische und konzeptuelle Verarbeitung in LEU II. Wissenschaftliches Zentrum, Institut für wissenschaftsbasierte Systeme, Report 119. Stuttgart, Hamburg: IBM.
 1990b *Position und Bewegung: Zur Semantik lokaler Verben*. Wissenschaftliches Zentrum, Institut für wissenschaftsbasierte Systeme, Report 138. Stuttgart, Hamburg: IBM.
 1991 Verbs of motion and position: On the optionality of the local argument. In: Herzog & Rollinger (eds.): *Text understanding in LILOG* 621–631.
 1995 Toward a compositional semantics for locative modifiers. *Proceedings from Semantics and Linguistic Theory V* 237–254. Ithaca (New York): Cornell University Publications.
 1996 *Situation und Lokation. Die Bedeutung lokaler Adjunkte von Verbalprojektionen*. Studien zur deutschen Grammatik 53. Tübingen: Stauffenburg.
 2001 On the position and interpretation of locative modifiers. *Natural Language Semantics* 9, 191–240.
- Maier-Meyer, Petra
 1995 *Lexikon und automatische Lemmatisierung*. CIS-Bericht-95-84. München.
<http://www.cis.uni-muenchen.de/pub/cis-berichte/CIS-Bericht-95-84.ps.gz>
- Maillat, Didier
 1999 *On the interpretation of directional expressions: empirical and theoretical considerations*. Ms. MPhil Thesis, University of Oxford.
 2001a Which template for *behind*? Empirical considerations of the meaning of directional PPs. *Oxford Working Papers in Linguistics* 6, 151–165. Oxford: University of Oxford.
<http://www.clp.ox.ac.uk/research/publications/owp/OWP2001.pdf>
 2001b Directional PPs and reference frames in DRT. *Proceedings of the ACL 2001 Workshop on Workshop on Temporal and Spatial Information Processing*.
<http://acl.ldc.upenn.edu/W/W01/W01-1304.pdf>
 2003a *The semantics and pragmatics of directionals: a case study in English and French*. PhD dissertation, University of Oxford.
 2003b Towards a universal DRT model for the interpretation of directional PPs within a reference frame. In: Jaszczolt, Katarzyna M.; Turner, Ken (eds.): *Meaning through language contrast*. Volume 1. Pragmatics and Beyond: New Series 99, 295–305. Amsterdam, Philadelphia: John Benjamins.
- Makhoul et al. [Makhoul, John; Kubala, Francis; Schwartz, Richard; Weischedel, Ralph]
 1999 Performance measures for information extraction. *Proceedings of DARPA Broadcast News Workshop, Herndon, VA, February 1999*.
<http://citeseer.ist.psu.edu/makhoul99performance.html>
- Mallchok, Friederike
 2005 *Automatic recognition of organization names in English business news*. Studien zur Informations- und Sprachverarbeitung 9. München: Centrum für Informations- und Sprachverarbeitung.

Mańczak, Witold

- 1995 Morphologie des noms: Règles de flexion, systèmes de flexion. In: Eichler et al. (eds.): *Namenforschung: ein internationales Handbuch zur Onomastik* 427–431.

Manias, Sven

- 1994 *Wörterbuch über geographische Namen in verschiedenen Sprachen*. Deutschland, Schweiz, Österreich. <http://www.sven-manias.de/geo2.dvi.gz>,
<http://www.sven-manias.de/geo2.ps.gz>

Manov et al. [Manov, Dimitar; Kiryakov, Atanas; Popov, Borislav; Bontcheva, Kalina; Maynard, Diana; Cunningham, Hamish]

- 2003 Experiments with geographic knowledge for information extraction. *Workshop on the Analysis of Geographic References, NAACL'03*.
<http://www.metacarta.com/kornai/NAACL/WS9/Conf/ws903.pdf>

Mark, David M.

- 1989 Cognitive and linguistic aspects of geographic space. National Center for Geographic Information and Analysis, Technical Report 1.
21999 Spatial representation: a cognitive view. In: Maguire, D. J.; Goodchild, M. F.; Rhind, D. W.; Longley, P. (eds.): *Geographical information systems: principles and application 1*, 81–89. London: Longmans Publishers.
<http://www.geog.buffalo.edu/~dmark/BB2chapter.html>

Mark, David M.; Smith, Barry

- 2003 Do mountains exist? Ontology of landforms and topography. *Environment & Planning B* 30/3, 411–427. <http://ontology.buffalo.edu/smith/articles/Mountains.pdf>

Mark, David M.; Smith, B.; Tversky, Barbara

- 1999 Ontology and geographic objects: an empirical study of cognitive category. In: Freksa, Christian; Mark, David M. (eds.): *Spatial information theory: a theoretical basis for GIS*. Lecture notes in computer science 1661, 283–298. Berlin: Springer.

Markert, Katja; Nissim, Malvina

- 2002 Towards a corpus annotated for metonymies: the case of location names. *Proceedings of the 3rd International Conference on Language Resources and Evaluation (LREC2002), Las Palmas, Canary Islands, 2002* 1385–1392.
<http://citeseer.ist.psu.edu/markert02towards.html>
2007a Metonymic proper names: A corpus-based account. In: Stefanowitsch, Anatol; Gries, Stefan Th. (eds.): *Corpus-based approaches to metaphor and metonymy* 152–174. Mouton de Gruyter. <http://www.comp.leeds.ac.uk/markert/Papers/metaphorbook2006.pdf>
2007b Metonymy resolution at SemEval I: guidelines for participants.
<http://www.comp.leeds.ac.uk/markert/guidelines.updated.pdf>

Marschner, Clemens

- 2007 Efficiently matching with local grammars using prefix overlay transducers. In: Holub, Jan; Zdárek, Jan (eds.): *Implementation and Application of Automata, 12th International Conference, CIAA 2007, Prague, Czech Republic, July 16-18, 2007, Revised Selected Papers*. Lecture Notes in Computer Science 4783, 314–316. Springer.

Marsh, Elaine; Perzanowski, Dennis

- 1998 MUC-7 evaluation of IE technology: overview of results. http://www-nlpir.nist.gov/related_projects/muc/proceedings/muc_7_proceedings/marsh_slides.pdf

Martin, Laura

- 1986 “Eskimo words for snow”: a case study in the genesis and decay of an anthropological example. *American Anthropologist* 88, 418–423.

- Martins, Bruno; Silva, Mário J.; Chaves, Marcirio Silveira
 2005 Challenges and resources for evaluating Geographical IR. *Proceedings of the Workshop on Geographic Information Retrieval at CIKM 2005*.
http://xldb.fc.ul.pt/data/Publications_attach/grease-evaluation.pdf
- Mathieu-Colas, Michel
 1998 Illustration d'une classe d'objets : les voies de communication. In: Le Pesant & Mathieu-Colas (eds.): *Les classes d'objets* 77–90.
- Matsumoto, Yo; Slobin, Dan I.
 2002 A bibliography of linguistic expressions for motion events. *Meiji Gakuin Review* 684, 83–158, Meiji Gakuin University.
<http://www.lit.kobe-u.ac.jp/~yomatsum/motionbiblio.html>
- Maurel, Denis
 2004 Les mots inconnus sont-ils des noms propres ? In: Purnelle, Gérald; Fairon, Cédric; Dister, Anne (eds.): *Le poids des mots*. Actes des 7es Journées internationales d'Analyse statistique des Données Textuelles 2, 776–784.
http://www.cavi.univ-paris3.fr/lexicometrica/jadt/jadt2004/pdf/JADT_074.pdf
- Maurel et al. [Maurel, Denis; Belleil, Claude; Piton, Odile; Eggert, Elmar]
 1997 Un dictionnaire électronique relationnel pour les noms propres. *Revue française de Linguistique appliquée* 2/1, 101–111.
- Maurel, Denis; Piton, Odile; Eggert, Elmar
 2000 Les relations entre noms propres : lieux et habitants dans le projet PROLEX. In: Maurel, Denis; Guenther, Franz (eds.): *Traitement automatique des noms propres*. Traitement automatique des langues 41/3, 623–641. Paris: ATALA/Hermes Sciences.
- Maurel et al. [Maurel, Denis; Tran, Mickaël; Vitas, Duško; Grass, Thierry; Savary, Agata]
 2004 Prolexbase : Proposition d'une ontologie multilingue des noms propres. Rapport interne du Laboratoire d'Informatique de l'Université de Tours 274.
http://tln.li.univ-tours.fr/Tln_Biblio/2004RapportOntologieProlexbase.zip
 2006 Prolex : Implantation d'une ontologie multilingue des noms propres. Rapport interne du Laboratoire d'Informatique de l'Université François-Rabelais de Tours 286.
http://tln.li.univ-tours.fr/Tln_Biblio/2006RapportImplantationOntologieProlex.zip
- McCormack, Richard W. B.
⁵2000 *Tief in Bayern*. Frankfurt am Main: Eichborn.
- McDonald, David D.
 1993 Internal and External Evidence in the Identification and Semantic Categorization of Proper Names. In: Boguraev, Branimir; Pustejovsky, James (eds.): *Acquisition of Lexical Knowledge from Text*. Proceedings of a Workshop Sponsored by the Special Interest Group on the Lexicon of the Association for Computational Linguistics 32–43. Columbus (Ohio). <http://ac1.ldc.upenn.edu/W/W93/W93-0104.pdf>
 1996 Internal and external evidence in the identification and semantic categorization of proper names. In: Boguraev, Branimir; Pustejovsky, James (eds.): *Corpus processing for lexical acquisition* 21–39. Cambridge (Massachusetts): MIT Press. [weitgehend identisch mit 1993]
- Mederle, Wolfgang
 2004 Automatische Adreßerkennung: Ein Ansatz für deutsche Adressen und seine Implementierung. Magisterarbeit im Studiengang Computerlinguistik, Centrum für Informations- und Sprachverarbeitung, Ludwig-Maximilians-Universität München.
<http://www.mederle.de/studium/clark.pdf>

Meibauer, Jörg

- 1995 Komplexe Präpositionen: Grammatikalisierung, Metapher, und division of pragmatic labour. In: Liedtke, Frank (ed.): *Implikaturen: Grammatische und pragmatische Analysen*. Linguistische Arbeiten 343, 47–74. Tübingen: Niemeyer.

Mel'čuk, Igor [= Mel'čuk, Igor' Aleksandrovič]

- 1974 *Opyt teorii lingvističeskich modelej «SMYSL ⇔ TEKST»*. Moskva.
- 1993–2000 *Cours de morphologie générale (théorique et descriptive)*. Montreal: Les Presses de l'Université de Montréal/Paris: CNRS Éditions.
- 1995 Phrasemes in language and phraseology in linguistics. In: Everaert, Martin; Linden, Erik-Jan van der; Schenk, André et al. (eds.): *Idioms: Structural and psychological perspectives* 167–232. Hillsdale: Lawrence Erlbaum Associates.
<http://www.uofaweb.ualberta.ca/linguistics/pdf/Melcuk2006.pdf>
- 1996 Lexical functions: a tool for the description of lexical relations in the lexicon. In: Wanner (ed.): *Lexical functions in lexicography and natural language processing* 37–102.
- 1997–2001 *Kurs obščej morfologii*. Moskva, Vena. [Übersetzung von 1993–2000]
- 2003 Levels of dependency in linguistic description: concepts and problems. In: Ágel et al. (eds.): *Dependenz und Valenz* 188–229.
<http://www.olst.umontreal.ca/FrEng/Dependency.pdf>
- 2004a Verbes supports sans peine. In: Gross, Gaston; Pontonx, Sophie de (eds.): *Verbes supports: nouvel état des lieux*. *Linguisticae Investigationes* 27/2, 203–217. Amsterdam, Philadelphia: John Benjamins.
- 2004b Actants in semantics and syntax I: Actants in semantics. *Linguistics* 42/1, 1–66.
- 2004c Actants in semantics and syntax II: Actants in syntax. *Linguistics* 42/2, 247–291.
- 2006 *Aspects of the theory of morphology*. Trends in linguistics 146. Berlin, New York: Mouton de Gruyter.

Mel'čuk, Igor A.; Pertsov, Nikolaj V.

- 1987 *Surface syntax of English*. A formal model within the Meaning-Text framework. Amsterdam, Philadelphia: John Benjamins.

Mel'čuk, Igor A.; Wanner, Leo

- 1996 Lexical functions and lexical inheritance for emotion lexemes in German. In: Wanner (ed.): *Lexical functions in lexicography and natural language processing* 209–278.

Meyers et al. [Meyers, Adam; Reeves, Ruth; Macleod, Catherine; Szekely, Rachel; Zielinska, Veronika; Young, Brian; Grishman, Ralph]

- 2004a Annotating noun argument structure for NomBank. *Proceedings of LREC-2004, Lisbon, Portugal*. <http://nlp.cs.nyu.edu/meyers/papers/nombank-ann.pdf>
- 2004b The NomBank project: an interim report. In: Meyers, Adam (ed.): *HLT-NAACL 2004 Workshop: Frontiers in Corpus Annotation, May 2 – May 7, Boston, Massachusetts, USA* 24–31. Association for Computational Linguistics.
<http://nlp.cs.nyu.edu/meyers/papers/nombank-pap.pdf>

Meynen, Emil (ed.) [= Ständiger Ausschuss für geographische Namen (StAGN)]

- 1966 *Duden Wörterbuch geographischer Namen*. Europa (ohne Sowjetunion). Mannheim: Bibliographisches Institut.

Mikheev, Andrei; Moens, Marc; Grover, Claire

- 1999 Named entity recognition without gazetteers. *Proceedings of EACL, 1999*.
<http://ucrel.lancs.ac.uk/acl/E/E99/E99-1001.pdf>

Milićević, Jasmina

- 2006 A short Guide to the Meaning-Text Linguistic Theory. *Journal of Koralex* 8, 187–233.
<http://www.olst.umontreal.ca/pdf/IntroMTTJM.pdf>

- Miller, George A.
 1956 The magical number seven, plus or minus two: some limits on our capacity for processing information. *The Psychological Review* 63, 81–97.
<http://www.musanim.com/miller1956/>
- Miller, George A.; Johnson-Laird, Philipp N.
 1976 *Language and perception*. Cambridge: Cambridge University Press.
- MINDSWAP:GEOONTOLOGIES
 Maryland Information and Network Dynamics Lab Semantic Web Agents Project
 (Mindswap): geoOntologies. <http://www.mindswap.org/2004/geo/geoOntologies.shtml>
- Mohri, Mehryar
 1994 Syntactic analysis by local grammars automata: an efficient algorithm. *Proceedings of COMPLEX '94*. <http://xxx.lanl.gov/pdf/cmp-lg/9407002>
- Mohri, Mehryar; Sproat, Richard
 2006 On a common fallacy in computational linguistics. *SKY Journal of Linguistics* 19.
http://www.ling.helsinki.fi/sky/julkaisut/SKY2006_1/1.6.5.%20MOHRI%20%20SPROAT.pdf
- Moilanen, Markku
 1979 *Statische lokative Präpositionen im heutigen Deutsch*. Wahrheits- und Gebrauchsbedingungen. *Linguistische Arbeiten* 70. Tübingen: Niemeyer.
- Morimoto et al. [Morimoto, Yasuhiko; Aono, Masaki; Houle, Michael E.; McCurley, Kevin S.]
 2003 Extracting spatial knowledge from the web. *Symposium on Applications and the Internet, January 27 - 31, 2003, Orlando, Florida* 326.
<http://www.almaden.ibm.com/cs/people/mccurley/pdfs/SAINT03.pdf>
- MSN:LOCAL
 MSN Local Live Search. <http://maps.live.de/LiveSearch.LocalLive>
- MUC-6
 1996: MUC-6. <http://cs.nyu.edu/cs/faculty/grishman/muc6.html>
- MUC-6-NE-TASK
 1995: Named Entity Task Definition.
http://cs.nyu.edu/cs/faculty/grishman/NEtask20.book_1.html
- MUC-7
 1998: Proceedings of the Seventh Message Understanding Conference. http://www.itl.nist.gov/iaui/894.02/related_projects/muc/proceedings/muc_7_toc.html,
http://www.itl.nist.gov/iaui/894.02/related_projects/muc/proceedings/ie_task.html
- Müller, Frank Henrik; Ule, Tylman
 2001 Satzklammer annotieren und Tags korrigieren. Ein mehrstufiges ‘Top-Down-Bottom-Up’-System zur flachen, robusten Annotierung von Sätzen im Deutschen. In: Lobin, Henning (ed.): *Sprach- und Texttechnologie in digitalen Medien. Proceedings der GLDV-Frühjahrstagung 2001* 225–234. Norderstedt.
- Müller, Joachim
²⁸2003 *Müllers großes deutsches Ortsbuch*. Wuppertal: Post- und Ortsbuchverlag Müller.
- Müller, Marina
 2006 Levels of abstraction in textual representations of geoinformation - example: cadastral descriptions of Brazil. *Proceedings of Joint Workshop on Multiple Representation and Interoperability of Spatial Data, Hannover, Febr. 2006*. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Science (IAPRSIS)

- 26/2/W40, 71–78.
<http://www.ipf.uni-karlsruhe.de/Personen/lucas/literatur/hannover.pdf>
- 2008 Transformations of Cadastral Descriptions with Incomplete Information into Maps. *Transactions in GIS* 12/1, 83–101.
- Müller, Marina; Bähr, Hans-Peter
- 2006 Beschreibungen in Text- und Kartenform – Grundlagen für einen formalen Vergleich. *Tagungsband der 41. Sitzung der Arbeitsgruppe Automation in der Kartographie AgA 2004 in Hamburg*. Mitteilungen des Bundesamtes für Kartographie und Geodäsie 34. Frankfurt am Main: Verlag des BKG. http://www.ipf.uni-karlsruhe.de/Personen/mueller/literatur/mueller_baehr_aga2004.pdf
- Müller, Stefan
- 1999 *Deutsche Syntax deklarativ*. Head-Driven Phrase Structure Grammar für das Deutsche. Linguistische Arbeiten 394. Tübingen: Max Niemeyer.
<http://www.cl.uni-bremen.de/~stefan/Pub/hpsg.html>
- Naaman et al. [Naaman, Mor; Song, Yee Jiun; Paepcke, Andreas; Garcia-Molina, Hector]
- 2004 Assigning textual names to sets of geographic coordinates. *SIGIR 2004*.
<http://www.geo.unizh.ch/~rsp/gir/abstracts/naaman.pdf>
- 2006 Assigning textual names to sets of geographic coordinates. *Computer, Environment and Urban Systems* 30/4 (Special Issue on Geographic Information Retrieval), 418–435.
- Nadeau, David
- 2005 Création de surcouche de documents hypertextes et traitement du langage naturel. *CLINE 05, 3rd Computational Linguistics in the North-East Workshop, Université du Québec en Outaouais, Gatineau (Québec), August 26th 2005*.
http://www.crtl.ca/cline05/cline05_papers/Nadeau.pdf
- 2007 Semi-supervised named entity recognition: learning to recognize 100 entity types with little supervision. Thesis submitted to the Faculty of Graduate and Postdoctoral Studies in partial fulfillment of the requirements for the PhD degree in Computer Science.
<http://cogprints.org/5859/1/Thesis-David-Nadeau.pdf>
- Nadeau, David; Sekine, Satoshi
- 2007 A survey of named entity recognition and classification. In: Sekine & Marques-Ranchhod (eds.): *Named Entities: Recognition, classification, and use* 3–26.
- NASA:SWEET
- NASA Jet Propulsion Laboratory o.J.: Semantic Web for Earth and Environmental Terminology (SWEET). <http://sweet.jpl.nasa.gov/ontology/>,
<http://sweet.jpl.nasa.gov/ontology/space.owl>
- NEGRA
- NEGRA Corpus. <http://www.coli.uni-saarland.de/projects/sfb378/negra-corpus/>
- Nerius, Dieter
- 1995 Schreibung der Namen: Prinzipien, Normen und Freiheiten. In: Eichler et al. (eds.): *Namenforschung: ein internationales Handbuch zur Onomastik* 414–419.
- Neumann, Günter; Piskorski, Jakub
- 2002 Shallow text processing core engine. *Computational Intelligence* 18/3, 451–476.
<http://citeseer.ist.psu.edu/neumann02shallow.html>
- Nevin, Bruce E. (ed.)
- 2002 *The legacy of Zellig Harris*. Language and information into the 21st century. Volume 1: Philosophy of science, syntax and semantics. *Current Issues in Linguistic Theory* 228.

Nevin, Bruce E.; Johnson, Stephen M. (eds.)

- 2002 *The legacy of Zellig Harris*. Language and information into the 21st century. Volume 2: Mathematics and computability of language. Current Issues in Linguistic Theory 229.

NGA:GNS

National Geospatial-Intelligence Agency (NGA): GEOnet Names Server (GNS): Names Files of Selected Countries. <http://earth-info.nima.mil/gns/html/>

NOMBANK

NomBank. <http://nlp.cs.nyu.edu/meyers/NomBank.html>

NOOJ

INTEX/NooJ. <http://www.nooj4nlp.net>

NOOJ-MANUAL

Silberztein, Max 2002–2008: *NooJ*. <http://www.nooj4nlp.net/NooJ%20Manual.pdf>

Nübling, Damaris

- 2004 Zum Proprialisierungsgrad von *die neuen Bundesländer*. In: Eichler (ed.): *Völkernamen, Ländernamen, Landschaftsnamen* 225–242.

O’Keefe, John

- 1996 The spatial prepositions in English: vector grammar, and the cognitive map theory. In: Bloom et al. (eds.): *Language and space* 277–316.
- 2003 Vector grammar, places, and the functional roles of the spatial prepositions in English. In: van der Zee & Slack (eds.): *Representing direction in language and space* 69–85.

Oliva et al. [Oliva, Karel; Hnátková, Milena; Petkevič, Vladimír; Květoň, Pavel]

- 2000 The linguistic basis of a rule-based tagger of Czech. In: Sojka, Petr; Ivan Kopaček; Karel Pala (eds.): *Text, Speech, Dialogue. Proceedings of the Third International Workshop, TSD 2000, Brno, Czech Republic, September 13-16, 2000*. Lecture Notes in Computer Science 1902, 3–8. Springer.
<http://ufal.mff.cuni.cz/czech-tagging/OlivaKvetonALTSD2000.pdf>

Olsen, Susan

- 1994 Lokalalternationen im Deutschen und Englischen. *Zeitschrift für Sprachwissenschaft* 13, 201–235.
- 1995 Alternative grammatische Realisierungen lokativer Komplemente. *Papiere zur Linguistik* 52, 3–26.
- 1997 Zur Kategorie Verbpartikel. *Beiträge zur Geschichte der deutschen Sprache und Literatur* 119, 1–32.
- 1999 Durch den Park durch, zum Bahnhof hin. Komplexe Präpositionalphrasen mit einfachem directionalem Kopf. In: Wegener, Heide (ed.): *Deutsch kontrastiv. Typologisch vergleichende Untersuchungen zur deutschen Grammatik*. Studien zur deutschen Grammatik 59, 111–134. Tübingen: Stauffenburg.

ONTOWORLD.ORG

Ontoworld.org. <http://ontoworld.org/>

OPENGIS-SQL

Open GIS Consortium OpenGIS 1999: Simple features specification for SQL, revision 1.1. <http://www.opengis.org/docs/99-049.pdf>

OPENSTREETMAP

OpenStreetMap – The Free Wiki World Map. <http://www.openstreetmap.org/>

Ormeling, Ferjan

2003a What is at issue? In: Ormeling; Stabe; Sievers (ed.): *Training course on toponymy* 11–15.

2003b Functions of geographical names for cartographic and non-cartographic purposes. In: Ormeling; Stabe; Sievers (ed.): *Training course on toponymy* 29–37.

2003c Exonyms in cartography. In: Ormeling; Stabe; Sievers (ed.): *Training course on toponymy* 119–125.

Ortner et al. [Ortner, L; Bollhagen-Müller, E.; Ortner, H.; Wellmann, H.; Pümpel-Mader, M.; Gärtner, H.]

1991 *Deutsche Wortbildung. Typen und Tendenzen in der Gegenwartssprache. Substantivkomposita*. Berlin, New York: Walter de Gruyter.

OSNOVENSISTEM

Makedonska Akademija na Naukite i Umetnostite, Meġunaroden Komitet na Slavistite, Komisija za Sloven. Onomastika 1983: *Osnoven sistem i terminologija na slovenskata onomastika*. Osnovna sistema i terminologija slavjanskoj onomastiki – Grundsystem und Terminologie der slawischen Onomastik. Skopje.

ÖSTERREICHISCHEBERGNAMEN

Pohl, Heinz Dieter o.J.: *Österreichische Bergnamen*.
<http://www.uni-klu.ac.at/spw/oenf/name4.htm>

Panevová, Jarmila

1994 Valency frames and the meaning of the sentence. In: Luelsdorff, Philip A. (ed.): *The Prague school of structural and functional linguistics. A short introduction*. Linguistic and literary studies in Eastern Europe 41, 223–243. Amsterdam: John Benjamins.

Pantel, Evelina

1993 *Valenz im ‚Smysl‘ ⇔ ‚Tekst‘-Modell*. Eine konfrontative Analyse russischer und polnischer Verben. Slavistische Beiträge 33. München: Otto Sagner.

Pasch et al. [Pasch, Renate; Brauße, Ursula; Breindl, Eva; Waßner, Ulrich Hermann]

2003 *Handbuch der deutschen Konnektoren*. Linguistische Grundlagen der Beschreibung und syntaktische Merkmale der deutschen Satzverknüper (Konjunktionen, Satzadverbien und Partikeln). Schriften des Instituts für Deutsche Sprache 9. Berlin, New York: de Gruyter.

Paumier, Sébastien

2002a Some remarks on the application of a lexicon-grammar. *Linguisticae Investigationes* 24/2, 245–256. Amsterdam, Philadelphia: John Benjamins.

2002b Manuel d'utilisation d'Unitex. <http://www-igm.univ-mlv.fr/~unitex/>

2004 Recursive Automata for Syntactic Grammars. *Proceedings of Lexicon-Grammar Workshop, Beijing, 16-17 oct, 2004*. <http://igm.univ-mlv.fr/~paumier/Recursive%20Automata%20for%20Syntactic%20Grammars.pdf>

2006 UNITEX 1.2, User Manual. <http://www-igm.univ-mlv.fr/~unitex/UnitexManual1.2.pdf>

PERL

The Perl Directory – perl.org. <http://www.perl.org/>

Peškovskij, Aleksandr Matveevič

⁷1956 [1914] *Russkij sintaksis v naučnom osveščanii*. Moskva: Učpedgiz.

Petkevič, Vladimír

2001 Grammatical agreement and automatic morphological disambiguation of inflectional languages. In: Matousek, Václav; Mautner, Pavel; Moucek, Roman; Tauser, Karel (eds.): *Text, Speech and Dialogue. Proceedings of the 4th International Conference, TSD 2001*,

Zelezna Ruda, Czech Republic, September 11-13, 2001. Lecture Notes in Computer Science 2166, 47–53. Springer.

Petras, Vivien; Gey, Fredric

2005 Berkeley2 at GeoCLEF: Cross-Language Geographic Information Retrieval of German and English Documents. *Working Notes for the CLEF 2005 Workshop, 21-23 September, Vienna, Austria*.

http://www.clef-campaign.org/2005/working_notes/workingnotes2005/gey205.pdf

Piskorski, Jakub; Neumann, Günter

2000 An intelligent text extraction and navigation system. *Proceedings of the 6th International Conference on Computer-Assisted Information Retrieval (RIAO-2000)*. Paris, April 2000. <http://citeseer.comp.nus.edu.sg/249769.html>,

<http://www.coli.uni-saarland.de/publikationen/softcopies/Piskorski:2000:ITE.pdf>

Piton, Odile; Maurel, Denis

1997 Le traitement informatique de la géographie politique internationale. *Bulag, numéro spécial. Colloque Franche-Comté Traitement automatique des langues (FRACTAL 97)*, Besançon, 10-12 décembre 321–328.

2000 Beijing frowns and Washington takes notice: computer processing of relations between geographical proper names in foreign affairs. *Fourth International Workshop on Applications of Natural Language to Data Bases (NLDB'00)*, Versailles, 28-30 juin 66–78. http://tln.li.univ-tours.fr/Tln_Biblio/2000nldb.zip

Pittner, Karin

1999 *Adverbiale im Deutschen*. Untersuchungen zu ihrer Stellung und Interpretation. Tübingen: Stauffenburg.

Plötz, Senta (ed.)

1972 *Transformationelle Analyse*. Die Transformationstheorie von Zellig Harris und ihre Entwicklung. Linguistische Forschungen 8. Frankfurt (Main): Athenäum.

Poibeau, Thierry

2006 Dealing with metonymic readings of Named Entities. *Proceedings of the Annual Conference of the Cognitive Science Society (COGSCI)*, Vancouver, Canada.

<http://arxiv.org/pdf/cs.AI/0607052>

POSTGIS

Geographic objects for PostgreSQL. <http://postgis.refractions.net/>

Pouliquen et al. [Pouliquen, Bruno; Kimler, Marco; Steinberger, Ralf; Ignat, Camelia; Oellinger, Tamara; Blackler, Ken; Fuart, Flavio; Zaghouani, Wajdi; Widiger, Anna; Forslund, Ann-Charlotte; Best, Clive]

2006 Geocoding multilingual texts: Recognition, disambiguation and visualisation. *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC'2006)*. Genoa, Italy, 24-26 May 2006.

http://langtech.jrc.it/Documents/0605_LREC_GeoCoding_Pouliquen-et-al.pdf

Pouliquen et al. [Pouliquen, Bruno; Steinberger, Ralf; Ignat, Camelia; Groeve, Tom De]

2004 Geographical information recognition and visualisation in texts written in various languages. *Proceedings of the 19th Annual ACM Symposium on Applied Computing (SAC'2004)*, Special Track on Information Access and Retrieval (SAC-IAR). Nicosia, Cyprus, 14 - 17 March 2004 2, 1051–1058.

http://langtech.jrc.it/Documents/ACM-SAC-2004_Pouliquen-Steinberger-et-al.pdf

Pribbenow, Simone

1991 Phenomena of localization. In: Herzog & Rollinger (eds.): *Text understanding in LILOG* 609–620.

1993 *Räumliche Konzepte in Wissens- und Sprachverarbeitung. Hybride Verarbeitung von Lokalisation*. Wiesbaden: DUV.

PROLEX

Laboratoire d'Informatique de l'Université de Tours: Site Prolex.

http://tln.li.univ-tours.fr/tln_prolex/prolex.php

PROLINTEX

Laboratoire d'Informatique de l'Université de Tours: Prolintex.

http://tln.li.univ-tours.fr/Tln_Prolintex.html

PROTÉGÉ

The Protégé Ontology Editor and Knowledge Acquisition System.

<http://protege.stanford.edu/>

Pullum, Geoffrey K.

1991 *The great Eskimo vocabulary hoax and other irreverent essays on the study of language*. Chicago, London: University of Chicago Press.

Purves, Ross; Clough, Paul; Joho, Hideo

2005 Identifying imprecise regions for geographic information retrieval using the web.

GISRUK 2005, Glasgow. <http://www.dcs.gla.ac.uk/~hideo/pub/gisuk05/gisuk05.pdf>

Purves et al. [Purves, Ross; Clough, Paul; Joho, Hideo; Jones, Christopher B.; Kreveld, Marc van]

2005 Modelling vague places with knowledge from the web. SPIRIT Public Deliverables (Reports) D24 3301. http://www.geo-spirit.org/publications/SPIRIT_D24_WP3.pdf

Quasthoff, Uwe; Biemann, Christian; Wolff, Christian

2002 Named entity learning and verification: expectation maximization in large corpora. In: Roth, Dan; Bosch, Antal van den (eds.): *Proceedings of CoNLL-2002* 8–14.

<http://www.cnts.ua.ac.be/con112002/pdf/00814qua.pdf>

Radden, Günter

1989 Das Bewegungskonzept: *to come* und *to go*. In: Habel; Herweg; Rehkamper (eds.): *Raumkonzepte in Verstehensprozessen* 228–248.

Randell, David A.; Cui, Zhan; Cohn, Anthony G.

1992 A spatial logic based on regions and connection. In: Nebel, Bernhard; Rich, Charles; Swartout, William (eds.): *Principles of knowledge representation and reasoning* 165–176.

Boston: Morgan Kaufmann. <http://citeseer.ist.psu.edu/randell192spatial.html>

Raper, Peter E.

1995 The United Nations Group of Experts on Geographical Names (UNGEGN). In: Eichler et al. (eds.): *Namenforschung: ein internationales Handbuch zur Onomastik* 281–287.

Rauh, Gisa

1981 On *coming* and *going* in English and German. *Papers and Studies in Contrastive Linguistics* 13, 53–68.

1996 *Zur Struktur von Präpositionalphrasen im Englischen*. Arbeiten des Sonderforschungsbereichs 282: Theorie des Lexikons 90. Wuppertal: Bergische Universität – Gesamthochschule Wuppertal.

1997 Lokale Präpositionen und referentielle Argumente. *Linguistische Berichte* 171, 415–42.

Rau, Lisa F.

1991 Extracting company names from text. *Proceedings of the 7th IEEE Conference on Artificial Intelligence Applications*.

Revuz, Dominique

- 1991 Dictionnaires et lexiques: méthodes et algorithmes. Thèse de doctorat, Paris: Université Paris 7 (Publication du LITP). <http://www-igm.univ-mlv.fr/~dr/thdr/>

RGN

Vasmer, Max et al. 1962–1989: *Russisches geographisches Namenbuch*. Wiesbaden: Harrassowitz.

Rijsbergen, C. V. “Keith” van

- ²1979 [1975] *Information Retrieval*. London, Boston: Butterworth.
<http://www.dcs.gla.ac.uk/Keith/Preface.html>

Riloff, Ellen; Jones, Rosie

- 1999 Learning dictionaries for information extraction by multi-level bootstrapping. *Proceedings of the Sixteenth National Conference on Artificial Intelligence* 474–479. <http://www.cs.cmu.edu/afs/cs.cmu.edu/project/theo-11/www/wkwb/aaai99iedict-rj.ps.gz>

Roche, Emmanuel

- 1992 Text disambiguation by finite-state automata: an algorithm and experiments on corpora. *COLING-92. Proceedings of the Conference, Nantes*.
<http://acl.ldc.upenn.edu/C/C92/C92-3153.pdf>
- 1993a Une représentation par automate fini des textes et des propriétés transformationnelles des verbes. *Linguisticae Investigationes* 17/1, 189–222. Amsterdam, Philadelphia: John Benjamins.
- 1993b Analyse syntaxique transformationnelle de français par transducteurs et lexique-grammaires. Thèse de Doctorat d’Informatique Fondamentale, Université Paris 7.
- 1993c Dictionary compression experiments. Technical Report IGM 93-5, Université de Marne la Vallée, Noisy le Grand, France.
- 1996 Parsing with finite-state transducers. Mitsubishi Electrical Research Laboratory, technical report 96-30 (MERL-TR-96-30). <http://www.merl.com/papers/docs/TR96-30.pdf>
- 1997 Parsing with finite-state transducers. In: Roche; Schabes (eds.): *Finite-state language processing* 241–281. [= Roche (1996)]

Roche, Emmanuel; Schabes, Yves

- 1995 Deterministic part-of-speech tagging with finite-state transducers. *Computational Linguistics* 21/2, 227–253. <http://citeseer.ist.psu.edu/roche95deterministic.html>
- 1996 Introduction to finite-state devices in natural language processing. Mitsubishi Electrical Research Laboratory, technical report 96-13 (MERL-TR-96-13).
<http://www.merl.com/reports/docs/TR96-13.pdf>
- 1997a Deterministic part-of-speech tagging with finite-state transducers. In: Roche; Schabes (eds.): *Finite-state language processing* 205–239. [= Roche & Schabes (1995)]
- 1997b Introduction. In: Roche; Schabes (eds.): *Finite-state language processing* . [= Roche & Schabes (1996)]

Rodríguez, M. Andrea

- 2000 Assessing semantic similarity among spatial entity classes. PhD thesis, University of Maine. http://www.spatial.maine.edu/Publications/phd_thesis/Rodriguez2000.pdf

Rodríguez, M. Andrea; Egenhofer, Max J.

- 2003 Determining semantic similarity among entity classes from different ontologies. *IEEE Transactions on Knowledge and Data Engineering* 15/2, 442–456.
<http://www.inf.udec.cl/~andrea/papers/TKDE03.pdf>
- 2004 Comparing geospatial entity classes: an asymmetric and context-dependent similarity measure. *International Journal of Geographic Information Science* 18/3, 229–256.
<http://www.inf.udec.cl/~andrea/papers/IJGIS04.pdf>

- Rodríguez, M. Andrea; Egenhofer, Max J.; Rugg, R.
 1999 Assessing semantic similarity among geospatial feature class definitions. In: Vckovski, Andrej; Brassel, Kurt E.; Schek, Hans-Jörg (eds.): *Interoperating Geographic Information Systems—Second International Conference, INTEROP'99, Zurich, Switzerland*. Lecture Notes in Computer Science 1580, 1–16. Berlin: Springer.
- Roget, Peter Mark
 1852 *Thesaurus of English words and phrases*. London.
- Room, Adrian
 1996 *An alphabetical guide to the language of name studies*. Lanham (Maryland): Scarecrow.
- Ross, Háj
 1995 The grammar of space. Unpublished paper, English Department, University of North Texas. Denton (Texas).
- Rössler, Marc
 2002 Using Markov models for named entity recognition in German newspapers. *Proceedings of the ESSLLI'02 Workshop on Machine Learning Approaches in Computational Linguistics, Trento*. <http://cl.informatik.uni-duisburg.de/roessler/esslli02.pdf>
 2004 Corpus-based learning of lexical resources for German named entity recognition. *Proceedings of LREC 2004, Lisboa, Portugal*. <http://cl.informatik.uni-duisburg.de/roessler/lrec2004.pdf>
 2007 Korpus-adaptive Eigennamenerkennung. Dissertation, Computerlinguistik, Universität Duisburg-Essen. http://duepublico.uni-duisburg-essen.de/servlets/DerivateServlet/Derivate-16089/diss_final2007_DS.pdf
- Rössler, Marc; Morik, Katharina
 2005 Using unlabeled texts for Named-Entity Recognition. In: Rüping, Stefan; Scheffer, Tobias (eds.): *Proceedings of the ICML 2005 Workshop on Learning with Multiple Views, Bonn, Germany, August 11th, 2005* 59–64. <http://www-ai.cs.uni-dortmund.de/MULTIVIEW2005/MultipleViews.pdf>
- Roth, Jeannette
 2002 Der Stand der Kunst in der Eigennamen-Erkennung. Mit einem Fokus auf Produktenamen-Erkennung. Lizentiatsarbeit der Philosophischen Fakultät der Universität Zürich. <http://www.ifi.unizh.ch/cl/study/lizarbeiten/lizjeannetteroth.pdf>
- Russ, Charles V. J.
 1991 Internationale Einwohnernamen und ihre Bildung im Deutschen. *Beiträge zur Namenforschung* 26, 5–23.
- Sablayrolles, Pierre
 1995 The semantics of motion. <http://acl.ldc.upenn.edu/E/E95/E95-1040.pdf>, <http://arxiv.org/pdf/cmp-lg/9503007>
- Sætre, Rune
 2006 GeneTUC: natural language understanding in medical text. Doctoral thesis for the degree of doktor ingeniør, Department of Computer and Information Science Norwegian University of Science and Technology (NTNU). <http://www.idi.ntnu.no/~satre/genetuc/genetuc.pdf>
- SAILLABS-NER
 SAIL LABS Technology: Named Entity Detection. http://www.sail-technology.com/index.html?mms/ned_demo_frameset.html

- Saint-Dizier, Patrick (ed.)
 2006 *Syntax and semantics of prepositions*. Text, Speech and Language Technology 29. Dordrecht: Springer.
- Sampson, Geoffrey
 2001 *Empirical linguistics*. Open linguistics. London: Continuum.
- Sanderson, Mark; Kohler, Janet
 2004 Analyzing geographic queries. *SIGIR 2004*.
<http://www.geo.unizh.ch/~rsp/gir/abstracts/sanderson.pdf>
- Sang, Erik F. Tjong Kim; De Meulder, Fien
 2003 Introduction to the CoNLL-2003 shared task: language-independent named entity recognition. *Proceedings of the seventh conference on Natural language learning at HLT-NAACL 2003, Edmonton, Canada 4*, 142–147.
<http://ifarm.nl/erikt/papers/conll2003a.pdf>, <http://arxiv.org/abs/cs.CL/0306050>
- Santos, Diana; Chaves, Marcirio Silveira
 2006 The place of place in geographical IR. *3rd Workshop on Geographic Information Retrieval, SIGIR'06* 5–8.
- Sarda, Laure
 1999 Contribution à l'étude de la sémantique de l'espace et du temps : analyse des verbes de déplacement transitifs directs du français. Thèse de Doctorat, Université de Toulouse-Le Mirail. <http://tel.archives-ouvertes.fr/tel-00067804>
- Savary, Agata
 2000 Recensement et description des mots composés - méthodes et applications. Thèse de doctorat en Informatique Fondamentale, Université de Marne-la-Vallée.
http://www.blois.univ-tours.fr/~savary/Papers/Savary_These_2000.pdf.zip,
http://tln.li.univ-tours.fr/Tln_Biblio/2000TheseSavary.zip
 2005 Towards a formalism for the computational morphology of multi-word units. In: Vetulani, Zygmunt (ed.): *Human language technologies as a challenge for computer science and linguistics. Proceedings of the 2nd Language & Technology Conference: Poznań, Poland, April 21-23, 2005*. http://www.blois.univ-tours.fr/~savary/Papers/Savary_LTC_2005.zip, http://tln.li.univ-tours.fr/Tln_Biblio/2005ltsavary.zip
- Scharl, Arno; Tochtermann, Klaus (Eds.)
 2007 *The Geospatial Web*. How geobrowsers, social software and the Web 2.0 are shaping the network society. London: Springer.
- Schilder, Frank; Versley, Yannick; Habel, Christopher
 2004 Extracting spatial information: grounding, classifying and linking spatial expressions. *SIGIR 2004*. <http://www.geo.unizh.ch/~rsp/gir/abstracts/schilder.pdf> [Extended abstract]
- Schmidtke et al. [Schmidtke, Hedda; Tschander, Ladina B.; Eschenbach, Carola; Habel, Christopher]
 2003 Change of orientation. In: van der Zee & Slack (eds.): *Representing direction in language and space* 166–190.
- Schmid, Wolfgang P.
 1990 Gewässernamenwörterbücher. In: Hausmann et al. (eds.): *Wörterbücher* 1284–1291.
- Schnetzler, Pia
 2004 Evaluation von Named-Entity Recognition-Software für das cross-linguale Information Retrieval.
<http://www.isi2004.ch/fileadmin/ISI/PDF-Dateien/StudBeitraege/Schnetzler.pdf>

Schnorbusch, Daniel

- 1999 *Einfache deutsche Verben*. Eine syntaktische und semantische Beschreibung der verbalen Simplizia für das elektronische Lexikonsystem CISLEX. Studien zur Informations- und Sprachverarbeitung 1. München: Centrum für Informations- und Sprachverarbeitung. [Dissertation Ludwig-Maximilians-Universität München]

Schpak-Dolt, Nikolaus

- 1991 Zur Semantik von transitiven Verben der Fortbewegung. *Zeitschrift für Sprachwissenschaft* 10, 191–219.

Schröder, Jochen

- 1990 [¹1986] *Lexikon deutscher Präpositionen*. Leipzig: Enzyklopädie.
1992 *Lexikon deutscher Präfixverben*. Berlin, Leipzig: Langenscheidt, Enzyklopädie.
1993 *Lexikon deutscher Verben der Fortbewegung*. Leipzig: Langenscheidt, Enzyklopädie.

Schulte im Walde, Sabine

- 2003 Experiments on the automatic induction of German semantic verb classes. PhD Thesis, Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart. AIMS Report 9/2.
2004 Automatic induction of semantic classes for German verbs. In: Schnorbusch, Daniel; Langer, Stefan (eds.): *Semantik im Lexikon*. Tübingen: Gunter Narr.
<http://www.ims.uni-stuttgart.de/~schulte/Publications/Chapter/dgfs-03.pdf>
2006 Experiments on the automatic induction of German semantic verb classes. *Computational Linguistics* 32/2, 159–194.
<http://www.ims.uni-stuttgart.de/~schulte/Publications/Journal/cl-06-vclass.pdf>

Schumacher, Helmut

- 2006 Deutschsprachige Valenzwörterbücher. In: Ágel et al. (eds.): *Dependenz und Valenz* 1396–1424.

Schumacher et al. [Schumacher, Helmut; Kubczak, Jacqueline; Schmidt, Renate; Ruitter, Vera de]

- 2004 *VALBU – Valenzwörterbuch deutscher Verben*. Studien zur deutschen Sprache 31. Tübingen: Gunter Narr.

Schwarze, Christoph

- 1991 *Hier oben, schräg gegenüber*. Komplexe Ortsadverbiale im Deutschen und Französischen. Arbeitspapier Fachgruppe Sprachwissenschaft der Universität Konstanz 32. Konstanz.

Schweickard, Wolfgang

- 1995 Morphologie der Namen: Ableitungen auf der Basis von Eigennamen. In: Eichler et al. (eds.): *Namenforschung: ein internationales Handbuch zur Onomastik* 431–435.

Schweisthal, Klaus G.

- 1971 *Präpositionen in der maschinellen Sprachbearbeitung*. Schriftenreihe zur kommunikativen Grammatik 1. Bonn: Dümmler.

SDTS

Spatial Data Transfer Standard. <http://mcmcweb.er.usgs.gov/sdts/>

SEKINEEXTNER

Sekine, Satoshi: Sekine's Extended Named Entity Hierarchy.
<http://nlp.cs.nyu.edu/ene/>

Sekine, Satoshi

- 2004 Named Entity: history and future.
<http://cs.nyu.edu/~sekine/papers/NEsurvey200402.pdf>

Sekine, Satoshi; Marques-Ranchhod, Elisabete (eds.)

- 2007 *Named Entities: Recognition, classification, and use*. *Lingvisticæ Investigationes* 30/1. Amsterdam, Philadelphia: John Benjamins.

- Sekine, Satoshi; Nobata, Chikashi
 2004 Definition, dictionary and tagger for Extended Named Entities. *LREC-04*.
<http://cs.nyu.edu/~sekine/papers/lrec04-65.pdf>
- Sekine, Satoshi; Sudo, Kiyoshi; Nobata, Chikashi
 2002 Extended Named Entity hierarchy. *Proceedings of Third International Conference on Language Resources and Evaluation (LREC-2002), Las Palmas, Spain*.
<http://nlp.cs.nyu.edu/publication/papers/sekine-lrec02.pdf>, <https://nats-www.informatik.uni-hamburg.de/intern/proceedings/2002/LREC/pdf/120.pdf>,
<http://citeseer.ist.psu.edu/sekine02extended.html>
- SEMANTIC-MEDIAWIKI
 Semantic MediaWiki. http://meta.wikimedia.org/wiki/Semantic_MediaWiki
- Senellart, Jean
 1998a Locating noun phrases with finite state transducers. *Proceedings of the 17th International Conference on Computational Linguistics (COLING-98)* 1212–1219. Montréal. <http://citeseer.nj.nec.com/senellart98locating.html>
 1998b Tools for locating noun phrases with finite state transducers. *The computational treatment of nominals. Proceedings of the Workshop, COLING-ACL'98* 80–84.
<http://acl.lldc.upenn.edu/W/W98/W98-0611.pdf>
- Senellart, Jean; Dienes, Péter; Váradi, Tamás
 2001 New generation Systran translation system. *MT Summit VIII Santiago de Compostela (Spain), 18 - 22 September 2001*.
http://www.systransoft.com/Technology/mtsummitviii_systranNG.pdf
- Senellart et al. [Senellart, Jean; Plitt, Mirko; Bailly, Christophe; Cardoso, Françoise]
 2001 Resource alignment and implicit transfer. *MT Summit VIII, Santiago de Compostela (Spain), 18 - 22 September 2001*.
http://www.systransoft.com/Technology/mtsummitviii_autodesk.pdf
- Sester, Monika; Heinzle, Frauke
 2004 Suchmaschinen mit räumlichem Bewusstsein. *Symposium Praktische Kartographie, Königslutter, Germany*. <http://www.geo-spirit.org/publications/sesterheinzle.pdf>
- Seuren, Pieter A. M.
 1969 *Operators and nucleus*. A contribution to the theory of grammar. Cambridge: Cambridge University Press.
 2002 Pseudoarguments and pseudocomplements. In: Nevin (ed.): *The legacy of Zellig Harris* 179–207.
- Shay, Erin; Seibert, Uwe (eds.)
 2003 *Motion, direction and location in language*. in honor of Zygmunt Frajzyngier. Typological Studies in Language 56. Amsterdam: John Benjamins.
- Sigurd, Bengt; Zlatev, Jordan
 2006 Motion event typology meets computational modelling. *SKY Journal of Linguistics* 19.
http://www.ling.helsinki.fi/sky/julkaisut/SKY2006_1/1.6.8.%20SIGURD%20&%20ZLATEV.pdf
- Silberztein, Max D.
 1990 Le dictionnaire électronique des mots composés. In: Courtois, Blandine, Silberztein, Max D. (eds.): *Dictionnaires électroniques du français*. Langue Française 87, 71–83. Paris: Larousse.
 1993 *Dictionnaires électroniques et analyse automatique de textes : le système INTEX*. Paris: Masson.

- 1994a INTEX: a corpus processing system. *COLING* 579–583.
<http://acl.ldc.upenn.edu/C/C94/C94-1095.pdf>
- 1994b INTEX and the processing of natural languages.
<http://msh.univ-fcomte.fr/intex/downloads/Notes.pdf>
- 1998–1999 Les graphes *INTEX*. In: Fairon, Cédric (ed.): *Analyse lexicale et syntaxique: le système INTEX* 3–29.
- 2005 NooJ's dictionaries. *2nd Language & Technology Conference: Human Language Technologies as a Challenge for Computer Science and Linguistics, April 21-23, 2005, Poznań, Poland*. <http://perso.wanadoo.fr/rosavram/LTC2005.pdf>
- Silva et al. [Silva, Mário J.; Martins, Bruno; Chaves, Marcirio Silveira; Cardoso, Nuno]
- 2004 Adding geographic scopes to web resources. *SIGIR 2004*.
<http://www.geo.unizh.ch/~rsp/gir/abstracts/silvia.pdf>
- Simmler, Franz
- 1998 *Morphologie des Deutschen*. Flexions- und Wortbildungsmorphologie. Germanistische Lehrbuchsammlung 4. Berlin: Weidler.
- Smart, Philip D.; Abdelmoty, Alia I.; Jones, Christopher B.
- 2004 An evaluation of geo-ontology representation languages for supporting web retrieval of geographic information. *Proceedings of the GIS Research UK 12th Annual Conference, Norwich, UK* 175–178.
<http://www.geo-spirit.org/publications/psmart-gisruk04-final.pdf>
- Smith, David A.; Crane, Gregory
- 2001 Disambiguating geographic names in a historical digital library. *Proceedings of ECDL, Darmstadt, 4-9 September 2001* 127–136.
<http://www.perseus.tufts.edu/Articles/geod101.pdf>
- Smith, David A.; Mann, Gideon S.
- 2003 Bootstrapping toponym classifiers. *Proceedings of the HLT-NAACL 2003 Workshop on Analysis of Geographic References*. <http://acl.ldc.upenn.edu/W/W03/W03-0107.pdf>
- Sommerfeldt, Karl-Ernst; Schreiber, Herbert
- ²1977 [¹1974] *Wörterbuch zur Valenz und Distribution deutscher Adjektive*. Leipzig: Bibliographisches Institut. [³1983]
- ³1983 [¹1977] *Wörterbuch zur Valenz und Distribution der Substantive*. Leipzig: Bibliographisches Institut. [3. unveränderte Auflage]
- Sonntag, Michael
- 2007 Erkennung von Toponymen in deutschsprachigen Webseiten unter besonderer Berücksichtigung von NLP-Ansätzen – Toponym recognition in German texts with emphasis on NLP approaches. Masterarbeit im Wahlpflichtmodul Datenverwaltung, Universität Bamberg. [unveröffentlicht]
- Sørensen, Holger Steen
- 1958 *Word-classes in modern English with special reference to proper names with an introductory theory of grammar, meaning and reference*. Copenhagen.
- Southall, Humphrey
- 2003 Defining and identifying the roles of geographic references within text. *Workshop on the Analysis of Geographic References, NAACL'03*.
<http://www.metacarta.com/kornai/NAACL/WS9/Conf/ws912.pdf>
- Sowa, John F.
- 1993 Lexical structures and conceptual structures. In: Pustejovsky, James (ed.): *Semantics and the lexicon*. Dordrecht: Kluwer.

SPATIALML

SpatialML. <http://sourceforge.net/projects/spatialml>

Spencer, Andrew

- o.J. Paradigm-based morphosyntax: the German genitive.
<http://privatewww.essex.ac.uk/~spena/papers/German%20gen.pdf>

Spiess, Ernst [= StAGN]

- 1994 Ausgewählte Exonyme der deutschen Sprache.
<http://www.ifag.de/Kartographie/Stagn/Exonymenliste.pdf>

SPIRIT

SPIRIT: Spatially-Aware Information Retrieval on the Internet.
<http://www.geo-spirit.com>

Šrámek, Rudolf

- 1995 Eigennamen im Rahmen einer Kommunikations- und Handlungstheorie. In: Eichler et al. (eds.): *Namenforschung: ein internationales Handbuch zur Onomastik* 380–383.

StAGN [= Ständiger Ausschuss für geographische Namen (StAGN)]

- 1981 *Geographisches Namenbuch Bundesrepublik Deutschland*. Frankfurt a. M.: Institut für angewandte Geodäsie. [gazetteer Federal Republic of Germany]
- ²2002 Ausgewählte Exonyme der deutschen Sprache. Deutsche Namen und ihre phonetische Umschriftung für geographische Objekte in Ländern oder Gebieten ohne deutsche Amtssprache. http://www.ifag.de/Kartographie/Stagn/Exonyme/f_Exonyme.htm,
http://webserver.bkg.bund.de/kartographie/stagn/Exonyme/f_Exonyme.htm,
http://webserver.bkg.bund.de/kartographie/stagn/020809%20EXOLISTE_hoch_RH_JS.pdf
- ⁸2004 *Liste der Staatennamen und ihrer Ableitungen im Deutschen*. Mit Anhang: Liste der Namen ausgewählter nichtselbständiger Gebiete.
<http://www.ifag.de/Kartographie/Stagn/Staatennamen.htm>,
<http://webserver.bkg.bund.de/kartographie/stagn/Staatennamen.htm>

STAGN:EXON-EMPF

StAGN [= Ständiger Ausschuss für geographische Namen] o.J.: Stellungnahme des StAGN zum Gebrauch von Exonymen.
http://www.ifag.de/Kartographie/Stagn/gebrauch_exonyme.htm,
http://webserver.bkg.bund.de/kartographie/stagn/gebrauch_exonyme.htm

Stechow, Arnim von; Dieter Wunderlich (eds.)

- 1991 *Semantik*. Handbücher zur Sprach- und Kommunikationswissenschaft 6. Berlin, New York: Walter de Gruyter.

Steinberger, Ralf; Pouliquen, Bruno

- 2007 Cross-lingual named entity recognition. In: Sekine & Marques-Ranchhod (eds.): *Named Entities: Recognition, classification, and use* 135–162.

Storrer, Angelika

- 2003 Ergänzungen und Angaben. In: Ágel et al. (eds.): *Dependenz und Valenz* 764–780.

Sturm, Afra

- 2005 Eigennamen als kontextabhängige und inhärent definite Ausdrücke. *Zeitschrift für Sprachwissenschaft* 24/1, 67–91.

Suchanek, Fabian M.; Ifrim, Georgiana; Weikum, Gerhard

- 2006a Combining linguistic and statistical analysis to extract relations from web documents.
<http://www.mpi-inf.mpg.de/~suchanek/publications/kdd2006.pdf>,
[http://domino.mpi-inf.mpg.de/internet/reports.nsf/c125634c000710cd80255ef200387b6e/3594aaf30f90fb41c125714800537d81/\\$FILE/MPI-I-2006-5-004.pdf](http://domino.mpi-inf.mpg.de/internet/reports.nsf/c125634c000710cd80255ef200387b6e/3594aaf30f90fb41c125714800537d81/$FILE/MPI-I-2006-5-004.pdf)

- 2006b LEILA: Learning to extract information by linguistic analysis.
<http://www.mpi-inf.mpg.de/~suchanek/publications/olp2006.pdf>
- Sugarewa, Tekla
 1974 Adjektivderivate zu Eigennamen und ihre Konkurrenz mit Substantivkomposita und syntaktischen Wortverbindungen. *Beiträge zur Geschichte der deutschen Sprache und Literatur* 94, 199–256.
- Summerell, Orrin F.
 1995 Philosophy of proper names. In: Eichler et al. (eds.): *Namenforschung: ein internationales Handbuch zur Onomastik* 368–372.
- Sundheim, Beth M.
 1995 Overview of results of the MUC-6 evaluation. *Sixth Message Understanding Conference (MUC-6): Proceedings of a Conference Held in Columbia, Maryland, November 6-8, 1995* 13–31. <http://acl.ldc.upenn.edu/M/M95/M95-1002.pdf> [= Sundheim (1996)]
 1996 Overview of results of the MUC-6 evaluation. *Annual Meeting of the ACL. Proceedings of a workshop on held at Vienna, Virginia: May 6-8, 1996* 423–442.
<http://acl.ldc.upenn.edu/X/X96/X96-1048.pdf> [= Sundheim (1995), anderer Textsatz]
- SYSTRAN
 SYSTRAN Language Translation Technology. <http://www.systransoft.com>
- Talmy, Leonard
 1996 Fictive motion in language and “ception”. In: Bloom et al. (eds.): *Language and space* 211–276.
- Tarvainen, Kalevi
 1981 *Einführung in die Dependenzgrammatik*. Tübingen: Niemeyer. [²2000 (unverändert)]
- Taylor, Holly A.; Tversky, Barbara
 1992 Spatial mental models derived from survey and route descriptions. *Journal of Memory and Language* 31, 261–292.
- TEI
 TEI: Text Encoding Initiative. <http://www.tei-c.org>
- TEI:P4
 Text Encoding Initiative 2004: *The XML version of the TEI guidelines. Names and dates, place names*. <http://www.tei-c.org/P4X/ND.html>
- TEI:P5
 TEI Consortium (eds.) 2007: *TEI P5: Guidelines for Electronic Text Encoding and Interchange. 13 – Names, Dates, People, and Places*.
<http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ND.html>
- Telljohann et al. [Telljohann, Heike; Hinrichs, Erhard W.; Kübler, Sandra; Zinsmeister, Heike]
 2006 Stylebook for the Tübingen Treebank of Written German (TüBa-D/Z).
<http://www.sfs.uni-tuebingen.de/resources/sty.pdf>
- Tesnière, Lucien
 1959 *Éléments de syntaxe structurale*. Paris: Klincksieck.
²1966 *Éléments de syntaxe structurale*. Paris: Klincksieck.
 1980 *Grundzüge der strukturalen Syntax*. Herausgegeben und übersetzt von Ulrich Engel. Stuttgart: Klett-Cotta. [= Übersetzung von Tesnière (1966) *Éléments de syntaxe structurale*. Paris: Klincksieck]

- Thielen, Christine
 1995 An approach to proper name tagging in German. *EACLSIGDAT-95*.
<http://citeseer.ist.psu.edu/thielen95approach.html>
- Tichelaar, Tjeerd
 2003a Toponymy and language. In: Ormeling; Stabe; Sievers (ed.): *Training course on toponymy* 47–57.
 2003b Field collection of names in multi-lingual and multi-cultural areas. In: Ormeling; Stabe; Sievers (ed.): *Training course on toponymy* 79–83.
- TIGER-CORPUS
 IMS, Uni Stuttgart: The TIGER Project.
<http://www.ims.uni-stuttgart.de/projekte/TIGER/TIGERCorpus/>
- TIPSTER
 TIPSTER Gazetteer. <http://crl.nmsu.edu/cgi-bin/Tools/CLR/clrinfo?GAZETEER>
- TKSR
 Mel'čuk, Igor Aleksandrovič; Žolkovskij, Aleksandr K. (eds.) 1984: *Tolkovo-kombinatornyj slovar' sovremennogo russkogo jazyka*. Opyty semantiko-sintaktičeskogo opisanija russkoj leksiki. Wiener Slawistischer Almanach Sonderband 14. Wien.
- TnT-TAGGER
 Brants, Thorsten: TnT – Statistical Part-of-Speech Tagging.
<http://www.coli.uni-saarland.de/~thorsten/tnt/>
- Tognini-Bonelli, Elena
 2001 *Corpus linguistics at work*. Amsterdam, Philadelphia: John Benjamins.
- Tomai, Eleni; Kavouras, Marinos
 2005 Qualitative linguistic terms and geographic concepts—quantifiers in definitions. *Transactions in GIS* 9/3, 277–290.
<http://www.iacm.forth.gr/regional/papers/transactioningis.pdf>
- Toral, Antonio; Muñoz, Rafael
 2006 A proposal to automatically build and maintain gazetteers for Named Entity Recognition by using Wikipedia. *11th Conference of the European Chapter of the Association for Computational Linguistics. Proceedings of the workshop on NEW TEXT Wikis and blogs and other dynamic text sources, April, 4, 2006 Trento, Italy* 56–61.
http://www.sics.se/jussi/newtext/working_notes/09_toral_munoz.pdf,
http://www.sics.se/jussi/newtext/working_notes/complete.pdf
- Toral, Antonio; Muñoz, Rafael; Montoyo, Andrés
 2005 Weak Named Entities Recognition using morphology and syntax. *Proceedings of the 5th International Conference on Recent Advances in Natural Language Processing (RANLP). Borovets (Bulgaria). September 2005*.
http://www.dlsi.ua.es/~atoral/publications/2005_ranlp_weakner_poster.pdf
- Traboulsi, Hayssam N.
 2004 A local grammar for proper names. MPhil to PhD transfer, University of Surrey.
<http://portal.surrey.ac.uk/pls/portal/url/ITEM/F3E6A8EF87037385E0340003BA296BDE>,
http://www.cs.surrey.ac.uk/pgr/tell_me/downloads/2004 TRABOULSI_TRANSFER.pdf
- Tran, Mickaël; Grass, Thierry; Maurel, Denis
 2004 An ontology for multilingual treatment of proper names. *Ontologies and Lexical Resources in Distributed Environments (OntoLex 2004), in Association with LREC2004, Lisboa, Portugal, 29 may* 75–78.
http://tln.li.univ-tours.fr/Tln_Biblio/2004ontolex.zip

Treig, Thomas

- 1977 Complétives en allemand. Classification. Rapport de Recherches du LADL 7, 39–203. Paris: Université Paris 7, LADL.

Trujillo, Arturo

- 1992 Spatial lexicalization in the translation of prepositional phrases. *Meeting of the Association for Computational Linguistics* 306–308.
<http://citeseer.nj.nec.com/article/trujillo92spatial.html>
- 1995a *Lexicalist machine translation of spatial prepositions*. PhD dissertation, University of Cambridge. <http://citeseer.nj.nec.com/trujillo95lexicalist.html>
- 1995b Towards a cross-linguistically valid classification of spatial prepositions. *Machine Translation* 10/1-2, 93–141.

Tschander, Ladina B.

- 1999 Bewegung und Bewegungsverben. In: Wachsmuth, I.; Jung, B. (eds.): *KogWis99: Proceedings der 4. Fachtagung der Gesellschaft für Kognitionswissenschaft, Bielefeld, 28. September - 1. Oktober 1999* 25–30. Sankt Augustin.
<http://bieson.ub.uni-bielefeld.de/volltexte/2003/388/html/tschander.pdf>

Tsuruga, Yoichiro

- 2005 A correspondance between N0-V-N1-*de*-N2 and N0-V-N2-*Loc*-N1 in French: The case of *planter*. In: Takagaki, Toshihiro; Zaima, Susumu; Tsuruga, Yoichiro; Moreno-Fernández, Francisco; Kawaguchi, Yuji (eds.): *Corpus-Based Approaches to Sentence Structures*. Usage-based linguistic informatics 2, 213–232. Amsterdam: John Benjamins.

Tversky, Barbara

- 1996 Spatial perspective in descriptions. In: Bloom et al. (eds.): *Language and space* 463–492.
- 2003 Places: Points, planes, paths, and portions. In: van der Zee & Slack (eds.): *Representing direction in language and space* 132–143.

Tversky, Barbara; Taylor, Holly A.

- 1998 Acquiring spatial and temporal knowledge from language. In: Egenhofer; Golledge (eds.): *Spatial and temporal reasoning in geographic information systems* 155–166.

Ulrich, Stefan

- 2002 *Syntax und Semantik prädikativer Nomen*. Studien zur Informations- und Sprachverarbeitung 5. München: Centrum für Informations- und Sprachverarbeitung.

UNGEGN

- United Nations Group of Experts on Geographical Names (UNGEGN).
<http://unstats.un.org/unsd/geoinfo/ungegn.htm>
- 1996 Towards categories of feature types for the concise gazetteer of Canada. Technical Papers of the 18th Session of the UNGEGN Working Paper 33.
<http://unstats.un.org/unsd/geoinfo/18th-GEGN-Docs/working-paper33.pdf>
- 2004 U.N. Work on Geographical Names. Technical Papers of the 22nd Session of the UNGEGN Working Paper 75. <http://unstats.un.org/unsd/geoinfo/gegn22wp75.pdf>
- 2006 *Manual for the national standardization of geographical names*. New York: United Nations. <http://unstats.un.org/unsd/geoinfo/geog%20names%20final.pdf>

UNITEX

Unitex – Corpus Processor. <http://www-igm.univ-mlv.fr/~unitex/>

UN/LOCODE

United Nations Economic Commission for Europe (UNECE): United Nations code for trade and transport locations (UN/LOCODE).
<http://www.unece.org/etrades/download/downloadindex.htm>

Uryupina, Olga [= Ourioupina, Olga]

- 2002 Extracting geographical knowledge from the internet. *Proceedings of the ICDM-AM International Workshop on Active Mining 2002*.
http://www.coli.uni-saarland.de/~ourioupi/our_i_fin.pdf
- 2003 Semi-supervised learning of geographical gazetteers from the internet. *Workshop on the Analysis of Geographic References, NAACL'03*.
<http://www.metacarta.com/kornai/NAACL/WS9/Conf/ws913.pdf>

USGS:GNIS

U.S. Department of the Interior: U.S. Geological Survey (USGS); U.S. Board on Geographic Names (BGN): Geographic Names Information System (GNIS).
<http://geonames.usgs.gov/index.html>

Vandeloise, Claude

- 1985a Description of space in French. San Diego, University of California, Diss., 1984. Ann Arbor (Michigan): Univ. Microfilms Internat.
- 1985b Au-delà des descriptions géométriques et logiques de l'espace: une description fonctionnelle. *Linguisticæ Investigaciones* 9/1, 109–129. Amsterdam, Philadelphia: John Benjamins.
- 1986 *L'espace en français*. Sémantique des prépositions spatiales. Paris: Seuil.
- 1991 *Spatial prepositions: a case study from French*. Chicago: University of Chicago Press.

van der Zee, Emile; Slack, John (eds.)

- 2003 *Representing direction in language and space*. Explorations in language and space. Oxford: Oxford University Press.

Vasmer, Max et al.

1961–1969 *Wörterbuch der russischen Gewässernamen*. Wiesbaden.

Vater, Heinz

2005 *Referenz-Linguistik*. UTB Uni-Taschenbücher 2685. München: Wilhelm Fink.

Velde, Marc van de

1995 Lokal- und Direktionalerweiterungen im Deutschen und im Niederländischen. In: Eichinger, Ludwig M.; Eroms, Hans-Werner (eds.): *Dependenz und Valenz*. Beiträge zur germanistischen Sprachwissenschaft 10, 319–335. Hamburg: Buske.

Vendler, Zeno

²1968 [1967] Verbs and times. *Linguistics in philosophy* 97–121. Ithaca (New York). [≈1957: Verbs and times. *The Philosophical Review* 66, 143–160.]

Vieu, Laure

1991 Sémantique des relations spatiales et inférences spatio-temporelles: Une contribution à l'étude des structures formelles de l'espace en langage naturel. Ph.D. thesis, Université Paul Sabatier.

Volk, Martin

- 2002 The automatic resolution of prepositional phrase-attachment ambiguities in German. Habilitation, Institut für Informatik, University of Zurich.
- 2006 German preposition and their kin. In: Saint-Dizier (ed.): *Syntax and semantics of prepositions* 83–99. <http://kracht.humnet.ucla.edu/marcus/html/direction.pdf>

Volk, Martin; Clematide, Simon

2001 Learn - Filter - Apply - Forget. *6th International Workshop on application of Natural Language for Information Systems (NLDB)*, Madrid, Spain. Mixed Approaches to Named Entity Recognition. http://www.ifi.unizh.ch/cl/volk/papers/Madrid_2001.pdf

Vorweg, Constanze; Rickheit, Gert

- 2000 Repräsentation und sprachliche Enkodierung räumlicher Relationen. In: Habel; Stutterheim (eds.): *Räumliche Konzepte und sprachliche Strukturen* 9–44.

WACkY

The WaCky Project. <http://wacky.sslmit.unibo.it/>

Walther, Hans

- 2003 *Namenkunde und geschichtliche Landeskunde*. Ein einführender Überblick, Erläuterungen namenkundlicher Fachbegriffe, Auswahlbiographie zur Namenkunde und Landeskunde Ostmitteleuropas. Leipzig: Leipziger Universitätsverlag.

Wanner, Leo (ed.)

- 1996 *Lexical functions in lexicography and natural language processing*. Studies in Language Companion Series (SLCS) 31. Amsterdam, Philadelphia: John Benjamins.

Watrin, Patrick

- 2003 Information extraction and lexicon-grammar. In: de Vries, Arjen P. (ed.): *Proceedings of the Dutch-Belgian Information Retrieval Workshop (DIR)* 16–21. Amsterdam.

<http://ilps.science.uva.nl/DIR/DIR2003proceedings.pdf#page=24>

- 2006 Une approche hybride de l'extraction d'information: sous-langages et lexique-grammaire. PhD thesis, Faculté de Philosophie et Lettres, Université Catholique de Louvain.

<http://patlab.org/~pat/watrin-phd.pdf>, <http://patlab.org/~pat/watrin-phd.ps.gz>

WBIMMOBILIEN

Pfeiffer, Elmar: Wörterbuch (DE - EN): Immobilienwirtschaft.

<http://www.stalys.de/data/wa.htm>

Weigel et al. [Weigel, Felix; Schulz, Klaus U.; Brunner, Levin; Torres-Schumann, Eduardo]

- 2006 Integrated document browsing and data acquisition for building large ontologies. *Proceedings of 10th International Conference on Knowledge-Based & Intelligent Information & Engineering Systems (KES2006), Bournemouth, UK, 9th - 11th October 2006*. Lecture notes in computer science 4253, 614–622.

<http://www.cis.uni-muenchen.de/~weigel/Literatur/weigel06integrated.pdf>

Weisgerber, Matthias

- 2005 Decomposing path shapes. *Sinn und Bedeutung 10, 10th annual meeting of the Gesellschaft für Semantik, 13 – 15 October 2005, Berlin*. http://www.zas.gwz-berlin.de/events/sub/abstracts/Matthias_Weisgerber_Decomposing_Path_Shapes.pdf

Wellmann, Hans

- 1975 *Deutsche Wortbildung: Das Substantiv*. Sprache der Gegenwart 32. Düsseldorf: Schwann.

WIKIPEDIA

Wikipedia. <http://wikipedia.org/>

WIKIPEDIA.DE

Wikipedia. <http://de.wikipedia.org/>

WIKITRAVEL.ORG

Wikitravel – der freie weltweite Reiseführer. <http://wikitravel.org>

WIKIVOYAGE.ORG

Wikivoyage. <http://www.wikivoyage.org>

Wimmer, Rainer

- 1995 Eigennamen im Rahmen einer allgemeinen Sprach- und Zeichentheorie. In: Eichler et al. (eds.): *Namenforschung: ein internationales Handbuch zur Onomastik* 372–379.

Winograd, Terry

1983 *Language as a cognitive process 1* (Syntax). Reading (Massachusetts): Addison-Wesley.

Woodman, Paul

2003 Exonyms: a structural classification and a fresh approach. *United Nations Group of Experts on Geographical Names, Working Group on Exonyms, Prague, September 24-26, 2003*. http://www.zrc-sazu.si/wge/Documents/Papers%20Prague/Woodman_Exonyms%203.pdf

WORDNET

Cognitive Science Laboratory, Princeton University: WordNet – a lexical database for the English language. <http://wordnet.princeton.edu/>

WORLD-GAZETTEER

Helders, Stefan: World Gazetteer. <http://www.world-gazetteer.com/>,
<http://bevoelkerungsstatistik.de/>

Wunderlich, Dieter

1982a Sprache und Raum I. *Studium Linguistik* 12, 1–19.

1982b Sprache und Raum II. *Studium Linguistik* 13, 37–59.

1983 On the compositionality of German prefix verbs. In: Bäerle, R.; Schwarze, Christoph; Stechow, Armin von (eds.): *Meaning, use, and interpretation of language* 452–465. Berlin: Walter de Gruyter.

1985a Raumkonzepte: Zur Semantik der lokalen Präpositionen. In: Ballmer, Thomas T.; Posner, Roland (ed.): *Nach-Chomskysche Linguistik*. Neuere Arbeiten von Berliner Linguisten 340–351. Berlin: de Gruyter.

1985b Raum, Zeit und das Lexikon. In: Schweizer, Harro (ed.): *Sprache und Raum* 66–89. Stuttgart.

1990 Ort und Ortswechsel. *Zeitschrift für Literaturwissenschaft und Linguistik* 20/78, 43–58.

1991 How do prepositional phrases fit into compositional syntax and semantics? *Linguistics* 29, 591–621.

1993 On German *um*: Semantic and conceptual aspects. *Linguistics* 31, 111–133.

Wunderlich, Dieter; Herweg, Michael

1991 Lokale und Direktionale. In: Stechow & Wunderlich (eds.): *Semantik* 758–785.

Wunderlich, Dieter; Kaufmann, Ingrid

1990 Lokale Verben und Präpositionen – semantische und konzeptuelle Aspekte. In: Felix, Sascha W.; Kanngießer, Siegfried; Rickheit, Gerd (eds.): *Sprache und Wissen*. Studien zur kognitiven Linguistik. Psycholinguistische Studien 223–252. Opladen: Westdeutscher Verlag.

Yli-Jyrä, Anssi Mikael

2005 Contributions to the Theory of Finite-State Based Linguistic Grammars. Ph.D. dissertation. Publications of the Department of General Linguistics, University of Helsinki 38. <http://ethesis.helsinki.fi/julkaisut/hum/yleis/vk/yli-jyra/contribu.pdf>,
<http://www.ling.helsinki.fi/~aylijyra/dissertation/contribu.pdf>

Zaenen, Annie

2006 Mark-up barking up the wrong tree. *Computational Linguistics* 32/4, 577–580.

Zelinsky, Wilbur

2002 Slouching toward a theory of names: a tentative taxonomic fix. *Names: a journal of onomastics* 50/4, 243–262. New York.

- Zhan, F. Benjamin
 1998 Approximate analysis of binary topological relations between geographic regions with indeterminate boundaries. *Soft Computing* 2, 28–34. Springer-Verlag.
http://uweb.txstate.edu/~fz01/Reprints/Zhan_1998_SC.PDF
- Zifonun, Gisela
 2006 Grundlagen der Valenz. In: Ágel et al. (eds.): *Dependenz und Valenz* 352–377.
- Zifonun, Gisela; Hoffmann, Ludger; Strecker, Bruno
 1997 *Grammatik der deutschen Sprache*. Schriften des Instituts für Deutsche Sprache 7. Berlin, New York: Walter de Gruyter.
- Zikmund, Hans (ed.) [= Ständiger Ausschuss für geographische Namen (StAGN)]
 2000 *Duden Wörterbuch geographischer Namen des Baltikums und der Gemeinschaft Unabhängiger Staaten (GUS)*. Mannheim: Duden.
- Žolkovskij, Aleksandr K.; Mel’čuk, Igor’ Aleksandrovič
 1965 O vozmožnom metode i instrumentach semantičeskogo sinteza. *Naučno-texničeskaja informacija* 5, 23–28.
- Zwarts, Joost
 1995 Lexical and functional direction. In: den Dikken, M.; Hengeveld, K. (eds.): *Linguistics in the Netherlands* 227–238. Amsterdam: John Benjamins.
 1997 Vectors as relative positions: A compositional semantics of modified PPs. *Journal of Semantics* 14, 57–86.
 2003 Vectors across spatial domains: From place to size, orientation, shape, and parts. In: van der Zee & Slack (eds.): *Representing direction in language and space* 39–68.
 2004 The case of prepositions: Government and compositionality in German PPs.
<http://www.let.uu.nl/users/Joost.Zwarts/personal/Papers/CofPs.pdf>
 2005 Prepositional aspect and the algebra of paths. *Linguistics and Philosophy* 28/6, 739–779.
- Zwarts, Joost; Winter, Yoad
 2000 Vector space semantics: a model-theoretic analysis of locative prepositions. *Journal of Logic, Language and Information* 9/2, 169–211.

Tabellenverzeichnis

2.1	Eigenschaften intrinsischer, absoluter und relativer Relatoren	23
2.2	Verbklassen mit lokativen Komplementen in der französischen Lexikongrammatik	49
2.3	Transitive lokative Verben, syntaktische Einteilung	51
2.4	Satzschemata lokativer Konstruktionen	51
2.5	Realisierungsmöglichkeiten von Source und Goal bei franz. Verben	52
2.7	Transitive deutsche Verben mit Direktionalergänzung (Caroli 1984b)	57
2.9	Klassen von Bewegungsverbren aus Eschenbach et al. (2000)	67
3.2	Kontextuelle syntaktische Klassifikation von Toponymen	105
3.3	Genus von Toponymen, Verteilung auf Toponymtypen	110
3.4	Artikelgebrauch bei Toponymen, Verteilung auf Toponymtypen	112
3.5	Erweiterungen der CISLEX-Flexionsklassen für Lokativa	115
3.6	Relationen zwischen Lokationen in ONTOWORLD.ORG	137
4.1	Topolog. Felder und Varianten des Satzes <i>Hans wohnt für ein Jahr in Dortmund</i>	155
4.2	Satzschemata lokativer Sätze, Überblick	173
5.1	GEO-Stopwörter (englisch, franz., deutsch)	178
5.2	Richtlinien zur Annotierung von Named Entities, Detailfragen	193
5.3	Performance des NER-Systems von Kühnlein (2003)	201
5.4	Performance des Systems zur Toponymerkennung von Sonntag (2007)	203
5.5	Performance von Systemen zur Toponymerkennung in deutschen Texten	203
5.6	Toponymerkennung (NER) mit lokalen Grammatiken, Evaluation	212
5.7	Homonymie und Allonymie von Toponymen im GETTY-Thesaurus	216
5.8	Ergebnisse Geokodierung (Auswahl)	218
A.2	Toponymtypen: onomastische Terminologie und Kennzeichnung im Lexikon	301
A.3	Suffixoperationen zur Bildung von Toponymderivatonen	317
C.1	Lexikongrammatik: Präpositionen, kombinatorische und semantische Eigenschaften	363
C.2	Lexikongrammatik: Präpositionen, Formen und morphosyntaktische Eigenschaften	375
C.3	Lexikongrammatik: Bewegungsverbren	395
C.4	Lexikongrammatik: statische lokative Verben „Mensch – Ort“	409
C.5	Lexikongrammatik: Relationen zwischen geographischen Entitäten (Verben)	411
C.6	Lexikongrammatik: Nomina Typ «Hauptstadt» (Geo-Relationen)	415
C.7	Lexikongrammatik: Adverbien und Adjektive $N_0 V_{sup} Loc N_1 loc Adv$	419
E.1	Bewertung des CONLL-2003-SHARED-TASK-Korpus	437
E.2	Größe der verwendeten Korpora	441

Abbildungsverzeichnis

1.1	Englische Zeitadverbien „rounded dates“, lokale Grammatik	4
1.2	Übersetzung von Zeitadverbien durch Transduktor (lokale Grammatik)	6
1.3	Englische Fachtermini CAD-Software, lokale Grammatik	7
1.4	Französische Fachtermini CAD-Software, lokale Grammatik	8
1.5	Textautomat des Satzes <i>Le pilote ferme la porte</i>	9
2.1	Relatives räumliches Bezugssystem	22
2.2	Eigenschaften intrinsischer räumlicher Relationen	22
2.3	Repräsentation des Ausdrucks <i>hoch im Norden</i> als lokale Grammatik und in tiefen-syntaktischer Darstellung mit lexikalischen Funktionen	30
2.4	Sätze des Typs <i>Nhum V Prép Nloc</i>	60
2.5	<i>Enklave/Exklave</i> , schematische Darstellung (ternäre Relation)	75
3.1	Lokativa: übergeordnete Klassen (Le Pesant)	82
3.2	GETTY-Thesaurus, Eintrag <i>Moskau</i>	128
3.3	Relationen von ‚Bayern‘ in der Ontologie	130
3.4	Geographische Hierarchie bei WIKIVOYAGE.ORG	137
3.5	<i>Sankt Petersburg</i> , historischen Namen in der Ontologie	138
3.6	Screenshot des Ontologiebrowsers PROTÉGÉ	139
4.1	<i>Hans wohnt für ein Jahr in Dortmund</i> , lokale Grammatik/Textautomat	154
5.1	System Toponymerkennung, Sonntag (2007)	202
5.2	Toponymerkennung (NER) mit lokalen Grammatiken, Module und Verarbeitungspipeline	206
5.3	Toponymerkennung (NER) mit lokalen Grammatiken, Lexika und Grammatiken	207
5.4	Vorkommen eines Toponyms mit Klassenbezeichner in Abhängigkeit zur Frequenz (Häufigkeitsklasse)	209
5.5	Toponymerkennung (NER) mit lokalen Grammatiken, Screenshot	213
A.1	OSNOVENSISTEM1983: Toponymklassen	333
A.2	Kamianets (2000): Klassifikation der Ortsnamen	334
A.3	Walther (2003): Klassifikation der Eigennamen	334
A.4	PROLEX: Taxonomiebaum Toponyme	335
A.5	Langer (1996): Taxonomie der Ortsbezeichnungen	336
A.6	WORDNET: Knoten COUNTY und COUNTRY	337
A.7	Taxonomie für deutsches Toponymlexikon/Ontologie	338
A.8	Definition der Klassen für eine Lexikon (Ontologie) deutscher Toponyme	339
A.9	DAML:GEOFILE: Taxonomiebaum	341
A.10	NASA:SWEET: Top-Hierarchie	342
A.11	NASA:SWEET: Auszug aus dem Taxonomiebaum	343
A.12	GEOFEATURES.OWL: Taxonomie toponymischer Objekte	344

A.13	CORINE: Taxonomie	345
A.14	SEKINEEXTNER: Überblick	346
A.15	SEKINEEXTNER: Location	347
A.16	SEKINEEXTNER: Facilities	348
A.17	ADL: Taxonomie, Teil 1	349
A.18	ADL: Taxonomie, Teil 2	350
A.19	BBNTAXON: Auszug aus der Taxonomie	351
A.20	KIM: Taxonomie	352
B.1	Phrasen <i>in der Nähe</i> , lokale Grammatik	355
B.2	Adverbiale mit der Bedeutung ‚nahe‘, lokale Grammatik	355
B.3	Phrasen des Typs <i>hoch im Norden</i> , lokale Grammatik	356
B.4	Adverbiale ‚in der Mitte‘, lokale Grammatik	356
B.5	Präpositionaladverbien mit Modifikatoren, lokale Grammatik	356
B.6	Höhenangaben, lokale Grammatik	357
B.7	Höhenangaben II (‚über dem Meeresspiegel‘), lokale Grammatik	357
B.8	Abhängigkeiten lokaler Grammatiken	357
B.9	Erkennung unbekannter Toponyme, lokale Grammatik	358
B.10	Erkennung von Hodonymen, lokale Grammatik	359
B.11	Sportvereine, lokale Grammatik	360

Index

- absolutes Bezugssystem 21, 22
Adjektivableitung *s.* Toponym, Derivation
Adressen 149
adverbiale Einschübe *s.* Adverbialsupplement
Adverbialsupplement 154
Aktant *s.* Komplement
 Bestimmung von 26, 32–35, 47
Allonym 94, 127
Angabe *s.* Supplement
Apposition 73, 92, 99, 100, 110, 122, 143, 144
appropriate noun 15, 69, 76, **85**
- Bewegungsverben *s.* lokative Verben
Bewohner *s.* Toponym, Derivation
Bezugssystem 21
- Chunk-Parsing 7
complément approprié *s.* appropriate noun
- Deixis 23, 24, 36, 65, 148
Deonomastika *s.* Toponym, Derivation
deonymisch *s.* Toponym, Derivation
Dependenz
 Dependenzgrammatik 12, 26, 64
 Harris 12
 Tesnière 12
destination *s.* Pfad, Zielpunkt
+Det *s.* Toponym, Artikelgebrauch
+DetZ *s.* Toponym, Artikelgebrauch
direktionale Präpositionen 42
disambiguation of geogr. names *s.* Geokodierung
- Eigenname
 Definition 94
 Eigenschaften 94–97
 geographischer *s.* Toponym
 vs. Gattungsname 97
Einwohner *s.* Toponym, Derivation
embedded named entities *s.* nested named entities
Endonym 100–102
Entitäten
 eingebettete/verschachtelte *s.* nested named entities
 Granularität der Klassifikation 189
Ergänzung *s.* Komplement
Ethnika 120
Evaluation
 NER 186–188
Exonym *s.* Toponym, Exonym
- faceted browsing *s.* faceted search
faceted search 222
fictive motion 66–67
finite-state
 automaton *s.* lokale Grammatik, endlicher Automat
 cascade *s.* lok. Gramm., Kaskadierung
 transducer *s.* lok. Gramm., Transduktor
Flexion 8, 101
F-measure *s.* F-Wert
frame of reference *s.* Bezugssystem
f-score *s.* F-Wert
Fugenform
 eines Toponyms *s.* Toponym, Fugenform
fuzzy regions *s.* unscharfe Regionen
F-Wert 186
- Gattungsname vs. Eigenname 97
Gazetteer 103
Gazetteer-Learning 219
gentilés *s.* Toponym, Wohnernamen
Genus 110
 unspezifiziert 110
 Wechsel 101
geo-coding *s.* Geokodierung
Geographical Markup 197
Geographic Information Retrieval 175
geographic scope 222, 223
Geographischer (Eigen)Name *s.* Toponym
geographisches Objekt 93
Geokodierung 93, 175, **215–218**, 219
geo-parsing *s.* Named Entity Recognition
georeferencing *s.* Geokodierung
GeoTagging *s.* Geokodierung
GIR *s.* Geographic Information Retrieval

- goal *s.* Pfad, Zielpunkt
grounding *s.* Geokodierung
Grundform *s.* Toponym, Grundform
- Homonymie
von Toponymen *s.* Toponym, Homonymie
- IE *s.* Informationsextraktion
imprecise regions *s.* unscharfe Regionen
Information Retrieval 175
Informationsextraktion 7, 175
Inserts *s.* Adverbialsupplement
intrinsisches Bezugssystem 22
IR *s.* Information Retrieval
- Komplement
angemessenes *s.* appropriate noun
lokatives 35–36
vs. Supplement 26, 32–35, 47
- Kompositum
mit Toponym *s.* Toponym, Fugenform
- Kontextfreiheit natürlicher Sprachen 18
Koordinatensystem *s.* Bezugssystem
Koordination 149
- LAF *s.* Learn-Apply-Forget
Learn-Apply-Forget 181, 185
Lemma *s.* Lexem
Lexem 109, 127
Grundform 109, 300
lexical lacunae 178
lexikalische Funktionen *s.* Meaning Text Th.
Lexikon
Klassenbezeichner von Toponymen 92
Ontologie *s.* Toponym, Ontologie
Toponyme 102–127, 300–317
- Lexikogrammatik **11–17**, 138
Adverbiale 61–62
Adverbien 170–172, 418–421
Instantiierung als lokale Grammatik 17, 358
lokative Prädikate 45–58, 61–62, 361–421
Nomina 167–170, 414–418
Präpositionen 61–62, 157–158, 361–394
Verben 45–58, 158–167, 394–414
- linguistische Situation 71, 73
lokale Grammatik **3–11**, **17–19**, 30, 138, 148, 154, 158, 355–361
Übersetzung 6–7
Adverbiale des Ortes 61–62, 355
Anwendungen 5–10
Eigenschaften 5–11
endlicher Automat 17
Implementierung 17–19
Intex 17, 19, 214
Nooj 19
Outilex 19
Unitex 18, 19, 140, 185, 205, 214
Kaskadierung 6, 7
lokative Sätze 58–59
Named Entity Recognition 184–186, 205–211
RTN 17
finite-state-Annäherung 18
Transduktor 5–7
- Lokalsätze 148
Lokativa *s.* Nomina, lokative
lokative Adjektive 75–76, 170–172
lokative Adverbien 76–77, 170–172
lokative Prädikate **36–37**
lokative prädikative Nomina *s.* Nomina
lokative Präpositionen **37–45**, 157–158
lokative Sätze 153–172, **173**
Definition 31, 156–157
lokale Grammatik 58–59
lokative Verben 62–73
Bewegungsverben 24–25, 44–45, 50, 53, 66–67, 158–163
deutsch 53–58, 158–167
französisch 45–52
statische 164–165
transitive Verben 53–56, 159, 163–166
Transportverben 65, 163
Valenz 64–65
wohnen 67–72, 84, 164, 426–431
- Meaning Text Theory 26–30
lexikalische Funktionen 27–30, 421–435
Wörterbuch (ECD) 26, 421–435
- Metonymie *s.* Polysemie, reguläre
Microformate 223
- named entity 90, 95, 175, 176, 186
Named Entity Recognition **175–215**
N_{app} *s.* appropriate noun
nested named entities 190
N_{loc} *s.* Nomina, lokative
nom approprié *s.* appropriate noun
Nomina
lokative 80, 92, 353
prädikative **73–74**, 167–170
noms ethniques *s.* Toponym, Bewohnernamen
- One-Sense-per-Discourse 181
One-Sense-per-Document 181

- Operator grammar 12
origin *s.* Pfad, Ausgangspunkt
- Paraphrase 6, 26, 27, 30, 33, 36–37, 44, 47, 69, 79, 90, 95, 120, 126, 146–147, 191
path *s.* Pfad
perlativ *s.* Pfad, Wegpunkt
Pfad 33, 52, 74, 151–152, 158, 166
 Ausgangspunkt 24, 42, 48, 54, 161
 fokussierter Pfadbestandteil 33, 157, 159, 161, 394
 Wegpunkt 24, 42, 48
 Zielpunkt 24, 42, 48, 54
Pfadangaben 151–152
Pluralia tantum *s.* Toponym, Pluralia tantum
Polysemie
 reguläre 73, 86–91
PP-attachment disambiguation 147
prédicats appropriés 74, 80, 81
Precision 186
Proprialisierung 97
- Quasikoordination 33, 149–151, 361
Query Expansion 225
- raumbezogenes Information Retrieval *s.* Geographic Information Retrieval
räumliche Präpositionen *s.* lokative Präp.
räumliche Relationen 21–24
Recall 186
reguläre Polysemie *s.* Polysemie, reguläre
Rekursion
 in natürlichen Sprachen 18
 RTN *s.* lokale Grammatik, RTN
relatives Bezugssystem 22
Routenbeschreibungen 25, 44, 45, 226
RTN (recursive transition network) 17
- Satzsegmentierung 8
semi-supervised 182, 183
source *s.* Pfad, Ausgangspunkt
spezifisches Prädikat *s.* prédicats appropriés
statische Präpositionen 41
Stützverb 15
String Analysis 13
String Grammar 13, 14
supervised 182
Supplement
 vs. Komplement *s.* Kompl. vs. Suppl.
support verb *s.* Stützverb
Synonym 5, 6, 94
- systematische Polysemie *s.* Polysemie, reguläre
- Tagging 8
Textalinierung 6
Textautomat 8, 154
Topologie 21
topologische Felder 153, 154
Toponym 23, **93–138**
 Abkürzungen 116
 Adjektivableitung *s.* Toponym, Derivation
 als Bestandteil komplexer Named Entities
 s. nested named entities
 Ambiguität 94, 108, 215
 Artikelgebrauch 101, 111
 Bewohnerbezeichnung *s.* Toponym, Derivation
 Bewohnernamen *s.* Toponym, Derivation
 Definition 93
 Derivation 120–125, 317–332
 Diakritika 101, 104
 Diakritika, Verlust von 116
 Disambiguierung 108
 Disambiguierung homonymer Toponyme *s.* Geokodierung
 Endonym 100–102
 Epitheton 118
 Exonym 95, 100–102, 106, 107
 Flexion 101, 113–116
 Fugenform 125–127, 145
 Genus 101, 110
 Grundform 109
 Homonymie 94, 108, 205, 215
 Lexikon 102–127, 300–317
 metonymische Verwendung *s.* Polysemie, reguläre
 Nickname 118, 301
 Ontologie 127–138
 Orthographie 96, 101, 104, 299–300
 Pluralia tantum 111
 Polysemie *s.* Polysemie, reguläre
 Taxonomie 131–136, 332–340
 Transliteration/Transkription 101
toponym resolution *s.* Geokodierung
trajectory *s.* Pfad
Transduktor *s.* lokale Gramm., Transduktor
 Kaskade *s.* lokale Gramm., Kaskadierung
Transformation 14
Transkription *s.* Toponym, Translit.
Transliteration *s.* Toponym, Translit.

trivalentes Nomen 74

:U s. Genus, unspezifiziert

Übersetzung

durch Transduktoren s. lokale Grammatik, Übersetzung

maschinelle 176

ÉTAP 26

Systran 7

von Toponymen s. Toponym, Exonym

unscharfe Regionen 21, 219

unsupervised 182, 183

V_{sup} s. Stützverb

Wortbildung s. Toponym, Derivation

Beschreib. mit lokalen Grammatiken 8

Anhang

A Beispiele zu den Lexika

A.1 Toponyme

A.1.1 Zur Orthographie der Toponyme

Die orthographischen Prinzipien bei der Schreibung von Toponymen im Deutschen wurden von Meynen ed. (1966: xxxvii-xxxviii) hervorragend beobachtet. Sie sind deshalb hier im Auszug wiedergegeben:

Geographische Namen unterliegen im allgemeinen nicht den Regeln der gültigen Rechtschreibung. Vielmehr gelten die amtlich festgelegten oder eingebürgerten Schreibweisen.

Die deutschen geographischen Namen richten sich jedoch in den nachstehend aufgeführten Punkten, soweit nicht eine amtliche Festlegung dem im Einzelfall entgegensteht, nach den allgemeinen Regeln der Rechtschreibung.

1. Groß- und Kleinschreibung

- a) Groß schreibt man Eigenschafts- und Mittelwörter (Adjektive und Partizipien) als Teile von geographischen Namen. Dies gilt auch für die Ableitungen auf -isch.

Beispiele: *Rotes Meer*; *Hohe Tauern*; *Großer Ozean*; *Atlantischer Ozean*; *Mittelländisches Meer*; *Holsteinische Schweiz*.

- b) Groß schreibt man die von geographischen Namen abgeleiteten Wortformen auf -er.

Beispiele: *Lüneburger Heide*; *Thüringer Wald*; *Millstätter See*.

2. Zusammen- oder Getrenntschreibung oder Bindestrich

- a) Geographische und andere Namen als Bestimmungswort

- aa) Zusammen schreibt man im allgemeinen Zusammensetzungen aus einfachem oder zusammengesetztem geographischen Namen oder aus Personennamen und Grundwort.

Beispiele: *Nildelta*; *Glocknergruppe*; *Beringmeer*; *Magellanstraße*.

- bb) Den Bindestrich setzt man oft, wenn dem geographischen Namen als Bestimmungswort eine zusammengesetztes Grundwort folgt, um die Übersichtlichkeit zu erhöhen, oder wenn der Name hervorgehoben werden soll.

Beispiele: *Rhein-Seitenkanal*.

Bleibt die Übersichtlichkeit gewahrt, dann schreibt man zusammen.

Beispiele: *Weserbergland*; *Alpenvorland*.

- cc) Bindestriche setzt man, wenn die Bestimmung zu dem Grundwort aus mehreren geographischen Namen besteht.

Beispiele: *Dortmund-Ems-Kanal*; *Saar-Nahe-Bergland*.

- dd) Bindestriche stehen auch, wenn die Bestimmung zu einem Grundwort aus mehreren Wörtern (auch Abkürzungen) besteht.

Beispiele: *Friedrich-Wilhelm-Lübke-Koog*;
Kaiser-Franz-Joseph-Land;
Rio-de-la-Plata-Bucht;
Sankt-Lorenz-Strom.

- ee) Der Bindestrich steht beim Zusammentreffen von drei gleichen Selbstlauten.

Beispiele: *Hawaii-Inseln*.

- b) Ableitungen auf -er

- aa) Getrennt schreibt man, wenn die Ableitungen auf -er von geographischen Namen die Lage bezeichnen.
Beispiele: *Walliser Alpen* (die Alpen im Wallis); *Schweizer Mittelland* (das in der Schweiz liegende Mittelland); *Glatzer Neiße* (die von Glatz kommende Neiße); aber auch schon: *Böhmerwald*, *Wienerwald*.
- bb) Es gibt geographische Namen, die auf -er enden und keine Ableitungen im obigen Sinne sind. Diese Namen werden nach 2,a zusammengeschrieben.
Beispiele: *Glocknergruppe*; *Brennerpaß*.
Zusammen schreibt man, wenn die Ableitungen geographischer Namen auf -er Personen bezeichnen.
- cc) Wird an einen geographischen Namen auf -ee die Silbe -er angehängt, dann schreibt man nur zwei „e“
Beispiele: *Tegernseer Alpen*; *Falkenseer Forst*.
- c) Eigenschaftswort als Bestimmungswort
 - aa) Zusammen schreibt man im allgemeinen Zusammensetzungen aus ungebeugten Eigenschaftswörtern wie „groß, klein, alt, neu“ usw., Himmelsrichtungen und geographischen Namen.
Beispiele: *Großbritannien*; *Kleinasien*; *Mittelfranken*; *Hinterindien*; *Ostafrika*; *Nordwesteuropa*.
Aber, weil amtlich festgelegt: *Groß Räschen*; *Groß-Gerau*.
 - bb) Der Bindestrich setzt man bei Zusammensetzungen aus endungslosen Eigenschaftswörtern auf -isch, die von Orts-, Länder- und Völkernamen abgeleitet sind, und geographischen Namen.
Beispiele: *Spanisch-Guinea*.
Aber, weil behördlich so vorgeschrieben: *Schwäbisch Gmünd*.
- d) Zusammensetzungen aus geographischen Namen
Den Bindestrich setzt man, wenn eine geographischer Name aus zwei geographischen Namen zusammengesetzt ist.
Beispiele: *Hamburg-Altona*; *Schleswig-Holstein*; *Trentino-Tiroler Etschland*.
Merke: Bei Ableitungen bleibt der Bindestrich erhalten.
Beispiele: *Schleswig-Holsteiner*; *schleswig-holsteinisch*.

3. Silbentrennung

Geographische Namen werden nach den allgemeinen Richtlinien getrennt

Beispiele: *Al-ster*; *Es-sen*; *Fel-ben*; *Norder-ney*; *Zwik-kau*.

Zusammengesetzte geographische Namen werden nach ihren Bestandteilen getrennt, wenn sich diese erkennen lassen.

Beispiele: *Neckar-sulm*; *Schwarz-ach*.

A.1.2 Beispiele für das Lexikon deutscher Toponyme

A.1.2.1 Syntax eines Wörterbucheintrags

Ein Eintrag im Grundformenlexikon hat folgende Syntax (siehe Paumier 2006: 31–4):

⟨Grundform⟩, ⟨Lemma⟩? . ⟨Wortart⟩ [+⟨klass. Merkmal⟩]* [: ⟨Genus⟩* ⟨Flexion⟩] / ⟨Kommentar⟩?

Sind Grundform und Lemma identisch, kann letzteres entfallen (vgl. Kap. 3.3.2.1 zur Unterscheidung zwischen Grundform und Lemma).

Die Syntax eines Eintrags im Vollformenlexikon ist ähnlich, und auch hier gilt: sind Vollform und Lemma identisch, wird letzteres nicht angegeben.

⟨Vollform⟩, ⟨Lemma⟩? . ⟨Wortart⟩ [+⟨klass. Merkmal⟩]* [: ⟨flekt. Merkmale⟩*] / ⟨Komm.⟩?

Kommen Metasymbole literal im Lexikoneintrag vor, werden diese durch \ geschützt:

Washington\, D.C., Washington, D\C\..EN+Topon+Oikon+...

Die Merkmale für die flektierten Formen folgen den Konventionen des CISLEX (siehe Maier-Meyer 1995: 67), als zusätzliches Merkmal markiert ‘S’ die markierten auf -s endenden Genitivformen (vgl. (3.97) und (3.98) auf S. 142).

Klassifikatorische Merkmale der Toponyme (und deren Ableitungen) sind:

- ‘+!’ für ambigüe Toponyme (*Weiden, Regen, Zug*, siehe S. 108)
- die Klasse des Toponyms bzw. der geographischen Entität (siehe Kap. 3.4.1.1 und 3.4.2.1). Klassenannotationen sind, um den einfachen Zugriff beim Schreiben von Grammatiken zu erleichtern, immer kumulativ mit allen (wichtigen) Hierarchieebenen annotiert, d.h. wenn der *Fehmarnbelt* als ‘+Topon+Hydron+Cl(Meer)+Cl(Strait)’ annotiert ist, gehört er der Klasse STRAIT/MEERENGE an, ist aber via Vererbung auch eine Entität der Klasse MEER, HYDRONYM/GEWÄSSER und nicht zuletzt ganz generell ein TOPONYM. Für eine Übersicht der wichtigsten Klassen siehe Tab. A.2. Die Angabe der Klasse (Feinklassifikation) innerhalb ‘+Cl(...)’ sind formal keine Grenzen gesetzt.
- Annotationen zur „Ontologisierung“ des text- und zeilenbasierten Lexikons (vgl. Kap. 3.4):

+cnt(Y)	Inklusionsrelation: Entität ist in Y räumlich enthalten, ist Teil von Y
+cap(Y)	X ist Hauptstadt von Y
+var(Y)	das Toponym ist eine Namensvariante eines anderen Toponyms Y, welches die gebräuchlichere Variante für dieselbe geographische Entität darstellt.
+trans(Y)	das Toponym ist Übersetzung eines anderen Toponyms Y. Im Gegensatz +var(...) kann +trans(...) auch auf „schlechtere“, da im Deutschen +cnt(Chile ungebräuchlichere Toponyme verweisen: das Exonym <i>Kap Hoorn</i> ist durch +trans(<i>Cabo de Hornos</i>) mit dem autochthonen Toponym <i>Cabo de Hornos</i> verlinkt.
+hist	historischer Name
+hist+var(Y)	historischer Name des Objekts mit Namen Y
+hist(Z)	der Gebrauch des Namens oder die Existenz der geographischen Entität ist auf den Zeitraum Z beschränkt
+id(...)	eine ID erlaubt, Namensvarianten zusammenzufassen und Homonyme einander zuzuordnen

- ‘+abbr’ kennzeichnet Abkürzungen, siehe Kap. 3.3.2.7
- ‘+nickname’ kennzeichnet Nicknames und Spitznamen, siehe Kap. 3.3.2.9
- ‘+adj’ kennzeichnet Toponyme, die adjektivisch flektieren, also über starke und schwache Formen verfügen

Bezeichnete Klasse von Objekten	Bezeichnung im Lexikon	Termini in der Onomastik
Ort, Örtlichkeit, topographisches Objekt	+Topon	Toponym, Ortsname, geographischer Name, Ortsbezeichnung, Geotoponym
Region/Ort, nicht zum Wohnen bestimmt		Anoikonum

Tabelle A.2: Bezeichnungen für Toponymtypen in der Onomastik und Kennzeichnung der Typen im Toponymlexikon

Bezeichnete Klasse von Objekten	Bezeichnung im Lexikon	Termini in der Onomastik
Siedlung	+Topon+Oikon	Oikononym, Siedlungsname, Ortsname, engl. eonym; Stadtname, Dorfname
Mikroobjekt innerhalb einer Siedlung	(+Topon+Oikon / +Topon+Aedif)	Urbanonym
Wüstung	(+Topon+Oikon+hist)	Wüstungsname
Stadt	(+Topon+Oikon)	Astionym
(Bauern)Hof	(+Topon+Oikon / +Topon+Aedif)	Hofname, Gehöftname, Aulonym
(einzelnes) Gebäude	+Topon+Aedif	Gebäudenname, Oikodomonym
Sakralbau, Kirche	+Topon+Aedif+Cl(Sakralbau)	Ekklesionym
Begräbnisstätte		Nekronym
Verkehrsweg	+Topon+Hodon	Hodonym, Dromonym; Straßenname
Platz, Marktplatz	+Topon+Hodon	Agoronym
Wasserweg	+Topon+Hydron+Hodon	
Region (bewohnte oder unbewohnte natürliche oder administrative Einheit)	+Topon+Choron	Choronym, Raumname
geopolitischen Einheit, Gebietseinheit, Verwaltungseinheit	+Topon+Choron+admin	administrativer Raumname
Internationale Organisation (mit Gebietseinheiten als Mitgliedern und Außengrenzen)	+Topon+ORG	
Landschaft, Region	+Topon+Choron	Landschaftsname, -bezeichnung
Gebirge	+Topon+Oron+Choron	Oronym, Gebirgsname
Berg	+Topon+Oron	Oronym, Bergname
Insel, Inselgruppe	+Topon+Choron+Cl(Insel)	Insulonym, Inselname
Tal	+Topon+Choron+Cl(Tal)	Talname, Koilonym
Sumpf, Moor	+Topon+Choron+Cl(Moor)	Helonym, Moorname, Sumpfname
Feld, Flur	(+Topon+Choron)	Flurname, Agronym, Feldname
Parzelle	(+Topon+Choron)	Parzellenname
Gewann	(+Topon+Choron)	Gewannname
Höhle, Grotte	+Topon+Cl(Höhle)	Speleonym
Gewässer	+Topon+Hydron	Hydronym, Gewässername
Meer	+Topon+Hydron+Cl(Meer)	Okeanonym, Pelagonym
Wasserlauf	+Topon+Hydron+Cl(Fluss)	Flussname, Potamonym, Bachname
See	+Topon+Hydron+Cl(See)	Limnonym, Seename; Teichname
Wasserweg	+Topon+Hydron+Hodon	
Ort/Region in Gewässer	(+Topon+Hydron)	Wasserflurname, Fischerflurname
Himmelskörper	-	Astronym, Sternen-, Planetenname
Bereich im Weltall, z.B. Sternbild	-	Kosmonym
Ort/Objekt auf der Oberfläche fremder Himmelskörper	-	Astrotoponym

Tabelle A.2: Bezeichnungen für Toponymtypen in der Onomastik und Kennzeichnung der Typen im Toponymlexikon

A.1.2.2 Grundformenlexikon

Im Folgenden sind die Einträge von 23 Toponymen im Grundformenwörterbuch wiedergegeben. Innerhalb von ‘{’ und ‘}’ steht der Flexionscode, anhand dessen die Expansion zum Vollformenwörterbuch erfolgt. ‘=’ steht per DELA-Konvention als Symbol für Leerzeichen oder Bindestrich.

```
Appenzell=Innerrhoden, Appenzell Innerrhoden. EN+Topon+Choron+admin+Cl (Halbkanton)+DetZ:N{NS2}
Dammvorstadt, .EN+Topon+Oikon+var (Słubice)+Det:F{NSO}
Essen, .EN+!+Topon+Oikon+DetZ:N{NS2}/<<0er. ADJ{ADJ00}, 0er. N+Hum:M{NS2;NP1}, 0erin. N+Hum:F{NSO;NP5}>>
Gera, .EN+Topon+Choron+admin+Cl (Bezirk)+oblCl:nxU
Gera, .EN+Topon+Hydron+Cl (Fluss)+Det:F{NSO}
Gera, .EN+Topon+Oikon+DetZ:N{NS2}/<<0er. ADJ{ADJ00}, 0er. N+Hum:M{NS2;NP1}, 0erin. N+Hum:F{NSO;NP5}>>
Große Arber, Großer Arber. EN+Topon+Oron+Det:M{ADJ+NS13}
Madrid, .EN+Topon+Oikon+DetZ:N{NS2}/<<0er. ADJ{ADJ00}, 0er. N+Hum:M{NS2;NP1}, 0erin. N+Hum:F{NSO;NP5}, 1lene. N+Hum:M {
  NS4;NP4}, 1lenin. N+Hum:F{NSO;NP5}, 1lenisch. ADJ{ADJ}, 1leño. N+Hum:F{NSO;NP6}, 1leño. N+Hum:M{NS2;NP6}>>
München, .EN+Topon+Choron+admin+Cl (Landkreis)+oblCl+DetZ:nxU
München, .EN+Topon+Oikon+DetZ:N{NS2}/<<0er. ADJ{ADJ00}, 0er. N+Hum:M{NS2;NP1}, 0erin. N+Hum:F{NSO;NP5},
  2ner. ADJ{ADJ00}, 2ner. N+Hum:M{NS2;NP1}, 2nerin. N+Hum:F{NSO;NP5}, 2nerisch. ADJ{ADJ}>>
Persische Golf, Persischer Golf. EN+Topon+Hydron+Cl (Meer)+Det:M{ADJ+NS1}
Pfaffenschlag bei Waidhofen an der Thaya, .EN+Topon+Oikon+DetZ:N{1:NS2}
Rügen, .EN+Topon+Choron+admin+Cl (Landkreis)+oblCl:nxU
Rügen, .EN+Topon+Choron+Cl (Insel)+DetZ:N{NS2}/<<0. FF, 0er. ADJ{ADJ00}, 0er. N+Hum:M{NS2;NP1},
  0erin. N+Hum:F{NSO;NP5}, 0sch. ADJ{ADJ}, 2er. ADJ{ADJ00}, 2aner. ADJ{ADJ00}, 2aner. N+Hum:M{NS2;NP1},
  2anerin. N+Hum:F{NSO;NP5}, 4ugianer. N+Hum:M{NS2;NP1}, 4ugianerin. N+Hum:F{NSO;NP5}, >>
Sankt Vincent und die Grenadinen, .EN+Topon+Choron+admin+Cl (Staat)+DetZ:m{0+NS2+0+DET4+NPO}
Schwarze Meer, Schwarzes Meer. EN+Topon+Hydron+Cl (Meer)+Det:N{ADJ+NS1}
Słubice, .EN+Topon+Oikon+DetZ:N{NS2}/<<1er. ADJ{ADJ00}, 1er. N+Hum:M{NS2;NP1}, 1erin. N+Hum:F{NSO;NP5}>>
Unna, .EN+Topon+Choron+admin+Cl (Landkreis)+oblCl:nxU
Unna, .EN+Topon+Oikon+DetZ:N{NS2}/<<0er. ADJ{ADJ00}, 0er. N+Hum:M{NS2;NP1}, 0erin. N+Hum:F{NSO;NP5},
  1eraner. ADJ{ADJ00}, 1eraner. N+Hum:M{NS2;NP1}, 1eranerin. N+Hum:F{NSO;NP5}>>
Wedding, .EN+Topon+Oikon+Det:M{NS13}/<<0er. ADJ{ADJ00}, 0er. N+Hum:M{NS2;NP1}, 0erin. N+Hum:F{NSO;NP5}>>
Wedding, .EN+Topon+Oikon+var (Wedding)+DetZ:N{NS2}
Wiener Neustadt, .EN+Topon+Oikon+Det:F{ADJ00+NSO}/ selbständige Stadt in der Nähe von Wien, kein Stadtteil!
Wiener Neustadt, .EN+Topon+Oikon+var (Wiener Neustadt)+DetZ:N{ADJ00+NS2}
```

A.1.2.3 Vollformenlexikon

Aus dem oben wiedergegebenen Grundformenlexikon werden alle Lemma dekliniert, einschließlich der abgeleiteten Adjektive und Bezeichnungen der Bewohner. Die Zahl der Einträge liegt nun bei knapp 300. In der Realität fällt dieser Zuwachs geringer aus. Hier wurden Einträge ausgewählt, die die Phänomene Flexion, Derivation und (orthographische) Variation demonstrieren. Die letzten beiden sind nur für einen kleinen Teil der im Lexikon enthaltenen Toponyme belegt oder beschrieben. Bei einer realistischen, repräsentativeren Auswahl vervierfacht sich der Umfang des Lexikons.

In der Darstellung wurden überlange Zeilen abgeschnitten. Im Fall von indeklinablen Adjektiven dient ‘:X’ als Platzhalter für alle Formen. Um das Schreiben von Grammatiken einfacher zu gestalten, sind im Lexikon jedoch alle Merkmalsbündel der homonymen Formen explizit aufgeführt. Anstelle von Münchner, .ADJ+Topon+Oikon+der (München):X muss also gelesen werden: Münchner, .ADJ+Topon+Oikon+der (München):X: aeFxp: aeFyp: aeFzp: aeMxp: aeMyp: aeMzp: aeNxp: aeNyp: aeNzp: amUxp: amUyp: amUzp: deFxp: deFyp: deFzp: deMxp: deMyp: deMzp: deNxp: deNyp: deNzp: dmUxp: dmUyp: dmUzp: geFxp: geFyp: geFzp: geMxp: geMyp: geMzp: geNxp: geNyp: geNzp: gmUxp: gmUyp: gmUzp: neFxp: neFyp: neFzp: neMxp: neMyp: neMzp: neNxp: neNyp: neNzp: nmUxp: nmUyp: nmUzp.

Appenzell-Innerrhoden, Appenzell Innerrhoden. EN+Topon+Choron+admin+Cl (Halbkanton)+DetZ: aeN: deN: neN
 Appenzell-Innerrhodens, Appenzell Innerrhoden. EN+Topon+Choron+admin+Cl (Halbkanton)+DetZ: geNS
 Appenzell Innerrhodens, Appenzell Innerrhoden. EN+Topon+Choron+admin+Cl (Halbkanton)+DetZ: geNS
 Appenzell Innerrhoden, . EN+Topon+Choron+admin+Cl (Halbkanton)+DetZ: aeN: deN: neN
 Dammvorstadt, . EN+Topon+Oikon+var (Stibice)+Det: aeF: deF: geF: neF
 Essens, Essen. EN+!+Topon+Oikon+DetZ: geNS
 Essen, . EN+!+Topon+Oikon+DetZ: aeN: deN: neN
 Essener, . ADJ+Topon+Oikon+der (Essen): X
 Essenern, Essener. N+Hum+Topon+Oikon+der (Essen): dmM
 Esseners, Essener. N+Hum+Topon+Oikon+der (Essen): geMS
 Essener, . N+Hum+Topon+Oikon+der (Essen): aeM: amM: deM: gmM: neM: nmM
 Essenerin, . N+Hum+Topon+Oikon+der (Essen): aeF: deF: geF: neF
 Essenerinnen, Essenerin. N+Hum+Topon+Oikon+der (Essen): amF: dmF: gmF: nmF
 Gera, . EN+Topon+Choron+admin+Cl (Bezirk)+oblCl: nxU
 Gera, . EN+Topon+Hydron+Cl (Fluss)+Det: aeF: deF: geF: neF
 gera, . FF+Topon+Hydron+Cl (Fluss)
 Gera, . EN+Topon+Oikon+DetZ: aeN: deN: neN
 Geras, Gera. EN+Topon+Oikon+DetZ: geNS
 Geraer, . ADJ+Topon+Oikon+der (Gera): X
 Geraer, . N+Hum+Topon+Oikon+der (Gera): aeM: amM: deM: gmM: neM: nmM
 Geraers, Geraer. N+Hum+Topon+Oikon+der (Gera): geMS
 Geraern, Geraer. N+Hum+Topon+Oikon+der (Gera): dmM
 Geraerinnen, Geraerin. N+Hum+Topon+Oikon+der (Gera): amF: dmF: gmF: nmF
 Geraerin, . N+Hum+Topon+Oikon+der (Gera): aeF: deF: geF: neF
 Große Arber, Großer Arber. EN+Topon+Oron+Det+adj: neMy
 Großer Arber, . EN+Topon+Oron+Det+adj: neMx: neMz
 Grossen Arber, Großer Arber. EN+Topon+Oron+Det+adj: aeMx: aeMy: aeMz: deMy: deMz: geMx: geMy: geMz
 Großem Arber, Großer Arber. EN+Topon+Oron+Det+adj: deMx
 Grosse Arber, Großer Arber. EN+Topon+Oron+Det+adj: neMy
 Grosser Arber, Großer Arber. EN+Topon+Oron+Det+adj: neMx: neMz
 Großen Arbers, Großer Arber. EN+Topon+Oron+Det+adj: geMx: geMy: geMz
 Grossen Arbers, Großer Arber. EN+Topon+Oron+Det+adj: geMx: geMy: geMz
 Großen Arber, Großer Arber. EN+Topon+Oron+Det+adj: aeMx: aeMy: aeMz: deMy: deMz: geMx: geMy: geMz
 Grossem Arber, Großer Arber. EN+Topon+Oron+Det+adj: deMx
 Madrid, . EN+Topon+Oikon+DetZ: aeN: deN: neN
 Madrids, Madrid. EN+Topon+Oikon+DetZ: geNS
 Madrider, . ADJ+Topon+Oikon+der (Madrid): X
 Madridern, Madrider. N+Hum+Topon+Oikon+der (Madrid): dmM
 Madriders, Madrider. N+Hum+Topon+Oikon+der (Madrid): geMS
 Madrider, . N+Hum+Topon+Oikon+der (Madrid): aeM: amM: deM: gmM: neM: nmM
 Madriderin, . N+Hum+Topon+Oikon+der (Madrid): aeF: deF: geF: neF
 Madriderinnen, Madriderin. N+Hum+Topon+Oikon+der (Madrid): amF: dmF: gmF: nmF
 Madrilene, . N+Hum+Topon+Oikon+der (Madrid): neM
 Madrilenen, Madrilene. N+Hum+Topon+Oikon+der (Madrid): aeM: amM: deM: dmM: geM: gmM: nmM
 Madrileneninnen, Madrilenin. N+Hum+Topon+Oikon+der (Madrid): amF: dmF: gmF: nmF
 Madrilenin, . N+Hum+Topon+Oikon+der (Madrid): aeF: deF: geF: neF
 madrilenischen, madrilenisches. ADJ+Topon+Oikon+der (Madrid): aeMxp: aeMyp: aeMzp: amUyp: amUzp: deFyp: deFzp: deMyp: ...
 madrilenisches, . ADJ+Topon+Oikon+der (Madrid): up
 madrilenischer, madrilenisches. ADJ+Topon+Oikon+der (Madrid): deFxp: geFxp: gmUxp: neMxp: neMzp
 madrilenisches, madrilenisches. ADJ+Topon+Oikon+der (Madrid): aeNxp: aeNzp: neNxp: neNzp
 madrilenisches, madrilenisches. ADJ+Topon+Oikon+der (Madrid): aeFxp: aeFyp: aeFzp: aeNyp: amUxp: neFxp: neFyp: neFzp: ...
 madrilenisches, madrilenisches. ADJ+Topon+Oikon+der (Madrid): deMxp: deNxp
 Madrileño, . N+Hum+Topon+Oikon+der (Madrid): aeF: deF: geF: neF
 Madrileño, Madrileño. N+Hum+Topon+Oikon+der (Madrid): aeF: deF: geF: neF
 Madrilenos, Madrileño. N+Hum+Topon+Oikon+der (Madrid): amF: dmF: gmF: nmF
 Madrileños, Madrileño. N+Hum+Topon+Oikon+der (Madrid): amF: dmF: gmF: nmF
 Madrileño, . N+Hum+Topon+Oikon+der (Madrid): aeM: deM: neM
 Madrileño, Madrileño. N+Hum+Topon+Oikon+der (Madrid): aeM: deM: neM
 Madrilenos, Madrileño. N+Hum+Topon+Oikon+der (Madrid): amM: dmM: geMS: gmM: nmM
 Madrileños, Madrileño. N+Hum+Topon+Oikon+der (Madrid): amM: dmM: geMS: gmM: nmM
 München, . EN+Topon+Choron+admin+Cl (Landkreis)+oblCl+DetZ: nxU
 Münchens, München. EN+Topon+Oikon+DetZ: geNS
 München, . EN+Topon+Oikon+DetZ: aeN: deN: neN
 Münchener, . ADJ+Topon+Oikon+der (München): X
 Münchenern, Münchener. N+Hum+Topon+Oikon+der (München): geMS
 Münchenern, Münchener. N+Hum+Topon+Oikon+der (München): dmM
 Münchener, . N+Hum+Topon+Oikon+der (München): aeM: amM: deM: gmM: neM: nmM
 Münchenerinnen, Münchenerin. N+Hum+Topon+Oikon+der (München): amF: dmF: gmF: nmF
 Münchenerin, . N+Hum+Topon+Oikon+der (München): aeF: deF: geF: neF
 Münchner, . ADJ+Topon+Oikon+der (München): X
 Münchner, . N+Hum+Topon+Oikon+der (München): aeM: amM: deM: gmM: neM: nmM
 Münchners, Münchner. N+Hum+Topon+Oikon+der (München): geMS

Münchnern, Münchner. N+Hum+Topon+Oikon+der (München) : dmM
Münchnerin, .N+Hum+Topon+Oikon+der (München) : aeF:deF:geF:neF
Münchnerinnen, Münchnerin. N+Hum+Topon+Oikon+der (München) : amF:dmF:gmF:nmF
münchnerisches, münchnerisch. ADJ+Topon+Oikon+der (München) : aeNxp:aeNzp:neNxp:neNzp
münchnerischer, münchnerisch. ADJ+Topon+Oikon+der (München) : deFxp:geFxp:gmUxp:neMxp:neMzp
münchnerische, münchnerisch. ADJ+Topon+Oikon+der (München) : aeFxp:aeFyp:aeFzp:aeNyp:amUxp:neFxp:neFyp:neFzp:...
münchnerischen, münchnerisch. ADJ+Topon+Oikon+der (München) : aeMxp:aeMyp:aeMzp:amUyp:amUzp:deFyp:deFzp:deMyp:...
münchnerischem, münchnerisch. ADJ+Topon+Oikon+der (München) : deMxp:deNxp
münchnerisch, .ADJ+Topon+Oikon+der (München) : up
Pers. Golfs, Persischer Golf. EN+Topon+Hydron+Cl (Meer)+Det+adj:geMx:geMy:geMz
Persische Golf, Persischer Golf. EN+Topon+Hydron+Cl (Meer)+Det+adj:neMy
Pers. Golfes, Persischer Golf. EN+Topon+Hydron+Cl (Meer)+Det+adj:geMx:geMy:geMz
Persischem Golf, Persischer Golf. EN+Topon+Hydron+Cl (Meer)+Det+adj:deMx
Pers. Golfs, Persischer Golf. EN+Topon+Hydron+Cl (Meer)+Det+adj:geMx:geMy:geMz
Pers. Golfe, Persischer Golf. EN+Topon+Hydron+Cl (Meer)+Det+adj:deMx:deMy:deMz
Persischen Golfe, Persischer Golf. EN+Topon+Hydron+Cl (Meer)+Det+adj:deMy:deMz
Persischer Golf, .EN+Topon+Hydron+Cl (Meer)+Det+adj:neMx:neMz
Persischen Golfs, Persischer Golf. EN+Topon+Hydron+Cl (Meer)+Det+adj:geMx:geMy:geMz
Pers. Golf, Persischer Golf. EN+Topon+Hydron+Cl (Meer)+Det+adj:aeMx:aeMy:aeMz:deMy:deMz:neMx:neMz:deMx
Pers. Golfe, Persischer Golf. EN+Topon+Hydron+Cl (Meer)+Det+adj:deMx:deMy:deMz
Persischen Golfes, Persischer Golf. EN+Topon+Hydron+Cl (Meer)+Det+adj:geMx:geMy:geMz
Pers. Golfes, Persischer Golf. EN+Topon+Hydron+Cl (Meer)+Det+adj:geMx:geMy:geMz
Persischem Golfe, Persischer Golf. EN+Topon+Hydron+Cl (Meer)+Det+adj:deMx
Pers. Golf, Persischer Golf. EN+Topon+Hydron+Cl (Meer)+Det+adj:aeMx:aeMy:aeMz:deMy:deMz:neMx:neMz:deMx
Persischen Golf, Persischer Golf. EN+Topon+Hydron+Cl (Meer)+Det+adj:aeMx:aeMy:aeMz:deMy:deMz
Pfaffenschlag b. Waidhofen\Thaya, Pfaffenschlag bei Waidhofen an der Thaya. EN+Topon+Oikon+DetZ:aeN:deN:...
Pfaffenschlags b. Waidhofen ad. T., Pfaffenschlag bei Waidhofen an der Thaya. EN+Topon+Oikon+DetZ:geN
Pfaffenschlag b. Waidhofen (Thaya), Pfaffenschlag bei Waidhofen an der Thaya. EN+Topon+Oikon+DetZ:aeN:deN:neN
Pfaffenschlags b. Waidhofen a.d. T., Pfaffenschlag bei Waidhofen an der Thaya. EN+Topon+Oikon+DetZ:geN
Pfaffenschlag bei Waidhofen (Th.), Pfaffenschlag bei Waidhofen an der Thaya. EN+Topon+Oikon+DetZ:aeN:deN:neN
Pfaffenschlag b. Waidhofen ad. T., Pfaffenschlag bei Waidhofen an der Thaya. EN+Topon+Oikon+DetZ:aeN:deN:neN
Pfaffenschlag b. Waidhofen ad. Thaya, Pfaffenschlag bei Waidhofen an der Thaya. EN+Topon+Oikon+DetZ:aeN:deN:neN
Pfaffenschlags b. Waidhofen\Th, Pfaffenschlag bei Waidhofen an der Thaya. EN+Topon+Oikon+DetZ:geN
Pfaffenschlags b. Waidhofen (Th.), Pfaffenschlag bei Waidhofen an der Thaya. EN+Topon+Oikon+DetZ:geN
Pfaffenschlags bei Waidhofen (T.), Pfaffenschlag bei Waidhofen an der Thaya. EN+Topon+Oikon+DetZ:geN
Pfaffenschlag bei Waidhofen ad. T., Pfaffenschlag bei Waidhofen an der Thaya. EN+Topon+Oikon+DetZ:aeN:deN:neN
Pfaffenschlag b. Waidhofen ad. T., Pfaffenschlag bei Waidhofen an der Thaya. EN+Topon+Oikon+DetZ:aeN:deN:neN
Pfaffenschlags b. Waidhofen a.d. Th., Pfaffenschlag bei Waidhofen an der Thaya. EN+Topon+Oikon+DetZ:geN
Pfaffenschlag b. Waidhofen an d. Th., Pfaffenschlag bei Waidhofen an der Thaya. EN+Topon+Oikon+DetZ:aeN:deN:neN
Pfaffenschlags bei Waidhofen ad. T., Pfaffenschlag bei Waidhofen an der Thaya. EN+Topon+Oikon+DetZ:geN
Pfaffenschlags b. Waidhofen a.d. Thaya, Pfaffenschlag bei Waidhofen an der Thaya. EN+Topon+Oikon+DetZ:geN
Pfaffenschlags bei Waidhofen a.d. T., Pfaffenschlag bei Waidhofen an der Thaya. EN+Topon+Oikon+DetZ:geN
Pfaffenschlag bei Waidhofen a.d. Thaya, Pfaffenschlag bei Waidhofen an der Thaya. EN+Topon+Oikon+DetZ:aeN:deN:neN
Pfaffenschlags b. Waidhofen a.d. Th., Pfaffenschlag bei Waidhofen an der Thaya. EN+Topon+Oikon+DetZ:geN
Pfaffenschlags b. Waidhofen a.d. Th., Pfaffenschlag bei Waidhofen an der Thaya. EN+Topon+Oikon+DetZ:geN
Pfaffenschlags b. Waidhofen\Thaya, Pfaffenschlag bei Waidhofen an der Thaya. EN+Topon+Oikon+DetZ:geN
Pfaffenschlags b. Waidhofen a.d. Thaya, Pfaffenschlag bei Waidhofen an der Thaya. EN+Topon+Oikon+DetZ:geN
Pfaffenschlags b. Waidhofen ad. Th., Pfaffenschlag bei Waidhofen an der Thaya. EN+Topon+Oikon+DetZ:geN
Pfaffenschlag b. Waidhofen an d. Th., Pfaffenschlag bei Waidhofen an der Thaya. EN+Topon+Oikon+DetZ:aeN:deN:neN
Pfaffenschlags b. Waidhofen a.d. T., Pfaffenschlag bei Waidhofen an der Thaya. EN+Topon+Oikon+DetZ:geN
Pfaffenschlag b. Waidhofen an d. Th., Pfaffenschlag bei Waidhofen an der Thaya. EN+Topon+Oikon+DetZ:aeN:deN:neN
Pfaffenschlags b. Waidhofen ad. Thaya, Pfaffenschlag bei Waidhofen an der Thaya. EN+Topon+Oikon+DetZ:geN
Pfaffenschlag b. Waidhofen\Th, Pfaffenschlag bei Waidhofen an der Thaya. EN+Topon+Oikon+DetZ:aeN:deN:neN
Pfaffenschlag b. Waidhofen an d. Th., Pfaffenschlag bei Waidhofen an der Thaya. EN+Topon+Oikon+DetZ:aeN:deN:neN
Pfaffenschlags b. Waidhofen ad. Th., Pfaffenschlag bei Waidhofen an der Thaya. EN+Topon+Oikon+DetZ:geN
Pfaffenschlag b. Waidhofen ad. T., Pfaffenschlag bei Waidhofen an der Thaya. EN+Topon+Oikon+DetZ:geN
Pfaffenschlags b. Waidhofen a.d. Thaya, Pfaffenschlag bei Waidhofen an der Thaya. EN+Topon+Oikon+DetZ:geN
Pfaffenschlags b. Waidhofen (Thaya), Pfaffenschlag bei Waidhofen an der Thaya. EN+Topon+Oikon+DetZ:geN
Pfaffenschlags bei Waidhofen\Th., Pfaffenschlag bei Waidhofen an der Thaya. EN+Topon+Oikon+DetZ:geN
Pfaffenschlag bei Waidhofen\T., Pfaffenschlag bei Waidhofen an der Thaya. EN+Topon+Oikon+DetZ:aeN:deN:neN
Pfaffenschlags b. Waidhofen a.d. Thaya, Pfaffenschlag bei Waidhofen an der Thaya. EN+Topon+Oikon+DetZ:geN
Pfaffenschlag bei Waidhofen (T.), Pfaffenschlag bei Waidhofen an der Thaya. EN+Topon+Oikon+DetZ:aeN:deN:neN
Pfaffenschlag b. Waidhofen an d. T., Pfaffenschlag bei Waidhofen an der Thaya. EN+Topon+Oikon+DetZ:aeN:deN:neN
Pfaffenschlags b. Waidhofen an der Th., Pfaffenschlag bei Waidhofen an der Thaya. EN+Topon+Oikon+DetZ:geN
Pfaffenschlags bei Waidhofen an d. Th., Pfaffenschlag bei Waidhofen an der Thaya. EN+Topon+Oikon+DetZ:geN

Pfaffenschlag b.Waidhofen (Th.),Pfaffenschlag bei Waidhofen an der Thaya.EN+Topon+Oikon+DetZ:aeN:deN:neN
 Pfaffenschlag b. Waidhofen ad T.,Pfaffenschlag bei Waidhofen an der Thaya.EN+Topon+Oikon+DetZ:aeN:deN:neN
 Pfaffenschlag b.Waidhofen ad Thaya,Pfaffenschlag bei Waidhofen an der Thaya.EN+Topon+Oikon+DetZ:aeN:deN:neN
 Pfaffenschlags b. Waidhofen (Thaya),Pfaffenschlag bei Waidhofen an der Thaya.EN+Topon+Oikon+DetZ:geN
 Pfaffenschlag b.Waidhofen (Thaya),Pfaffenschlag bei Waidhofen an der Thaya.EN+Topon+Oikon+DetZ:aeN:deN:neN
 Pfaffenschlags bei Waidhofen ad.Thaya,Pfaffenschlag bei Waidhofen an der Thaya.EN+Topon+Oikon+DetZ:geN
 Pfaffenschlags b.Waidhofen\T.,Pfaffenschlag bei Waidhofen an der Thaya.EN+Topon+Oikon+DetZ:geN
 Pfaffenschlags b. Waidhofen\Th.,Pfaffenschlag bei Waidhofen an der Thaya.EN+Topon+Oikon+DetZ:geN
 Pfaffenschlag b.Waidhofen\T.,Pfaffenschlag bei Waidhofen an der Thaya.EN+Topon+Oikon+DetZ:aeN:deN:neN
 Pfaffenschlags bei Waidhofen an d. Th.,Pfaffenschlag bei Waidhofen an der Thaya.EN+Topon+Oikon+DetZ:geN
 Pfaffenschlag bei Waidhofen ad.Thaya,Pfaffenschlag bei Waidhofen an der Thaya.EN+Topon+Oikon+DetZ:aeN:deN:neN
 Pfaffenschlags bei Waidhofen a.d. T.,Pfaffenschlag bei Waidhofen an der Thaya.EN+Topon+Oikon+DetZ:geN
 Pfaffenschlags b. Waidhofen ad.T.,Pfaffenschlag bei Waidhofen an der Thaya.EN+Topon+Oikon+DetZ:geN
 Pfaffenschlag bei Waidhofen an der Thaya,.EN+Topon+Oikon+DetZ:aeN:deN:neN
 Pfaffenschlags b.Waidhofen\Thaya,Pfaffenschlag bei Waidhofen an der Thaya.EN+Topon+Oikon+DetZ:geN
 Pfaffenschlags bei Waidhofen ad.Th.,Pfaffenschlag bei Waidhofen an der Thaya.EN+Topon+Oikon+DetZ:geN
 Pfaffenschlags b. Waidhofen an d. Th.,Pfaffenschlag bei Waidhofen an der Thaya.EN+Topon+Oikon+DetZ:geN
 Pfaffenschlag b.Waidhofen a.d.T.,Pfaffenschlag bei Waidhofen an der Thaya.EN+Topon+Oikon+DetZ:aeN:deN:neN
 Pfaffenschlags bei Waidhofen ad T.,Pfaffenschlag bei Waidhofen an der Thaya.EN+Topon+Oikon+DetZ:geN
 Pfaffenschlag b. Waidhofen ad.Th.,Pfaffenschlag bei Waidhofen an der Thaya.EN+Topon+Oikon+DetZ:aeN:deN:neN
 Pfaffenschlag b. Waidhofen a.d. Thaya,Pfaffenschlag bei Waidhofen an der Thaya.EN+Topon+Oikon+DetZ:aeN:deN:neN
 Pfaffenschlags b.Waidhofen\Th.,Pfaffenschlag bei Waidhofen an der Thaya.EN+Topon+Oikon+DetZ:geN
 Pfaffenschlag b. Waidhofen a.d. T.,Pfaffenschlag bei Waidhofen an der Thaya.EN+Topon+Oikon+DetZ:aeN:deN:neN
 Pfaffenschlag b.Waidhofen an d.T.,Pfaffenschlag bei Waidhofen an der Thaya.EN+Topon+Oikon+DetZ:aeN:deN:neN
 Pfaffenschlag bei Waidhofen (Th),Pfaffenschlag bei Waidhofen an der Thaya.EN+Topon+Oikon+DetZ:aeN:deN:neN
 Pfaffenschlags bei Waidhofen a.d.Th.,Pfaffenschlag bei Waidhofen an der Thaya.EN+Topon+Oikon+DetZ:geN
 Pfaffenschlag b. Waidhofen a.d.T.,Pfaffenschlag bei Waidhofen an der Thaya.EN+Topon+Oikon+DetZ:aeN:deN:neN
 Pfaffenschlags b.Waidhofen (Th),Pfaffenschlag bei Waidhofen an der Thaya.EN+Topon+Oikon+DetZ:geN
 Pfaffenschlags bei Waidhofen (Th.),Pfaffenschlag bei Waidhofen an der Thaya.EN+Topon+Oikon+DetZ:geN
 Pfaffenschlag b. Waidhofen ad.Thaya,Pfaffenschlag bei Waidhofen an der Thaya.EN+Topon+Oikon+DetZ:aeN:deN:neN
 Pfaffenschlag b.Waidhofen (Th),Pfaffenschlag bei Waidhofen an der Thaya.EN+Topon+Oikon+DetZ:aeN:deN:neN
 Pfaffenschlag b. Waidhofen ad. Thaya,Pfaffenschlag bei Waidhofen an der Thaya.EN+Topon+Oikon+DetZ:aeN:deN:neN
 Pfaffenschlag bei Waidhofen ad.Th.,Pfaffenschlag bei Waidhofen an der Thaya.EN+Topon+Oikon+DetZ:aeN:deN:neN
 Pfaffenschlags bei Waidhofen an der Thaya,Pfaffenschlag bei Waidhofen an der Thaya.EN+Topon+Oikon+DetZ:geN
 Pfaffenschlags b. Waidhofen an der Th.,Pfaffenschlag bei Waidhofen an der Thaya.EN+Topon+Oikon+DetZ:geN
 Pfaffenschlag b.Waidhofen a.d. Th.,Pfaffenschlag bei Waidhofen an der Thaya.EN+Topon+Oikon+DetZ:aeN:deN:neN
 Pfaffenschlag b. Waidhofen (Th),Pfaffenschlag bei Waidhofen an der Thaya.EN+Topon+Oikon+DetZ:aeN:deN:neN
 Pfaffenschlag b.Waidhofen a.d. Thaya,Pfaffenschlag bei Waidhofen an der Thaya.EN+Topon+Oikon+DetZ:aeN:deN:neN
 Pfaffenschlag b.Waidhofen ad. Thaya,Pfaffenschlag bei Waidhofen an der Thaya.EN+Topon+Oikon+DetZ:aeN:deN:neN
 Pfaffenschlag b. Waidhofen ad Th,Pfaffenschlag bei Waidhofen an der Thaya.EN+Topon+Oikon+DetZ:aeN:deN:neN
 Pfaffenschlag b.Waidhofen a.d.Th.,Pfaffenschlag bei Waidhofen an der Thaya.EN+Topon+Oikon+DetZ:aeN:deN:neN
 Pfaffenschlag bei Waidhofen an der Th.,Pfaffenschlag bei Waidhofen an der Thaya.EN+Topon+Oikon+DetZ:aeN:deN:neN
 Pfaffenschlag b.Waidhofen an d. T.,Pfaffenschlag bei Waidhofen an der Thaya.EN+Topon+Oikon+DetZ:aeN:deN:neN
 Pfaffenschlags bei Waidhofen an d. T.,Pfaffenschlag bei Waidhofen an der Thaya.EN+Topon+Oikon+DetZ:geN
 Pfaffenschlag bei Waidhofen a.d.Th.,Pfaffenschlag bei Waidhofen an der Thaya.EN+Topon+Oikon+DetZ:aeN:deN:neN
 Pfaffenschlag b.Waidhofen an d.Th.,Pfaffenschlag bei Waidhofen an der Thaya.EN+Topon+Oikon+DetZ:aeN:deN:neN
 Pfaffenschlags b.Waidhofen ad.T.,Pfaffenschlag bei Waidhofen an der Thaya.EN+Topon+Oikon+DetZ:geN
 Pfaffenschlag b.Waidhofen an der Thaya,Pfaffenschlag bei Waidhofen an der Thaya.EN+Topon+Oikon+DetZ:aeN:deN:neN
 Pfaffenschlag b.Waidhofen ad Th,Pfaffenschlag bei Waidhofen an der Thaya.EN+Topon+Oikon+DetZ:aeN:deN:neN
 Pfaffenschlags b.Waidhofen ad Thaya,Pfaffenschlag bei Waidhofen an der Thaya.EN+Topon+Oikon+DetZ:geN
 Pfaffenschlags bei Waidhofen\Thaya,Pfaffenschlag bei Waidhofen an der Thaya.EN+Topon+Oikon+DetZ:geN
 Pfaffenschlags b.Waidhofen a.d.T.,Pfaffenschlag bei Waidhofen an der Thaya.EN+Topon+Oikon+DetZ:geN
 Pfaffenschlag bei Waidhofen a.d. T.,Pfaffenschlag bei Waidhofen an der Thaya.EN+Topon+Oikon+DetZ:aeN:deN:neN
 Pfaffenschlag b. Waidhofen (Th.),Pfaffenschlag bei Waidhofen an der Thaya.EN+Topon+Oikon+DetZ:aeN:deN:neN
 Rügen, .EN+Topon+Choron+admin+Cl (Landkreis)+oblCl:nxU
 Rügen,Rügen.EN+Topon+Choron+Cl (Insel)+DetZ:geNS
 Rügen, .EN+Topon+Choron+Cl (Insel)+DetZ:aeN:deN:neN
 rügen, .FF+Topon+Choron+Cl (Insel)
 Rügener, .ADJ+Topon+Choron+Cl (Insel)+der (Rügen):X
 Rügener, .N+Hum+Topon+Choron+Cl (Insel)+der (Rügen):aeM:amM:deM:gmM:neM:nmM
 Rügerner,Rügener.N+Hum+Topon+Choron+Cl (Insel)+der (Rügen):dmM
 Rügeners,Rügener.N+Hum+Topon+Choron+Cl (Insel)+der (Rügen):geMS
 Rügenerin, .N+Hum+Topon+Choron+Cl (Insel)+der (Rügen):aeF:deF:geF:neF
 Rügenerinnen,Rügenerin.N+Hum+Topon+Choron+Cl (Insel)+der (Rügen):amF:dmF:gmF:nmF
 rügensche,rügensch.ADJ+Topon+Choron+Cl (Insel)+der (Rügen):aeFxp:aeFyp:aeFzp:aeNyp:amUxp:neFxp:neFyp:neFzp: . . .
 rügenschen,rügensch.ADJ+Topon+Choron+Cl (Insel)+der (Rügen):aeMxp:aeMyp:aeMzp:amUyp:amUzp:deFyp:deFzp:deMyp: . . .
 rügenscher,rügensch.ADJ+Topon+Choron+Cl (Insel)+der (Rügen):deFxp:geFxp:gmUxp:neMxp:neMzp
 rügensch, .ADJ+Topon+Choron+Cl (Insel)+der (Rügen):up
 rügenschem,rügensch.ADJ+Topon+Choron+Cl (Insel)+der (Rügen):deMxp:deNxp
 rügensches,rügensch.ADJ+Topon+Choron+Cl (Insel)+der (Rügen):aeNxp:aeNzp:neNxp:neNzp
 Rüger, .ADJ+Topon+Choron+Cl (Insel)+der (Rügen):X

Rüganer, .ADJ+Topon+Choron+Cl(Insel)+der(Rügen):X
Rüganers,Rüganer.N+Hum+Topon+Choron+Cl(Insel)+der(Rügen):geMS
Rüganern,Rüganer.N+Hum+Topon+Choron+Cl(Insel)+der(Rügen):dmM
Rüganer, .N+Hum+Topon+Choron+Cl(Insel)+der(Rügen):aeM:amM:deM:gmM:neM:nmM
Rüganerin, .N+Hum+Topon+Choron+Cl(Insel)+der(Rügen):aeF:deF:geF:neF
Rüganerinnen,Rüganerin.N+Hum+Topon+Choron+Cl(Insel)+der(Rügen):amF:dmF:gmF:nmF
Rugianer, .N+Hum+Topon+Choron+Cl(Insel)+der(Rügen):aeM:amM:deM:gmM:neM:nmM
Rugianers,Rugianer.N+Hum+Topon+Choron+Cl(Insel)+der(Rügen):geMS
Rugianern,Rugianer.N+Hum+Topon+Choron+Cl(Insel)+der(Rügen):dmM
Rugianerinnen,Rugianerin.N+Hum+Topon+Choron+Cl(Insel)+der(Rügen):amF:dmF:gmF:nmF
Rugianerin, .N+Hum+Topon+Choron+Cl(Insel)+der(Rügen):aeF:deF:geF:neF
St. Vincents und der Grenadinen,Sankt Vincent und die Grenadinen.EN+Topon+Choron+admin+Cl(Staat)+DetZ:gmUy
Sankt Vincent und den Grenadinen,Sankt Vincent und die Grenadinen.EN+Topon+Choron+admin+Cl(Staat)+DetZ:dmUy
St. Vincents & Grenadinen,Sankt Vincent und die Grenadinen.EN+Topon+Choron+admin+Cl(Staat)+DetZ:gmUy
Sankt Vincent und die Grenadinen, .EN+Topon+Choron+admin+Cl(Staat)+DetZ:amUy:nmUy
Sankt Vincents und der Grenadinen,Sankt Vincent und die Grenadinen.EN+Topon+Choron+admin+Cl(Staat)+DetZ:gmUy
St.Vincent und die Grenadinen,Sankt Vincent und die Grenadinen.EN+Topon+Choron+admin+Cl(Staat)+DetZ:amUy:nmUy
St. Vincent & Grenadinen,Sankt Vincent und die Grenadinen.EN+Topon+Choron+admin+Cl(Staat)+DetZ:amUy:nmUy:dmUy
St. Vincent und den Grenadinen,Sankt Vincent und die Grenadinen.EN+Topon+Choron+admin+Cl(Staat)+DetZ:dmUy
St.Vincents und der Grenadinen,Sankt Vincent und die Grenadinen.EN+Topon+Choron+admin+Cl(Staat)+DetZ:gmUy
St.Vincents & Grenadinen,Sankt Vincent und die Grenadinen.EN+Topon+Choron+admin+Cl(Staat)+DetZ:gmUy
St Vincent & Grenadinen,Sankt Vincent und die Grenadinen.EN+Topon+Choron+admin+Cl(Staat)+DetZ:dmUy:amUy:nmUy
St.Vincent & Grenadinen,Sankt Vincent und die Grenadinen.EN+Topon+Choron+admin+Cl(Staat)+DetZ:amUy:nmUy:dmUy
St Vincents und der Grenadinen,Sankt Vincent und die Grenadinen.EN+Topon+Choron+admin+Cl(Staat)+DetZ:gmUy
St. Vincent und die Grenadinen,Sankt Vincent und die Grenadinen.EN+Topon+Choron+admin+Cl(Staat)+DetZ:amUy:nmUy
St Vincents & Grenadinen,Sankt Vincent und die Grenadinen.EN+Topon+Choron+admin+Cl(Staat)+DetZ:gmUy
Sankt Vincent & Grenadinen,Sankt Vincent und die Grenadinen.EN+Topon+Choron+admin+Cl(Staat)+DetZ:amUy:nmUy:dmUy
Sankt Vincents & Grenadinen,Sankt Vincent und die Grenadinen.EN+Topon+Choron+admin+Cl(Staat)+DetZ:gmUy
St.Vincent und den Grenadinen,Sankt Vincent und die Grenadinen.EN+Topon+Choron+admin+Cl(Staat)+DetZ:dmUy
St Vincent und die Grenadinen,Sankt Vincent und die Grenadinen.EN+Topon+Choron+admin+Cl(Staat)+DetZ:amUy:nmUy
St Vincent und den Grenadinen,Sankt Vincent und die Grenadinen.EN+Topon+Choron+admin+Cl(Staat)+DetZ:dmUy
Schwarzes Meer, .EN+Topon+Hydron+Cl(Meer)+Det+adj:aeNx:aeNz:neNx:neNz
Schwarzen Meer,Schwarzes Meer.EN+Topon+Hydron+Cl(Meer)+Det+adj:deNy:deNz
Schwarzen Meere,Schwarzes Meer.EN+Topon+Hydron+Cl(Meer)+Det+adj:deNy:deNz
Schwarzem Meer,Schwarzes Meer.EN+Topon+Hydron+Cl(Meer)+Det+adj:deNx
Schwarzem Meere,Schwarzes Meer.EN+Topon+Hydron+Cl(Meer)+Det+adj:deNx
Schwarzen Meers,Schwarzes Meer.EN+Topon+Hydron+Cl(Meer)+Det+adj:geNx:geNy:geNz
Schwarzen Meeres,Schwarzes Meer.EN+Topon+Hydron+Cl(Meer)+Det+adj:geNx:geNy:geNz
Schwarze Meer,Schwarzes Meer.EN+Topon+Hydron+Cl(Meer)+Det+adj:aeNy:neNy
Slubice,Słubice.EN+Topon+Oikon+DetZ:aeN:deN:neN
Slubices,Słubice.EN+Topon+Oikon+DetZ:geNS
Słubices,Słubice.EN+Topon+Oikon+DetZ:geNS
Słubice, .EN+Topon+Oikon+DetZ:aeN:deN:neN
Słubicer, .ADJ+Topon+Oikon+der(Słubice):X
Slubicer,Słubicer.ADJ+Topon+Oikon+der(Słubice):X
Słubicer, .N+Hum+Topon+Oikon+der(Słubice):aeM:amM:deM:gmM:neM:nmM
Slubicern,Słubicer.N+Hum+Topon+Oikon+der(Słubice):dmM
Slubicer,Słubicer.N+Hum+Topon+Oikon+der(Słubice):aeM:amM:deM:gmM:neM:nmM
Słubicers,Słubicer.N+Hum+Topon+Oikon+der(Słubice):geMS
Slubicers,Słubicer.N+Hum+Topon+Oikon+der(Słubice):geMS
Slubicern,Słubicer.N+Hum+Topon+Oikon+der(Słubice):dmM
Słubicerinnen,Słubicerin.N+Hum+Topon+Oikon+der(Słubice):amF:dmF:gmF:nmF
Słubicerin, .N+Hum+Topon+Oikon+der(Słubice):aeF:deF:geF:neF
Slubicerin,Słubicerin.N+Hum+Topon+Oikon+der(Słubice):aeF:deF:geF:neF
Slubicerinnen,Słubicerin.N+Hum+Topon+Oikon+der(Słubice):amF:dmF:gmF:nmF
Unna, .EN+Topon+Choron+admin+Cl(Landkreis)+oblCl:nxU
Unnas,Unna.EN+Topon+Oikon+DetZ:geNS
Unna, .EN+Topon+Oikon+DetZ:aeN:deN:neN
Unnaer, .ADJ+Topon+Oikon+der(Unna):X
Unnaern,Unnaer.N+Hum+Topon+Oikon+der(Unna):dmM
Unnaers,Unnaer.N+Hum+Topon+Oikon+der(Unna):geMS
Unnaer, .N+Hum+Topon+Oikon+der(Unna):aeM:amM:deM:gmM:neM:nmM
Unnaerinnen,Unnaerin.N+Hum+Topon+Oikon+der(Unna):amF:dmF:gmF:nmF
Unnaerin, .N+Hum+Topon+Oikon+der(Unna):aeF:deF:geF:neF
Unneraner, .ADJ+Topon+Oikon+der(Unna):X
Unneraner, .N+Hum+Topon+Oikon+der(Unna):aeM:amM:deM:gmM:neM:nmM
Unneranern,Unneraner.N+Hum+Topon+Oikon+der(Unna):dmM
Unneraners,Unneraner.N+Hum+Topon+Oikon+der(Unna):geMS
Unneranerin, .N+Hum+Topon+Oikon+der(Unna):aeF:deF:geF:neF
Unneranerinnen,Unneranerin.N+Hum+Topon+Oikon+der(Unna):amF:dmF:gmF:nmF
Wedding, .EN+Topon+Oikon+Det:aeM:deM:geM:neM
Weddings,Wedding.EN+Topon+Oikon+Det:geMS

Weddinger, .ADJ+Topon+Oikon+der(Wedding):X
 Weddinger, .N+Hum+Topon+Oikon+der(Wedding):aeM:amM:deM:gmM:neM:nmM
 Weddingers,Weddinger.N+Hum+Topon+Oikon+der(Wedding):geMS
 Weddingern,Weddinger.N+Hum+Topon+Oikon+der(Wedding):dmM
 Weddingerinnen,Weddingerin.N+Hum+Topon+Oikon+der(Wedding):amF:dmF:gmF:nmF
 Weddingerin, .N+Hum+Topon+Oikon+der(Wedding):aeF:deF:geF:neF
 Wedding, .EN+Topon+Oikon+var(Wedding)+DetZ:aeN:deN:neN
 Weddings,Wedding.EN+Topon+Oikon+var(Wedding)+DetZ:geNS
 Wiener Neustadt, .EN+Topon+Oikon+Det:aeF:deF:geF:neF/ selbständige Stadt in der Nähe von Wien, kein Stadtteil!
 Wiener Neustadts,Wiener Neustadt.EN+Topon+Oikon+var(Wiener Neustadt)+DetZ:geN
 Wiener Neustadt, .EN+Topon+Oikon+var(Wiener Neustadt)+DetZ:aeN:deN:neN

A.1.3 Regeln und Pattern zur Abkürzung

Die Ableitung der abgekürzten Formen und Varianten (siehe Kap. 3.3.2.7 auf S. 116) erfolgt ausgehend von den Flexionsformen eines Toponyms. Die dazu notwendigen Regeln sind als geordnete Menge realisiert, d.h. mehrere Regeln können nacheinander, aber in wohldefinierter Reihenfolge, auf die Form eines Toponyms angewendet werden. Dies ermöglicht, die Menge der Ersetzungsregeln klein zu halten, auch dann, wenn viele Varianten zu beschreiben sind. So können in *Sankt Johann in Tirol* die folgenden Bestandteile auf z.T. unterschiedliche Weise abgekürzt werden: *Sankt, in, Tirol*. Ein Leerzeichen nach dem Abkürzungspunkt ist dabei optional. Darüberhinaus ist es möglich *in* durch die Symbole *'/'* und *'(...)'* zu substituieren. In einem deutschen Korpus mit knapp 6 Milliarden Wortformen (siehe Tab. E.2 auf S. 441) sind von den mind. 48 durch die Kombination der Teilabkürzungen möglichen Formen die folgenden Formen und Abkürzungen auch mit folgender Häufigkeit belegt:

720	St. Johann in Tirol
244	St.Johann i.T.
186	St.Johann in Tirol
107	St. Johann/Tirol
101	St. Johann i. T.
61	Sankt Johann in Tirol
59	St. Johann i.T.
45	St. Johann/T.
36	St.Johann/Tirol
21	St Johann in Tirol
20	St.Johann/T.
16	St. Johann i. Tirol
10	St.Johann (Tirol)
7	St. Johann (Tirol)
2	Sankt Johann/Tirol
2	St. Johann in T.
2	St.Johann in T.
2	St. Johann i.Tirol
1	Sankt Johann (Tirol)
1	St.Johann i. T.
1	St.Johann i. Tirol
1	St.Johann i.Tirol

Die Abkürzungsregeln sind eine geordnete Menge von Regelblöcken, jeweils bestehend aus einem Pattern und Ersetzungsregeln, die wie folgt angewendet werden:

- für jeden Regelblock $r = \langle \text{Pattern}, \text{Ersetzungsregeln} \rangle$
 - für jede Form f eines Toponyms (aus der aktuellen Menge aller Formen, die durch die Anwendung der Ersetzungsregeln nach jedem Regelblock größer werden kann)
 - * teste, ob das Pattern von r auf f zutrifft (matcht). Wenn ja:
 - für jede der zugehörigen Ersetzungsregeln (von r): wende die Ersetzungsregel an und füge die resultierende neue Form zur Menge der Formen hinzu

Zwei Ersetzungsregeln werden dabei separat nach allen anderen Regeln angewendet:

- ersetze '=' durch Bindestrich bzw. Leerzeichen ('=' wird als Metazeichen gemäß der DELA-Konvention verwendet)
- entferne auf einen Punkt folgendes Leerzeichen (.) oder Bindestrich

Diese beiden Ersetzungsregeln werden nicht global auf alle erkannten/gematchten Positionen angewendet, sondern für jede Position einzeln. Die Ersetzung erfolgt aber in allen möglichen Kombinationen. Die nicht durchgängige Ersetzung spiegelt die Realität in Texten wieder, vgl. die in deutschen Texten belegten Vorkommen (mit Frequenz) des Toponyms *Saint-Germain-en-Laye*:

474	Saint-Germain-en-Laye
48	St. Germain-en-Laye
38	St. Germain en Laye
17	St.-Germain-en-Laye
16	St-Germain-en-Laye
10	Saint Germain en Laye
7	Saint Germain-en-Laye
6	St.Germain-en-Laye
5	St Germain en Laye
4	St-Germain en Laye
3	Saint-Germain-en Laye
2	St Germain-en-Laye
2	St.-Germain en Laye
1	Saint-Germain en Laye
1	St.-Germain-en Laye

Solche Regeln erzeugen v^n Formen (mit v = Anzahl der Varianten je Position und n = Anzahl der Positionen): im obigen Fall $2^3 = 8$ Varianten für die Regel 'ersetze = durch Leerzeichen oder Bindestrich'. Durch die Kombination mit weiteren Ersetzungsregeln kann die Anzahl der kombinatorisch möglichen Formen dabei schnell sehr groß werden. Für *Saint-Germain-en-Laye* mit den drei Regeln 'ersetze =', '*Saint* → *St.* / *St*' und 'entferne Leerzeichen/Bindestrich nach Punkt' sind damit 28 Formen kombinatorisch möglich, wovon 15 auch im Korpus (siehe Tab. E.2 auf S. 441) belegt sind – die Flexionsformen und die Kurzform *Saint-Germain* mit ihren Varianten nicht mitgezählt.

A.1.3.1 Liste der Pattern und Ersetzungsregeln

Die Ersetzungsregeln sind als reguläre Ausdrücke mittels des Substitute-Operators der Programmiersprache PERL angegeben:

```
s/<Pattern>/<Ersetzung>/g;
```

(Sind mehrere Ersetzungen für ein Pattern möglich, ist das Pattern aus Gründen der Übersichtlichkeit nur einmal gegeben.)

1. *Deutschland/West* ↔ *Deutschland (West)*

```
s/\/(.+)$/ ($1)/;
```

2. *St. Gallen* ↔ *Sankt Gallen*

St. Louis ↔ *Saint Louis*

St. Maarten ↔ *Sint Maarten*

St.-Lorenz-Strom ↔ *Sankt-Lorenz-Strom*

```
s/\bS(?:ank|ain|in)t( [= -])/St.$1/g;  
/St$1/g;
```

3. *Sault Ste. Marie* ↔ *Sault Sainte Marie*

Ste.-Croix ↔ *Sainte-Croix*

```
s/\bSainte( [= -])/Ste.$1/g;  
/Ste$1/g;
```

4. *St. Johann iT* ↔ *St. Johann i. T.* ↔ *St. Johann in Tirol*

Matrei i. O. ↔ *Matrei in Osttirol*

Weilheim (Oberbayern) ↔ *Weilheim in Oberbayern*

```
s/ (i)n ((([:upper:]))[:lower:]+)$/ $1n $3./;  
/ $1 $3./;  
/ $1. $2/;  
/ $1. $3./;  
/ ($2)/;  
/ ($3.)/;  
/ $1$3/;
```

5. *Alzenau (UFr)* ↔ *Alzenau (Unterfranken)* ↔ *Alzenau in Unterfranken*

Hofheim i. UFr ↔ *Hofheim in Unterfranken*

Heiligenstadt i.OFr. ↔ *Heiligenstadt in Oberfranken*

```
s/ (i)n ([UOM])(?:nter|ber|ittel)franken$/ $1n $2Fr/;  
/ $1. $2Fr/;  
/ $1. $2Fr./;  
/ ($2Fr)/;  
/ ($2Fr.)/;
```

6. *Neumarkt (OPf) ↔ Neumarkt (Oberpfalz) ↔ Neumarkt in der Oberpfalz*
Weiden id Opf. ↔ Weiden i.d.OPf. ↔ Weiden in der Oberpfalz

s/ (i)n der Oberpfalz\$/ \$1n der OPf/;
/ \$1n d. OPf/;
/ \$1. d. OPf/;
/ \$1. d. OPf./;
/ \$1d OPf/;
/ \$1d OPf./;
/ (OPf)/;
/ (OPf.)/;

7. *Weilheim (Obb.) ↔ Weilheim (Oberbayern) ↔ Weilheim in Oberbayern*
Weilheim i. OB ↔ Weilheim in Oberbayern

s/ (i)n Oberbayern\$/ \$1n OB/;
/ \$1. OB/;
/ \$1n Obb./;
/ \$1. Obb./;
/ \$1n Obb/;
/ \$1. Obb/;
/ (OB)/;
/ (Obb.)/;
/ (Obb)/;

8. *Reichenbach OL ↔ Reichenbach in der Oberlausitz*
Rothenburg (OL) ↔ Rothenburg in der Oberlausitz

s/ in der Oberlausitz\$/ \$1n der OL/;
/ \$1. d. OL/;
/ OL/;
/, OL/;
/ (OL)/;
/ (O.L.)/;
/ (Ob. Laus.)/;
/ (Ob.-Laus.)/;

9. *Frankfurt aM ↔ Frankfurt a. M. ↔ Frankfurt am Main*
Bad Berneck i. Fichtelgebirge ↔ Bad Berneck im Fichtelgebirge
Bad Grund (Harz) ↔ Bad Grund im Harz
Lindau (B) ↔ Lindau am Bodensee

s/ ([ia]m ([[[:upper:]]][[:lower:]]+)\$/ \$1m \$3./;
/ \$1. \$2/;
/ \$1. \$3./;
/ (\$2)/;
/ (\$3.)/;
/ (\$3)/;
/ \$1\$3/;

10. *Freiburg i. Br.* ↔ *Freiburg i. Breisg.* ↔ *Freiburg im Breisg.* ↔ *Freiburg im Breisgau*
Gutach i. Br. ↔ *Gutach im Breisgau*

s/ im Breisgau\$/ im Breisg./;
/ im Br./;
/ i. Breisg./;
/ i. Br./;
/ i. Br/;
/ (Breisg.)/;
/ (Br.)/;
/ (Br)/;

11. *Hilter a. T. W.* ↔ *Hilter am TW* ↔ *Hilter am Teutoburger Wald*
Dissen aTW ↔ *Dissen am Teutoburger Wald*

s/ ([ia]m ([[[:upper:]]][[:lower:]]+)) ((W)alde?s?)\$/ \$1m \$3. \$5./;
/ \$1. \$2 \$4/;
/ \$1. \$3. \$5./;
/ \$1m \$3\$5/;
/ \$1\$3\$5/;
/ (\$2 \$4)/;
/ (\$3. \$5.)/;

12. *Zimmern o. R.* ↔ *Zimmern ob Rottweil*
Neuhausen o. E. ↔ *Neuhausen ob Eck*

s/ (o)b ([[[:upper:]]][[:lower:]]+)\$/ \$1b \$3./;
/ \$1. \$2/;
/ \$1. \$3./;
/ (\$2)/;
/ (\$3.)/;

13. *Haibach o.d.D.* ↔ *Haibach ob der Donau*
Rothenburg o.d. Tauber ↔ *Rothenburg ob der Tauber*

s/ (o)b der ([[[:upper:]]][[:lower:]]+)\$/ \$1b d. \$3./;
/ \$1.d. \$2/;
/ \$1.d. \$3./;
/ (\$2)/;
/ (\$3.)/;

14. *Frankfurt a.d. Oder* ↔ *Frankfurt an der Oder*
Frankfurt (Oder) ↔ *Frankfurt an der Oder*
Bad Neustadt (Saale) ↔ *Bad Neustadt an der Saale*
Halle (S.) ↔ *Halle an der Saale*
St. Michael id Wachau ↔ *Sankt Michael in der Wachau*
Au i.d. Hallertau ↔ *Au in der Hallertau*
Altenberg ad. Rax ↔ *Altenberg an der Rax*

s/ ([ia])n der (([[:upper:]] [[:lower:]]+)\$/ \$1n d. \$3./;
 / \$1.d. \$2/;
 / \$1.d. \$3./;
 / \$1d. \$3./;
 / \$1d. \$2/;
 / \$1d \$2/;
 / \$1d \$3./;
 / (\$2)/;
 / (\$3.)/;

15. *Waidhofen (Th.)* ↔ *Waidhofen an der Thaya*

Ellikon a.d.Th. ↔ *Ellikon a.d. Thur* ↔ *Ellikon an der Thur*

s/ an der (Th[[:lower:]]+)\$/ an der Th./;
 / an d. Th./;
 / a.d. Th./;
 / ad. Th./;
 / ad Th/;
 / (Th.)/;
 / (Th)/;

16. *Wyk/Föhr* ↔ *Wyk auf Föhr*

Burg a. F. ↔ *Burg a. Fehmarn* ↔ *Burg auf Fehmarn*

Kölpinsee (Usedom) ↔ *Kölpinsee auf Usedom*

s/ (a)uf (([[:upper:]] [[:lower:]]+)\$/ \$1 \$3./;
 / \$1. \$2/;
 / \$1. \$3./;
 / (\$2)/;
 / (\$3.)/;

17. *Bernstadt adE* ↔ *Bernstadt a.d.E.* ↔ *Bernstadt a. d. Eigen* ↔ *Bernstadt auf dem Eigen*

Bernstadt (Eigen) ↔ *Bernstadt auf dem Eigen*

Ruit adF ↔ *Ruit a. d. Fildern* ↔ *Ruit auf den Fildern*

s/ (a)uf de[mn] (([[:upper:]] [[:lower:]]+)\$/ auf d. \$3./;
 / \$1.d. \$2/;
 / \$1.d. \$3./;
 / \$1d. \$3./;
 / \$1d \$2/;
 / \$1d \$3./;
 / (\$2)/;
 / (\$3.)/;
 / \$1d\$3/;

18. *Kirchheim (Teck)* ↔ *Kirchheim u. Teck* ↔ *Kirchheim unter Teck*

Eningen u. A. ↔ *Eningen unter Achalm*

s/ (u)nter (([[:upper:]] [[:lower:]]+)\$ / \$1 \$3./;
 / \$1. \$2/;
 / \$1. \$3./;
 / (\$2)/;
 / (\$3.)/;

19. *Neukirchen v. Wald* ↔ *Neukirchen vorm Wald*

Aicha v. W. ↔ *Aicha v. Wald* ↔ *Aicha vorm Wald*

s/ (v)orm (([[:upper:]] [[:lower:]]+)\$ / \$1. \$2/;
 / \$1. \$3./;
 / (\$2)/;
 / (\$3.)/;

20. *Garching b. München* ↔ *Garching bei München*

Wentorf b. Hamburg ↔ *Wentorf bei Hamburg*

s/ bei / b. /g;

21. *Antigua & Barbuda* ↔ *Antigua und Barbuda*

Saint Vincent & Grenadinen ↔ *Saint Vincent und die Grenadinen*

s/ und (? :d(? :e[rmns]|ie|as))? / & /g;

22. *Bayrischer Wald* ↔ *Bayerischer Wald*

Bayrischer Plöckenstein ↔ *Bayerischer Plöckenstein*

s/(Bayerisch)(?=(? :e[mnrs])?[= -])/Bayrisch/g;

23. *Oberberg. Kreis* ↔ *Oberbergischer Kreis*

Rhein.-Berg. Kreis ↔ *Rheinisch-Bergischer Kreis*

Bayer. Wald ↔ *Bayerischer Wald*

Märk. Kreis ↔ *Märkischer Kreis*

s/isch(? :e[mnrs])?(?=[= -])/./g;

24. *Wrtbg* ↔ *Wttbg.* ↔ *Württbg* ↔ *Württembg* ↔ *Württemberg*

Baden-Wbg ↔ *Baden-Württemberg*

Maulbronn (Wttbg.) ↔ *Maulbronn (Württemberg)*

s/([Ww])ürttemberg\b/\$1ürttembg./g;
 /\$1ürttembg/g;
 /\$1ürttbg./g;
 /\$1ürttbg/g;
 /\$1ürttbg./g;
 /\$1ttbg/g;
 /\$1ttbg./g;
 /\$1rtbg/g;
 /\$1rtbg./g;
 /\$1bg./g;
 /\$1bg/g;

25. *Bd.-Wttbg.* ↔ *Baden-Wttbg.*
 $s/\text{Baden}(?=[-] [\text{Ww}]\ddot{u}?[rt]*bg)\b/Bd./g;$
 $\quad\quad\quad /Bd/g;$
26. *AR Krim* ↔ *Autonome Republik Krim*
AR Nachitschewan ↔ *Autonome Republik (AR) Nachitschewan* ↔ *Autonome Republik Nachitschewan*
 $s/\backslash\text{autonome}[rn]? \text{Republik}\b/AR/ig;$
 $\quad\quad\quad /Autonome \text{Republik (AR)}/ig;$
27. *DR Kongo* ↔ *Demokratische Republik Kongo*
DR Timor-Leste ↔ *Demokratische Republik Timor-Leste*
 $s/\backslash\text{demokratische}[rn]? \text{Republik}\b/DR/ig;$
28. *Adscharische ASSR* ↔ *Adscharische Autonome Sozialistische Sowjetrepublik*
 $s/\backslash\text{autonome}[rn]? \text{sozialistische}[rn]? \text{Sowjetrepublik}\b/ASSR/ig;$
29. *VR China* ↔ *Volksrepublik China*
Demokratische VR Korea ↔ *Demokratische Volksrepublik Korea*
 $s/\backslash\text{Volksrepublik}\b/VR/g;$
30. *DVR Korea* ↔ *Demokratische Volksrepublik Korea*
 $s/\backslash\text{Demokratische}[rn]? \text{Volksrepublik}\b/DVR/g;$
31. *Frankfurt/Main* ↔ *Frankfurt (Main)*
Halle/S. ↔ *Halle (S.)*
 $s/\backslash s*\backslash((.+?)\backslash)\$/\backslash \$1/;$
 $\quad\quad\quad / \backslash / \$1/;$

A.1.3.2 Beispiele für die sequentielle Anwendung mehrerer Regeln

Die sequentielle Anwendung von Regeln erlaubt eine große Menge an Varianten mit verhältnismäßig wenigen Regeln zu beschreiben. Die Ergebnisse aus einer Regel können fakultativ durch weitere Regeln verändert werden, das Ergebnis wird der Menge der Varianten hinzugefügt. *Sault Ste. Marie* ↔ *Sault Ste. Marie* ↔ *Sault Sainte Marie*

St. Johann i. T. ↔ *St. Johann i. T.* ↔ *St. Johann in Tirol* ↔ *Sankt Johann in Tirol*

Weilheim/Obb ↔ *Weilheim (Obb)* ↔ *Weilheim in Oberbayern*

Rennersdorf/O.L. ↔ *Rennersdorf (O.L.)* ↔ *Rennersdorf in der Oberlausitz*

Frankfurt a.M. ↔ *Frankfurt a. M.* ↔ *Frankfurt am Main*

Frankfurt/M. ↔ *Frankfurt (M.)* ↔ *Frankfurt (Main)* ↔ *Frankfurt am Main*

St. Ruprecht/Murau ↔ *St. Ruprecht (Murau)* ↔ *St. Ruprecht ob Murau* ↔ *Sankt Ruprecht ob Murau*

Frankfurt a.d.O. ⇔ Frankfurt a.d. Oder ↔ Frankfurt an der Oder

Frankfurt/O. ⇔ Frankfurt (O.) ↔ Frankfurt an der Oder

Bayr. Wald ⇔ Bayrischer Wald ↔ Bayerischer Wald

Bd.-Württembg. ↔ Baden-Württembg. ↔ Baden-Württemberg

Bd.-Wttbg. ⇔ Baden-Wttbg. ↔ Baden-Württemberg

A.1.4 Derivationsuffixe für Adjektive und Bewohner

Die folgende Tabelle listet alle Suffixoperationen, mittels derer von einem gegebenen Toponym die Bezeichnungen für Adjektive und Bewohner abgeleitet werden können. Zu den Details siehe Kap. 3.3.2.11.

Suffixop.	Typ	Flexion	Frq.	gem. Sfx.	Beispiele
0 er	Adj.	ADJ00	2462		Berliner, Schweizer, Hamburger
0 er	Bew.	NS2;NP1	2414		Berliner, Schweizer, Hamburger
0 erin	B-in	NS0;NP5	2398		Berlinerin, Schweizerin, Münchenerin
2 er	Adj.	ADJ00	273	/[alnrs]\$/	Bremer, Thüringer, Saarbrücker
2 erin	B-in	NS0;NP5	251	/[anrs]\$/	Norwegerin, Bremerin, Thüringerin
2 er	Bew.	NS2;NP1	250	/[anrs]\$/	Bremer, Thüringer, Saarbrücker
1 er	Bew.	NS2;NP1	150	/[aehinos]\$/	Karlsruher, Berner (Berne), Auer
1 erin	B-in	NS0;NP5	148	/[aehinos]\$/	Ukrainerin, Karlsruherin, Bernerin (Berne)
0 isch	Adj.	ADJ	126		italienisch, japanisch, berlinisch
3 isch	Adj.	ADJ	116	/[adns]\$/	englisch, spanisch, arabisch (Arabien)
1 er	Adj.	ADJ00	110	/[aehnos]\$/	Karlsruher, Berner (Berne), Auer
" 0 er	Bew.	NS2;NP1	104	/[adkmstz]\$/	Römer, Europäer, Engländer
" 0 erin	B-in	NS0;NP5	104	/[adkmstz]\$/	Engländerin, Europäerin, Holländerin
3 ier	Bew.	NS2;NP1	77	/i(en/us)\$/	Spanier, Australier, Belgier
3 ierin	B-in	NS0;NP5	77	/i(en/us)\$/	Spanierin, Australierin, Belgierin
1 isch	Adj.	ADJ	74	/[aeinosuá]\$/	bayerisch, ungarisch, ukrainisch
2 nerin	B-in	NS0;NP5	73	/[um(en/us)\$/	Münchenerin, Dresdnerin, Taufkirchnerin
" 0 er	Adj.	ADJ00	70	/[dkstz]\$/	Darmstädter, Mailänder, Ingolstädter
2 ner	Bew.	NS2;NP1	69	/[um(en/us)\$/	Münchner, Dresdner, Kärntner
2 ner	Adj.	ADJ00	68	/[en/us]\$/	Münchner, Dresdner, Kärntner
2 isch	Adj.	ADJ	67	/[aiknors]\$/	türkisch, schwedisch, slowakisch
" 0 isch	Adj.	ADJ	64	/[adkmrstz]\$/	europäisch, römisch, holländisch
0 nerin	B-in	NS0;NP5	59	/[aen]\$/	Amerikanerin, Afrikanerin, Südafrikanerin
0 ner	Bew.	NS2;NP1	58	/[aen]\$/	Amerikaner, Afrikaner, Südafrikaner

Tabelle A.3: Suffixoperationen zur Bildung von Toponymderivativen

Suffixop.	Typ	Flexion	Frq.	gem. Sfx.	Beispiele
0nisch	Adj.	ADJ	51	/[aeo]\$/	amerikanisch, spartanisch, afrikanisch
0aner	Bew.	NS2;NP1	47	/[eiklɪrtuɪ]\$/	Hannoveraner, Münsteraner, Peruaner
0anerin	B-in	NS0;NP5	47	/[eiklɪrtuɪ]\$/	Hannoveranerin, Münsteranerin, Bundesrepublikanerin (Bundesrepublik)
2ler	Bew.	NS2;NP1	33	/e[ln]\$/	Basler, Kochler, Waldviertler (Waldviertel)
2lerin	B-in	NS0;NP5	33	/e[ln]\$/	Baslerin, Kochlerin, Südstaatlerin (Südstaaten)
3e	Bew.	NS4;NP4	33	/i(ka/en)\$/	Kroate, Tscheche, Serbe
2e	Bew.	NS4;NP4	32	/e(i ik/en)\$/	Hesse, Böhme, Pole
2ler	Adj.	ADJ00	32	/e[ln]\$/	Basler, Kochler, Waldviertler (Waldviertel)
3in	B-in	NS0;NP5	28	/i(ka/en)\$/	Tschechin, Rumänin, Kroatin
0aner	Adj.	ADJ00	25	/e[lr]\$/	Hannoveraner, Münsteraner, Kasseler
2in	B-in	NS0;NP5	24	/e(i ik/en)\$/	Türkin, Schwedin, Polin
" 2er	Adj.	ADJ00	23	-en	Recklinghäuser, Mühlhäuser, Sachsenhäuser
" 2er	Bew.	NS2;NP1	23	-en	Recklinghäuser, Mühlhäuser, Sachsenhäuser
" 2erin	B-in	NS0;NP5	23	-en	Recklinghäuserin, Sachsenhäuserin, Mühlhäuserin (Mühlhausen)
4isch	Adj.	ADJ	20	/[dkmn]\$/	russisch, finnisch, irisch
1esisch	Adj.	ADJ	19	/[aio]\$/	chinesisch, locarnesisch (Locarno), maltesisch
3erin	B-in	NS0;NP5	19	-n	Inderin, Albanerin, Pommerin (Pommern)
1ese	Bew.	NS4;NP4	18	/[aio]\$/	Chinese, Veronese, Maltese
1esin	B-in	NS0;NP5	18	/[aio]\$/	Chinesin, Locarnesin (Locarno), Barcelonesin (Barcelona)
3er	Bew.	NS2;NP1	18	-n	Albaner, Araber (Arabien), Inder
4e	Bew.	NS4;NP4	18	/[nd um an]\$/	Russe, Schotte, Finne
4in	B-in	NS0;NP5	18	/[nd um an]\$/	Russin, Irin, Finnin
0anisch	Adj.	ADJ	16	/[eiklɪrtuɪ]\$/	hannoveranisch (Hannover), bundesrepublikanisch, peruanisch
0erisch	Adj.	ADJ	16	/[glɪnsuz]\$/	berlinerisch, schweizerisch, bernerisch (Bern)
0esisch	Adj.	ADJ	16	/[lɪmrt]\$/	vietnamesisch, kantonesisch, siamesisch (Siam)
0esin	B-in	NS0;NP5	14	/[lɪmrt]\$/	Vietnamesin, Kantonessin (Kanton), Taiwanessin
0ese	Bew.	NS4;NP4	13	/[lɪmrt]\$/	Vietnamese, Sudanese, Siamese (Siam)
1in	B-in	NS0;NP5	12	-rn	Ungarin, Bayerin, Niederbayerin
" 2isch	Adj.	ADJ	12	-en	schwäbisch, sächsisch, niedersächsisch

Tabelle A.3: Suffixoperationen zur Bildung von Toponymderivationen

Suffixop.	Typ	Flexion	Frq.	gem. Sfx.	Beispiele
3 risch	Adj.	ADJ	12	-ern	bayrisch, bairisch (Baiern), oberbayrisch (Oberbayern)
" 2 in	B-in	NS0;NP5	11	-en	Schwäbin, Sächsin, Niedersächsin
5 isch	Adj.	ADJ	11	/[ɛknt]\$/	dänisch, sowjetisch, kurdisch
0 i	Bew.	NS13;NP6	10	/[bhklnrt]\$/	Israeli, Pakistani, Nepali
0 sch	Adj.	ADJ	10	/[egnru]\$/	hannoversch, münstersch (Münster), wiesbadensch (Wiesbaden)
1	Bew.	NS4;NP4	10	-n	Baier, Niederbayer, Oberbayer
3 ischerseits	Adv.		10	/[and/iɛn]\$/	spanischerseits (Spanien), englischerseits (England), tschechischerseits (Tschechien)
4	Adj.	ADJ	10	-deutschland	deutsch, norddeutsch, ostdeutsch
0 i	B-in	NS0;NP6	9	/[bhklnrt]\$/	Israeli, Pakistani, Nepali
5 in	B-in	NS0;NP5	9	/[rk][ae]n]\$/	Kurdin, Dänin, Afghanin (Afghanistan)
1 iner	Bew.	NS2;NP1	8	/[aeo]\$/	Montenegriner, Kairiner (Kairo), Grenadiner (Grenada)
1 inerin	B-in	NS0;NP5	8	/[aeo]\$/	Montenegrinerin, Kairinerin (Kairo), Grenadinerin (Grenada)
1 inisch	Adj.	ADJ	8	/[aeo]\$/	kairinisch (Kairo), montenegrinisch (Montenegro), grenadinisch (Grenada)
2 aner	Bew.	NS2;NP1	8	-en	Brasilianer, Kolumbianer, Sizilianer
2 anerin	B-in	NS0;NP5	8	-en	Brasilianerin, Kolumbianerin, Sizilianerin
5 e	Bew.	NS4;NP4	8	/[rk][ae]n]\$/	Däne, Kurde, Afghane
0 ler	Bew.	NS2;NP1	7	/[DRWdm]\$/	DDRler, NRWler, BRDler
0 lerin	B-in	NS0;NP5	7	/[DRWdm]\$/	DDRlerin, NRWlerin (NRW), BRDlerin (BRD)
1 aner	Bew.	NS2;NP1	7	/[nosá]\$/	Mexikaner, Marokkaner, Fehmaraner (Fehmarn)
1 anerin	B-in	NS0;NP5	7	/[nosá]\$/	Mexikanerin, Marokkanerin, Pommeranerin (Pommern)
1 eser	Bew.	NS2;NP1	7	/[ai]\$/	Malteser, Genueser, Veroneser
1 eserin	B-in	NS0;NP5	7	/[ai]\$/	Malteserin, Genueserin, Veroneserin (Verona)
4 e	B-in	ADJ	7	-deutschland	Deutsche, Süddeutsche, Westdeutsche
4 e	Bew.	ADJ	7	-deutschland	Deutsche, Süddeutsche, Westdeutsche
0 itisch	Adj.	ADJ	6	/[ɛlɪmn]\$/	israelitisch, jemenitisch, bethlehemitisch (Bethlehem)
1 eser	Adj.	ADJ00	6	/[ai]\$/	Genueser, Veroneser, Sieneser
2 anisch	Adj.	ADJ	6	-en	brasilianisch, sizilianisch, kolumbianisch
2 atisch	Adj.	ADJ	6	-asien	asiatisch, ostasiatisch, südostasiatisch (Südostasien)

Tabelle A.3: Suffixoperationen zur Bildung von Toponymderivativen

Suffixop.	Typ	Flexion	Frq.	gem. Sfx.	Beispiele
2 ier	Bew.	NS2;NP1	6	/[ia/en/os]\$/	Malaysier, Namibier, Kapverdier (Kapverden)
2 ierin	B-in	NS0;NP5	6	/[ia/en/os]\$/	Malaysierin, Namibierin (Namibia), Barbadierin (Barbados)
3 rer	Bew.	NS2;NP1	6	-ern	Benediktbeurer, Zyprer, Reichersbeurer (Reichersbeuern)
3 rerin	B-in	NS0;NP5	6	-ern	Benediktbeurerin (Benediktbeuern), Zyprerin (Zypern), Reichersbeurerin (Reichersbeuern)
0 in	B-in	NS0;NP5	5	/[anr]\$/	Altomünsterin, Altonain (Altona), Omanin (Oman)
0 ischerseits	Adv.		5	/[hln]\$/	österreichischerseits (Österreich), italienischerseits (Italien), japanischerseits (Japan)
0 ite	Bew.	NS4;NP4	5	/[elmn]\$/	Israelite, Jemenite, Bethlehemite (Bethlehem)
0 itin	B-in	NS0;NP5	5	/[emn]\$/	Jemenitin, Bethlehemitin (Bethlehem), Kanaanitin (Kanaan)
0 ner	Adj.	ADJ00	5	/[aen]\$/	Lissabonner, Grevener (Greve), Paduaner
" 1 er	Bew.	NS2;NP1	5	/[dβ]e/ds]\$/	Niederländer, Falkländer (Falklands), Wernigeröder (Wernigerode)
" 1 erin	B-in	NS0;NP5	5	/[dβ]e/ds]\$/	Niederländerin, Wernigeröderin (Wernigerode), Osteröderin (Osterode)
1 enser	Bew.	NS2;NP1	5	/[na/[ln]e]\$/	Palästinenser, Jenenser, Wernenser (Werne)
1 enserin	B-in	NS0;NP5	5	/[na/[ln]e]\$/	Palästinenserin, Jenenserin, Wernenserin (Werne)
1 iner	Adj.	ADJ00	5	/[aεo]\$/	Kairiner (Kairo), Mallorkiner (Mallorka), Menorciner (Menorca)
2 at	Bew.	NS3;NP3	5	-asien	Asiat, Ostasiat, Südasiat (Südostasien)
2 ate	Bew.	NS4;NP4	5	-asien	Asiate, Südasiatische (Südostasien), Ostasiatische (Ostasien)
2 atin	B-in	NS0;NP5	5	-asien	Asiatin, Südasiatin (Südostasien), Ostasiatin (Ostasien)
2 ischerseits	Adv.		5	/e[in]\$/	türkischerseits (Türkei), schwedischerseits (Schweden), norwegischerseits (Norwegen)
3 rer	Adj.	ADJ00	5	-beuern	Benediktbeurer, Reichersbeurer (Reichersbeuern), Neubeurer (Neubeuern)
" 0 ler	Bew.	NS2;NP1	4	-wald	Schwarzwäldler (Schwarzwald), Bayerwäldler (Bayerwald), Frankenwäldler (Frankenwald)

Tabelle A.3: Suffixoperationen zur Bildung von Toponymderivationen

Suffixop.	Typ	Flexion	Frq.	gem. Sfx.	Beispiele
" 0lerin	<i>B-in</i>	NS0;NP5	4	<i>-wald</i>	<i>Schwarzwäldlerin (Schwarzwald), Spreewäldlerin (Spreewald), Bayerwäldlerin (Bayerwald)</i>
0erseits	<i>Adv.</i>		4	<i>/[gnz]\$/</i>	<i>berlinerseits (Berlin), schweizerseits (Schweiz), freiburgerseits (Freiburg)</i>
0it	<i>Bew.</i>	NS3;NP3	4	<i>/[ve em en]\$/</i>	<i>Jemenit, Bethlehemit (Bethlehem), Niniveit (Ninive)</i>
0ser	<i>Adj.</i>	ADJ00	4	<i>-us</i>	<i>Cottbusser (Cottbus), Putbusser (Putbus), Petkusser (Petkus)</i>
0ser	<i>Bew.</i>	NS2;NP1	4	<i>-us</i>	<i>Cottbusser (Cottbus), Putbusser (Putbus), Petkusser (Petkus)</i>
0serin	<i>B-in</i>	NS0;NP5	4	<i>-us</i>	<i>Cottbusserin (Cottbus), Putbusserin (Putbus), Petkusserin (Petkus)</i>
" 1er	<i>Adj.</i>	ADJ00	4	<i>/([dβ]e ds)\$/</i>	<i>Falkländer (Falklands), Wernigeröder (Wernigerode), Osteröder (Osterode)</i>
1enser	<i>Adj.</i>	ADJ00	4	<i>/[na ln]e)\$/</i>	<i>Jenenser, Wernenser (Werne), Hallenser</i>
2n	<i>B-in</i>	NS0;NP5	4	<i>-ien</i>	<i>Germanin (Germanien), Wallonin (Wallonien), Tschuwaschin (Tschuwaschien)</i>
2nerisch	<i>Adj.</i>	ADJ	4	<i>/[en us]\$/</i>	<i>münchnerisch, kärntnerisch, graubündnerisch (Graubünden)</i>
2rerin	<i>B-in</i>	NS0;NP5	4	<i>-er</i>	<i>Speyrerin (Speyer), Hohenkammrerin (Hohenkammer), Nigrerin (Niger)</i>
" 0ischerseits	<i>Adv.</i>		3	<i>/[pa om lz]\$/</i>	<i>europäischerseits (Europa), römischerseits (Rom), pfälzischerseits (Pfalz)</i>
0eser	<i>Bew.</i>	NS2;NP1	3	<i>/o[nr]\$/</i>	<i>Lyoneser, Avignoneser (Avignon), Timoreser (Timor)</i>
0ianer	<i>Bew.</i>	NS2;NP1	3	<i>-r</i>	<i>Ecuadorianer, Singapurianer, Ekuadorianer (Ekuador)</i>
0ianerin	<i>B-in</i>	NS0;NP5	3	<i>-r</i>	<i>Ecuadorianerin, Singapurianerin (Singapur), Ekuadorianerin (Ekuador)</i>
0ianisch	<i>Adj.</i>	ADJ	3	<i>-r</i>	<i>singapurianisch (Singapur), ecuadorianisch, ekuadorianisch (Ekuador)</i>
0ier	<i>Bew.</i>	NS2;NP1	3	<i>/[on er es]\$/</i>	<i>Babylonier, Peloponnesier (Peloponnes), Nigerier (Niger)</i>
0ierin	<i>B-in</i>	NS0;NP5	3	<i>/[on er es]\$/</i>	<i>Babylonierin (Babylon), Nigerierin (Niger), Peloponnesierin (Peloponnes)</i>
0inerin	<i>B-in</i>	NS0;NP5	3	<i>/h[reb alt]\$/</i>	<i>Sachsen-Anhaltinerin (Sachsen-Anhalt), Anhaltinerin (Anhalt), Maghrebinerin (Maghreb)</i>

Tabelle A.3: Suffixoperationen zur Bildung von Toponymderivationen

Suffixop.	Typ	Flexion	Frq.	gem. Sfx.	Beispiele
0 inisch	Adj.	ADJ	3	/h(reb alt)\$/	sachsen-anhaltinisch (Sachsen-Anhalt), anhaltinisch (Anhalt), maghrebisch (Maghreb)
0 nese	Bew.	NS4;NP4	3	/a(va li)\$/	Balinese, Malinese (Mali), Javanese
0 nesin	B-in	NS0;NP5	3	/a(va li)\$/	Balinesin, Malinesin (Mali), Javanesin (Java)
0 nesisch	Adj.	ADJ	3	/a(va li)\$/	balinesisch (Bali), malinesisch (Mali), javanesisch (Java)
1 aner	Adj.	ADJ00	3	/(rn ns tá)\$/	Fehmaraner (Fehmarn), Orléananer (Orléans), Bogotaner (Bogotá)
1 ier	Bew.	NS2;NP1	3	/(da [tw]i)\$/	Kanadier, Malawier (Malawi), Kiribatier (Kiribati)
1 ierin	B-in	NS0;NP5	3	/(da [tw]i)\$/	Kanadierin, Malawierin (Malawi), Kiribatierin (Kiribati)
1 ler	Bew.	NS2;NP1	3	-e	Erzgebirgler, Fichtelgebirgler (Fichtelgebirge), Elfenbeinküstler (Elfenbeinküste)
1 sch	Adj.	ADJ	3	-n	kölsch, pommersch (Pommern), fallerslebesch (Fallersleben)
2 nisch	Adj.	ADJ	3	/(um en)\$/	polnisch, meißnisch (Meißen), latinisch (Latium)
2 rer	Bew.	NS2;NP1	3	-er	Speyrer, Nigrer, Traismaurer (Traismauer)
6 isch	Adj.	ADJ	3	/(and eln)\$/	griechisch, baskisch (Baskenland), marshallisch (Marshallinseln)
14 isch	Adj.	ADJ	3	-nische republik	hellenisch (Hellenische Republik), zentralafrikanisch (Zentralafrikanische Republik), dominikanisch (Dominikanische Republik)
0	Bew.	NPO	2	/(nda mey)\$/	Dahomey (Dahomey), Venda (Venda)
0 eser	Adj.	ADJ00	2	-on	Lyoneser, Avignoneser (Avignon)
0 eserin	B-in	NS0;NP5	2	-on	Lyoneserin (Lyon), Avignoneserin (Avignon)
0 iner	Adj.	ADJ00	2	-Anhalt	Sachsen-Anhaltiner, Anhaltiner
0 lese	Bew.	NS4;NP4	2	-go	Kongolese, Togolese
0 lesin	B-in	NS0;NP5	2	-go	Kongolesin, Togolesin
0 lesisch	Adj.	ADJ	2	-go	kongolesisch, togolesisch (Togo)
0 ne	Bew.	NS4;NP4	2	/(rma ile)\$/	Chilene, Birmane
0 nin	B-in	NS0;NP5	2	/(rma ile)\$/	Chilenin, Birmanin (Birma)
0 ter	Adj.	ADJ00	2	/(nna kio)\$/	Tokioter, Ravennater (Ravenna)
1 anisch	Adj.	ADJ	2	-ko	mexikanisch, marokkanisch
1 ensisch	Adj.	ADJ	2	-na	palästinensisch, jenensisch (Jena)
1 eraner	Adj.	ADJ00	2	-a	Unneraner, Vechteraner (Vechta)
1 eraner	Bew.	NS2;NP1	2	-a	Unneraner, Vechteraner (Vechta)
1 eranerin	B-in	NS0;NP5	2	-a	Unneranerin (Unna), Vechteranerin (Vechta)

Tabelle A.3: Suffixoperationen zur Bildung von Toponymderivationen

Suffixop.	Typ	Flexion	Frq.	gem. Sfx.	Beispiele
1 eña	<i>B-in</i>	NS0;NP6	2	/i la/ote\$/\$/	Manileña (Manila), Lanzaroteña (Lanzarote)
1 eño	<i>Bew.</i>	NS2;NP6	2	/i la/ote\$/\$/	Manileño (Manila), Lanzaroteño (Lanzarote)
1 ianer	<i>Bew.</i>	NS2;NP1	2	/upe/aii\$/\$/	Hawaiianer, Guadeloupiener (Guadeloupe)
1 ianerin	<i>B-in</i>	NS0;NP5	2	/upe/aii\$/\$/	Hawaiianerin, Guadeloupienerin (Guadeloupe)
1 ischerseits	<i>Adv.</i>		2	-rn	bayerischerseits (Bayern), ungarischerseits (Ungarn)
1 llerin	<i>B-in</i>	NS0;NP5	2	-gebirge	Erzgebirglerin (Erzgebirge), Fichtelgebirglerin (Fichtelgebirge)
1 ner	<i>Bew.</i>	NS2;NP1	2	-as	Texaner, Honduraner
1 nerin	<i>B-in</i>	NS0;NP5	2	-as	Texanerin, Honduranerin (Honduras)
1 nisch	<i>Adj.</i>	ADJ	2	-as	texanisch, honduranisch (Honduras)
1 o	<i>Bew.</i>	NS2;NP6	2	-a	Gomero (Gomera), Gran Canaria (Gran Canaria)
1 oter	<i>Bew.</i>	NS2;NP1	2	/o(fia/kio)\$/\$/	Tokioter, Sofioter
1 oterin	<i>B-in</i>	NS0;NP5	2	/o(fia/kio)\$/\$/	Tokioterin (Tokio), Sofioterin (Sofia)
1 r	<i>Bew.</i>	NS2;NP1	2	-ien	Babylonier, Numidier (Numidien)
1 rin	<i>B-in</i>	NS0;NP5	2	-ien	Babylonierin (Babylonien), Numidierin (Numidien)
1 tiner	<i>Bew.</i>	NS2;NP1	2	-nz	Florentiner, Byzantiner (Byzanz)
1 tinerin	<i>B-in</i>	NS0;NP5	2	-nz	Florentinerin, Byzantinerin (Byzanz)
1 tinisch	<i>Adj.</i>	ADJ	2	-nz	byzantinisch (Byzanz), florentinisch
1 ueño	<i>Bew.</i>	NS2;NP6	2	/ag[ao]\$/\$/	Santiagoueño (Santiago), Málagaueño (Málaga)
2 ese	<i>Bew.</i>	NS4;NP4	2	-n	Libanese, Gabese (Gabun)
2 esin	<i>B-in</i>	NS0;NP5	2	-n	Libanesin, Gabesin (Gabun)
2 esisch	<i>Adj.</i>	ADJ	2	-n	libanesisch, gabesisch (Gabun)
2 rer	<i>Adj.</i>	ADJ00	2	-er	Speyrer, Traismaurer (Traismauer)
3 ese	<i>Bew.</i>	NS4;NP4	2	-ien	Albanese, Kalabrese (Kalabrien)
3 esin	<i>B-in</i>	NS0;NP5	2	-ien	Albanesin (Albanien), Kalabresin (Kalabrien)
3 esisch	<i>Adj.</i>	ADJ	2	-ien	albanesisch (Albanien), kalabresisch (Kalabrien)
3 iake	<i>Bew.</i>	NS4;NP4	2	-ien	Bosniake (Bosnien), Sibiriake (Sibirien)
3 iakin	<i>B-in</i>	NS0;NP5	2	-ien	Bosniakin (Bosnien), Sibiriakin (Sibirien)
3 ianisch	<i>Adj.</i>	ADJ	2	-ien	kastilianisch (Kastilien), asturianisch (Asturien)
4	<i>B-in</i>	NS0;NP6	2	-iland	Swazi (Swasiland), Swasi (Swasiland)

Tabelle A.3: Suffixoperationen zur Bildung von Toponymderivationen

Suffixop.	Typ	Flexion	Frq.	gem. Sfx.	Beispiele
4	Bew.	NS13;NP6	2	-iland	Swazi (Swaziland), Swasi (Swasiland)
4 ischerseits	Adv.		2	-russland	russischerseits (Russland), belarussischerseits (Belarussland)
5 ischerseits	Adv.		2	/(ark/ion)\$/	dänischerseits (Dänemark), sowjetischerseits (Sowjetunion)
6 e	Bew.	NS4;NP4	2	-enland	Griechen, Baske (Baskenland)
6 in	B-in	NS0;NP5	2	-enland	Griechin, Baskin (Baskenland)
6 weltlich	Adj.	ADJ	2	-e welt	altweltlich (Alte Welt), neuweltlich (Neue Welt)
6 zösisch	Adj.	ADJ	2	-frankreich	französisch, nordfranzösisch (Nordfrankreich)
9 schweizer	Adj.	ADJ00	2	-ische schweiz	Französischschweizer (Französische Schweiz), Italienischschweizer (Italienische Schweiz)
9 schweizer	Bew.	NS2;NP1	2	-ische schweiz	Französischschweizer (Französische Schweiz), Italienischschweizer (Italienische Schweiz)
9 schweizerin	B-in	NS0;NP5	2	-ische schweiz	Französischschweizerin (Französische Schweiz), Italienischschweizerin (Italienische Schweiz)
12 si	Bew.	NS2;NP6	2	-stdeutschland	Ossi, Wessi (Westdeutschland)
14 er	Bew.	NS2;NP1	2	-ikanische republik	Zentralafrikaner (Zentralafrikanische Republik), Dominikaner (Dominikanische Republik)
14 erin	B-in	NS0;NP5	2	-ikanische republik	Zentralafrikanerin (Zentralafrikanische Republik), Dominikanerin (Dominikanische Republik)
" 0 lerisch	Adj.	ADJ	1		schwarzwäldlerisch (Schwarzwald)
" 0 sch	Adj.	ADJ	1		hannöversch
0 -Araber	Bew.	NS2;NP1	1		Bahrain-Araber (Bahrain)
0 -Araberin	B-in	NS0;NP5	1		Bahrain-Araberin (Bahrain)
0 ale	Bew.	NS4;NP4	1		Orientale
0 alin	B-in	NS0;NP5	1		Orientalin
0 alisch	Adj.	ADJ	1		orientalisch
0 e	Bew.	NS4;NP4	1		Myanmare
0 erischerseits	Adv.		1		schweizerischerseits (Schweiz)
0 eser	Bew.	NS4;NP4	1		Piémonteser (Piémont)
0 eño	Bew.	NS2;NP6	1		Gibraltareño (Gibraltar)
0 in	B-in	NS0;NP3	1		Israelin
0 iner	Bew.	NS2;NP1	1		Maghrebiner (Maghreb)
0 iter	Bew.	NS2;NP1	1		Kanaaniter (Kanaan)
0 iterin	B-in	NS0;NP5	1		Kanaaniterin (Kanaan)
0 itin	B-in	NS0;NP3	1		Israelitin
0 lich	Adj.	ADJ	1		basellandschaftlich (Basellandschaft)

Tabelle A.3: Suffixoperationen zur Bildung von Toponymderivationen

Suffixop.	Typ	Flexion	Frq.	gem. Sfx.	Beispiele
0nischerseits	Adv.		1		<i>amerikanischerseits (Amerika)</i>
0ssier	Bew.	NS2;NP1	1		<i>Jurassier (Jura)</i>
0ssierin	B-in	NS0;NP5	1		<i>Jurassierin (Jura)</i>
0ssisch	Adj.	ADJ	1		<i>jurassisch (Jura)</i>
0te	Bew.	NS4;NP4	1		<i>Ravennate (Ravenna)</i>
0tisch	Adj.	ADJ	1		<i>ravennatisch (Ravenna)</i>
0uí	B-in	NS0;NP6	1		<i>Saharauí (Sahara)</i>
0uí	Bew.	NS13;NP6	1		<i>Saharauí (Sahara)</i>
0äisch	Adj.	ADJ	1		<i>kanaanäisch (Kanaan)</i>
0äner	Adj.	ADJ00	1		<i>Kasseläner (Kassel)</i>
0äner	Bew.	NS2;NP1	1		<i>Kasseläner (Kassel)</i>
0änerin	B-in	NS0;NP5	1		<i>Kasselänerin (Kassel)</i>
" 1isch	Adj.	ADJ	1		<i>niederländisch</i>
" 1ischerseits	Adv.		1		<i>niederländischerseits (Niederlande)</i>
" 1ler	Bew.	NS2;NP1	1		<i>Bergsträßler (Bergstraße)</i>
" 1lerin	B-in	NS0;NP5	1		<i>Bergsträßlerin (Bergstraße)</i>
1 -Amerikaner	Bew.	NS2;NP1	1		<i>US-Amerikaner</i>
1 -Amerikanerin	B-in	NS0;NP5	1		<i>US-Amerikanerin</i>
1 -amerikanisch	Adj.	ADJ	1		<i>US-amerikanisch</i>
1	Bew.	NS11;NP4	1		<i>Ungar</i>
1	Bew.	NS4;NP4 NS2	1		<i>Bayer</i>
1ale	Bew.	NS4;NP4	1		<i>Provencale (Provence)</i>
1alin	B-in	NS0;NP5	1		<i>Provençalín (Provence)</i>
1alisch	Adj.	ADJ	1		<i>provençalisch (Provence)</i>
1an	Bew.	NS2;NP6	1		<i>San Franciscan (San Francisco)</i>
1are	Bew.	NS4;NP4	1		<i>Kosovare</i>
1arin	B-in	NS0;NP5	1		<i>Kosovarin</i>
1arisch	Adj.	ADJ	1		<i>kosovarisch (Kosovo)</i>
1asker	Bew.	NS2;NP1	1		<i>Bergamasker (Bergamo)</i>
1askerin	B-in	NS0;NP5	1		<i>Bergamaskerin (Bergamo)</i>
1askisch	Adj.	ADJ	1		<i>bergamaskisch (Bergamo)</i>
1de	Bew.	NS4;NP4	1		<i>Atlantide (Atlantis)</i>
1din	B-in	NS0;NP5	1		<i>Atlantidin (Atlantis)</i>
1encisch	Adj.	ADJ	1		<i>ibizencisch (Ibiza)</i>
1enco	B-in	NS0;NP6	1		<i>Ibizenco (Ibiza)</i>
1enco	Bew.	NS2;NP6	1		<i>Ibizenco (Ibiza)</i>
1enkisch	Adj.	ADJ	1		<i>ibizenkisch (Ibiza)</i>
1enko	B-in	NS0;NP6	1		<i>Ibizenko (Ibiza)</i>
1enko	Bew.	NS2;NP6	1		<i>Ibizenko (Ibiza)</i>
1ensischerseits	Adv.		1		<i>palästinensischerseits (Palästina)</i>
1esischerseits	Adv.		1		<i>chinesischerseits (China)</i>
1eña	B-in	NS0;NP6 NPO	1		<i>Atacameña (Atacama)</i>
1eño	Bew.	NS2;NP6 NPO	1		<i>Atacameño (Atacama)</i>
1ianisch	Adj.	ADJ	1		<i>hawaiianisch</i>
1inienser	Bew.	NS2;NP1	1		<i>Karthaginienser (Karthago)</i>
1inienserin	B-in	NS0;NP5	1		<i>Karthaginienserin (Karthago)</i>
1iot	Bew.	NS3;NP3	1		<i>Korfiot (Korfu)</i>

Tabelle A.3: Suffixoperationen zur Bildung von Toponymderivationen

Suffixop.	Typ	Flexion	Frq.	gem. Sfx.	Beispiele
1 iote	Bew.	NS4;NP4	1		<i>Korfiote (Korfu)</i>
1 iotin	B-in	NS0;NP5	1		<i>Korfiotin (Korfu)</i>
1 iotisch	Adj.	ADJ	1		<i>korfiotisch (Korfu)</i>
1 it	Bew.	NS3;NP3	1		<i>Odessit (Odessa)</i>
1 itaner	Adj.	ADJ00	1		<i>Palermitaner</i>
1 itaner	Bew.	NS2;NP1	1		<i>Palermitaner</i>
1 itanerin	B-in	NS0;NP5	1		<i>Palermitanerin (Palermo)</i>
1 itanisch	Adj.	ADJ	1		<i>palermitanisch (Palermo)</i>
1 itin	B-in	NS0;NP5	1		<i>Odessitin (Odessa)</i>
1 itisch	Adj.	ADJ	1		<i>odessitisch (Odessa)</i>
1 jer	Bew.	NS2;NP1	1		<i>Altajer (Altai)</i>
1 jerin	B-in	NS0;NP5	1		<i>Altajerin (Altai)</i>
1 lene	Bew.	NS4;NP4	1		<i>Madrilene</i>
1 lenin	B-in	NS0;NP5	1		<i>Madrilenin (Madrid)</i>
1 lenisch	Adj.	ADJ	1		<i>madrilenisch (Madrid)</i>
1 leña	B-in	NS0;NP6	1		<i>Madrileña</i>
1 leño	Bew.	NS2;NP6	1		<i>Madrileño</i>
1 mer	Bew.	NS2;NP1	1		<i>Bademer</i>
1 merin	B-in	NS0;NP5	1		<i>Bademerin (Baden)</i>
1 ore	Bew.	NS4;NP4	1		<i>Hallore (Halle)</i>
1 orin	B-in	NS0;NP5	1		<i>Hallorin (Halle)</i>
1 oter	Adj.	ADJ00	1		<i>Sofioter</i>
1 otisch	Adj.	ADJ	1		<i>sofiotisch (Sofia)</i>
1 sier	Bew.	NS2;NP1	1		<i>Milesier (Milet)</i>
1 sierin	B-in	NS0;NP5	1		<i>Milesierin (Milet)</i>
1 sisch	Adj.	ADJ	1		<i>milesisch (Milet)</i>
1 taner	Adj.	ADJ00	1		<i>Cáditaner (Cádiz)</i>
1 taner	Bew.	NS2;NP1	1		<i>Cáditaner (Cádiz)</i>
1 tanerin	B-in	NS0;NP5	1		<i>Cáditanerin (Cádiz)</i>
1 te	Bew.	NS4;NP4	1		<i>Laote</i>
1 teke	Bew.	NS4;NP4	1		<i>Guatemalteke</i>
1 tekin	B-in	NS0;NP5	1		<i>Guatemaltekin</i>
1 tekisch	Adj.	ADJ	1		<i>guatemalteckisch</i>
1 tin	B-in	NS0;NP5	1		<i>Laotin</i>
1 tisch	Adj.	ADJ	1		<i>laotisch</i>
1 unke	Bew.	NS4;NP4	1		<i>Hallunke</i>
1 unkin	B-in	NS0;NP5	1		<i>Hallunkin (Halle)</i>
1 äer	Bew.	NS2;NP1	1		<i>Urartäer (Urartu)</i>
1 äerin	B-in	NS0;NP5	1		<i>Urartäerin (Urartu)</i>
1 äisch	Adj.	ADJ	1		<i>urartäisch (Urartu)</i>
" 2 ischerseits	Adv.		1		<i>sächsischerseits (Sachsen)</i>
2 aner	Adj.	ADJ00	1		<i>Rüganer</i>
2 arde	Bew.	NS4;NP4	1		<i>Savoyarde (Savoyen)</i>
2 ardin	B-in	NS0;NP5	1		<i>Savoyardin (Savoyen)</i>
2 ardisch	Adj.	ADJ	1		<i>savoyardisch (Savoyen)</i>
2 enisch	Adj.	ADJ	1		<i>hellenisch (Hellas)</i>
2 enser	Adj.	ADJ00	1		<i>Badenser</i>

Tabelle A.3: Suffixoperationen zur Bildung von Toponymderivationen

Suffixop.	Typ	Flexion	Frq.	gem. Sfx.	Beispiele
2enser	Bew.	NS2;NP1	1		<i>Badenser</i>
2enserin	B-in	NS0;NP5	1		<i>Badenserin</i>
2ensisch	Adj.	ADJ	1		<i>badensisch (Baden)</i>
2erseits	Adv.		1		<i>bremerseits (Bremen)</i>
2eño	Bew.	NS2;NP6	1		<i>Hondureño (Honduras)</i>
2frau	B-in	NS0;NP3	1		<i>Halligfrau (Halligen)</i>
2i	B-in	NS0;NP6	1		<i>Somali (Somalia)</i>
2i	Bew.	NS13;NP6	1		<i>Somali (Somalia)</i>
2ianer	Bew.	NS2;NP1	1		<i>Kapverdianer (Kapverden)</i>
2ianerin	B-in	NS0;NP5	1		<i>Kapverdianerin (Kapverden)</i>
2iese	Bew.	NS4;NP4	1		<i>Portugiese</i>
2iesin	B-in	NS0;NP5	1		<i>Portugiesin</i>
2iesisch	Adj.	ADJ	1		<i>portugiesisch</i>
2iser	Bew.	NS2;NP1	1		<i>Waliser</i>
2iserin	B-in	NS0;NP5	1		<i>Waliserin</i>
2isisch	Adj.	ADJ	1		<i>walisisch</i>
2itaner	Bew.	NS2;NP1	1		<i>Tripolitaner (Tripolis)</i>
2itanerin	B-in	NS0;NP5	1		<i>Tripolitanerin (Tripolis)</i>
2itanisch	Adj.	ADJ	1		<i>tripolitanisch (Tripolis)</i>
2lerisch	Adj.	ADJ	1		<i>baslerisch</i>
2leute	Bew.	NP1	1		<i>Halligleute (Halligen)</i>
2mann	Bew.	NS1;NP14	1		<i>Halligmann (Halligen)</i>
2nischerseits	Adv.		1		<i>polnischerseits (Polen)</i>
2o	B-in	NS0;NP6	1		<i>Philippino</i>
2o	Bew.	NS2;NP6	1		<i>Philippino</i>
2olitaner	Adj.	ADJ00	1		<i>Neapolitaner</i>
2olitaner	Bew.	NS2;NP1	1		<i>Neapolitaner</i>
2olitanerin	B-in	NS0;NP5	1		<i>Neapolitanerin</i>
2olitanisch	Adj.	ADJ	1		<i>neapolitanisch</i>
2ot	Bew.	NS3;NP3	1		<i>Epirot (Epirus)</i>
2ote	Bew.	NS4;NP4	1		<i>Epirote (Epirus)</i>
2otin	B-in	NS0;NP5	1		<i>Epirotin (Epirus)</i>
2otisch	Adj.	ADJ	1		<i>epirotisch (Epirus)</i>
2ovese	Bew.	NS4;NP4	1		<i>Genovese</i>
2oveser	Adj.	ADJ00	1		<i>Genoveser (Genua)</i>
2oveser	Bew.	NS2;NP1	1		<i>Genoveser (Genua)</i>
2oveserin	B-in	NS0;NP5	1		<i>Genoveserin (Genua)</i>
2ovesisch	Adj.	ADJ	1		<i>genovesisch (Genua)</i>
2owite	Bew.	NS4;NP4	1		<i>Moskowite (Moskau)</i>
2owiter	Bew.	NS2;NP1	1		<i>Moskowiter (Moskau)</i>
2owiterin	B-in	NS0;NP5	1		<i>Moskowiterin (Moskau)</i>
2owitin	B-in	NS0;NP5	1		<i>Moskowitin (Moskau)</i>
2owitisch	Adj.	ADJ	1		<i>moskowitisch</i>
2queña	B-in	NS0;NP6	1		<i>Oaxaqueña (Oaxaca)</i>
2queño	Bew.	NS2;NP6	1		<i>Oaxaqueño (Oaxaca)</i>
2quiner	Bew.	NS2;NP1	1		<i>Mallorquiner</i>
2quinerin	B-in	NS0;NP5	1		<i>Mallorquinerin</i>

Tabelle A.3: Suffixoperationen zur Bildung von Toponymderivationen

Suffixop.	Typ	Flexion	Frq.	gem. Sfx.	Beispiele
2quinisch	Adj.	ADJ	1		<i>mallorquinisch</i>
2risch	Adj.	ADJ	1		<i>nigrisch (Niger)</i>
2sch	Adj.	ADJ	1		<i>libysch (Libyen)</i>
2ueño	Bew.	NS2;NP6	1		<i>Galápagueño (Galápagos)</i>
2zale	Bew.	NS4;NP4	1		<i>Provenzale (Provence)</i>
2zalin	B-in	NS0;NP5	1		<i>Provençalín (Provence)</i>
2zalisch	Adj.	ADJ	1		<i>provençalisch (Provence)</i>
2çale	Bew.	NS4;NP4	1		<i>Provençale</i>
2çalin	B-in	NS0;NP5	1		<i>Provençalín (Provence)</i>
2çalisch	Adj.	ADJ	1		<i>provençalisch</i>
2ñero	B-in	NS0;NP6	1		<i>Habañero (Habana)</i>
2ñero	Bew.	NS2;NP6	1		<i>Habañero (Habana)</i>
3	B-in	NS0;NP6	1		<i>Mandschu (Mandschukuo)</i>
3	Bew.	NS13;NP6	1		<i>Mandschu (Mandschukuo)</i>
3aner	Bew.	NS2;NP1	1		<i>Martinianer (Martinique)</i>
3anerin	B-in	NS0;NP5	1		<i>Martinianerin (Martinique)</i>
3anisch	Adj.	ADJ	1		<i>martinianisch (Martinique)</i>
3cher	Adj.	ADJ00	1		<i>Zürcher</i>
3cher	Bew.	NS2;NP1	1		<i>Zürcher</i>
3cherin	B-in	NS0;NP5	1		<i>Zürcherin</i>
3cherisch	Adj.	ADJ	1		<i>zürcherisch (Zürich)</i>
3dentiner	Adj.	ADJ00	1		<i>Tridentiner (Trient)</i>
3dentinisch	Adj.	ADJ	1		<i>tridentinisch (Trient)</i>
3egasse	Bew.	NS4;NP4	1		<i>Monegasse (Monaco)</i>
3egassin	B-in	NS0;NP5	1		<i>Monegassin (Monaco)</i>
3egassisch	Adj.	ADJ	1		<i>monegassisch (Monaco)</i>
3ener	Bew.	NS2;NP1	1		<i>Nazarener (Nazareth)</i>
3enerin	B-in	NS0;NP5	1		<i>Nazarenerin (Nazareth)</i>
3enisch	Adj.	ADJ	1		<i>nazarenisch (Nazareth)</i>
3eraner	Adj.	ADJ00	1		<i>Fehmeraner (Fehmarn)</i>
3eraner	Bew.	NS2;NP1	1		<i>Fehmeraner (Fehmarn)</i>
3eranerin	B-in	NS0;NP5	1		<i>Fehmeranerin (Fehmarn)</i>
3ianer	Bew.	NS2;NP1	1		<i>Asturianer (Asturien)</i>
3ianerin	B-in	NS0;NP5	1		<i>Asturianerin (Asturien)</i>
3iner	Bew.	NS2;NP1	1		<i>Dalmatiner (Dalmatien)</i>
3inerin	B-in	NS0;NP5	1		<i>Dalmatinerin (Dalmatien)</i>
3inger	Bew.	NS2;NP1	1		<i>Färingner (Färöer)</i>
3ingerin	B-in	NS0;NP5	1		<i>Färingnerin (Färöer)</i>
3ingisch	Adj.	ADJ	1		<i>färingisch (Färöer)</i>
3inisch	Adj.	ADJ	1		<i>dalmatinisch (Dalmatien)</i>
3jake	Bew.	NS4;NP4	1		<i>Sibirjake (Sibirien)</i>
3jakin	B-in	NS0;NP5	1		<i>Sibirjakin (Sibirien)</i>
3ne	Bew.	NS4;NP4	1		<i>Normanne (Normandie)</i>
3nenhofer	Adj.	ADJ00	1		<i>Königinnenhofer (Königinhof)</i>
3nin	B-in	NS0;NP5	1		<i>Normannin (Normandie)</i>
3nisch	Adj.	ADJ	1		<i>normannisch (Normandie)</i>
3queño	Bew.	NS2;NP6	1		<i>Caraqueño (Caracas)</i>

Tabelle A.3: Suffixoperationen zur Bildung von Toponymderivationen

Suffixop.	Typ	Flexion	Frq.	gem. Sfx.	Beispiele
3rier	Bew.	NS2;NP1	1		Zyprier (Zypern)
3rierin	B-in	NS0;NP5	1		Zyprierin (Zypern)
3riot	Bew.	NS3;NP3	1		Zypriot (Zypern)
3riote	Bew.	NS4;NP4	1		Zypriote (Zypern)
3riotin	B-in	NS0;NP5	1		Zypriotin (Zypern)
3riotisch	Adj.	ADJ	1		zypriotisch (Zypern)
3se	Bew.	NS4;NP4	1		Madagasse (Madagaskar)
3sin	B-in	NS0;NP5	1		Madagassin (Madagaskar)
3sisch	Adj.	ADJ	1		madagassisch (Madagaskar)
3tianer	Adj.	ADJ00	1		Venetianer
3tianerin	B-in	NS0;NP5	1		Venetianerin (Venedig)
3tianisch	Adj.	ADJ	1		venetianisch (Venedig)
3zener	Adj.	ADJ00	1		Damaszener
3zener	Bew.	NS2;NP1	1		Damaszener
3zenerin	B-in	NS0;NP5	1		Damaszenerin (Damaskus)
3zenisch	Adj.	ADJ	1		damaszenisch (Damaskus)
3zianer	Adj.	ADJ00	1		Venezianer
3zianer	Bew.	NS2;NP1	1		Venezianer
3zianerin	B-in	NS0;NP5	1		Venezianerin (Venedig)
3zianisch	Adj.	ADJ	1		venezianisch
3zäisch	Adj.	ADJ	1		nizäisch (Nikäa)
4	Bew.	NS4;NP4	1		Ingusche (Inguschetien)
4er	Adj.	ADJ00	1		Steierer (Steiermark)
4er	Bew.	NS2;NP1	1		Steierer (Steiermark)
4erin	B-in	NS0;NP5	1		Steiererin (Steiermark)
4erseits	Adv.		1		deutscherseits (Deutschland)
4ie	Bew.	NS4;NP4	1		Malaie
4ierin	B-in	NS0;NP5	1		Malaierin (Malaysia)
4iisch	Adj.	ADJ	1		malaiisch
4nisch	Adj.	ADJ	1		estnisch
4olaner	Bew.	NS2;NP1	1		Venezolaner
4olanerin	B-in	NS0;NP5	1		Venezolanerin (Venezuela)
4olanisch	Adj.	ADJ	1		venezolanisch (Venezuela)
4one	Bew.	NS4;NP4	1		Bretone (Bretagne)
4onin	B-in	NS0;NP5	1		Bretonin (Bretagne)
4onisch	Adj.	ADJ	1		bretonisch (Bretagne)
4ugianer	Bew.	NS2;NP1	1		Rugianer (Rügen)
4ugianerin	B-in	NS0;NP5	1		Rugianerin (Rügen)
4zisch	Adj.	ADJ	1		maurizisch (Mauritius)
5-Ricaner	Bew.	NS2;NP1	1		Costa-Ricaner (Costa Rica)
5-Ricanerin	B-in	NS0;NP5	1		Costa-Ricanerin (Costa Rica)
5-ricanisch	Adj.	ADJ	1		costa-ricanisch (Costa Rica)
5ane	Bew.	NS4;NP4	1		Katalane (Katalonien)
5anin	B-in	NS0;NP5	1		Katalanin (Katalonien)
5anisch	Adj.	ADJ	1		katalanisch (Katalonien)
5apolitaner	Adj.	ADJ00	1		Napolitaner (Neapel)
5apolitaner	Bew.	NS2;NP1	1		Napolitaner (Neapel)

Tabelle A.3: Suffixoperationen zur Bildung von Toponymderivationen

Suffixop.	Typ	Flexion	Frq.	gem. Sfx.	Beispiele
5 apolitanerin	<i>B-in</i>	NS0;NP5	1		<i>Napolitanerin (Neapel)</i>
5 eña	<i>B-in</i>	NS0;NP6	1		<i>Extremeña (Extremadura)</i>
5 eño	<i>Bew.</i>	NS2;NP6	1		<i>Extremeño (Extremadura)</i>
5 i	<i>Bew.</i>	NS13;NP6	1		<i>Ami</i>
5 iker	<i>Bew.</i>	NS2;NP1	1		<i>Phöniker (Phönizien)</i>
5 ikerin	<i>B-in</i>	NS0;NP5	1		<i>Phönikerin (Phönizien)</i>
5 me	<i>Bew.</i>	NS4;NP4	1		<i>Flame (Flandern)</i>
5 ricaner	<i>Bew.</i>	NS2;NP1	1		<i>Costaricaner (Costa Rica)</i>
5 ricanerin	<i>B-in</i>	NS0;NP5	1		<i>Costaricanerin (Costa Rica)</i>
5 ricanisch	<i>Adj.</i>	ADJ	1		<i>costaricanisch (Costa Rica)</i>
5 rikaner	<i>Bew.</i>	NS2;NP1	1		<i>Costarikaner (Costa Rica)</i>
5 rikanerin	<i>B-in</i>	NS0;NP5	1		<i>Costarikanerin (Costa Rica)</i>
5 rikanisch	<i>Adj.</i>	ADJ	1		<i>costarikanisch (Costa Rica)</i>
6 Magyare	<i>Bew.</i>	NS4;NP4	1		<i>Magyare (Ungarn)</i>
6 Magyarin	<i>B-in</i>	NS0;NP5	1		<i>Magyarin (Ungarn)</i>
6 aner	<i>Adj.</i>	ADJ00	1		<i>Moselaner (Moselregion)</i>
6 aner	<i>Bew.</i>	NS2;NP1	1		<i>Moselaner (Moselregion)</i>
6 anerin	<i>B-in</i>	NS0;NP5	1		<i>Moselanerin (Moselregion)</i>
6 er	<i>Bew.</i>	NS2;NP1	1		<i>Marshaller (Marshallinseln)</i>
6 erin	<i>B-in</i>	NS0;NP5	1		<i>Marshallerin (Marshallinseln)</i>
6 ischerseits	<i>Adv.</i>		1		<i>griechischerseits (Griechenland)</i>
6 laner	<i>Adj.</i>	ADJ00	1		<i>Mosellaner (Moselregion)</i>
6 laner	<i>Bew.</i>	NS2;NP1	1		<i>Mosellaner (Moselregion)</i>
6 lanerin	<i>B-in</i>	NS0;NP5	1		<i>Mosellanerin (Moselregion)</i>
6 madjarisch	<i>Adj.</i>	ADJ	1		<i>madjarisch (Ungarn)</i>
6 magyarisch	<i>Adj.</i>	ADJ	1		<i>magyarisch (Ungarn)</i>
6 rer	<i>Adj.</i>	ADJ00	1		<i>Steirer (Steiermark)</i>
6 rer	<i>Bew.</i>	NS2;NP1	1		<i>Steirer (Steiermark)</i>
6 rerin	<i>B-in</i>	NS0;NP5	1		<i>Steirerin (Steiermark)</i>
6 risch	<i>Adj.</i>	ADJ	1		<i>steirisch (Steiermark)</i>
6 städtisch	<i>Adj.</i>	ADJ	1		<i>friedrich-wilhelmstädtisch (Friedrich-Wilhelm-Stadt)</i>
6 zose	<i>Bew.</i>	NS4;NP4	1		<i>Franzose</i>
6 zösin	<i>B-in</i>	NS0;NP5	1		<i>Französín</i>
6 zösischerseits	<i>Adv.</i>		1		<i>französischerseits (Frankreich)</i>
6 ämin	<i>B-in</i>	NS0;NP5	1		<i>Flämin (Flandern)</i>
6 ämisch	<i>Adj.</i>	ADJ	1		<i>flämisch (Flandern)</i>
" 7 in	<i>B-in</i>	NS0;NP5	1		<i>Eidgenossein (Eidgenossenschaft)</i>
" 7 isch	<i>Adj.</i>	ADJ	1		<i>eidgenosseisch (Eidgenossenschaft)</i>
7 Canario	<i>Bew.</i>	NS2;NP6	1		<i>Canario (Kanaren)</i>
7 Caribeña	<i>B-in</i>	NS0;NP6	1		<i>Caribeña (Karibik)</i>
7 Caribeño	<i>Bew.</i>	NS2;NP6	1		<i>Caribeño (Karibik)</i>
7 e	<i>Bew.</i>	NS4;NP4	1		<i>Eidgenossee (Eidgenossenschaft)</i>
7 gallisch	<i>Adj.</i>	ADJ	1		<i>sanktgallisch (Sankt Gallen)</i>
7 in	<i>B-in</i>	NS0;NP5	1		<i>Eidgenossein (Eidgenossenschaft)</i>
7 isch	<i>Adj.</i>	ADJ	1		<i>maurisch (Mauretanien)</i>
7 nesisch	<i>Adj.</i>	ADJ	1		<i>sundanesisch (Sunda-Inseln)</i>

Tabelle A.3: Suffixoperationen zur Bildung von Toponymderivationen

Suffixop.	Typ	Flexion	Frq.	gem. Sfx.	Beispiele
8 Palmero	Bew.	NS2;NP6	1		<i>Palmero (La Palma)</i>
9 Herreña	B-in	NS0;NP6	1		<i>Herreña (El Hierro)</i>
9 Herreño	Bew.	NS2;NP6	1		<i>Herreño (El Hierro)</i>
9 Tinerfeño	Bew.	NS2;NP6	1		<i>Tinerfeño (Teneriffa)</i>
9 i	B-in	NS0;NP6	1		<i>Aseri (Aserbaidshan)</i>
9 i	Bew.	NS13;NP6	1		<i>Aseri (Aserbaidshan)</i>
9 in	B-in	NS0;NP5	1		<i>Aserin (Aserbaidshan)</i>
10 hebrider	Bew.	NS2;NP1	1		<i>Neuhebrider (Neue Hebriden)</i>
10 hebriderin	B-in	NS0;NP5	1		<i>Neuhebriderin (Neue Hebriden)</i>
10 hebridisch	Adj.	ADJ	1		<i>neuhebridisch (Neue Hebriden)</i>
10 toreño	Bew.	NS2;NP6	1		<i>Bocatoreño (Bocas del Toro)</i>
11 Extremeño	Bew.	NS2;NP6	1		<i>Extremeño (Estremadura)</i>
11 Filipino	B-in	NS0;NP6	1		<i>Filipino</i>
11 Filipino	Bew.	NS2;NP6	1		<i>Filipino</i>
11 Salvadoreña	B-in	NS0;NP6	1		<i>Salvadoreña (El Salvador)</i>
11 Salvadoreño	Bew.	NS2;NP6	1		<i>Salvadoreño (El Salvador)</i>
11 Salvadorianer	Bew.	NS2;NP1	1		<i>Salvadorianer (El Salvador)</i>
11 Salvadorianerin	B-in	NS0;NP5	1		<i>Salvadorianerin (El Salvador)</i>
11 salvadorianisch	Adj.	ADJ	1		<i>salvadorianisch (El Salvador)</i>
12 Porteña	B-in	NS0;NP6	1		<i>Porteña (Buenos Aires)</i>
12 Porteño	Bew.	NS2;NP6	1		<i>Porteño (Buenos Aires)</i>
12 ner	Bew.	NS2;NP1	1		<i>Antiguaner (Antigua und Barbuda)</i>
12 nerin	B-in	NS0;NP5	1		<i>Antiguanerin (Antigua und Barbuda)</i>
12 nisch	Adj.	ADJ	1		<i>antiguanisch (Antigua und Barbuda)</i>
13 Ivorer	Bew.	NS2;NP1	1		<i>Ivorer (Côte d'Ivoire)</i>
13 Ivorerin	B-in	NS0;NP5	1		<i>Ivorerin (Côte d'Ivoire)</i>
13 ivorisch	Adj.	ADJ	1		<i>ivorisch (Côte d'Ivoire)</i>
14 Brite	Bew.	NS4;NP4	1		<i>Brite</i>
14 Britin	B-in	NS0;NP5	1		<i>Britin</i>
14 britisch	Adj.	ADJ	1		<i>britisch</i>
14 britischerseits	Adv.		1		<i>britischerseits (Großbritannien)</i>
14 mediterran	Adj.	ADJ	1		<i>mediterran (Mittelmeerraum)</i>
16 Heidjer	Adj.	ADJ00	1		<i>Heidjer (Lüneburger Heide)</i>
16 Heidjer	Bew.	NS2;NP1	1		<i>Heidjer (Lüneburger Heide)</i>
16 Heidjerin	B-in	NS0;NP5	1		<i>Heidjerin (Lüneburger Heide)</i>
16 heidjerisch	Adj.	ADJ	1		<i>heidjerisch (Lüneburger Heide)</i>
18 -Toméer	Bew.	NS2;NP1	1		<i>São-Toméer (São Tomé und Príncipe)</i>
18 -Toméerin	B-in	NS0;NP5	1		<i>São-Toméerin (São Tomé und Príncipe)</i>
18 -toméisch	Adj.	ADJ	1		<i>são-toméisch (São Tomé und Príncipe)</i>
18 isch-montenegrinisch	Adj.	ADJ	1		<i>serbisch-montenegrinisch (Serbien und Montenegro)</i>
20 deutsch	Adj.	ADJ	1		<i>bundesdeutsch (Bundesrepublik Deutschland)</i>

Tabelle A.3: Suffixoperationen zur Bildung von Toponymderivativen

Suffixop.	Typ	Flexion	Frq.	gem. Sfx.	Beispiele
20 deutsche	<i>B-in</i>	NS0;NP5	1		<i>Bundesdeutsche (Bundesrepublik Deutschland)</i>
20 deutsche	<i>Bew.</i>	NS4;NP4	1		<i>Bundesdeutsche (Bundesrepublik Deutschland)</i>
20 deutscherseits	<i>Adv.</i>		1		<i>bundesdeutscherseits (Bundesrepublik Deutschland)</i>
21 Santomeer	<i>Bew.</i>	NS2;NP1	1		<i>Santomeer (São Tomé und Príncipe)</i>
21 Santomeerin	<i>B-in</i>	NS0;NP5	1		<i>Santomeerin (São Tomé und Príncipe)</i>
21 santomeisch	<i>Adj.</i>	ADJ	1		<i>santomeisch (São Tomé und Príncipe)</i>
32 Vincenter	<i>Bew.</i>	NS2;NP1	1		<i>Vincenter (Saint Vincent und die Grenadinen)</i>
32 Vincenterin	<i>B-in</i>	NS0;NP5	1		<i>Vincenterin (Saint Vincent und die Grenadinen)</i>
32 vincentisch	<i>Adj.</i>	ADJ	1		<i>vincentisch (Saint Vincent und die Grenadinen)</i>

Tabelle A.3: Suffixoperationen zur Bildung von Toponymderivationen

A.1.5 Taxonomien geographischer Entitäten

An dieser Stelle sind einige Taxonomien in Baumdarstellung oder in Form tabellarischer Listen wiedergegeben, zu deren Darstellung in Kap. 3.4.2.1 nicht der notwendige Platz vorhanden gewesen wäre.

A.1.5.1 Linguistisch orientierte Klassifikationssysteme

Die hier exemplarisch vorgestellten linguistisch orientierten Klassifikationssysteme und Taxonomien sind entweder der onomastischen Tradition verpflichtet oder sie wollen den Wortschatz in semantische oder Selektionsklassen organisieren.

- OSNOVENSYSTEM1983, nur Toponyme (Abb. A.1 auf der nächsten Seite)
- Kamianets (2000), nur Toponyme (Abb. A.2 auf S. 334)
- Walther (2003), alle Eigennamen (Abb. A.3 auf S. 334)
- ein Auszug aus dem Taxonomiebaum von Langer (1996), Abb. A.5 auf S. 336: enthalten sind nur lokative Klassen
- WORDNET (Abb. A.6 auf S. 337): die beiden Knoten COUNTY und COUNTRY im Kontext
- Eigennamenklassen des Projekts PROLEX (Abb. A.4)
- für ein Lexikon und Ontologie deutscher Toponyme erarbeitete Taxonomie (Abb. A.7 auf S. 338 und A.8 auf S. 339)

Toponym, geographischer Name, Geotoponym – Name eines unbelebten natürlichen Objektes auf der Erde (Berg, Wasser, Insel, Niederung, Wald, Brachland etc.) und eines solchen von Menschen geschaffenen Objektes auf der Erde, das in der Landschaft fest verankert ist (Stadt, Dorf, bearbeiteter oder aufgeforsteter Boden, Verkehrsweg etc.)

└ **Choronym, Raumname** – Name einer größeren bewohnten oder unbewohnten natürlichen oder administrativen Einheit hinsichtlich ihrer horizontalen Gliederung

└ **geographischer Raumname, natürlicher Raumname** – Name einer größeren bewohnten oder unbewohnten natürlichen Einheit hinsichtlich ihrer horizontalen Gliederung (also eines Erdteils oder Teils desselben, einer größeren Insel oder Halbinsel, einer Landschaft, einer Wüste etc.) – Europa, Sächsische Schweiz

└ **administrativer Raumname** – Name einer größeren administrativen Einheit hinsichtlich ihrer horizontalen Gliederung (also eines Staates, Landes, einer Provinz, eines Bezirkes, Kreises etc.)

└ **Oikonym, Ortsname** – Name einer Siedlung (d.h. eines bewohnten Ortes) und zwar auch einer verlassenen oder untergegangenen:

- a) Name einer Stadt eines Dorfes und ihrer Teile (Viertel, Vorstädte etc.),
- b) Name eines sogenannten Ortsteiles oder Wohnplatzes, d.h. einer von der eigentlichen Siedlung getrennt liegenden Gruppe von Häusern oder eines solchen Hauses (Einzelhof, Wirtschaftshof, Forsthaus, Ziegelei, Bergwerkshütte, Sägewerk, Mühle, Burg, Jagdschloss, Ruine, Touristenunterkunft etc.),
- c) Name eines einzelnen Objektes mit Siedlungscharakter in der Stadt oder auf dem Dorf (Einzelhaus oder Einzelwirtschaft, Schloss, Hotel, Gaststätte, Apotheke, Theater oder Kino, Turm, Teil einer Befestigung, Kirche, Kloster etc.),
- d) Name einer Örtlichkeit in einem Objekt mit Siedlungscharakter (bewohnte Räume, Halle, Vortragssaal, Musiksaal etc.).

└ **Anoikonym** – Name eines unbelebten natürlichen Objektes oder eines solchen von Menschen geschaffenen Objektes, das nicht zum Wohnen bestimmt ist und in der Landschaft fest verankert ist.

└ **Hydronym, Gewässername** – Name eines Gewässers (Meer, Meerbusen, Meerenge, See, Sumpf, Torfmoor, Mörgegend, Fischteich, Stausee, Feuerlöschteich, Brunnen, Quelle, Fluss, Bach, Stromschnelle, Wasserfall, Flussbett, Kanal etc.)

└ **Oronym, Bergname** – Name einer Formation der vertikalen Gliederung der Erdoberfläche oder des Meeresgrundes (orographisches Gefüge, Gebirge, Bergkette, Hochland, Bergland, Hügelland, Bergrücken, Tal, Niederung, Geländeeinschnitt, Bodensenke, Pass, Bergsattel, Abgrund, Schlucht, Tiefland, Ebene, Plateau, Becken etc.).

└ **Speleonym** – Name einer Höhle, Grotte

└ **Parzellenname** – Name mehrerer zusammengehöriger Parzellen (Grundstücke) oder einer einzelnen Parzelle:

- a) Name eines bearbeiteten Grundstücks (Feld, Wiese, Koppel, Weide, Garten, Weinberg etc.),
- b) Name einer Waldparzelle (Waldrevier, Jagen, Wald, Gehölz, Wildgehege etc.),
- c) Name eines unbearbeiteten waldlosen Grundstückes.

└ **Hodonym** – Name eines Verkehrsweges (Strasse, Platz (nicht Sportplatz o. ä.), Uferweg, Mole, Autobahn, Landstrasse, Weg (Fussweg, Viehweg, Feldweg), Tunnel, Furt, Brücke, Steg, Fähre, Eisenbahnlinie, Drahtseilbahn etc.).

Kosmonym – Name eines unbelebten natürlichen Objekts oder Erscheinung außerhalb der Erde.

Abbildung A.1: Untergliederung der Toponyme nach OSNOVENSYSTEM1983

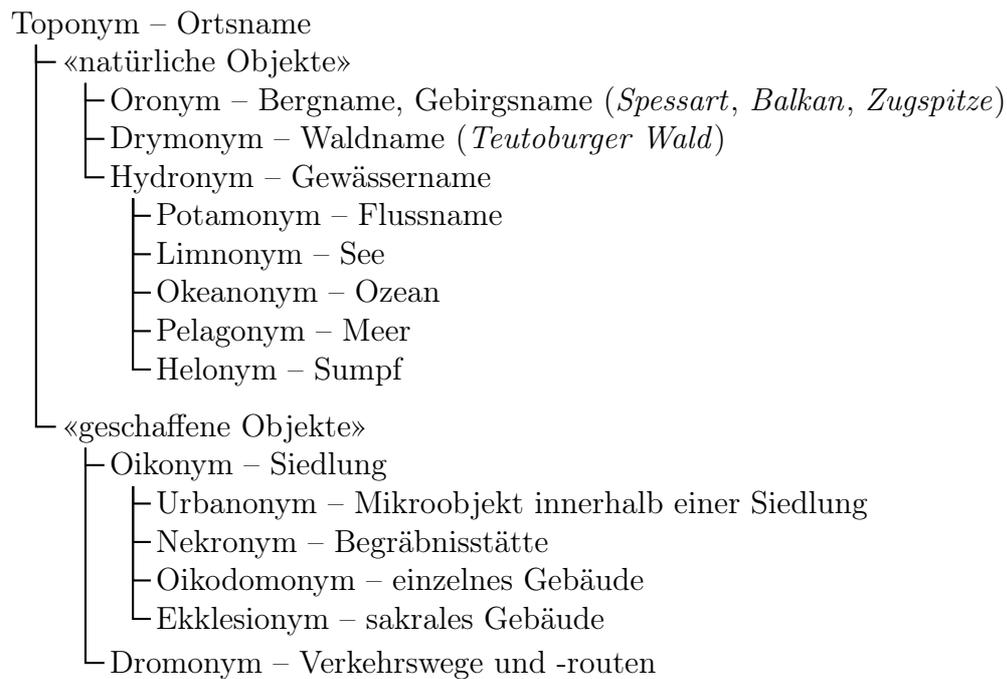


Abbildung A.2: Klassifikation der Ortsnamen nach Kamianets (2000)



Abbildung A.3: Eigennamenklassen nach Walther (2003)

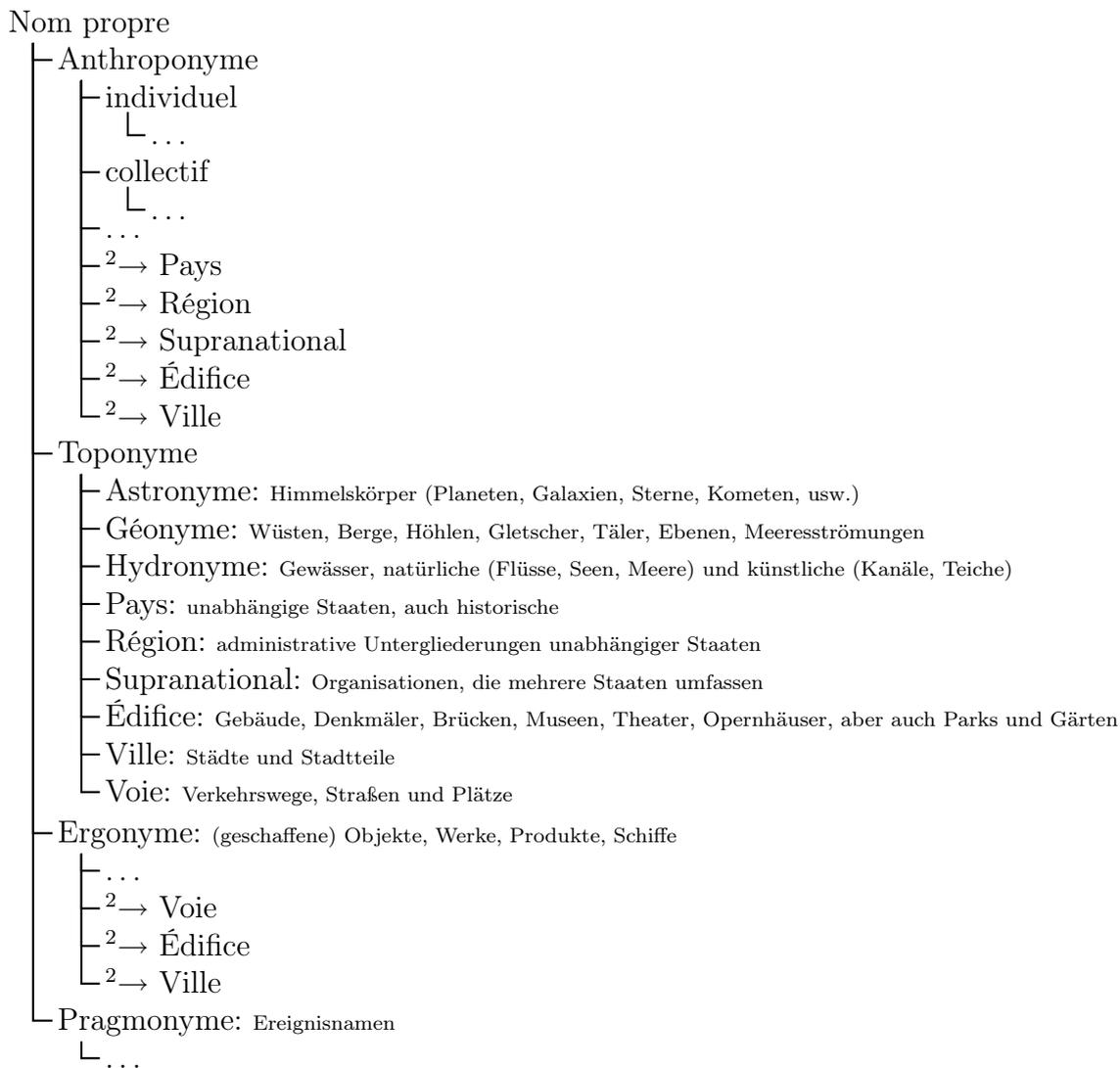


Abbildung A.4: Teilbaum mit Toponymklassen des Projektes PROLEX (nach Maurel et al. 2006: 29–30). “²→” steht für sekundäre Hyperonyme.

ort Ortsbezeichnungen

- |geb Gebäude
 - └ger Kultgebäude
- |vkw Verkehrswege
- |ora Räume
- |osc Schalter/Tresen/Ausgaben/Ausleihen
(Orte, an denen etwas ausgegeben etc. wird)
- |oas Ansiedlung
- |old Land
- |oln Landschaftsformen und -elemente
 - └olb Landschaftserhebungen
 - └olt Täler, Senken
 - └olw Wald
 - └oli Inseln
- |ogw Gewässer
 - └ogf Fließgewässer
- |ori Richtungs-/Lageangaben
- |ohr Himmelsrichtungen
- |ote Objektteile
 - └beg Begrenzungen
- |omy Mythischer Ort

knk Konkreta

- └kon Konkrete Individualnomina
 - └ste Gestirne: Sonne, Mond und Sterne
 - └uni Bezeichnungen für die Welt
 - └vkw Verkehrswege (→ s.o.)
 - └oln Landschaftsformen und -elemente (→ s.o.)
 - └art Artefakte (Von Menschenhand geschaffene Gegenstände)
 - └geb Gebäude (→ s.o.)
 - └beg Begrenzungen (→ s.o.)

itu Institutionen

- └oit Institutionen mit lokativen Implikationen
 - └ovk Verkaufsort
 - └ofa Fabrik
 - └ovw Verwaltungs-/Administrationsbetriebe
 - └opa Parlamente
 - └olo Lokal
 - └oho Unterbringungsbetriebe
 - └osx Bordelle
 - └ovb Vorführorte - Bühnen
 - └oab Ausbildungsstätten
 - └oms Museen, Ausstellungsorte

Abbildung A.5: „Ortsbezeichnungen“ und untergeordnete Klassen bei Langer (1996), vgl. http://www.cis.uni-muenchen.de/people/langer/sem_html/

entity

- └ physical entity
 - └ object, physical object
 - └ location: a point or extent in space
 - └ region, part: the extended spatial location of something: *the farming regions of France*
 - └ atmosphere, air
 - └ biosphere
 - └ depth
 - └ interplanetary space
 - └ interstellar space
 - └ **county**: (United Kingdom) a region created by territorial division for the purpose of local government: *the county has a population of 12,345 people*
 - └ (instance) Avon: a county in southwestern England
 - └ (instance) Berkshire: a county in southern England
 - └ (instance) Cornwall: a hilly county in southwestern England
 - └ (instance) . . .
 - └ (instance) Wight, Isle of Wight: an isle and county of southern England in the English Channel
 - └ Eden, paradise, nirvana, heaven, promised land, Shangri-la
 - └ hell, hell on earth, hellhole, snake pit, the pits, inferno
 - └ inside, interior
 - └ . . .
 - └ vacuum, vacuity
 - └ top
 - └ zodiac
 - └ region: a large indefinite location on the surface of the Earth: *penguins inhabit the polar regions*
 - └ district, territory, territorial dominion, dominion: a region marked off for administrative or other purposes
 - └ administrative division, administrative division, territorial division: a district defined for administrative purposes
 - └ borough: one of the administrative divisions of a large city
 - └ canton: a small administrative division of a country
 - └ city
 - └ commune: the smallest administrative district of several European countries
 - └ country, state, land
 - └ **county**: (United States) the largest administrative district within a state: *the county plans to build a new road*
 - └ . . .

(entity)

- └ (physical entity)
 - └ (object, physical object)
 - └ (location: a point or extent in space)
 - └ region: a large indefinite location on the surface of the Earth: *penguins inhabit the polar regions*
 - └ district, territory, territorial dominion, dominion: a region marked off for administrative or other purposes
 - └ administrative division, administrative division, territorial division: a district defined for administrative purposes
 - └ borough
 - └ canton
 - └ city
 - └ commune
 - └ **country, state, land**: the territory occupied by a nation: *he returned to the land of his birth; he visited several European countries*
 - └ banana republic
 - └ fatherland, homeland, motherland, mother country, country of origin, native land
 - └ buffer state, buffer country
 - └ kingdom
 - └ sultanate
 - └ tax haven
 - └ European country, European nation
 - └ African country, African nation
 - └ . . .
 - └ (instance) Antigua and Barbuda
 - └ (instance) Cape Verde, Republic of Cape Verde
 - └ (instance) . . .
 - └ county: (United States)
 - └ → . . .
 - └ prefecture: the district administered by a prefect (as in France or Japan or the Roman Empire)
 - └ state, province
 - └ township, town
 - └ ward: a district into which a city or town is divided for the purpose of administration and elections
 - └ (instance) Lower Egypt: one of the two main administrative districts of Egypt; consists of the Nile delta
 - └ (instance) Upper Egypt: one of the two main administrative districts of Egypt; extends south from Cairo to Sudan
 - └ (instance) Khabarovsk: an administrative territory in Russia on the eastern coast of Siberia

Abbildung A.6: Auszug aus WORDNET: die Knoten COUNTY und COUNTRY im Kontext

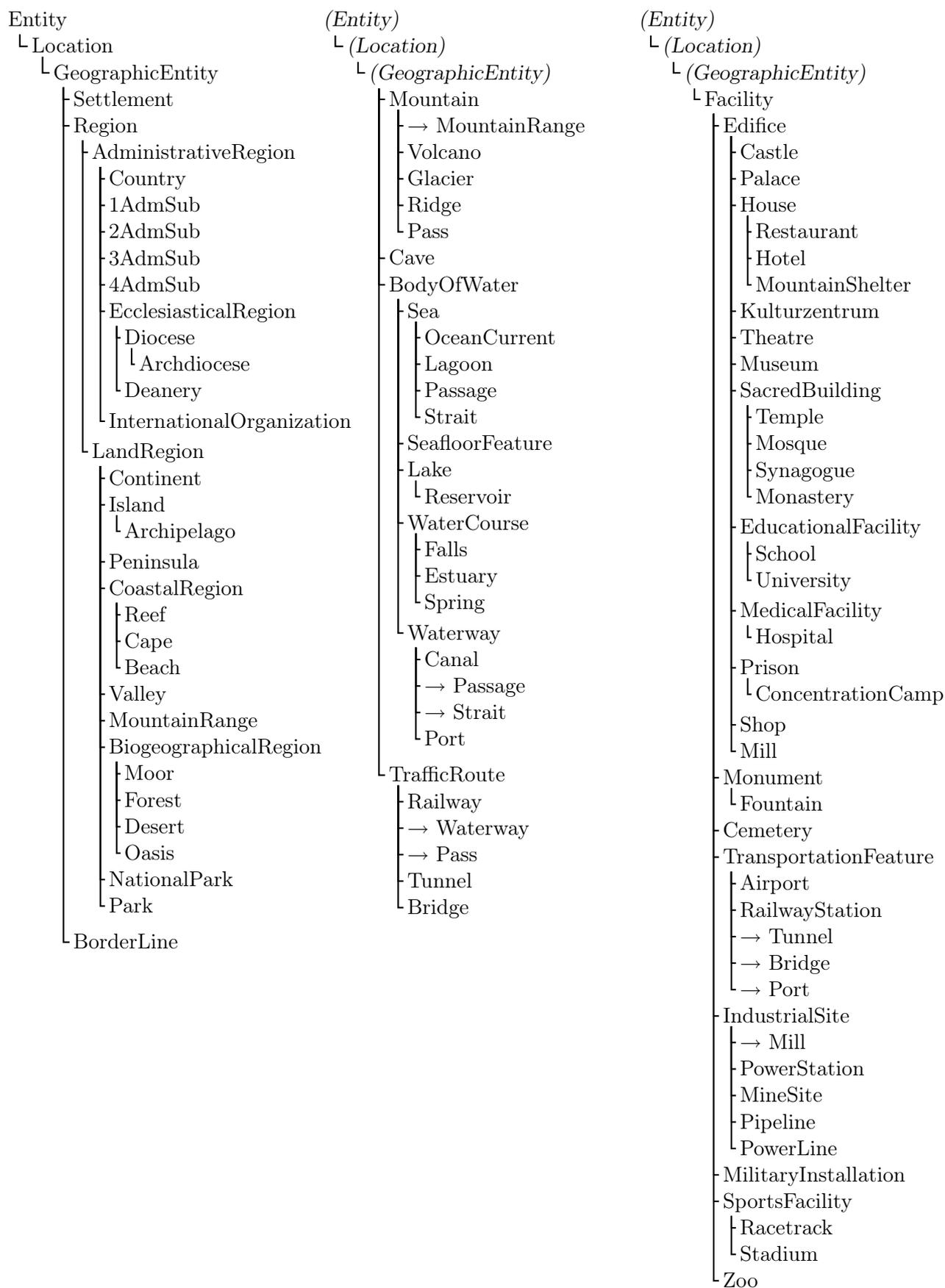


Abbildung A.7: Taxonomie, erarbeitet für ein Lexikon (Ontologie) deutscher Toponyme

SETTLEMENT	Siedlung, Stadt, Dorf, Stadtteil, Ortslagen	STRAIT	Meerenge
COUNTRY	Unabhängiger Staat	PASS	Pass
1ADM SUB	Administrative Untergliederung der ersten Ebene	FACILITY	Bauwerk, Bebauung, Anlage
2ADM SUB	Administrative Untergliederung der zweiten Ebene	EDIFICE	Gebäude, Gebäudekomplex
3ADM SUB	Administrative Untergliederung der dritten Ebene	CASTLE	Burg, Schloss
4ADM SUB	Administrative Untergliederung der vierten Ebene	PALACE	Palast, Schloss
ECCLESIASTICAL REGION	Kirchliche Territorialeinheit	HOUSE	Haus, Wohnhaus, Bürohaus
DIocese	Diözese	RESTAURANT	Gasthof, Gaststätte, Restaurant
ARCHDIOCESE	Erzdiözese	HOTEL	Hotel, Gästehaus, Herberge
DEANERY	Dekanat	MOUNTAIN SHELTER	Berghütte
LAND REGION	Landfläche, naturräumliche Region	KULTURZENTRUM	Kulturzentrum
CONTINENT	Kontinent	THEATRE	Vorführibetrieb: Theater, Kino, Oper, usw
ISLAND	Insel	MUSEUM	Museum
ARCHIPELAGO	Inselgruppe, Archipel	SACRED BUILDING	Sakrallbau
PENINSULA	Halbinsel	TEMPLE	Kirche
REEF	Riff	MOSQUE	Moschee
CAPE	Kap	SYNAGOGUE	Synagoge
VALLEY	Tal	MONASTERY	Kloster
MOOR	Moor	EDUCATIONAL FACILITY	Bildungseinrichtung
FOREST	Forst, Wald	MEDICAL FACILITY	Medizinische Einrichtung
DESERT	Wüste	HOSPITAL	Krankenhaus, Klinik
OASIS	Oase	PRISON	Gefängnis
NATIONAL PARK	Nationalpark	CONCENTRATION CAMP	Konzentrationslager
PARK	Park	SHOP	Geschäft
BORDER LINE	Grenze, administrative Grenze	MONUMENT	Monument
MOUNTAIN	Berg, Hügel, Gipfel, Anhöhe, alle Arten von Erhebungen	FOUNTAIN	Brunnen
MOUNTAIN RANGE	Gebirge, Gebirgszug, Bergkette	CEMETERY	Friedhof, Bestattungsplatz
VOLCANO	Vulkan	TRANSPORTATION FEATURE	Verkehrsbauwerk, baulicher Komplex benutzt für Verkehr
GLACIER	Gletscher	AIRPORT	Flughafen
RIDGE	Grat	RAILWAY STATION	Bahnhof
CAVE	Höhle	TUNNEL	Tunnel
BODY OF WATER	Gewässer	BRIDGE	Brücke
SEA	Meer	PORT	Hafen
OCEAN CURRENT	Meeresströmung	INDUSTRIAL SITE	Produktionsstätte, industrieller Komplex, industrielle Infrastruktur
LAGOON	Lagune, Haff	MILL	Mühle
SEAFLOOR FEATURE	Untermeerische geologische Formation	POWER STATION	Kraftwerk
LAKE	See, Teich, stehendes Binnengewässer	MINE SITE	Bergwerk, Bergbau, Mine
RESERVOIR	Stausee	PIPELINE	Pipeline
WATER COURSE	Wasserlauf, Fließgewässer (Fluss, Bach)	POWER LINE (Stark)	Stromleitung, Freileitung
FALLS	Wasserfall	MILITARY INSTALLATION	Militärische Einrichtung
ESTUARY	Flussmündung, Delta	SPORTS FACILITY	Sportstätte
SPRING	Quelle	RACE TRACK	Rennstrecke
TRAFFIC ROUTE	Verkehrswege	STADIUM	Stadion
RAILWAY	Eisenbahnlinie	ZOO	Zoo
WATERWAY	Wasserweg	INTERNATIONAL ORGANIZATION	Internationale Organisation
CANAL	Kanal		
PASSAGE	Passage, Seeweg		

Abbildung A.8: Definition der Klassen für eine Lexikon (Ontologie) deutscher Toponyme

A.1.5.2 Geographisch orientierte Taxonomien

- DAML:GEOFILE, vollständiger Taxonomiebaum (Abb. A.9 auf S. 341). Auf die (exemplarische) Wiedergabe von den Klassen zugeordneten Entitäten wurde verzichtet.
- NASA:SWEET (Abb. A.11 auf S. 343). Die Klassifikation nach physischen Eigenschaften der Objekte ist klar zu erkennen. Alle Klassen sind (mittelbar) von abstrakten Klassen abgeleitet, siehe Abb. A.10 auf S. 342.
- GEOFEATURES.OWL (Abb. A.12 auf S. 344)
- CORINE (Abb. A.13 auf S. 345): „Das europaweite Projekt CORINE Land Cover (CLC) hat die Bereitstellung von einheitlichen und damit vergleichbaren Daten der Bodenbedeckung für Europa zum Ziel. Es ist Teil des Programms CORINE (Coordination of Information on the Environment) der Europäischen Union. Die Kartierung der Bodenbedeckung und Landnutzung wurde europaweit auf der Basis von Satellitendaten im Maßstab 1:100.000 durchgeführt. Die Ersterfassung (CLC1990) erfolgte einheitlich nach 44 Landnutzungs-klassen, von denen 37 Klassen in Deutschland relevant sind.“ (CORINE)
- Sekine, Sudo & Nobata (2002), Sekine & Nobata (2004), SekineExtNER: Abb. A.14 auf S. 346 zeigt alle Klassen im Überblick, Abb. A.15 und A.16 die lokativen Klassen im Detail.
- ADL Feature Type Thesaurus (Hill 2000, Hill, Goodchild & Janée 2004), Abb. A.17 auf S. 349 und A.18 auf S. 350, siehe auch die Diskussion auf S. 135
- KIM, Auszug aus dem Taxonomiebaum (Abb. A.20 auf S. 352): Enthalten sind alle untergeordneten Klassen (Hyponyme) der Klasse LOCATION, sowie zur Orientierung die übergeordneten und benachbarten Klassen (Hyperonyme und Kohyponyme).
- BBNTAXON, Auszug aus der Taxonomie (Abb. A.19 auf S. 351): Enthalten sind alle lokativen Klassen, außerdem die Klassen ORGANIZATION NAME und NORP (“nationality, religion, organization, political”), da sie Klassen enthalten, die deutliche Überschneidungen zu den Lokativa aufweisen, siehe die Diskussion auf S. 135).



Abbildung A.9: Vollständiger Taxonomiebaum von DAML:GEOFILE (Klassen ohne Entitäten)

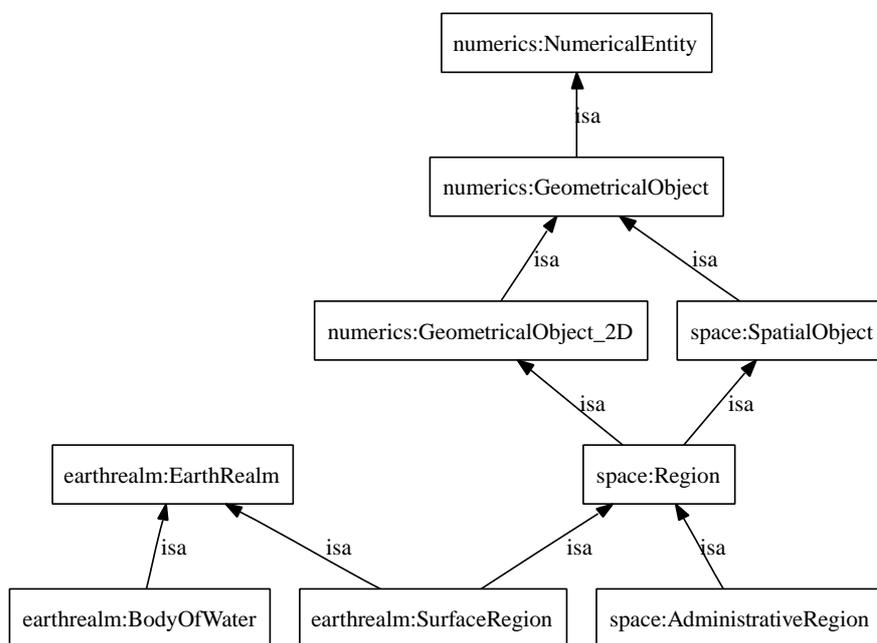


Abbildung A.10: Die übergeordnete Hierarchie der Klassen ‘body of water’, ‘surface region’ und ‘administrative region’ der Ontologie NASA:SWEET



Abbildung A.11: Auszug aus NASA:SWEET: Der Übersichtlichkeit halber wurde auf einen Teil der hierarchischen Beziehungen und die Angabe der Namensräume 'space' und 'earthrealm' verzichtet (vgl. Abb. A.10).

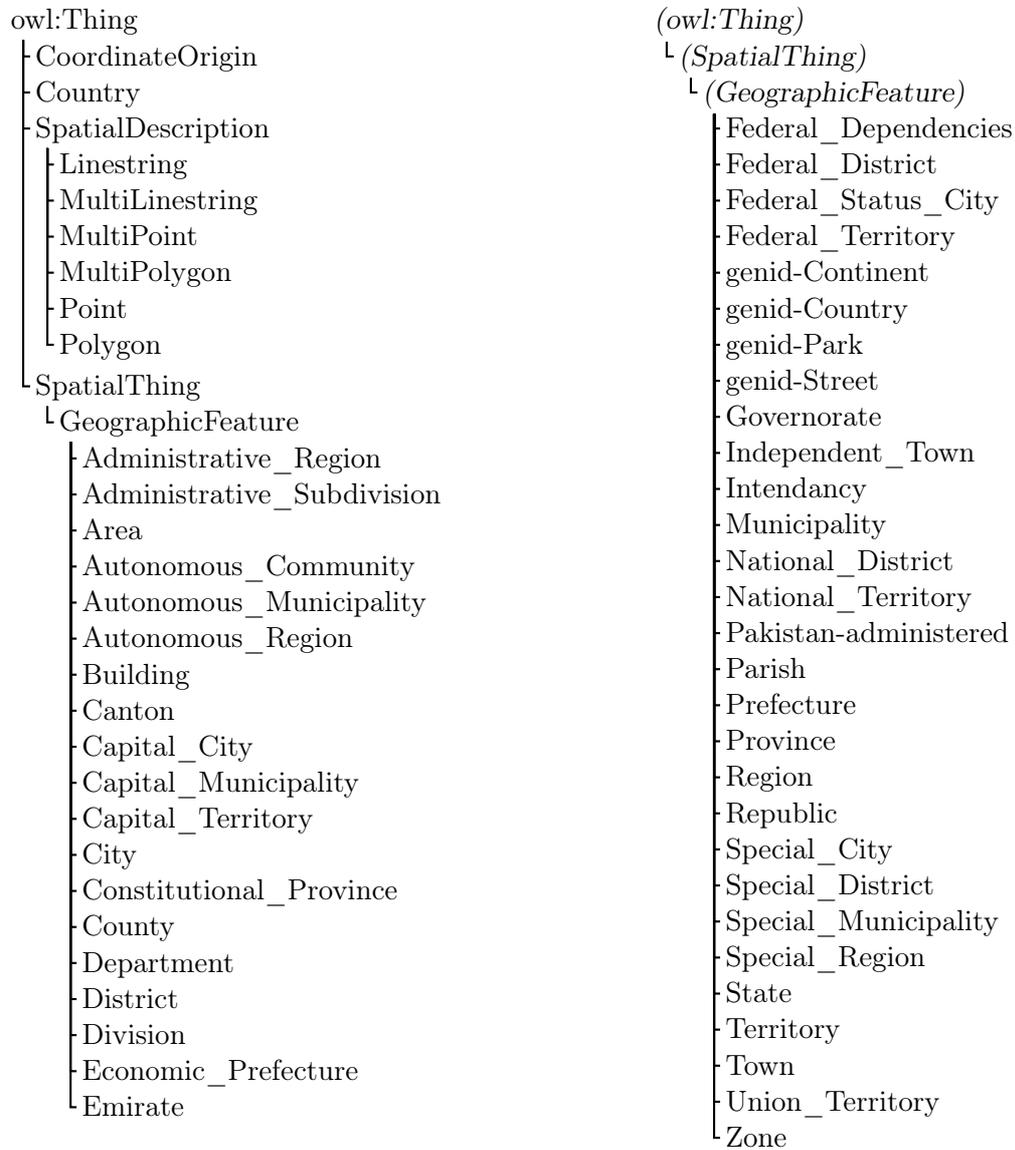


Abbildung A.12: „Taxonomie“ toponymischer Objekte GEOFEATURES.OWL

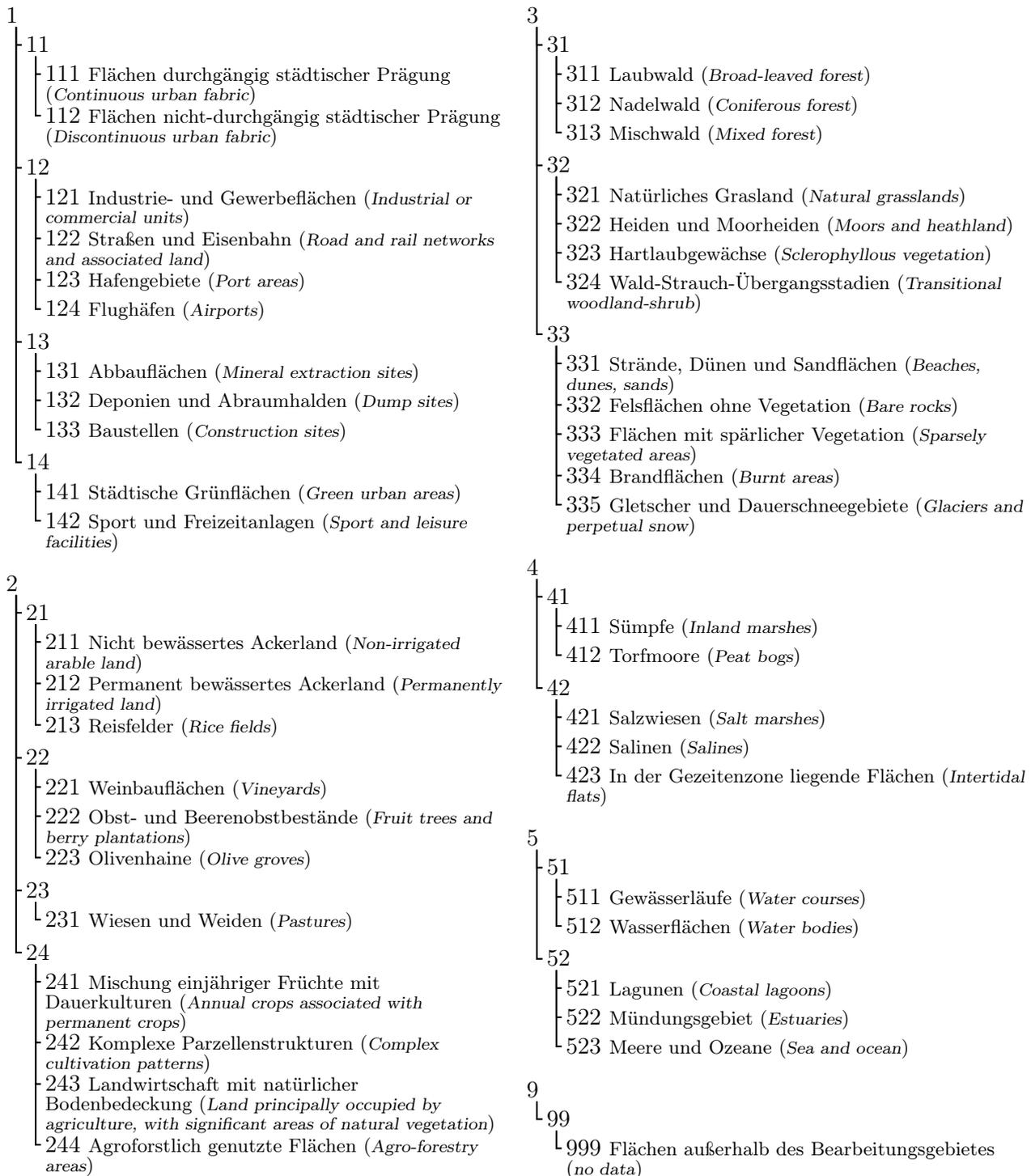


Abbildung A.13: Die Taxonomie von CORINE (deutsch/englisch) (Quellen: http://www.corine.dfd.dlr.de/media/download/clc_lut_de.pdf bzw. [clc_lut_en.pdf](http://www.corine.dfd.dlr.de/media/download/clc_lut_en.pdf))

Extended Named Entity (ver. 6.1.2)

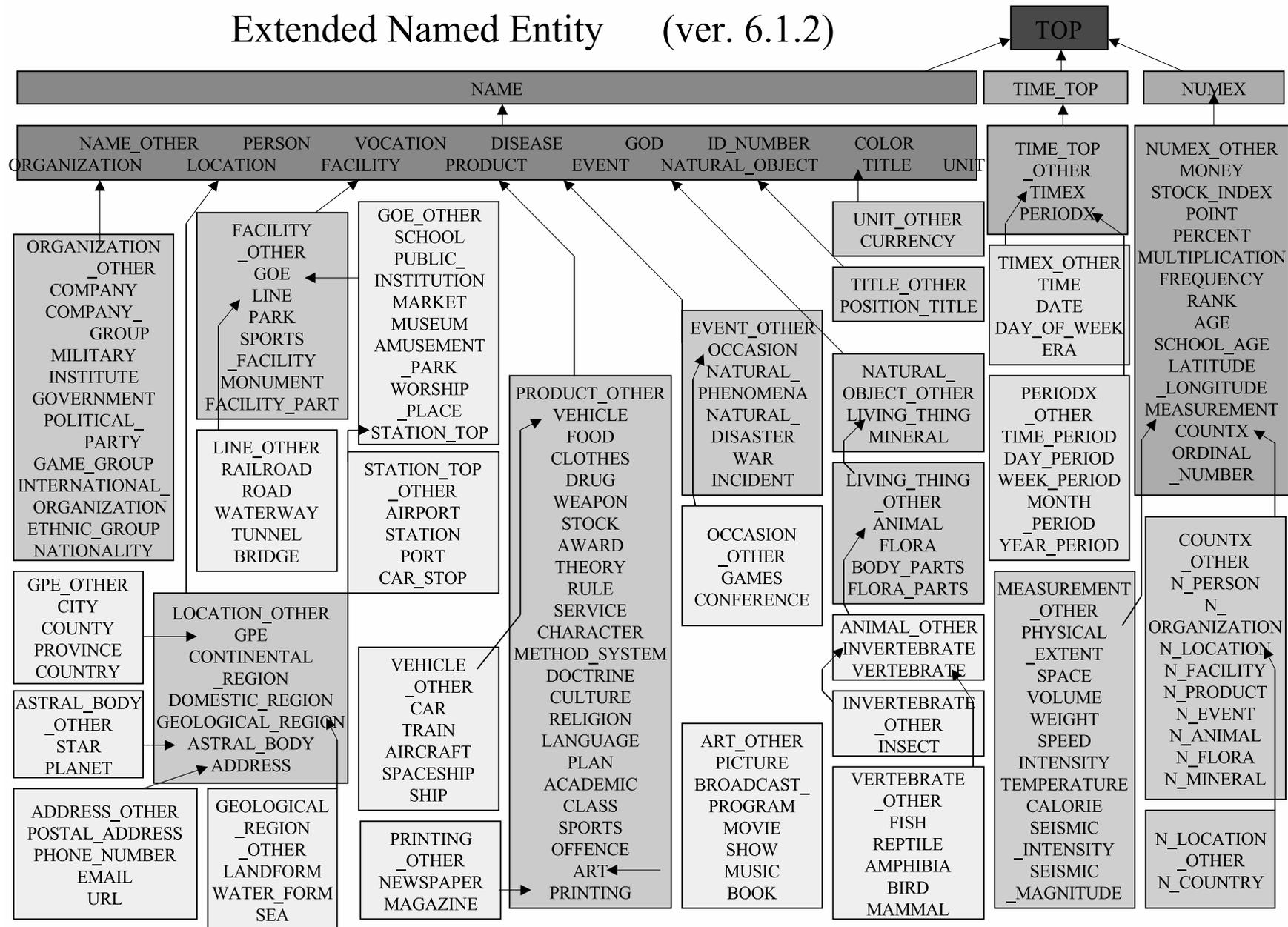


Abbildung A.14: Taxonomie von Sekine, Gesamtüberblick (Quelle: http://nlp.cs.nyu.edu/ene/zentaizu6_1_2eng.jpg)

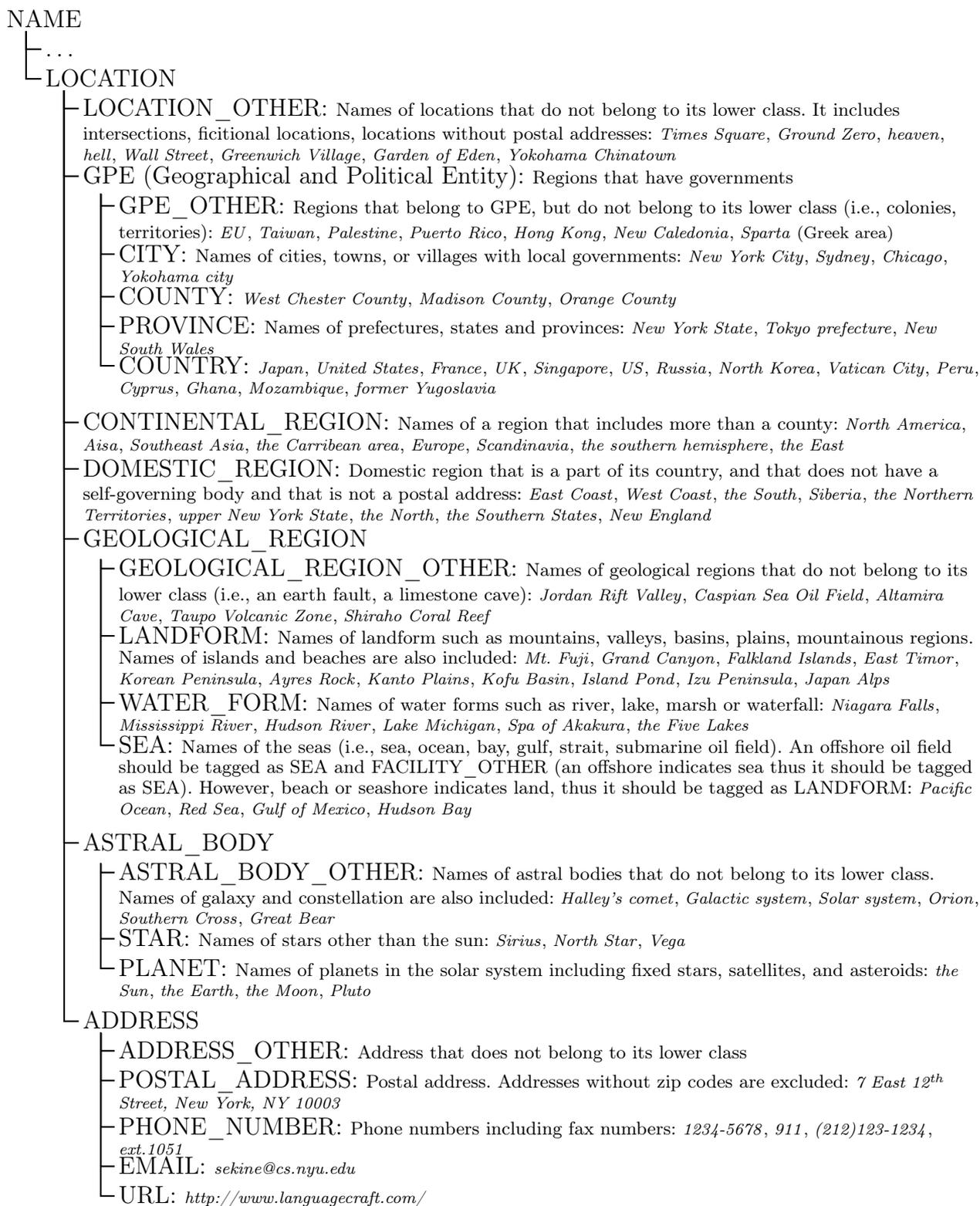


Abbildung A.15: Taxonomie von Sekine, Teil I: Locations (Quelle: SEKINEEXTNER)

(NAME)

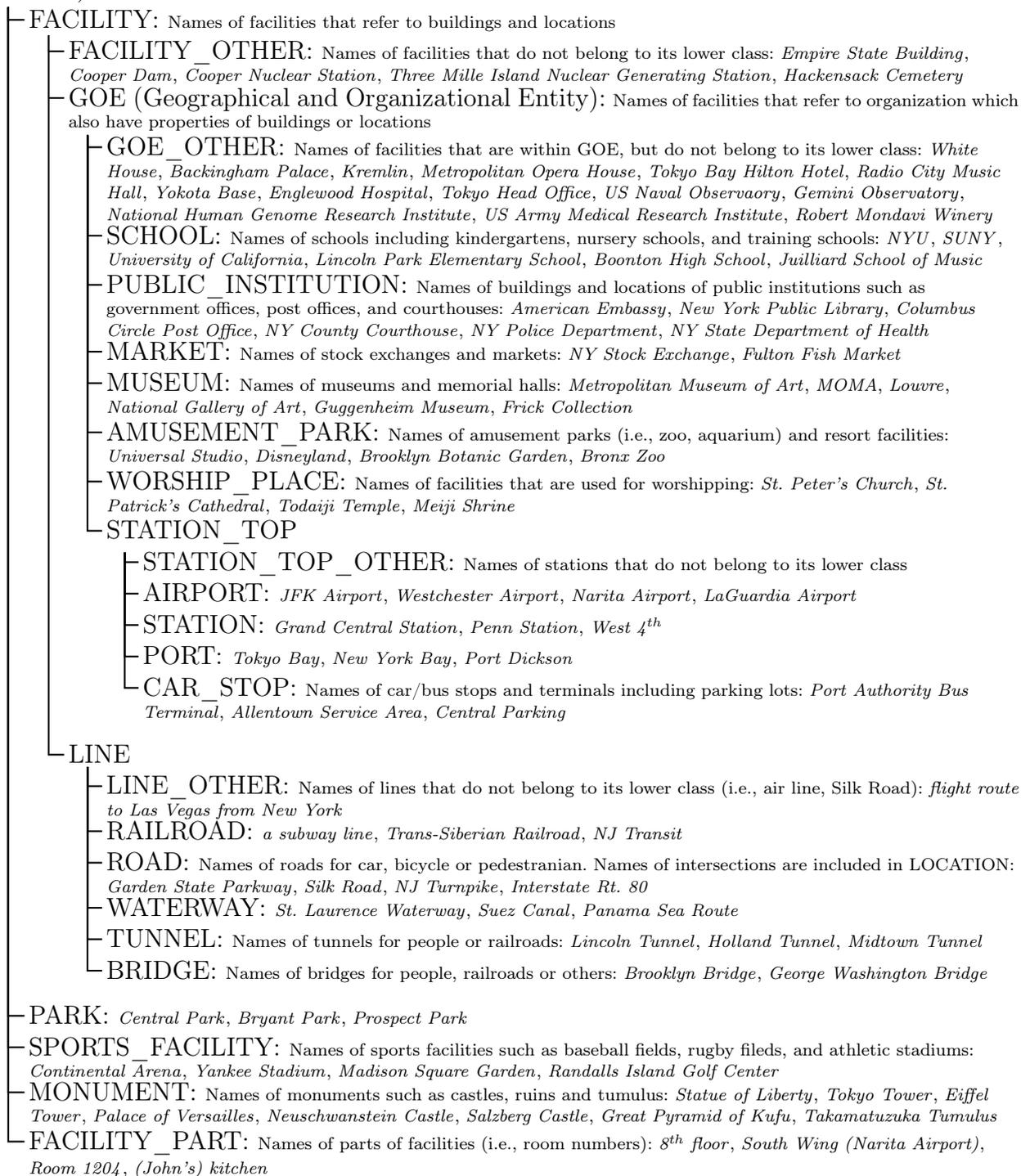


Abbildung A.16: Taxonomie von Sekine, Teil II: Facilities (Quelle: SEKINEEXTNER)

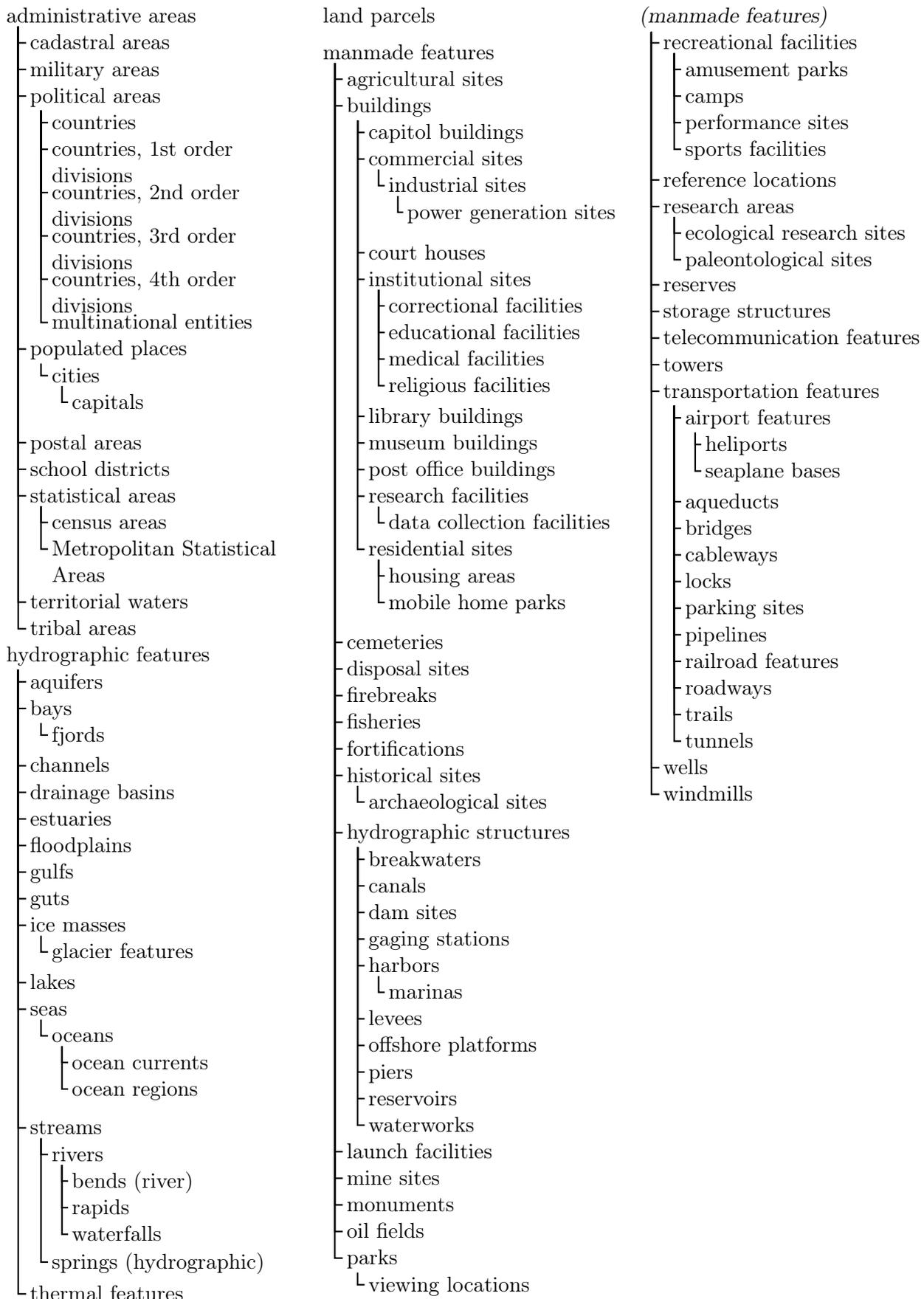


Abbildung A.17: Der ADL Feature Type Thesaurus, Teil 1
 (Quelle: <http://www.alexandria.ucsb.edu/gazetteer/FeatureTypes/ver070302/top.htm>)

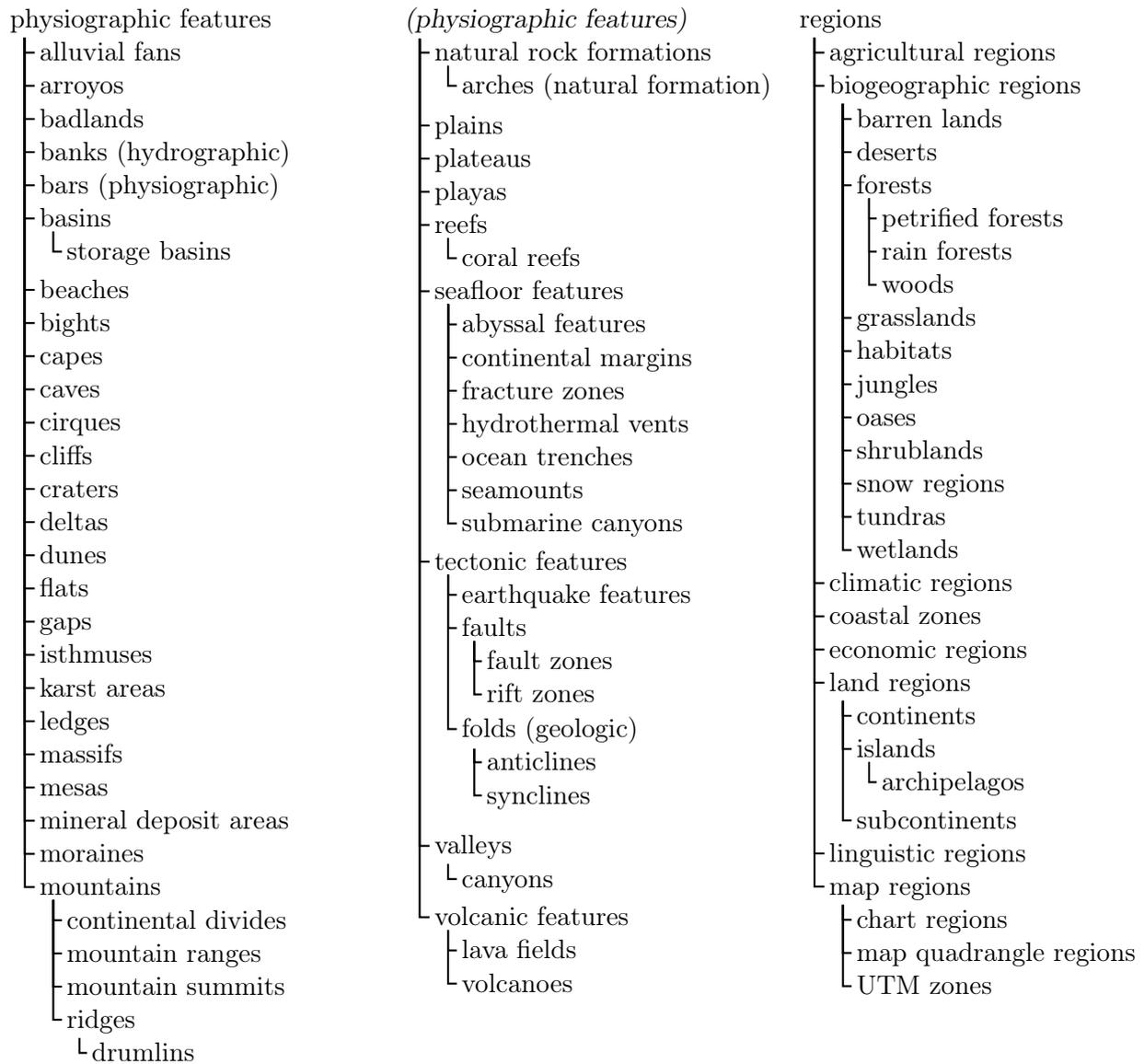


Abbildung A.18: Der ADL Feature Type Thesaurus, Teil 2
 (Quelle: <http://www.alexandria.ucsb.edu/gazetteer/FeatureTypes/ver070302/top.htm>)

NORP

NATIONALITY: Adjectival form of GPE/Location modifier, (*an American, Chinese soil, European organization*, etc.) or head word referring to a person by the name of the GPE or Location they are from.
...

ORGANIZATION NAME: Markable: Names of companies, government agencies, educational institutions and other institutions: sport teams, hospitals, museums, and libraries. *White House, Pentagon, Kremlin* are ORG names as is *Capitol Hill*. Adjectival forms of organization names are also included: Congressional efforts, etc. ...

GOVERNMENT: ...

CORPORATION: ...

EDUCATIONAL: schools, universities

RELIGIOUS: churches, mosques, etc.

POLITICAL: political parties: *Contra(s), Democratic Party, PLO, Sandanista, Taliban*

MUSEUM: museums

HOTEL: hotels, resorts

HOSPITAL: hospitals

OTHER: libraries, unions, environmental agencies, professional associations, health associations (Planned Parenthood), musical groups

Names: city or country name referring to a sports team, i.e. "*Boston shortstop Nomar Garciapara*", *IMF, National Sports Fund, Planned Parenthood, Professional Organizations (Health Ins. Association of America, etc.), Red Sox*, (city names preceding teams are considered part of the team name, as in *New York Yankees*), *World Bank, Yankees*

GPE: Names of countries, cities, states, provinces, municipalities. In cases like *southern California, California* is marked as a GPE name and there is no other markup.

COUNTRY: countries, nations

CITY: cities, towns, villages, *Washington & DC* are tagged separately

STATE/PROVINCE: states, provinces

OTHER: Boroughs, islands that have governments

GPE DESCRIPTOR: Heads of descriptors of countries, cities, states, provinces, municipalities: See GPE Name for subtypes.

FACILITY NAME: Names of man-made structures, including infrastructure, buildings, monuments, etc.

BUILDING: buildings, cathedrals, laboratories, monasteries, plants, refineries, stadiums, synagogues, towers, vineyards, etc.

BRIDGE: bridges: *G Street Bridge, Golden Gate*

AIRPORT: airports, parts of an airport including terminals, runways, and towers, air force bases

HIGHWAY_STREET: avenues, boardwalks, highways, plazas, streets, exits: *Interstate 64, Madison Avenue, Exit 10*

ATTRACTION: amusement parks, tourist attractions, aquariums: *Statue of Liberty*

OTHER: courtyards, farms, (gas) stations, Space Centers, (Labor) camps, mines, oil fields, ports, (subway) stations, golf courses

FACILITY DESCRIPTOR: Head noun of a noun phrase describing buildings, bridges, airports, highways, attractions. . . : *Roads and bridges in the Bay area, traffic intersections, 1.5 million-gallon saltwater aquarium*

[GLEICHE SUBKATEGORIEN wie FACILITY NAME]

LOCATION: Names of locations other than GPE's, e.g. mountain ranges, coasts, borders, planets, geocoordinates, bodies of water. Also included in this category are named regions such as *the Middle East*, areas, neighborhoods, continents and regions of continents such as *Eastern Europe*.

Unmarkable: Do not mark deictics or other non-proper nouns; here, there, everywhere, etc. Directional modifiers such as *western* are not marked.

RIVER

LAKE_SEA_OCEAN: Bay, Gulf

BORDER: named border regions: *the famous Four Corners intersection*

REGION: Named areas larger than a city; *Baltic, Bay Area, Bekka Valley, Benelux, Central America, East bloc, Indochina, Latin America, Long Island, Scandinavia, Silicon Valley, Urals, Western Hemisphere, New England, Central America, the Northeast*

LATITUDE-LONGITUDE

CONTINENT

OTHER: *Brighton Beach, Micro-rayon, Named districts of a city, School district: Earth*

Abbildung A.19: Auszug aus der Taxonomie von BBN-TAXON.



Abbildung A.20: Die Taxonomie von KIM. Die Einbettung der Lokationen in den Kontext weiterer nicht-lokativer Klassen ist hier gegeben, die weitere Untergliederung dieser Klassen wird aber nur teilweise angedeutet. Quelle: <http://www.ontotext.com/kim/kimo.rdfs>

A.2 Lokative Nomina und Klassenbezeichner der Toponyme

Das Lexikon der lokativen Nomina (Kap. 3.1.6), das auch und vor allem Klassenbezeichner für Toponyme enthält, folgt in der Kodierung den Konventionen des CISLEX und des Toponymlexikons (Kap. A.1.2). Hier einige Beispiele aus dem Grundformenlexikon:

Alpentransversale, .N+loc+geoCl+Topon+Hodon:F{NS0;NP4}
Bachlauf, .N+loc+geoCl+Topon+Hydron+Cl(Fluss):M{NS1;NP12}
Badeort, .N+loc+geoCl+Topon+Oikon:M{NS1;NP2}
Bergkette, .N+loc+geoCl+Topon+Choron+Oron:F{NS0;NP4}
Kleinststaat, .N+loc+geoCl+Topon+Choron+admin+Cl(Staat):M{NS1;NP3}
kreisfreie Stadt, .N+loc+geoCl+Topon+Oikon:F{ADJ+NS0;NP12}
Meerenge, .N+loc+geoCl+Topon+Hydron+Cl(Meer)+Cl(Strait):F{NS0;NP4}
Ortschaft, .N+loc+geoCl+Topon+Oikon:F{NS0;NP3}
Powiat, .N+loc+geoCl+Topon+Choron+admin+cl_slot(adm_subj_of):M{NS13;NP6|NP120}/2nd order adm. subdiv. Poland
Ría, .N+loc+geoCl+Topon+Choron:F{NS0;NP6}/Küstenform mit tiefen Buchten
Sonderverwaltungszone, .N+loc+geoCl+Topon+Choron+admin:F{NS0;NP4}
Stadtvilla, .N+loc+geoCl+Topon+Aedif:F{NS0;NP18}
Sultanat, .N+loc+geoCl+Topon+Choron+admin:N{NS1;NP2}
SVZ, Sonderverwaltungszone, .N+loc+geoCl+Topon+Choron+admin:F{NS0}
Villenkolonie, .N+loc+geoCl+Topon+Oikon:F{NS0;NP4}
Weinberg, .N+loc+geoCl+Topon+Choron:M{NS1;NP2}

A.3 Lokative Adjektive und Adverbien

Lokative Adjektive, Adverbien sowie Kombinationen aus Adverbien (Kap. 3.5.1.1 und 3.5.3) wurden teils in ein Lexikon aufgenommen, teils aber auch mit lokalen Grammatiken beschrieben (siehe Kap. B.1 für Beispiele). Falls die Adjektive typischerweise mit bestimmten Toponymklassen stehen, sind diese markiert. Die Adverbien sind als statische (+STAT) oder direktionale (+DIR) annotiert, letztere mit den Unterklassen Ausgangspunkt/Source (+SRC), Zielpunkt/Goal ('destination' = +DST) und Wegpunkte ('passage' = +PASS).

achtstöckig, .ADJ+loc+geoCl+Topon+Aedif{ADJ}
äquatorwärts, .ADV+loc+DIR+DST
auf weite Strecken hin, .ADV+loc+STAT
fast ganz rechts, .ADV+loc+STAT
fußläufig, .ADJ{ADJ}
ganz rechts, .ADV+loc+STAT
gewässerfern, .ADJ+loc+geoCl{ADJ}
gipfelwärts, .ADV+loc+DIR+DST
grenznah, .ADJ+loc+geoCl{ADJ}
hier wie dort, .ADV+loc+STAT/=überall
nach ganz rechts, .ADV+loc+DIR+DST
schneesicher, .ADJ+loc+geoCl+Topon{ADJO}
stadtauswärts, .ADV+loc+DIR+DST
stadteinwärts, .ADV+loc+DIR+DST
über Berg und Tal, .ADV+loc+DIR+PASS
von ganz rechts, .ADV+loc+DIR+SRC
waldreich, .ADJ+loc+geoCl+Topon+Choron{ADJO}
weithin sichtbar, .ADJ+loc{ADJ}

B Lokale Grammatiken

B.1 Komplexe Adverbien und Präpositionen

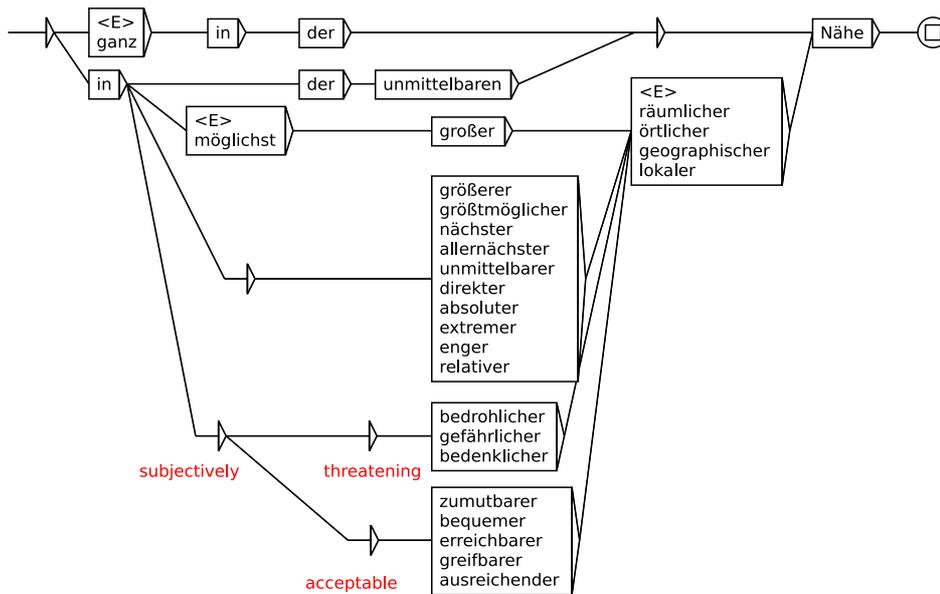


Abbildung B.1: Phrasen mit der Bedeutung ‚in der Nähe‘, als Präposition und Adverbial (siehe Abb. B.2)

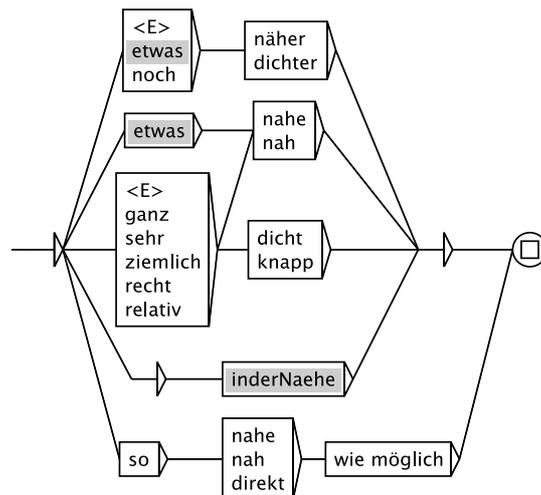


Abbildung B.2: Adverbiale mit der Bedeutung ‚nahe‘. Die Grammatik verwendet als Teilgrammatik Abb. B.1.

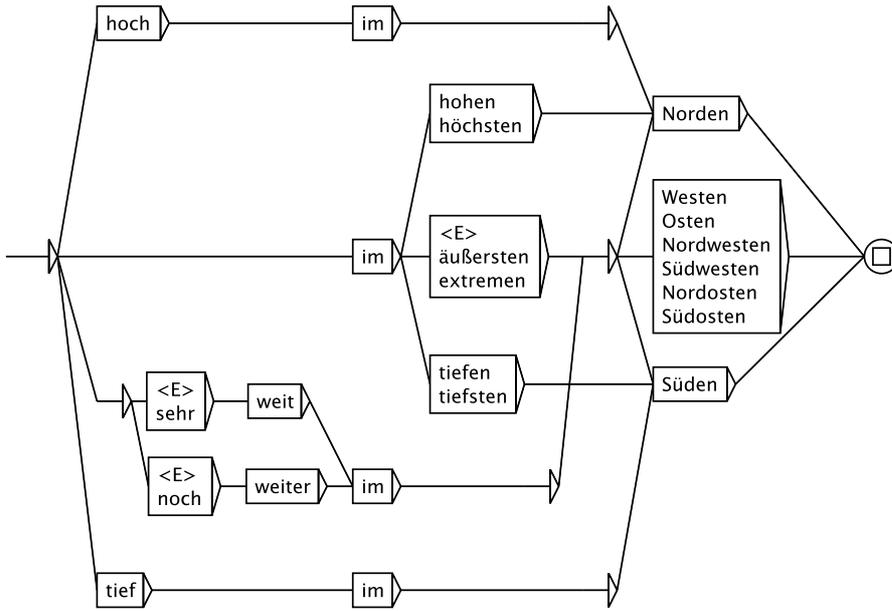


Abbildung B.3: Phrasen des Typs *hoch im Norden*, als Adverbial und Präposition *im Norden von*, vgl. Tab. C.1 und C.2

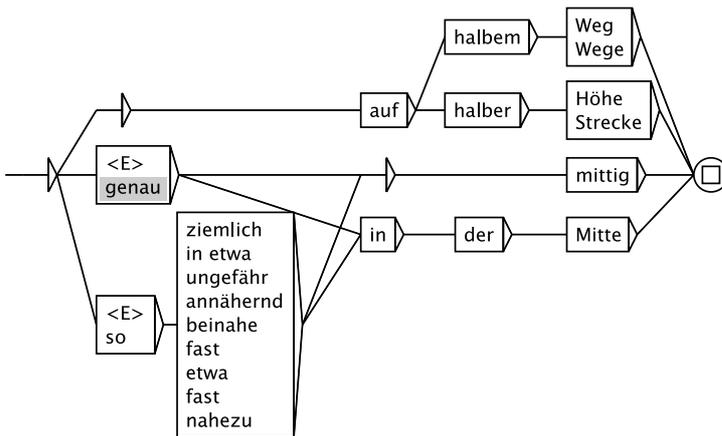


Abbildung B.4: Adverbiale mit der Bedeutung ‚in der Mitte‘

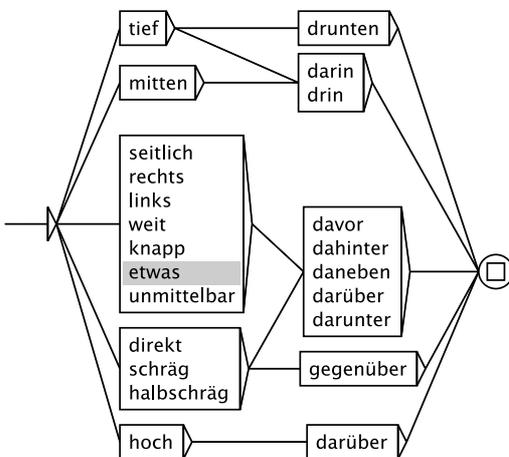


Abbildung B.5: Präpositionaladverbien mit Modifikatoren

phrases expressing height of a location or geographic area
 «1200 m über dem Meeresspiegel»

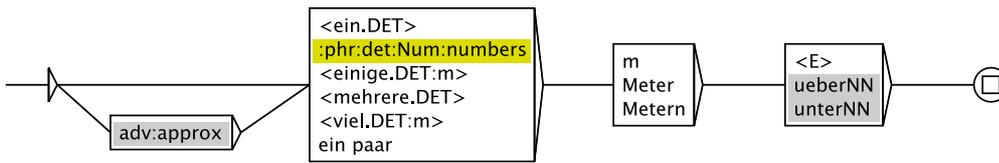


Abbildung B.6: Höhenangaben, siehe auch Abb. B.7 und B.8

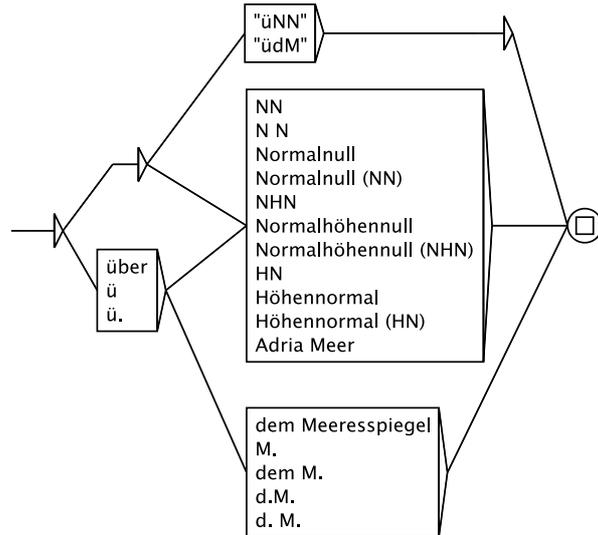


Abbildung B.7: Ausdrücke mit der Bedeutung ‚über dem Meeresspiegel‘ als Teil einer lokalen Grammatik der Höhenangaben (Abb. B.6)

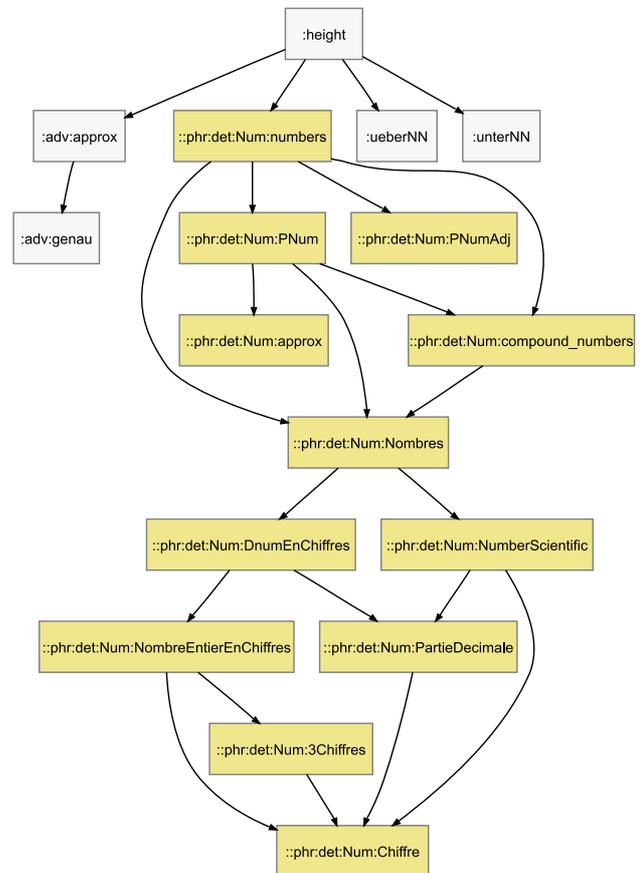


Abbildung B.8: Abhängigkeiten (Abhängigkeiten) des Graphen :height (Abb. B.6) und seiner Subgraphen. Subgraphennamen, die mit '::' beginnen, verweisen auf Graphen einer Graphenbibliothek zur Beschreibung von Zahlen in Worten und Ziffern, die wiederum Grammatiken von Constant (2003b: 43–6) nutzen.

B.3 Antigrammatiken

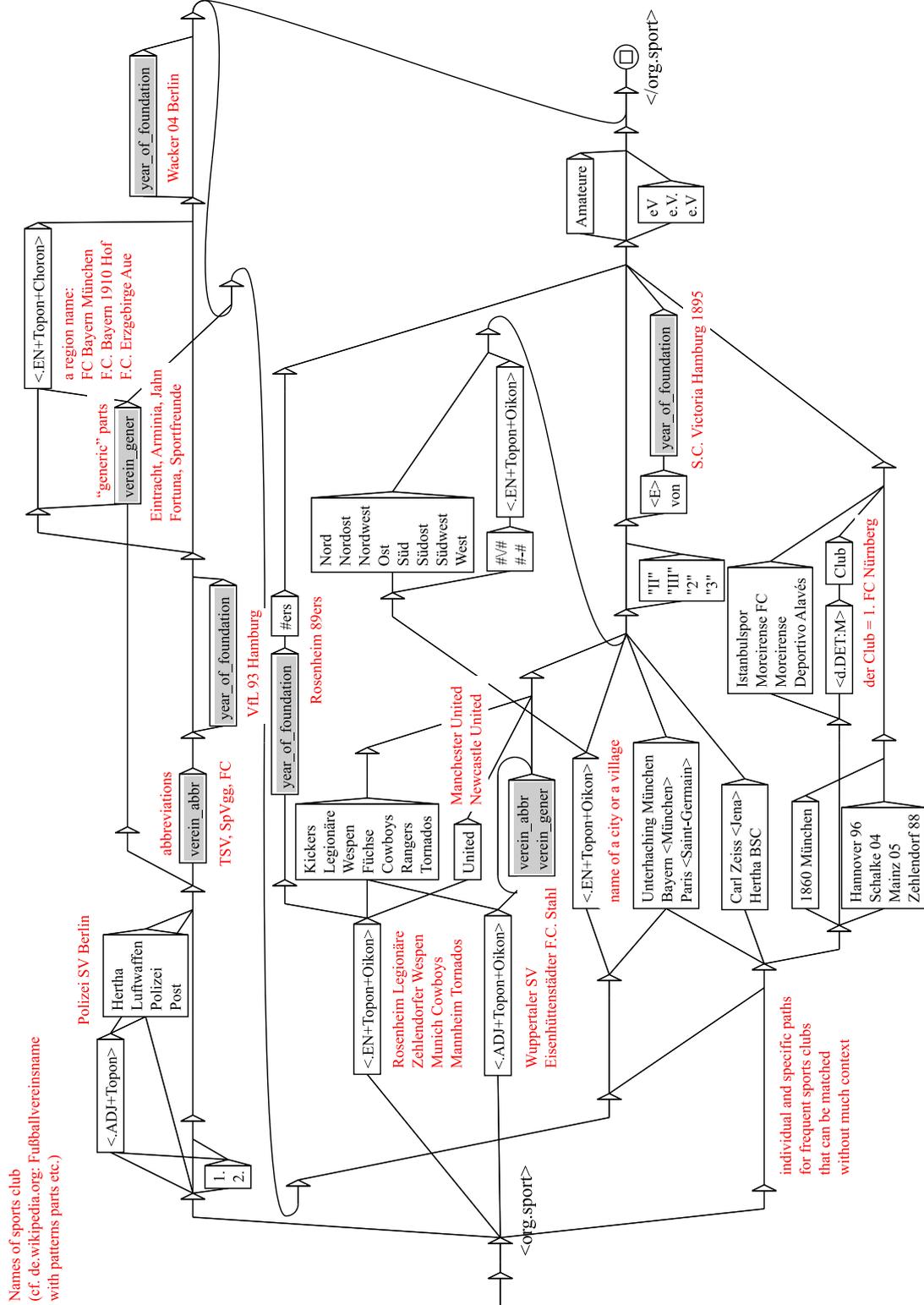


Abbildung B.11: Lokale Grammatik zur Erkennung von Namen von Sportvereinen. Die Grammatik ist angewiesen auf die Subgraphen, die u.a. große Listen von Abkürzungen (*TSV*, *FC*, *EHC*) und Namensrelikte (*Fortuna*, *Viktoria*, *Sportivo*) enthalten.

C Tabellen der Lexikongrammatik

C.1 Präpositionen (Adpositionen)

Räumliche Präposition sind in zwei Tabellen beschrieben, die einmal Form, Wortart und Rektion, zum anderen semantische Eigenschaften und Selektionsrestriktionen beschreiben. Eine ID und das Lemma (Zitierform) ermöglichen die in Tab. C.1 und Tab. C.2 beschriebenen Eigenschaften zusammenzuführen. Die Gründe für diese Aufteilung sind die hohe Spaltenzahl (inklusive Kommentar und Beispielen) und die Eigenschaft von Präpositionen, dass einem Lemma mehrere Formen (Verschmelzungen mit Artikel) zugeordnet sein können, die sich semantische und kombinatorische Eigenschaften teilen. Beispiele in Form vollständiger Sätze sind ab S. 388 vor allem für selten verwendete Präpositionen gegeben.

Syntaktische und semantische Eigenschaften

Tab. C.1 enthält syntaktisch-kombinatorische und semantische Eigenschaften räumlicher Präpositionen:

- mögliche Modifikatoren (siehe auch S. 42). Als Modifikatoren der Präposition gelten Wortformen oder Konstruktionen, die vor der Präposition in einem elementaren Satz des Typs $N_0 V_{sup} Loc N_1 loc$ stehen, aber nicht als selbständige Adverbiale fungieren können:

- (1) a. *Das Hotel liegt (E+direkt+gleich+schräg) gegenüber der Oper.*
b. **Das Hotel liegt (direkt+gleich+schräg).*
c. *Das Hotel liegt in der Innenstadt gegenüber der Oper.*
d. *Das Hotel liegt in der Innenstadt.*

Schräg (bzw. *gleich* oder *direkt*) ist damit Modifikator der Präposition, während *in der Innenstadt* selbständiges Adverbiale in Quasikoordination ist.

Modifikatoren stehen vor der Präposition (1a). Bei Postpositionen stehen sie meist vor der regierten Nominalphrase (2a), manchmal auch zwischen Nominalphrase und Postposition (2b).

- (2) a. *Ein Stück die Straße entlang stehen die Gebäude der neuen Landwirtschaftsschule.*
b. *Die Straße ein Stück entlang stehen ...*

- der semantische Typ des regierten Nomens (N_1). Kann N_1 als Toponym realisiert werden (Oiko-, Choro-, Oro-, Hydro-, Hodonym oder Gebäude)?
- der semantische Typ der Präposition: STATisch, DIRektional (Ausgangs-, Ziel-, Wegpunkt: SouRCE, DeSTination, PASSage), siehe S. 41.
- ist die Präposition adverbial verwendbar, d.h. kann sie auch ohne ein regiertes Nomen stehen, siehe (2.49) auf S. 38.

Morphosyntaktischen Eigenschaften

Tab. C.2 enthält die morphosyntaktischen Eigenschaften räumlicher Präpositionen, nämlich:

- die orthographische Form, für Zirkumpositionen auch postponierte Formbestandteile
 - ein Lemma kann mehrere Formen haben, und jede Form dabei andere Eigenschaften haben, z.B.
 - * eine andere Zuordnung zu den Typen Prä- und Postposition: *die Fahrt den Rhein flussaufwärts*, *die Fahrt flussaufwärts des Rheins*
 - * orthographische Varianten: *flußaufwärts* und *flussaufwärts*
 - * verschiedene mit dem Artikel kontrahierte Formen: *zu*, *zum* (mask., neutr.), *zur*Sind Lemma und Form identisch, wird letztere nicht angegeben.
 - Einige Adpositionen werden durch Graphen repräsentiert (z.B. *im Norden*, *in der Nähe*, vgl. Abb. B.1 und B.3), um Varianten, einfache Analogien (*im Norden*, *Süden*, ..., *Südosten*) und Modifikationen (*in (der+unmittelbarer+...) Nähe*) nicht als eigene Einträge der Lexikogrammatik führen zu müssen.
- die „Wortart“ der Adposition
 - in Abhängigkeit von der Stellung: PREP – Prä-, POSTP – Post- und CIRCP – Zirkumpositionen.
 - für Zusammenziehungen aus Präposition und bestimmtem Artikel: PDET (*im*), CIRCPDET (*zum ... hin*), beide mit Angabe der Flexion, z.B. *im* als :deM:deN (Dativ Sg. mask./neutr.). Die nachfolgenden Angaben zur Rektion können damit entfallen.
- Kasusrektion oder regierte Präposition. Traditionell werden Präpositionen als die drei obliquen Kasus regierend beschrieben, hier kommen aber als Möglichkeiten der Rektion hinzu:
 - Ersetzung des Genitivs durch die Präposition *von* mit Dativ. Dies ist nicht immer möglich (*entlang der Küste* vs. ?*entlang von der Küste*), muss also als Merkmal kodiert werden.
 - in Kombination mit einigen komplexen Präpositionen kann der Anschluss über die Präposition *zu* erfolgen: *in Sichtweite zum Meer*
 - Rektion des Nominativs bei einigen wenigen Präpositionen (*via*, *im Raum*, *Richtung*): *Sie gingen Richtung erster Bezirk*, *via Bayrischer Wald nach Karlsbad*, *im Raum Stuttgart* vs. **im Raum Stuttgarts* (vgl. Di Meola 2000: 99–100)
- Numerusrektion (Plural) der Adpositionen *zwischen* und *zwischen ... hindurch*, siehe S. 41
- Anforderungen an die regierte NP bzgl. Artikel und pränominalen Modifikatoren:
 - obligatorischer Artikel und/oder Modifikator: *in das kurfürstliche München* (direktional), sonst *nach München*
 - immer ohne Artikel und/oder Modifikator: *nach München*, (**/≠nach dem kurfürstlichen München*), *im Raum Stuttgart*, veraltet auch *nach dem nahen Osten*

Tabelle C.1: Syntaktisch-kombinatorische und semantische Eigenschaften räumlicher Präpositionen

id	Lemma	Modifikator	N ₁ =Toponym	N ₁ =Oikonym	N ₁ =Choronym	N ₁ =Hydronym	N ₁ =Oronym	N ₁ =Bauwerk	N ₁ =Hodonym	STATisch	DIRektonal	SouRCe	DeSTination	PASSage	ADV	Kommentar	Beispiel
1	ab		+	+				+	+	-	+	+	-	-	-		<i>(der Zug geht) ab Berlin, ab Ostbahnhof, ab Koburger Brücke, ab Halleschem Tor</i>
2	abseits	:approx	+	+	+	+	+		+	+	-	-	-	-	+		<i>abseits der großen Straßen</i>
3	abwärts	:distance	+			+				+	-	-	-	-	+	statisch	<i>rund 850 m abwärts der Eisenbahnbrücke über die Mur</i>
4	abwärts		+							-	+	-	-	+	+	direktional	<i>den Fluss abwärts</i>
5	am Fuß	direkt+gleich+unmittelbar	-	-	-	-	+	+		+	-	-	-	-	-		<i>am Fuße des Mount Everest</i>
6	:amRand	direkt+gleich+unmittelbar	+	+	+	-	-		+	+	-	-	-	-	+		<i>am Rande des Allgäu, am Rand von Dresden</i>
7	:amUfer	:near+direkt+unmittelbar	+			+				+	-	-	-	-	+		<i>am Ufer der Donau</i>
8	an	:bis+direkt	-	-	-	+	+			-	+	-	+	-	-	direktional, mit Akk.	<i>an den Rhein</i>
9	an	:near+direkt+gleich+unmittelbar	+	+	+	+	+	+	+	+	-	-	-	-	-	statisch, mit Dativ	<i>am Rhein</i>
10	an		+							-	+	-	-	+	-	direktional, prolativ (transitory), nur bei präfigierten Verben der Gruppe N ₀ VPrepN ₁ pass	<i>An der Erde flog ein Asteroid vorbei.</i>
11	an der Grenze zu	:near+direkt+gleich+unmittelbar	+		+					+	-	-	-	-	+	Bed. auch 'nahe'	<i>an der Grenze zur Mecklenburger Seenplatte</i>
12	an ... entlang	:distance+ :near+direkt+unmittelbar+weiter	+			+			+	-	+	-	-	+	-		<i>an der Sieg entlang, an der Ismaninger Straße entlang, am Flußlauf der Alster entlang</i>
13	an ... heran	:bis+direkt+unmittelbar	+			+				-	+	-	+	-	-		<i>an den Nordpol heran</i>
14	:ansUfer	:bis+ :near+direkt+unmittelbar	+			+				-	+	-	+	-	+		<i>ans Ufer der Spree</i>
15	an ... vorbei	hart+knapp+links+rechts	+	+	+	+	+	+	+	-	+	-	-	+	-		<i>an den Kapverdischen Inseln vorbei</i>
16	an ... vorüber	:near+direkt+unmittelbar	+	+		+	+	+	+	-	+	-	-	+	-		<i>an der Johanneskirche vorüber, an einem winzigen Weiler vorbei</i>
17	auf	:bis+direkt+unmittelbar	+	+	+	+	+	+	+	-	+	-	+	-	-	direktional, mit Akk.	<i>auf die Zugspitze</i>
18	auf beiden Seiten		+	-	-	+	+		+	+	-	-	-	-	+	=beiderseits	<i>auf beiden Seiten des Rheins</i>
19	auf		+	-	-	+	+	+	+	+	-	-	-	-	-	statisch, mit Dativ	<i>auf Kuba, auf den Philippinen</i>

Tabelle C.1: Syntaktisch-kombinatorische und semantische Eigenschaften räumlicher Präpositionen (Fortsetzung)

id	Lemma	Modifikator	N ₁ =Toponym	N ₁ =Oikonym	N ₁ =Choronym	N ₁ =Hydronym	N ₁ =Oronym	N ₁ =Bauwerk	N ₁ =Hodonym	STATisch	DIREktional	SouRce	DeSTination	PASSage	ADV	Kommentar	Beispiel
20	auf ... hinauf	: <i>bis</i> + <i>direkt</i> + <i>unmittelbar</i>	+					+	+	-	+	-	+	-	-		<i>auf den Domberg hinauf, auf die Zugspitze hinauf</i>
21	: <i>aufHoehe</i>	<i>direkt</i> + <i>gleich</i> + <i>unmittelbar</i>	+	+	+	+	+	+	+	+	-	-	-	-	-		<i>auf Höhe von Bad Wurzach, auf halber Höhe des Turmes</i>
22	aufwärts		+	-	-	-	+	-		+	-	-	-	-	+	statisch	<i>aufwärts des Finkenbachs</i>
23	aufwärts	: <i>distance</i> + <i>weiter</i>	+							-	+	-	-	+	+	direktional	<i>den Pfad aufwärts</i>
24	: <i>ausallenEcken</i>		+	+	+	-	-		-	-	+	+	-	-	+		<i>aus allen Ecken der Schweiz, aus verschiedenen Ecken der Slowakei</i>
25	: <i>ausallenTeilen</i>		+	+	+	-	-		-	-	+	+	-	-	+		<i>aus allen Teilen von Bosnien und Herzegowina</i>
26	aus	<i>direkt</i> + <i>unmittelbar</i>	+	+	+	+	+	+	+	-	+	+	-	-	-		<i>aus München</i>
27	aus dem Raum		+	+	+	-	-			-	+	+	-	-	-		<i>aus dem Raum Stuttgart</i>
28	aus der Nähe	<i>direkt</i> + <i>gleich</i> + <i>unmittelbar</i>	+	+	+	+	+	+	+	-	+	+	-	-	+		<i>aus der Nähe von Hamburg</i>
29	ausgehend		+	+	+	+	+			-	+	+	-	-	-	meist Zirkumposition von ... ausgehend	<i>vom Tal ausgehend</i>
30	aus ... heraus		+							-	+	+	-	-	-		<i>aus dem Irak heraus, aus dem engen niederbayerischen Plattling heraus</i>
31	aus ... her		-							-	+	+	-	-	-		<i>aus der Schweiz her</i>
32	aus ... hinaus		+					+		-	+	+	-	-	-		<i>aus Jerusalem hinaus</i>
33	aus Richtung		+	+	+	+	+	+	+	-	+	+	-	-	-		<i>aus Richtung Nordwest, aus Richtung Hannover</i>
34	außer		-							-	+	+	-	-	-	sehr eingeschränkte Kontexte	<i>außer Haus / außer Sicht / außer Landes gehen</i>
35	außerhalb	: <i>approx</i> +: <i>etwas</i> + <i>direkt</i> + <i>gleich</i> + <i>unmittelbar</i>	+	+	+					+	-	-	-	-	+		<i>außerhalb der Stadt, außerhalb Münchens</i>
36	auswärts	: <i>approx</i>	+							+	-	-	-	-	+		<i>auswärts der Stadt</i>
37	bachabwärts	: <i>approx</i> +: <i>distance</i> + <i>leicht</i> + <i>nur wenig</i> + <i>weit</i>	+	+				+		+	-	-	-	-	+		<i>300 m bachabwärts der Dorschenmühle</i>
38	bachaufwärts	: <i>approx</i> +: <i>distance</i> + <i>leicht</i> + <i>nur wenig</i> + <i>weit</i>	+	+				+		+	-	-	-	-	+		<i>300 m bachaufwärts des Weilers Paznaun</i>
39	bei	: <i>near</i>	+	+	-	-	+	+	-	+	-	-	-	-	-		<i>bei Leipzig</i>

Tabelle C.1: Syntaktisch-kombinatorische und semantische Eigenschaften räumlicher Präpositionen (Fortsetzung)

id	Lemma	Modifikator	N ₁ =Toponym	N ₁ =Oikonym	N ₁ =Choronym	N ₁ =Hydronym	N ₁ =Oronym	N ₁ =Bauwerk	N ₁ =Hodonym	STATisch	DIRektional	SouRCe	DeSTination	PASSage	ADV	Kommentar	Beispiel
40	beiderseits		+			+	+	+	+	+	-	-	-	-	-		<i>beiderseits der A 9</i>
41	beidseits		+			+	+	+	+	+	-	-	-	-	-		<i>Wandertouren beidseits der Elbe</i>
42	bergseits	<i>:approx+:distance+leicht +nurwenig+weit</i>	+	+					+	+	-	-	-	-	+		<i>bergseits des Bahnhofs</i>
43	bis	<i>direkt</i>	+	+	+	+	+	+	+	-	+	-	+	-	-	falls mit Artikel/ Determinativ, dann kom- biniert mit anderer Prä- position (<i>bis zu</i>)	<i>Er fährt bis Berlin mit dem Zug.</i>
44	bis ... hin		+					+	+	-	+	-	+	-	-		<i>bis Damaskus hin, bis zum Fluss hin, bis Eferding und Grieskirchen hin</i>
45	diesseits	<i>gleich</i>	+	+	+	+	+	+	+	+	-	-	-	-	+		<i>diesseits des Rheins, diesseits der Grenze</i>
46	dies- und jenseits		+	+	+	+	+	+	+	+	-	-	-	-	+		<i>dies- und jenseits des Atlantiks</i>
47	durch	<i>mitten+quer</i>	+	+	+	+	+	+	+	-	+	-	-	+	-		<i>durch die Alpen</i>
48	durch ... hindurch	<i>mitten</i>	+	+	+	+	+	+	+	-	+	-	-	+	-		<i>durch das Rheinische Schiefergebirge hindurch</i>
49	Ecke		+						+	+	-	-	-	-	-		<i>Ecke Maximilianstraße</i>
50	ennet		+		+					+	-	-	-	-	-	schweizerisch 'jenseits'	<i>ennet des Rheins</i>
51	entgegen		+							-	+	-	+	-	-		<i>der Küste entgegen</i>
52	entlang	<i>:distance+direkt+unmittelbar +weiter</i>	+		+				+	+	+	-	-	+	-		<i>(ein Stück) die Straße entlang</i>
53	fernab		+	+	+	+	+	+	+	+	-	-	-	-	+		<i>fernab der Alpen</i>
54	fern		+	+	+	+	+	+	+	+	-	-	-	-	+		<i>fern dem Kaukasus, fern von Medi- na</i>
55	flussabwärts	<i>:approx+:distance+leicht +nurwenig+weit</i>	+	+				+		+	-	-	-	-	+	statisch	<i>15 Meter flussabwärts der Linden- bachbrücke, einige hundert Meter flussabwärts des Klosters</i>
56	flussabwärts	<i>:distance+weiter</i>	+	-	-	+		-	-	-	+	-	-	+	+	direktional	<i>den Rhein flussabwärts nach St. Katharinental</i>

Tabelle C.1: Syntaktisch-kombinatorische und semantische Eigenschaften räumlicher Präpositionen (Fortsetzung)

id	Lemma	Modifikator	N ₁ =Toponym	N ₁ =Oikonym	N ₁ =Choronym	N ₁ =Hydronym	N ₁ =Oronym	N ₁ =Bauwerk	N ₁ =Hodonym	STATisch	DIREktional	SouRce	DeSTination	PASSage	ADV	Kommentar	Beispiel
57	flussauf- und -abwärts		+	+				+		+	-	-	-	-	+	statisch	<i>die malerischen Orte flussauf- und abwärts der Moselscheife</i>
58	flussaufwärts	: <i>approx</i> +: <i>distance</i> + <i>leicht</i> + <i>nur wenig</i> + <i>weit</i>	+	+				+		+	-	-	-	-	+	statisch	<i>etwa 20 Kilometer flussaufwärts von Bonn</i>
59	flussaufwärts	: <i>distance</i> + <i>weiter</i>	+	-	-	+		-	-	-	+	-	-	+	+	direktional	<i>bei der Weiterfahrt flussaufwärts des Kemijoki</i>
60	gegen	: <i>bis</i> + <i>direkt</i>	+	-	-	-	+	+	-	-	+	-	+	-	-	direktional	<i>gegen die Wand, bis gegen Tirschenreuth</i>
61	gegen ... hin		-							-	+	-	+	-	-	kaum mit Toponymen oder veraltet	<i>gegen Kyrene hin, gegen Norden hin</i>
62	gegenüber	<i>direkt</i> + <i>gleich</i> + <i>schräg</i> + <i>unmittelbar</i>	+	+	+	+	+	+	+	+	-	-	-	-	+		<i>der Oper gegenüber, gegenüber des Teterower Gartenmarktes, gegenüber des Sears Tower</i>
63	gegen ... zu		-							-	+	-	+	-	-	kaum mit Toponymen	<i>gegen Süden zu</i>
64	gen		+	+	+					-	+	-	-	-	-		<i>gen Rom, gen Mekka</i>
65	gen ... hin		-							-	+	-	+	-	-	kaum mit Toponymen	<i>gen Himmel hin, gen Westen hin</i>
66	gen ... zu		-							+	+	-	+	+	-		<i>gen Süden zu</i>
67	herwärts		+			+	+			+	-	-	-	-	+		<i>die Chubs im Kanton St. Gallen herwärts des Rickens</i>
68	hinab	: <i>distance</i> + <i>weit</i>	+	-	-	+	+	+		+	+	-	+	+	+		<i>den Hang hinab</i>
69	hinauf	: <i>distance</i> + <i>weit</i>	+	-	-	+	+	+		+	+	-	+	+	+		<i>den Berg hinauf</i>
70	hinseits		+							+	-	-	-	-	-	selten	<i>hinseits der Grenze</i>
71	hinter	: <i>bis</i> +: <i>distance</i> +: <i>near</i> + <i>direkt</i> + <i>gleich</i> + <i>links</i> + <i>rechts</i> + <i>schräg</i> + <i>seitlich</i> + <i>unmittelbar</i> + <i>weit</i>	+	+	+	+	+	+	+	-	+	-	+	-	-	direktional, mit Akk.	<i>hinter das Haus</i>
72	hinter	: <i>distance</i> +: <i>near</i> + <i>direkt</i> + <i>gleich</i> + <i>links</i> + <i>rechts</i> + <i>schräg</i> + <i>seitlich</i> + <i>unmittelbar</i> + <i>weit</i>	+	+	+	+	+	+	+	+	-	-	-	-	-	statisch, mit Dativ	<i>hinter dem Haus</i>
73	hinter ... her		-							-	+	-	+	-	-	=folgen, verfolgen	
74	hinter ... hervor		-							-	+	-	+	-	-	kaum mit Toponymen	<i>hinter der Hecke hervor</i>
75	hinter ... hin		-							-	+	-	+	-	-	kaum mit Toponymen	

Tabelle C.1: Syntaktisch-kombinatorische und semantische Eigenschaften räumlicher Präpositionen (Fortsetzung)

id	Lemma	Modifikator	N ₁ =Toponym	N ₁ =Oikonym	N ₁ =Choronym	N ₁ =Hydronym	N ₁ =Oronym	N ₁ =Bauwerk	N ₁ =Hodonym	STATisch	DIRektonal	SouRCe	DeSTination	PASSage	ADV	Kommentar	Beispiel
76	hinterwärts		+	+						+	-	-	-	-	-	selten	<i>hinterwärts von Temesvar</i>
77	hinunter	: <i>distance</i> + <i>weit</i>	+	-	-	+	+		+	+	-	+	+	+			<i>den Berg hinunter</i>
78	hinwärts		-							+	-	-	-	-	+	sehr selten, kaum mit Ortsnamen	<i>hinwärts von dir</i>
79	Höhe		+	+	+			+	+	+	-	-	-	-	-		<i>in Bochum, Höhe Haarstraße</i>
80	im Bereich	<i>direkt</i> + <i>unmittelbar</i>	+							+	-	-	-	-	+		<i>im Bereich von Berlin (West)</i>
81	im Herzen	<i>direkt</i> + <i>unmittelbar</i>	+	+	+	-	-			+	-	-	-	-	-		<i>im Herzen Münchens</i>
82	: <i>imHintWinkel</i>		+	+	+	-	+	-		+	-	-	-	-	+		<i>im hintersten Winkel Bayerns</i>
83	: <i>imLandesinneren</i>	: <i>distance</i> + <i>tief</i> + <i>weit</i> + <i>weiter</i>	+	+	+	-	-	-	-	+	-	-	-	-	+		<i>300 km im Landesinneren der Insel</i>
84	: <i>imNorden</i>		+	+	+	+	+			+	-	-	-	-	+		<i>im Norden von Bayern</i>
85	im Raum		+							+	-	-	-	-	-		<i>im Raum Hannover</i>
86	: <i>imUmfeld</i>	<i>direkt</i> + <i>unmittelbar</i>	+	+	+			+	+	+	-	-	-	-	+		<i>im Umfeld des Flughafens München II, im östlichen Umfeld von Berlin</i>
87	: <i>imUmkreis</i>		+	+				+		+	-	-	-	-	+	vgl. <i>um</i>	<i>im Umkreis von Erfurt</i>
88	: <i>imUmland</i>		+	+	-	-	-	-	-	+	-	-	-	-	+		<i>im Umland von Berlin</i>
89	in	: <i>bis</i> +: <i>deep</i> +: <i>high</i> + <i>direkt</i> + <i>mitten</i> + <i>unmittelbar</i>	+	+	+	+	+	+	+	-	+	-	+	-	-	direktional, mit Akk.	<i>in den Libanon</i>
90	: <i>inallenEcken</i>		+	+	+	-	-	-		+	-	-	-	-	+		<i>in allen Ecken der Stadt</i>
91	: <i>inallenTeilen</i>		+	+	+	-	-	+	-	+	-	-	-	-	+		<i>in allen Teilen von NRW, in allen Teilen der Türkei</i>
92	: <i>inalleTeile</i>		+	+	+	-	-	+	-	-	+	-	+	-	+		<i>in alle Teile Europas</i>
93	in	: <i>deep</i> +: <i>high</i> + <i>direkt</i> + <i>gleich</i> + <i>mitten</i> + <i>tief</i> + <i>unmittelbar</i> + <i>weit</i>	+	+	+	+	+	+	+	+	-	-	-	-	-	statisch, mit Dativ	<i>im Libanon, in der Schweiz</i>
94	: <i>indenNorden</i>	: <i>bis</i>	+	+	+	+	+			+	-	-	-	-	+		<i>in den Westen der USA</i>
95	in den Raum		+	+	-	-	-	-	-	-	+	-	+	-	-		<i>in den Raum Hannover</i>
96	in der Mitte	<i>direkt</i> + <i>gleich</i>	+	+	+	+				+	-	-	-	-	+		<i>in der Mitte von Berlin</i>

Tabelle C.1: Syntaktisch-kombinatorische und semantische Eigenschaften räumlicher Präpositionen (Fortsetzung)

id	Lemma	Modifikator	N ₁ =Toponym	N ₁ =Oikonym	N ₁ =Choronym	N ₁ =Hydronym	N ₁ =Oronym	N ₁ =Bauwerk	N ₁ =Hodonym	STATisch	DIREktional	SouRce	DeSTination	PASSage	ADV	Kommentar	Beispiel
97	:inderNaehe	<i>fast+gleich</i>	+	+	+	+	+	+	+	+	-	-	-	-	+	z.T. auch mit "inkorporiertem" Aktanten: <i>in Hafennähe, in Bahnhofsnähe, in Innenstadtnähe, in Ostseenähe</i>	<i>in der Nähe von Paris</i>
98	in die Mitte	<i>direkt+gleich</i>	+	+	+	+				-	+	-	+	-	+		<i>in die Mitte der Stadt</i>
99	in ... herein	<i>:bis+direkt</i>	+					+		-	+	-	+	-	-		<i>in den Indischen Ozean herein</i>
100	in ... hinab		+	+	+					-	+	-	+	-	-		<i>in die Rheinebene hinab</i>
101	in ... hinein	<i>:bis+direkt</i>	+	+	+	+	+	+	-	-	+	-	+	-	-		<i>bis weit in das Münsterland hinein</i>
102	in ... hinunter		+	+						-	+	-	+	-	-		<i>in das Tal der Mangfall hinunter</i>
103	inmitten		+	+	+	+				+	-	-	-	-	-		<i>inmitten ausgedehnter Parkanlagen</i>
104	:inNachbarschaft	<i>direkt+gleich</i>	+	+	+			+	+	+	-	-	-	-	+		<i>in der Nachbarschaft Berlins</i>
105	innerhalb		+	+	+					+	-	-	-	-	+		<i>innerhalb Leipzigs, innerhalb von Leipzig</i>
106	innert		+							+	-	-	-	-	-	schwyzerdütsch 'binnen, innerhalb', meist temporal	<i>innert der Türkei</i>
107	inner- und außerhalb		+	+						+	-	-	-	-	-		<i>inner- und außerhalb Berlins</i>
108	in Richtung auf	<i>direkt+exakt</i>	+	+						-	+	-	+	-	-		<i>in Richtung auf die Stadt</i>
109	in Richtung auf ... zu	<i>direkt+exakt</i>	+	+						-	+	-	+	-	-		<i>in Richtung auf den Berg zu</i>
110	in Richtung	<i>direkt+exakt</i>	+	+	+					-	+	-	+	-	-		<i>in Richtung des Waldes, in Richtung Berlin, Richtung Berlin-Mitte</i>
111	ins Herz	<i>direkt+unmittelbar</i>	+	+	+	-	-			-	+	-	+	-	-		<i>von der Atlantikküste ins Herz von Paris</i>
112	:inSichtweite	<i>direkt+kaum+unmittelbar</i>	+	+	+	+	+	+	+	+	-	-	-	-	+		<i>in Sichtweite der Martinskirche, in Sichtweite zur Freiheitsstatue</i>
113	:inderUmgebung	<i>direkt</i>	+	+	+	+	+	+	+	+	-	-	-	-	+		<i>in der näheren Umgebung der Stadt Ueckermünde</i>
114	:inderUmgegend	<i>direkt</i>	+	+	+			+	+	+	-	-	-	-	+		<i>in der Umgegend von Frankfurt a. M., in der näheren Umgegend Bilbaos</i>

Tabelle C.1: Syntaktisch-kombinatorische und semantische Eigenschaften räumlicher Präpositionen (Fortsetzung)

id	Lemma	Modifikator	N ₁ =Toponym	N ₁ =Oikonym	N ₁ =Choronym	N ₁ =Hydronym	N ₁ =Oronym	N ₁ =Bauwerk	N ₁ =Hodonym	STATisch	DIRektional	SouRCe	DeSTination	PASSage	ADV	Kommentar	Beispiel
115	in, um und um ... herum		+	+						+	-	-	-	-	-	vom Zungenbrecher <i>in Ulm und um Ulm und um Ulm herum</i> , jetzt aber auch mit anderen Ortsnamen möglich	<i>in, um und um Ulm herum</i>
116	in und außerhalb		+							+	-	-	-	-	-		<i>in und ausserhalb der Dominikanischen Republik, in- und außerhalb Kölns</i>
117	in und um		+	+	+	+		+		+	-	-	-	-	-		<i>in und um Berlin</i>
118	inwärts		+							+	-	-	-	-	+	selten, veraltet	<i>inwärts der Thüre</i>
119	jenseit		+							+	-	-	-	-	-	veraltet	<i>jenseit der Berge, jenseit der Donau</i>
120	jenseits	: <i>distance</i> + <i>weit</i>	+	+	+	+	+	+	+	+	-	-	-	-	+		<i>jenseits der Saale</i>
121	längsseit		+							+	-	-	-	-	+		<i>längsseit der Küste</i>
122	längsseits		+							+	-	-	-	-	+		<i>längsseits der Brigg</i>
123	längs		+		+					+	-	-	-	-	-	statisch	<i>längs der Straße parken Autos</i>
124	längs	<i>weiter</i>	+		+					-	+	-	-	+	-	direktional	<i>reizvolle Uferwanderung längs der Aare</i>
125	landseits	: <i>approx</i> +: <i>distance</i> + <i>leicht</i> + <i>nur wenig</i> + <i>weit</i>	+							+	-	-	-	-	+		<i>landseits der Uferlinie, landseits des Hochwasserdeiches</i>
126	lang		+		+					-	+	-	-	-	-	direktional	<i>die Staße lang</i>
127	langseit		-							+	-	-	-	-	+	selten, veraltet	
128	linker Hand		+	-	-	+	+		+	+	-	-	-	-	+		<i>linker Hand des Weges</i>
129	links	: <i>approx</i> +: <i>distance</i> + <i>leicht</i> + <i>nur wenig</i> + <i>weit</i>	+		+					+	-	-	-	-	+		<i>links der Donau</i>
130	mang		+							+	-	-	-	-	-	norddt., berlin. ‘mitten unter’	
131	mittenmang		+							+	-	-	-	-	-	norddt., berlin. ‘mitten unter’	<i>mittenmang der Gropius-Stadt</i>
132	nach ... hinein	: <i>bis</i> + <i>direkt</i>	+	+	+	-	-	-	-	-	+	-	+	-	-		<i>bis weit nach Ostfriesland hinein</i>
133	nach ... hin	: <i>bis</i> + <i>direkt</i>	+							+	+	-	+	-	-		<i>bis nach Zoar hin</i>
134	nach	: <i>bis</i> + <i>direkt</i>	+	+	+	-	-	-	-	-	+	-	+	-	-		<i>nach Afrika, nach dem nahen Osten</i>

Tabelle C.1: Syntaktisch-kombinatorische und semantische Eigenschaften räumlicher Präpositionen (Fortsetzung)

id	Lemma	Modifikator	N ₁ =Toponym	N ₁ =Oikonym	N ₁ =Choronym	N ₁ =Hydronym	N ₁ =Oronym	N ₁ =Bauwerk	N ₁ =Hodonym	STATisch	DIREktional	SouRCE	DeSTination	PASSage	ADV	Kommentar	Beispiel
135	nach ... zu	:bis+direkt	+							-	+	-	-	-	-	veraltet	nach dem Norden zu, nach dem Rhein zu
136	Nähe		+	+	+			+	+	+	-	-	-	-	-		Nähe Katakomben San Giovanni
137	nahe	ganz	+	+	+	+	+	+	+	+	-	-	-	-	+		nahe der Elbe
138	neben	:bis+:near+direkt+gleich+links +rechts+schräg+seitlich+unmittelbar	+	-						-	+	-	+	-	-	direktional, mit Akk.	neben den Baum
139	neben	:near+direkt+gleich+links+rechts +schräg+seitlich+unmittelbar	+	+	+	+	+			+	-	-	-	-	-	statisch, mit Dativ	neben dem Haus
140	neben ... her		-							-	+	-	-	+	-		neben der Straße her
141	neben ... hin		-							-	+	-	+	-	-	selten	neben dem Pferdebahngleise hin
142	nid		+							+	-	-	-	-	-	oberdt. 'unterhalb', praktisch nur in festen Wendungen und Namen	Unterwalden nid dem Wald
143	niederwärts		+							+	-	-	-	-	+	selten, meist als Adverb	die Havel niederwärts
144	:noerdlich		+	+	+	+	+	+	+	+	-	-	-	-	+		nördlich von Berlin, nördlich des Polarkreises
145	:nordwaerts	:approx+:distanceWeiter+leicht +weit	+	+	+	+	+	+	+	+	-	-	-	-	+		südwärts der Ruhr, nord- und südwärts der Alpen
146	oberhalb	:approx+:distance+leicht +nurwenig+weit	+							+	-	-	-	-	+		oberhalb des Spitzingsees
147	ober		+							+	-	-	-	-	-	praktisch nur in festen Wendungen und Namen	Latschach ober dem Faaker See
148	oberwärts		+							+	-	-	-	-	-	selten	oberwärts der Stadt
149	oberwasserseits	:approx+:distance+leicht +nurwenig+weit	+							+	-	-	-	-	+	=(fluss)aufwärts, selten	oberwasserseits der Stadt Birecik
150	ob		+		+					+	-	-	-	-	-	praktisch nur in festen Wendungen und Namen	Unterwalden ob dem Wald, Rothenburg ob der Tauber
151	rauf	:distance+weit	+	-	-	+	+		+	+	+	-	+	+	+		die Straße rauf
152	rechter Hand		+	-	-	+	+		+	+	-	-	-	-	+		rechter Hand des Nagoldtales
153	rechts	:approx+:distance+leicht +nurwenig+weit	+		+					+	-	-	-	-	+		rechts der Isar

Tabelle C.1: Syntaktisch-kombinatorische und semantische Eigenschaften räumlicher Präpositionen (Fortsetzung)

id	Lemma	Modifikator	N ₁ =Toponym	N ₁ =Oikonym	N ₁ =Choronym	N ₁ =Hydronym	N ₁ =Oronym	N ₁ =Bauwerk	N ₁ =Hodonym	STATisch	DIRektonal	SouRce	DeSTination	PASSage	ADV	Kommentar	Beispiel
154	rechts und links		+	-	-	+	+		+	+	-	-	-	-	+	=beiderseits	<i>rechts und links des Rheins, rechts und links des Lahnhöhenweges</i>
155	ringsum	:distance	+							+	-	-	-	-	+		<i>ringsum den Hügel, ringsum den Faaker See</i>
156	rückwärts		+							+	-	-	-	-	+		<i>in meinem rückwärts des Hauses gelegenen Obstgarten</i>
157	runter	:distance+weit	+	-	-	+	+		+	+	-	+	+	+			<i>den Broadway runter</i>
158	seeseits	:approx+:distance+leicht +nur wenig+weit	+							+	-	-	-	-	+		<i>seeseits der Deiche, seeseits der Murtenstrasse</i>
159	seitab		+							+	-	-	-	-	+		<i>seitab der Straße</i>
160	seitlich	:etwas	+							+	-	-	-	-	+		<i>Bushaltestelle seitlich des Hauptbahnhofs</i>
161	seitwärts	:etwas	+							+	-	-	-	-	+		<i>seitwärts der vielbefahrenen Route</i>
162	stromabwärts	:approx+:distance+leicht +nur wenig+weit	+	+				+		+	-	-	-	-	+	statisch	<i>stromabwärts des Staudamms</i>
163	stromabwärts	:distance+weiter	+	-	-	+		-	-	-	+	-	-	+	+	direktional	<i>den Rhein stromabwärts</i>
164	stromaufwärts	:approx+:distance+leicht +nur wenig+weit	+	+				+		+	-	-	-	-	+	statisch	<i>stromaufwärts der Anlegestelle</i>
165	stromaufwärts	:distance+weiter	+	-	-	+		-	-	-	+	-	-	+	+	direktional	<i>den Niger stromaufwärts</i>
166	talseits	:approx+:distance+leicht +nur wenig+weit	+							+	-	-	-	-	+		<i>talseits des Weges</i>
167	talwärts		+	+						+	-	-	-	-	+		<i>talwärts des Ortsteils Monterenzio</i>
168	über	:distance+:high+:near+links +rechts+schräg+seitlich+weit	-							-	+	-	+	-	-	direktional, illativ (goal), mit Akk.	<i>Häng das Bild über die Tür!</i>
169	über	:distance	+	+	+	+	+	+	+	-	+	-	-	+	-	direktional, prolativ (transitory), mit Akk.	<i>über den Kirchturm</i>
170	über	:distance+:high+:near+links +rechts+schräg+seitlich+weit	+	+	+	+	+	+	+	+	-	-	-	-	-	statisch, mit Dativ	<i>über München, über dem Turm</i>
171	über ... hinaus	:bis+weit	-							-	+	-	+	-	-	selten mit Toponymen, häufig übertragen	<i>über die Landesgrenzen hinaus</i>
172	über ... hin		-							-	+	-	-	+	-	selten	<i>bis weit über den Zürichsee hin</i>

Tabelle C.1: Syntaktisch-kombinatorische und semantische Eigenschaften räumlicher Präpositionen (Fortsetzung)

id	Lemma	Modifikator	N ₁ =Toponym	N ₁ =Oikonym	N ₁ =Choronym	N ₁ =Hydronym	N ₁ =Oronym	N ₁ =Bauwerk	N ₁ =Hodonym	STATisch	DIREktional	SouRce	DeSTination	PASSage	ADV	Kommentar	Beispiel
173	über ... hinweg		+	+	+	+	+	+	+	-	+	-	-	+	-		<i>übers Rheintal hinweg, über das Schneefernerhaus hinweg, über die Inseln Martinique, St. Eustatius und Barbados hinweg</i>
174	über ... weg		-							-	+	-	-	+	-		<i>über den Fluß weg</i>
175	um ... her		+	+						+	-	-	-	-	-	veraltet	<i>um Jerusalem her</i>
176	um	<i>:distance+ :imUmkreisvonDistance+rings+rund</i>	+	+	+	+	+	+	+	+	-	-	-	+	-		<i>rund um Berlin, rund um die Nürnberger Altstadt</i>
177	unfern		+	+	+	+				+	-	-	-	-	+		<i>unfern der Stadt</i>
178	unter	<i>:bis+ :deep+ :distance+ :near+ direkt+ exakt+ genau+ gleich+ schräg+ seitlich+ unmittelbar+ weit</i>	+	+	+	+	+	+	+	-	+	-	+	-	-	direktional, mit Akk.	<i>unter den Baum</i>
179	unter	<i>:deep+ :distance+ :near+ direkt+ exakt+ gleich+ schräg+ seitlich+ unmittelbar+ weit</i>	+	+	+	+	+	+	+	+	-	-	-	-	-	statisch, mit Dativ	<i>unter dem Baum</i>
180	unterhalb	<i>:distance+ :near+ direkt+ gleich+ leicht+ nur wenig+ schräg+ seitlich+ unmittelbar+ weit</i>	+							+	-	-	-	-	+		<i>unterhalb der Brücke</i>
181	unter ... her		-							-	+	-	-	+	-		<i>unter der Bahnlinie her</i>
182	unter ... hindurch	<i>direkt</i>	+					+	+	-	+	-	-	+	-		<i>unter dem Ärmelkanal hindurch, unter den Gleisen hindurch</i>
183	unterkant		-							+	-	-	-	-	-	praktisch nicht mit Toponymen	<i>unterkant des Fensters</i>
184	unterwärts		+							+	-	-	-	-	+	selten	<i>das Städtchen unterwärts des Flusses</i>
185	unterwasserseits	<i>:approx+ :distance+ leicht+ nur wenig+ weit</i>	+							+	-	-	-	-	+	=(fluss)abwärts	<i>unterwasserseits der Brücke</i>
186	unweit		+	+	+			+	+	+	-	-	-	-	+		<i>unweit der Kaulsdorfer Seen, unweit der Mosel</i>
187	via		+	+						-	+	-	-	-	-		<i>via Frankfurt</i>
188	vis-à-vis		+							+	-	-	-	-	+		<i>vis-a-vis des Bahnhofs</i>
189	vom Ufer	<i>direkt+ unmittelbar</i>	+		+					-	+	+	-	-	+		<i>vom Ufer des Sees</i>

Tabelle C.1: Syntaktisch-kombinatorische und semantische Eigenschaften räumlicher Präpositionen (Fortsetzung)

id	Lemma	Modifikator	N ₁ =Toponym	N ₁ =Oikonym	N ₁ =Choronym	N ₁ =Hydronym	N ₁ =Oronym	N ₁ =Bauwerk	N ₁ =Hodonym	STATisch	DIRektonal	SouRCe	DeSTination	PASSage	ADV	Kommentar	Beispiel
190	von ... ab	<i>direkt+gleich+unmittelbar</i>	+	+						-	+	+	-	-	-		<i>diese von Skandinavien ab ostwärts vorkommende Vogelart</i>
191	von ... an	<i>direkt+gleich+unmittelbar</i>	+	+	+	+	+			-	+	+	-	-	-	=ab hier	<i>von der Brücke an</i>
192	von ... aus		+	+	+	+	+	+	+	-	+	+	-	-	-		<i>von Frankfurt aus</i>
193	von beiden Seiten		+	-	-	+	+		+	-	+	+	-	-	+	=von beiderseits	<i>von beiden Seiten der Mauer</i>
194	von beiderseits		+	-	-	+	+		+	-	+	+	-	-	-		<i>von beiderseits des Rheins</i>
195	von ... her		+	+	+	+	+			-	+	+	-	-	-		<i>vom Bayerischen Wald und Berchtesgadener Land her</i>
196	von ... aus		+	+	+	+	+	+	+	-	+	+	-	-	-		<i>Georg Franz hatte von hinter dem Tor aus Oliver Vöst bedient.</i>
197	von hinter	<i>direkt+unmittelbar</i>	+	+	+	+	+			-	+	+	-	-	-		<i>von hinter der Bühne</i>
198	von jenseits		+	-	+	+	+			-	+	+	-	-	+		<i>von jenseits des Atlantiks</i>
199	von neben		+	+	+	+	+			-	+	+	-	-	-		<i>Blick von neben dem Haus auf den Strand</i>
200	von und nach		+	+	+	-	-	-	-	-	+	+	+	-	-		<i>3000 Sonderzüge von und nach Nürnberg</i>
201	von		+	+	+	+	+			-	+	+	-	-	-		<i>von Rügen nach Hiddensee</i>
202	von vor		+	+	+	+	+			-	+	+	-	-	-		<i>von vor Rostock über Heiligendamm bis nach Kühlungsborn</i>
203	von ... weg		+	+	+	+	+			-	+	+	-	-	-		<i>von den Kanarischen Inseln weg, von der A4 weg</i>
204	vor	<i>:bis+ :distance+ :near+ direkt+ exakt+ gleich+ links+ rechts+ schräg+ seitlich+ unmittelbar+ weit</i>	+	+	+	+	+			-	+	-	+	-	-	direktional, mit Akk.	<i>vor das Haus</i>
205	vor	<i>:distance+ :near+ direkt+ gleich+ links+ rechts+ schräg+ seitlich+ unmittelbar+ weit</i>	+	+	+	+	+			+	-	-	-	-	-	statisch, mit Dativ	<i>der Platz vor dem Bahnhof</i>
206	vor den Toren	<i>:distance+ direkt+ gleich+ unmittelbar+ weit</i>	+	+						+	-	-	-	-	+		<i>vor den Toren Berlins</i>
207	vor die Tore	<i>:bis+ :distance+ direkt+ gleich+ unmittelbar</i>	+	+						-	+	-	+	-	+		<i>vor die Tore Wiens</i>

Tabelle C.1: Syntaktisch-kombinatorische und semantische Eigenschaften räumlicher Präpositionen (Fortsetzung)

id	Lemma	Modifikator	N ₁ =Toponym	N ₁ =Oikonym	N ₁ =Choronym	N ₁ =Hydronym	N ₁ =Oronym	N ₁ =Bauwerk	N ₁ =Hodonym	STATisch	DIREktional	SouRce	DeSTination	PASSage	ADV	Kommentar	Beispiel
208	vorwärts		+	+	+					+	-	-	-	-	+	selten, milit.	<i>vorwärts des Bodzapasses, Infanteriefeuer schlug alsbald aus unsichtbaren Stellungen vorwärts des Forts herüber</i>
209	wasserseits	: <i>approx</i> +: <i>distance</i> + <i>leicht</i> + <i>nur wenig</i> + <i>weit</i>	+							+	-	-	-	-	+		<i>wasserseits des Rheindeiches</i>
210	zu ... her		-							-	+	-	+	-	-	seltenst	<i>zur Krippe her kommt</i>
211	zu ... hin	: <i>bis</i> + <i>direkt</i> + <i>unmittelbar</i>	+							-	+	-	+	-	-		<i>zur Straße hin, zu den Rocky Mountains hin</i>
212	zunächst		+							+	-	-	-	-	+	selten, veraltet, unsicher	
213	zuseiten		-							+	-	-	-	-	-	selten	<i>zuseiten des Weges</i>
214	zu	: <i>bis</i> + <i>bis hin</i> + <i>direkt</i> + <i>hin</i> + <i>unmittelbar</i>	+			+	+			-	+	-	+	-	-		<i>Das Schiff fährt zur Insel Helgoland.</i>
215	zwischen	<i>in die Mitte</i>	+	+	+	+	+	+	+	-	+	-	+	-	-	direktional, mit Akk.	<i>zwischen die Häuser</i>
216	zwischen	: <i>mittig</i>	+	+	+	+	+	+	+	+	-	-	-	-	-	statisch, mit Dativ	<i>zwischen den Häusern, zwischen München und Augsburg</i>
217	zwischen ... hindurch		+					+		-	+	-	-	+	-		<i>zwischen den Bäumen hindurch</i>

Tabelle C.2: Formen und morphosyntaktische Eigenschaften räumlicher Präpositionen

id	Lemma	Form	postpon. Form	PREP	POSTP	CIRCP	PDET	CIRCPDET	Flexion	Akk.	Dat.	Gen.	von	Nom.	Plural	zu	oblig. mit Det.	ohne Det.	Kommentar
1	ab			+						-	+	-	-	-	-	-	-	-	
2	abseits			+						-	-	+	+	-	-	-	-	-	
3	abwärts			+						-	-	+	+	-	-	-	-	-	
4	abwärts				+					+	-	-	-	-	-	-	-	-	
4	abwärts			+						-	-	+	-	-	-	-	-	-	
5	am Fuß	<i>am FuÙe</i>		+						-	-	+	+	-	-	-	-	-	
5	am Fuß			+						-	-	+	+	-	-	-	-	-	
6	:amRand	::ent:misc:space:adv:amRand		+						-	-	+	+	-	-	-	-	-	
7	:amUfer	:am_Ufer		+						-	-	+	+	-	-	-	-	-	
8	an			+						+	-	-	-	-	-	-	-	-	
8	an	<i>ans</i>					+		aeN								-	-	
9	an	<i>am</i>					+		deM										
9	an	<i>am</i>					+		deN										
9	an			+						-	+	-	-	-	-	-	-	-	
10	an	<i>am</i>					+		deM										
10	an	<i>am</i>					+		deN										
10	an			+						-	+	-	-	-	-	-	-	-	
11	an der Grenze zu	<i>an der Grenze zum</i>					+		deM										
11	an der Grenze zu	<i>an der Grenze zum</i>					+		deN										
11	an der Grenze zu	<i>an der Grenze zur</i>					+		deF										
11	an der Grenze zu			+						-	+	-	-	-	-	-	-	-	
12	an ... entlang	<i>am</i>	<i>entlang</i>					+	deM										
12	an ... entlang	<i>am</i>	<i>entlang</i>					+	deN										
12	an ... entlang	<i>an</i>	<i>entlang</i>				+			-	+	-	-	-	-	-	-	-	
13	an ... heran	<i>an</i>	<i>heran</i>				+			+	-	-	-	-	-	-	-	-	
13	an ... heran	<i>ans</i>	<i>heran</i>					+	aeN										
14	:ansUfer	:ans_Ufer		+						-	-	+	+	-	-	-	-	-	
15	an ... vorbei	<i>am</i>	<i>vorbei</i>					+	deM										
15	an ... vorbei	<i>am</i>	<i>vorbei</i>					+	deN										

Tabelle C.2: Formen und morphosyntaktische Eigenschaften räumlicher Präpositionen (Fortsetzung)

id	Lemma	Form	postpon. Form	PREP	POSTP	CIRCP	PDET	CIRCPDET	Flexion	Akk.	Dat.	Gen.	von	Nom.	Plural	zu	oblig. mit Det.	ohne Det.	Kommentar
27	aus dem Raum			+						-	-	-	-	+	-	-	-	-	
28	aus der Nähe			+						-	-	+	+	-	-	-	-	-	
29	ausgehend			+						-	-	-	+	-	-	-	-	-	<i>ausgehend von Westfriesland, als Präp. meist "abstrakt" von Salzburg ausgehend</i>
29	ausgehend	<i>von</i>	<i>ausgehend</i>			+				-	+	-	-	-	-	-	-	-	
30	aus ... heraus	<i>aus</i>	<i>heraus</i>			+				-	+	-	-	-	-	-	-	-	
30	aus ... heraus	<i>aus'm</i>	<i>heraus</i>					+	deM										
30	aus ... heraus	<i>ausm</i>	<i>heraus</i>					+	deM										
30	aus ... heraus	<i>aus'm</i>	<i>heraus</i>					+	deN										
30	aus ... heraus	<i>ausm</i>	<i>heraus</i>					+	deN										
31	aus ... her	<i>aus</i>	<i>her</i>			+				-	+	-	-	-	-	-	-	-	
31	aus ... her	<i>aus'm</i>	<i>her</i>					+	deM										
31	aus ... her	<i>ausm</i>	<i>her</i>					+	deM										
31	aus ... her	<i>aus'm</i>	<i>her</i>					+	deN										
31	aus ... her	<i>ausm</i>	<i>her</i>					+	deN										
32	aus ... hinaus	<i>aus</i>	<i>hinaus</i>			+				-	+	-	-	-	-	-	-	-	
32	aus ... hinaus	<i>aus'm</i>	<i>hinaus</i>					+	deM										
32	aus ... hinaus	<i>ausm</i>	<i>hinaus</i>					+	deM										
32	aus ... hinaus	<i>aus'm</i>	<i>hinaus</i>					+	deN										
32	aus ... hinaus	<i>ausm</i>	<i>hinaus</i>					+	deN										
33	aus Richtung			+						-	-	-	-	+	-	-	-	-	+
33	aus Richtung			+						-	-	+	+	-	-	-	-	-	
34	außer			+						-	-	+	+	-	-	-	-	-	+
35	außerhalb			+						-	-	+	+	-	-	-	-	-	
36	auswärts			+						-	-	+	+	-	-	-	-	-	
37	bachabwärts			+						-	-	+	+	-	-	-	-	-	
38	bachaufwärts			+						-	-	+	+	-	-	-	-	-	
39	bei			+						-	+	-	-	-	-	-	-	-	
39	bei	<i>beim</i>					+		deM										

Tabelle C.2: Formen und morphosyntaktische Eigenschaften räumlicher Präpositionen (Fortsetzung)

id	Lemma	Form	postpon. Form	PREP	POSTP	CIRCP	PDET	CIRCPDET	Flexion	Akk.	Dat.	Gen.	von	Nom.	Plural	zu	oblig. mit Det.	ohne Det.	Kommentar
39	bei	<i>beim</i>					+		deN										
40	beiderseits			+						-	-	+	+	-	-	-	-	-	
41	beidseits			+						-	-	+	+	-	-	-	-	-	
42	bergseits			+						-	-	+	+	-	-	-	-	-	
43	bis			+						+	-	-	-	-	-	-	-	-	+
44	bis ... hin	<i>bis</i>	<i>hin</i>			+				+	-	-	-	-	-	-	-	-	+
45	diesseits			+						-	-	+	+	-	-	-	-	-	
46	dies- und jenseits			+						-	-	+	+	-	-	-	-	-	
47	durch			+						+	-	-	-	-	-	-	-	-	
47	durch	<i>durchs</i>						+	aeN										
48	durch ... hindurch	<i>durch</i>	<i>hindurch</i>			+				+	-	-	-	-	-	-	-	-	
48	durch ... hindurch	<i>durchs</i>	<i>hindurch</i>					+	aeN										
49	Ecke			+						-	-	-	-	+	-	-	-	-	+
50	ennet			+						-	+	+	-	-	-	-	-	-	
51	entgegen			+	+					-	+	-	-	-	-	-	-	-	
52	entlang				+					+	-	-	-	-	-	-	-	-	<i>ein Stück die Straße entlang</i>
52	entlang			+						-	+	+	-	-	-	-	-	-	<i>entlang der Küste, entlang dem Ufer, Reise entlang des Mains</i>
53	fernab			+						-	-	+	+	-	-	-	-	-	
54	fern			+						-	+	+	+	-	-	-	-	-	
55	flussabwärts			+						-	-	+	+	-	-	-	-	-	
55	flussabwärts	<i>flußabwärts</i>		+						-	-	+	+	-	-	-	-	-	
56	flussabwärts				+					+	-	-	-	-	-	-	-	-	
56	flussabwärts			+						-	-	+	+	-	-	-	-	-	
56	flussabwärts	<i>flußabwärts</i>			+					+	-	-	-	-	-	-	-	-	
56	flussabwärts	<i>flußabwärts</i>		+						-	-	+	+	-	-	-	-	-	
57	flussauf- und -abwärts	<i>flußauf- und -abwärts</i>		+						-	-	+	+	-	-	-	-	-	
57	flussauf- und -abwärts	<i>flußauf- und abwärts</i>		+						-	-	+	+	-	-	-	-	-	
57	flussauf- und -abwärts			+						-	-	+	+	-	-	-	-	-	

Tabelle C.2: Formen und morphosyntaktische Eigenschaften räumlicher Präpositionen (Fortsetzung)

id	Lemma	Form	postpon. Form	PREP	POSTP	CIRCP	PDET	CIRCPDET	Flexion	Akk.	Dat.	Gen.	von	Nom.	Plural	zu	oblig. mit Det.	ohne Det.	Kommentar
57	flussauf- und -abwärts	<i>flussauf- und abwärts</i>		+						-	-	+	+	-	-	-	-	-	
58	flussaufwärts			+						-	-	+	+	-	-	-	-	-	
58	flussaufwärts	<i>flußaufwärts</i>		+						-	-	+	+	-	-	-	-	-	
59	flussaufwärts				+					+	-	-	-	-	-	-	-	-	
59	flussaufwärts			+						-	-	+	+	-	-	-	-	-	
59	flussaufwärts	<i>flußaufwärts</i>			+					+	-	-	-	-	-	-	-	-	
59	flussaufwärts	<i>flußaufwärts</i>		+						-	-	+	+	-	-	-	-	-	
60	gegen			+						+	-	-	-	-	-	-	-	-	
61	gegen ... hin	<i>gegen</i>	<i>hin</i>			+				+	-	-	-	-	-	-	-	-	
62	gegenüber			+	+					-	+	+	-	-	-	-	-	-	
63	gegen ... zu	<i>gegen</i>	<i>zu</i>			+				+	-	-	-	-	-	-	-	-	
64	gen			+						+	-	-	-	-	-	-	-	-	+
65	gen ... hin	<i>gen</i>	<i>hin</i>			+				+	-	-	-	-	-	-	-	-	+
66	gen ... zu	<i>gen</i>	<i>zu</i>			+				+	-	-	-	-	-	-	-	-	+
67	herwärts			+						-	-	+	+	-	-	-	-	-	
68	hinab				+					+	-	-	-	-	-	-	-	-	
69	hinauf				+					+	-	-	-	-	-	-	-	-	
70	hinseits			+						-	-	+	+	-	-	-	-	-	
71	hinter	<i>hintern</i>					+		aeM										
71	hinter	<i>hinters</i>					+		aeM										
71	hinter	<i>hinters</i>					+		aeN										
71	hinter			+						+	-	-	-	-	-	-	-	-	
72	hinter	<i>hinterm</i>					+		deM										
72	hinter	<i>hinterm</i>					+		deN										
72	hinter			+						-	+	-	-	-	-	-	-	-	
73	hinter ... her	<i>hinter</i>	<i>her</i>			+				-	+	-	-	-	-	-	-	-	
73	hinter ... her	<i>hinterm</i>	<i>her</i>					+	deM										
73	hinter ... her	<i>hinterm</i>	<i>her</i>					+	deN										
74	hinter ... hervor	<i>hinter</i>	<i>hervor</i>			+				-	+	-	-	-	-	-	-	-	

Tabelle C.2: Formen und morphosyntaktische Eigenschaften räumlicher Präpositionen (Fortsetzung)

id	Lemma	Form	postpon. Form	PREP	POSTP	CIRCP	PDET	CIRCPDET	Flexion	Akk.	Dat.	Gen.	von	Nom.	Plural	zu	oblig. mit Det.	ohne Det.	Kommentar
94	:indenNorden	::ent:misc:space:adv:CindenNorden		+						-	-	+	+	-	-	-	-	-	
95	in den Raum			+						-	-	-	-	+	-	-	-	+	
96	in der Mitte			+						-	-	+	+	-	-	-	-	-	<i>in Münchens Mitte</i>
96	in der Mitte	<i>in</i>	<i>Mitte</i>			+				-	-	+	-	-	-	-	-	+	
97	:inderNaeh	::ent:misc:space:adv:inderNaeh		+						-	-	+	+	-	-	+	-	-	
98	in die Mitte			+						-	-	+	+	-	-	-	-	-	
98	in die Mitte	<i>in</i>	<i>Mitte</i>			+				-	-	+	-	-	-	-	-	+	
99	in ... herein	<i>in</i>	<i>herein</i>			+				+	-	-	-	-	-	-	-	-	
99	in ... herein	<i>ins</i>	<i>herein</i>					+	aeN										
100	in ... hinab	<i>in</i>	<i>hinab</i>			+				+	-	-	-	-	-	-	-	-	
100	in ... hinab	<i>ins</i>	<i>hinab</i>					+	aeN										
101	in ... hinein	<i>in</i>	<i>hinein</i>			+				+	-	-	-	-	-	-	-	-	
101	in ... hinein	<i>ins</i>	<i>hinein</i>					+	aeN										
102	in ... hinunter	<i>in</i>	<i>hinunter</i>			+				+	-	-	-	-	-	-	-	-	
102	in ... hinunter	<i>ins</i>	<i>hinunter</i>					+	aeN										
103	inmitten			+						-	-	+	+	-	-	-	-	-	
104	:inNachbarschaft	:in_Nachbarschaft		+						-	-	+	+	-	-	+	-	-	
105	innerhalb			+						-	-	+	+	-	-	-	-	-	
106	innert			+						-	-	+	-	-	-	-	-	-	
107	inner- und außerhalb			+						-	-	+	+	-	-	-	-	-	
108	in Richtung auf			+						+	-	-	-	-	-	-	-	-	
109	in Richtung auf ... zu	<i>in Richtung auf</i>	<i>zu</i>			+				+	-	-	-	-	-	-	-	-	
110	in Richtung			+						-	-	-	-	+	-	-	-	+	
110	in Richtung			+						-	-	+	+	-	-	-	-	-	
110	in Richtung	<i>Richtung</i>		+						-	-	-	-	+	-	-	-	+	
111	ins Herz	<i>in</i>	<i>Herz</i>			+				-	-	+	-	-	-	-	-	+	<i>in Münchens Herz</i>
111	ins Herz			+						-	-	+	+	-	-	-	-	-	
112	:inSichtweite	::ent:misc:space:adv:inSichtweite		+						-	-	+	+	-	-	+	-	-	
113	:inderUmgebung	:in_Umgebung		+						-	-	+	+	-	-	-	-	-	

Tabelle C.2: Formen und morphosyntaktische Eigenschaften räumlicher Präpositionen (Fortsetzung)

id	Lemma	Form	postpon. Form	PREP	POSTP	CIRCP	PDET	CIRCPDET	Flexion	Akk.	Dat.	Gen.	von	Nom.	Plural	zu	oblig. mit Det.	ohne Det.	Kommentar
135	nach ... zu	<i>nach</i>	<i>zu</i>			+				-	+	-	-	-	-	-	-	-	
136	Nähe			+						-	-	-	-	+	-	-	-	+	
137	nahe			+						-	+	-	-	-	-	-	-	-	
137	nahe	<i>nah</i>		+						-	+	-	-	-	-	-	-	-	
138	neben			+						+	-	-	-	-	-	-	-	-	
139	neben			+						-	+	-	-	-	-	-	-	-	
140	neben ... her	<i>neben</i>	<i>her</i>			+				-	+	-	-	-	-	-	-	-	
141	neben ... hin	<i>neben</i>	<i>hin</i>				+			-	+	-	-	-	-	-	-	-	
142	nid			+						-	+	-	-	-	-	-	-	-	
143	niederwärts				+					+	-	-	-	-	-	-	-	-	
143	niederwärts			+						-	-	+	+	-	-	-	-	-	
144	:noerdlich	::ent:misc:space:adv:noerdlich		+						-	-	+	+	-	-	-	-	-	
145	:nordwaerts	::ent:misc:space:adv:nordwaerts		+						-	-	+	+	-	-	-	-	-	
146	oberhalb			+						-	-	+	+	-	-	-	-	-	
147	ober			+						-	+	-	-	-	-	-	-	-	
148	oberwärts			+						-	-	+	+	-	-	-	-	-	
149	oberwasserseits			+						-	-	+	+	-	-	-	-	-	
150	ob			+						-	+	-	-	-	-	-	-	-	
151	rauf				+					+	-	-	-	-	-	-	-	-	
152	rechter Hand	<i>rechterhand</i>		+						-	-	+	+	-	-	-	-	-	
152	rechter Hand			+						-	-	+	+	-	-	-	-	-	
153	rechts			+						-	-	+	+	-	-	-	-	-	
154	rechts und links	<i>links und rechts</i>		+						-	-	+	+	-	-	-	-	-	
154	rechts und links			+						-	-	+	+	-	-	-	-	-	
155	ringsum			+						+	-	-	-	-	-	-	-	-	
156	rückwärts			+						-	-	+	+	-	-	-	-	-	
157	runter				+					+	-	-	-	-	-	-	-	-	
158	seeseits			+						-	-	+	+	-	-	-	-	-	
159	seitab			+						-	-	+	+	-	-	-	-	-	

Tabelle C.2: Formen und morphosyntaktische Eigenschaften räumlicher Präpositionen (Fortsetzung)

id	Lemma	Form	postpon. Form	PREP	POSTP	CIRCP	PDET	CIRCPDET	Flexion	Akk.	Dat.	Gen.	von	Nom.	Plural	zu	oblig. mit Det.	ohne Det.	Kommentar
175	um ... her	<i>um</i>	<i>her</i>			+				+	-	-	-	-	-	-	-	-	
175	um ... her	<i>ums</i>	<i>her</i>					+	aeN										
176	um			+						+	-	-	-	-	-	-	-	-	
176	um	<i>ums</i>				+			aeN										
177	unfern			+								+	+		-	-	-	-	
178	unter			+						+	-	-	-	-	-	-	-	-	
178	unter	<i>untern</i>					+		aeM										
178	unter	<i>unters</i>					+		aeN										
179	unter			+						-	+	-	-	-	-	-	-	-	
179	unter	<i>unterm</i>					+		deM										
179	unter	<i>unterm</i>					+		deN										
180	unterhalb			+						-	-	+	+	-	-	-	-	-	
181	unter ... her	<i>unter</i>	<i>her</i>			+				-	+	-	-	-	-	-	-	-	
181	unter ... her	<i>unterm</i>	<i>her</i>					+	deM										
181	unter ... her	<i>unterm</i>	<i>her</i>					+	deN										
182	unter ... hindurch	<i>unter</i>	<i>hindurch</i>			+				-	+	-	-	-	-	-	-	-	
182	unter ... hindurch	<i>unterm</i>	<i>hindurch</i>					+	deM										
182	unter ... hindurch	<i>unterm</i>	<i>hindurch</i>					+	deN										
183	unterkant			+						-	-	+	-	-	-	-	-	-	
184	unterwärts			+						-	-	+	+	-	-	-	-	-	
185	unterwasserseits			+						-	-	+	+	-	-	-	-	-	
186	unweit			+						-	-	+	+	-	-	-	-	-	
187	via			+						-	+	-	-	+	-	-	-	-	
188	vis-à-vis	<i>vis-a-vis</i>		+						-	+	-	+	-	-	-	-	-	
188	vis-à-vis			+						-	+	-	+	-	-	-	-	-	
189	vom Ufer			+						-	-	+	+	-	-	-	-	-	
190	von ... ab	<i>vom</i>	<i>ab</i>					+	deM										
190	von ... ab	<i>vom</i>	<i>ab</i>					+	deN										
190	von ... ab	<i>von</i>	<i>ab</i>			+				-	+	-	-	-	-	-	-	-	

Tabelle C.2: Formen und morphosyntaktische Eigenschaften räumlicher Präpositionen (Fortsetzung)

id	Lemma	Form	postpon. Form	PREP	POSTP	CIRCP	PDET	CIRCPDET	Flexion	Akk.	Dat.	Gen.	von	Nom.	Plural	zu	oblig. mit Det.	ohne Det.	Kommentar
204	vor			+						+	-	-	-	-	-	-	-	-	
204	vor	<i>vorn</i>					+		aeM										
204	vor	<i>vors</i>					+		aeN										
205	vor	<i>vorm</i>					+		deM										
205	vor	<i>vorm</i>					+		deN										
205	vor			+						-	+	-	-	-	-	-	-	-	
206	vor den Toren			+						-	-	+	+	-	-	-	-	-	
206	vor den Toren	<i>vor</i>	<i>Toren</i>			+				-	-	+	-	-	-	-	-	+	<i>vor Augsburgs Toren</i>
207	vor die Tore			+						-	-	+	+	-	-	-	-	-	
208	vorwärts			+						-	-	+	+	-	-	-	-	-	
209	wasserseits			+						-	-	+	+	-	-	-	-	-	
210	zu ... her	<i>zu</i>	<i>her</i>			+				-	+	-	-	-	-	-	-	-	
210	zu ... her	<i>zum</i>	<i>her</i>					+	deM										
210	zu ... her	<i>zum</i>	<i>her</i>					+	deN										
210	zu ... her	<i>zur</i>	<i>her</i>					+	deF										
211	zu ... hin	<i>zu</i>	<i>hin</i>			+				-	+	-	-	-	-	-	-	-	
211	zu ... hin	<i>zum</i>	<i>hin</i>					+	deM										
211	zu ... hin	<i>zum</i>	<i>hin</i>					+	deN										
211	zu ... hin	<i>zur</i>	<i>hin</i>					+	deF										
212	zunächst				+					-	+	-	-	-	-	-	-	-	
212	zunächst			+						-	-	+	-	-	-	-	-	-	
213	zuseiten			+						-	-	+	+	-	-	-	-	-	
214	zu			+						-	+	-	-	-	-	-	-	-	
214	zu	<i>zum</i>					+		deM										
214	zu	<i>zum</i>					+		deN										
214	zu	<i>zur</i>					+		deF										
215	zwischen			+						+	-	-	-	-	+	-	-	-	
216	zwischen			+						-	+	-	-	-	+	-	-	-	
217	zwischen ... hindurch	<i>zwischen</i>	<i>hindurch</i>			+				+	+	-	-	-	+	-	-	-	

- 1 ab : Die Anreise erfolgt **ab** Kamenz mit Bus 182 Richtung Bischofswerda Bahnhof bis zur Haltestelle "Abzweig nach Kindisch".
- 2 abseits : **Abseits** der großen Verkehrsadern geht es von der Küste ins bergige Iblei-Land hinauf.
- 3 abwärts : Die dritte, die Nieder- oder Buschmühle, findet man 2 km **abwärts** der Lomnitzer Kirche im Wiesental der Kleinen Röder zwischen dem Hainbusch im Süden, der Mittelheide und dem Bachbusch im Norden.
- 4 abwärts : Die Flussfahrt **abwärts** des großen Stroms Ayeyarwaddy beginnt abenteuerlich Blick Isar **abwärts** auf die Maximiliansbrücke. Nach kurzem Bustransfer (ca.1 Stunde) steigen wir auf die Boote um und paddeln **abwärts** des schönen Drawa Flusses.
- 11 an der Grenze zu : **An der Grenze zur** Gemarkung Modelwitz (Stadt Schkeuditz, Pegel: 97,1 m üNN) verlässt die Weiße Elster schließlich die Stadt Leipzig in westliche Richtung. Laa an der Thaya ist eine Kleinstadt im nördlichen Weinviertel in Niederösterreich, unmittelbar **an der Grenze zur** Tschechischen Republik gelegen.
- 12 an ... entlang : Die Strecke **am Rhein entlang** ist die kürzeste. **An** den Schienen **entlang** stehen kleine Häuser.
- 14 : **ansUfer** : Eine Etappe von Châtaigneraie in der Haut-Bocage der Vendée **bis ans Ufer** der Loire.
- 16 an ... vorüber : Hinter Hilpertsau dem rechten Ufer folgend, **an** der Ziegelhütte **vorüber** nach Reichenthal. Auf dem Selketalstieg weiter **am** ehemaligen Forsthaus Scheerenstieg **vorüber**.
- 18 auf beiden Seiten : Die kommunalen Krankenhäuser **auf beiden Seiten** der Weser wollen künftig enger zusammenarbeiten.
- 21 : **aufHöhe** : Bauarbeiten an der Stauffenbergallee **auf Höhe** der Prießnitzbrücke. Vor vielen Jahren saß André Mielke, der Autor dieser Kolumne, in einem Personenzug, der **auf Höhe** des Betriebsbahnhofs Rummelsburg entgleiste.
- 22 aufwärts : Ihre Provinz umfasst neben einem Teil des Beckens des Cismon, Zufluss des Brenta im Westen, außerdem das Becken des Brenta **aufwärts** des Quero. Der Ort diente der Grenzsicherung der Republik Florenz zu ihrem **aufwärts** des Arno gelegenen Nachbarn Arezzo.
- 23 aufwärts : Danach geht's den Neckar **aufwärts** bis Nürtingen, wo wir in das Aichtal abbiegen. Vom Parkplatz gehen wir die Straße **aufwärts** in den Ort.
- 24 : **ausallenEcken** : **Aus allen Ecken und Winkeln** des Landes strömen sie in ihre Hauptstadt, um die Volksfront-Regierung gegen Angriffe der OMON zu verteidigen.
- 26 aus : Lena Gräfin von Mahlenberg reist **aus** Berlin in ihre Heimat Ostpreußen.
- 28 aus der Nähe : Mehrere Lastwagen von Triquis und Einwohnern **aus der Nähe** der Stadt Oaxaca haben sich mit ihr getroffen.
- 29 ausgehend : In acht Tagesetappen (Gehzeit zwischen fünf und acht Stunden) können ambitionierte Wanderer den Tennengau von St. Martin **ausgehend** den Gosaukamm entlang über Postalm und Osterhorngruppe bis Puch im Salzbachtal durchstreifen. In der Wintersaison gibt es auch nur eine regelmäßige Schiffs-Verbindung zum Festland, diese stellt derzeit, **ausgehend** von Cuxhaven, die Funny Girl mehrmals wöchentlich sicher.
- 33 aus Richtung : Mit Hilfe dieser Anfahrtbeschreibung, wollen wir unseren Besuchern, die **aus Richtung** Deggendorf und somit evtl. von den Autobahnen A92 (aus Richtung München) oder A3 (aus Richtung Regensburg) kommen, die optimale Anreise zum Einkaufspark Reggen beschreiben.
- 36 auswärts : Der Internationale Flughafen von Belfast befindet sich etwas **auswärts** der Stadt und verfügt über ... Das Schloß Boros befindet sich in Westungarn, **auswärts** von Fonyód.
- 37 bachabwärts : 300 m **bachabwärts** der Dorschenmühle ist der Bach aufgestaut.

- 38 bachaufwärts : *Etwa einen Kilometer **bachaufwärts** der Schierenbeeke liegt das gleichnamige Naturschutzgebiet.*
- 41 beidseits : *Thraker und Kelten siedelten **beidseits** der Karpaten.*
- 42 bergseits : *Danach fällt er über einen Abhang in die Wälder unmittelbar **bergseits** von Serpiano auf 700 m ab. In der Tat werden demnächst rund 25 Bauplätze im "Under Feld/ Under Chrüzli" inmitten des Dorfes, **bergseits** des Gemeindehauses, erschlossen.*
- 49 Ecke : *Wir wohnen **Ecke** Herderstraße/Bachstraße.*
- 50 ennet : ***Ennet** der Glatt befindet sich das dritte Dorf, Ennet- oder heute verkürzt Endhöri geheissen.*
- 52 entlang : ***Entlang** der Fränkischen Saale verläuft die Saaletalbahn genannte Bahnlinie.*
- 53 fernab : *Im Hinterland Niersteins, **fernab** des Rheins, bietet das wellige Hügelland ein ganz anderes landschaftliches Bild: ...*
- 55 flussabwärts : *Der Unterspreewald beginnt **flussabwärts** der Stadt Lübben.*
- 56 flussabwärts : *Iwans Armee segelte die Wolga **flußabwärts** mit dem Ziel, Kasan einzunehmen. Auf dem Weg **flussabwärts** des Wanganui River erhebt sich die Landschaft steil nach allen Seiten und erschafft so die Illusion, tief im Herzen des Waldes eingebettet zu sein.*
- 57 flussauf- und -abwärts : *Über 100 Schiffe, meist Bulkcarrier und Containerfrachter, aber auch drei grosse Kreuzfahrer, lagen **flussauf- und abwärts** der Unglücksstelle fest, ... Mit dieser Urkunde übertrug Graf Konrad dem Kloster die Wasser- und Fischereirechte **flußauf- und abwärts** der Alzette von der Sankt-Ulrich-Brücke bis zum Mor-Felsen sowie die an diesem Flußabschnitt gelegenen Mühlen und Öfen. Daß die Schlösser und Burgen zu beiden Seiten des Hochrheins letztlich weniger besungen wurden als jene **flußauf- und -abwärts** der Loreley, mag daran liegen, ...*
- 58 flussaufwärts : *Rund 1000 Meter **flussaufwärts** des Stadtzentrums im Ortsteil Kues befindet sich der größte Wasserhafen der Mittelmosel, zugleich Heimathafen zahlreicher Passagierschiffe.*
- 59 flussaufwärts : *Townshend sollte mit seinen Einheiten den Tigris **flußaufwärts** vorstoßen und Bagdad einnehmen. Die Fahrt mit einem Drachenboot führt uns **flußaufwärts** des Parfüm-Flusses.*
- 61 gegen ... hin : *als sie aber **gegen** Mysien **hin** kamen, versuchten sie nach Bithynien zu reisen, und der Geist Jesu erlaubte es ihnen nicht.*
- 62 gegenüber : *Calais liegt Dover **gegenüber**.*
- 65 gen ... hin : *Das Vinschgau ist charakterisiert durch seine relativ breite Talsohle, welche **gen** Osten **hin** am sogenannten Sonnenberg endet.*
- 68 hinab : *Schiffahrt den Rhein **hinab** vorbei am Loreleyfelsen nach St. Goar*
- 71 hinter : *Mit dem gestohlenen Gut war der Dieb inzwischen schon bis **hinter** den Berg Isel gekommen.*
- 72 hinter : *Der Rhein bildet hier die Grenze Deutschlands zu Frankreich, direkt **hinter** dem Rhein liegt das Elsass.*
- 76 hinterwärts : *Will man eine Ahnung davon gewinnen, so muß man entweder vom höchsten Stadteile Borg'unto aus, oder **hinterwärts** der Kathedrale den Weg überm Mugnone zurückgehen und dann, abseits von der Trambahnstraße, den alten Fahrweg zwischen den großen unteren Villengartenmauern bis nach Florenz verfolgen.*
- 80 im Bereich : *Die Waltersdorfer Chaussee gehört **im Bereich** von Berlin Neukölln-Rudow zur Bundesstrasse B179. Im Verlauf heisst Sie auch Neuköllnerstrasse, Rudowerstrasse ...*
- 81 im Herzen : *Günstige Eigentumswohnungen **im Herzen** der Stadt Essen! Das Wohnheim befindet mitten **im Herzen** von Charlottenburg.*

- 84 :imNorden : *Der Prozess der Landbildung geht **im Osten** der Halbinsel auch heute weiter.*
- 87 :imUmkreis : *Gemeinde warnt: Keine Silvesterraketen **im Umkreis** von 200 Metern von Reetdachhäusern.*
- 92 :inalleTeile : *Aufgrund der günstigen Busverbindungen können bequem Ausflüge **in alle Teile** der Insel unternommen werden.*
- 94 :indenNorden : *Die nördlichen Rocky Mountains erstrecken sich vom nördlichen Idaho bis **in den Westen** von Montana.*
- 95 in den Raum : *Die Gletscher drangen damals bis **in den Raum** Hamburg vor.*
- 97 :inderNaehel : *Der Osterfjord liegt **in der Nähe** von Bergen und ist einer der schönsten Fjorde Westnorwegens. Unser "Landkrog Strübbel" liegt am Rande eines anmutigen Bauerndorfes von Weiden und Feldern umgeben, **in der Nähe** der Eiderabdämmung an der Nordsee.*
- 98 in die Mitte : *Nach einer Umrundung des Gr. Müggelsee führt unser Fahrradweg am nächsten Tag **in die Mitte** von Berlin.*
- 104 :inNachbarschaft : ***In unmittelbarer Nachbarschaft** von Naundorf befindet sich Zitzschewig.*
- 107 inner- und außerhalb : *Die Freizeitmannschaft nimmt an Berliner Freizeittligen/ -runden und an Turnieren **inner- und außerhalb** von Berlin teil.*
- 108 in Richtung auf : *Der Weg biegt nach 25 Min. nach links **in Richtung auf** die Straße.*
- 109 in Richtung auf ... zu : *Der »Sangweg« führt uns zum Wegetreffpunkt »Sieben Wegweiser« und von dort **in Richtung auf** die Eisenberge **zu**, die man jedoch nicht erreicht. Die „Polflucht“, das heißt, die durch die Erdrotation erzeugte Fliehkraft, sollte die Kontinentalmassen langsam **in Richtung auf** den Äquator **zu** bewegen.*
- 110 in Richtung : *Gemeinsam setzten wir die Fahrt **Richtung** Moskau fort. Als ein Fehler im Militärcomputer ein Geschwader von SAC-Bombern **in Richtung** Moskau beordert, ... Aufgrund eines technischen Fehlers wird ein Bombengeschwader der USA mit Atombomben an Bord **Richtung** Moskau losgeschickt, ...*
- 114 :inderUmgegend : ***In der ganzen Umgegend** von Glasgow nimmt die Arbeitslosigkeit täglich zu.*
- 115 in, um und um ... herum : *Dank des verwirrenden Strassensystems **in, um und um** Perth **herum** schafften wir es, dort ca 2 Stunden rumzukurven.*
- 121 längsseit : *Die USS Monitor zog jedoch nicht ab, sondern bezog wieder Position **längsseit** der USS Minnesota. Die **längsseit** der Weser gelegene Kirchenweide von Stephani wird in ihrer Bedeutung von der bremischen Wirtschaft entdeckt.*
- 122 längsseits : *Bis zur Weiterreise liegt ‚Merlin‘ **längsseits** der Mole im ‚Old Port‘ von Korfu-Stadt gut geschützt. Ein Tanker mit 1500 Tonnen Heizöl hatte sich **längsseits** der Steinheimer Brücke gelegt.*
- 124 längs : *Von Dresden nach Prag fährt man **längs** der Elbe. Neu herangeführte Divisionen und zahlreiche Panzerwagen warfen sich **längs** den von Bapaume - auf Cambrai und Peronne - führenden Straßen unseren vorwärtsdringenden Truppen entgegen.*
- 125 landseits : *Telijawila liegt etwas nördlich **landseits** von Weligama. Das Untersuchungsgebiet liegt im Süden von Rheinland-Pfalz und umfasst die gesamten Rheinauen zwischen Leimersheim und Sondernheim **landseits** des Rheinhauptdeiches bis zum Hochgestade. Hierzu soll **landseits** des Hachmannkais eine neue Kaimauer hergestellt werden.*
- 126 lang : *So sieht es die ganze Straße **lang** aus.*
- 128 linker Hand : *Rund 100 m weiter, **linkerhand** des Gräfinthaler Weges, schaut der Kopf eines offenbar eingesunkenen Zwischengrenzsteines heraus.*
- 129 links : *Die Stadt Weißenthurm liegt **links** des Rheins im "Neuwieder Becken".*

- 131 mittenmang : *Sie tummelten sich **mittenmang** der Bäume, genossen die Blütenpracht hautnah und erfreuten sich wieder an dem Bild, das in jedem ...*
- 133 nach ... hin : *Im Norden der Grafschaft Ruppin, hart an der mecklenburgischen Grenze, zieht sich von den Städtchen Gransee bis **nach** Rheinsberg **hin** eine mehrere Meilen lange Seenkette durch eine menschenarme, nur hier und da mit ein paar alten Dörfern, sonst aber ausschließlich mit Förstereien, Glas- und Teeröfen besetzte Waldung.*
- 135 nach ... zu : *Auf den gebahntesten Wegen steigt man eine Schlucht herauf, die das Wasser **nach** dem Inn **zu** sendet, ...*
- 140 neben ... her : *Dann führt die Trasse bewachsen **neben** der Straße **her**. Im Walkersbacher Tal wandern wir dann links **neben** der Straße **her** das Tal hinab, passieren bald den Edenhof und erreichen in Kürze ...*
- 141 neben ... hin : *... und nach wenig mehr als fünf Minuten rollte die Droschke **neben** dem Pferdebahngleise **hin** in die Dorotheenstraße hinein und auf die Schadowstraße zu, ...*
- 143 niederwärts : *Sie fuhren den Hügel **niederwärts** in die Stadt.*
- 145 :**nordwaerts** : *Von dort folgt die Grenze weiter **südwärts** der Isel bis zum Ort Huben. Die Steppe **südwärts** des Eric und Ontario, so wie des Michigan-Sees, zeigt in ihrer ganzen Ausdehnung eine, bisher in Amerika noch nicht gehörig gewürdigte ...*
- 146 oberhalb : *Das Haus liegt in einer ruhigen Seitenstraße der Dinnendahlstraße **oberhalb** des Siepentals.*
- 148 oberwärts : *Der Fluß Aurajocky hat schlechtes Wasser und ist von unten nur bis an die Brücke schiffbar; **oberwärts** der Brücke gehen nur kleine Kahne.*
- 149 oberwasserseits : *Die im Südwesten der Türkei gelegene Wasserkraftanlage Birecik ist am Fluß Euphrat, 60 km unterwasserseits des Atatürk Dammes und 8 km **oberwasserseits** der Stadt Birecik situiert.*
- 151 rauf : *..., das einen Kilometer die Straße **rauf** liegt*
- 152 rechter Hand : *Wir schlagen den gelb markierten Weg **rechter Hand** des Parkplatzes ein, der uns entlang herrlicher Fichtenwälder in wenigen Minuten zum ... **Rechter Hand** des Weges liegt eine ehemalige Sandentnahmestelle, die heute mit ihrem Gebüschbestand und ihren kleinen Tümpeln einen wichtigen Lebensraum für ...*
- 153 rechts : *Während sich heute der größte Teil der Stadt **rechts** der Donau befindet, nahm die Stadt Mühlheim ihren Anfang links der Donau, in der so genannten Altstadt.*
- 154 rechts und links : *Die Kieferngehölze auf den südexponierten Hängen **rechts und links** des Weges sind ein Beispiel für die Entwicklung ehemaliger Weideflächen, ...*
- 155 ringsum : ***Ringsum** den Trümmerberg waren Heckenrosen gepflanzt wurden.*
- 156 rückwärts : *Der Garten liegt überwiegend geschützt seitwärts und **rückwärts** des Hauses.*
- 157 runter : *Ein paar hundert Meter die Strasse **runter** liegt der Hudson Bay.*
- 158 seeseits : *Der Nationalpark Niedersächsisches Wattenmeer umfasst Watt- und Wasserflächen, Salzwiesen, Strände und Dünen vor den Deichen der niedersächsischen Nordseeküste bis zu einer Linie **seeseits** der Ostfriesischen Inseln beziehungsweise der Platen und Sandbänke an der Wesermündung. **Seeseits** der Geest konnten sich später, in Abhängigkeit vom Stand des Meeresspiegels, die großen fruchtbaren Marschgebiete entwickeln. Teilaufhebung des Bebauungsplanes Nr. 16a für ein Gebiet **seeseits** der Strandallee*

- 162 stromabwärts : Als Vorgängerbau der steinernen Werrabrücke läßt sich eine hölzerne Brücke in einer Pfahljochkonstruktion **stromabwärts** der heutigen Brücke anhand des Stadtgrundrisses nachweisen. Die Uferschäden erstreckten sich bis **stromabwärts** der Bahnbrücke der BD AG.
- 163 stromabwärts : Ich muß bei der Flößerei vom Schwarzwald den Rhein **stromabwärts** eigentlich vor allem an die Beschreibung die Wilhelm Hauff in den Märchen "Das kalte Herz" von der Schwurwaldflößerei gibt
- 164 stromaufwärts : Die Brücke wurde ab Oktober 1969 47 Meter **stromaufwärts** der schon vorhandenen, provisorischen Brücke gebaut. Da ihr Fischereigebiet nur wenige Kilometer **stromaufwärts** des Projektgebietes liege, seien auch hier negative Konsequenzen zu erwarten.
- 166 talseits : Die **talseits** der Ließemer Straße zwischen den Hausgrundstücken Nr. 58 und 74 gelegenen Freiflächen werden durch den Bebauungsplan als öffentliche Grünfläche mit der Zweckbestimmung Parkanlage festgesetzt. **Talseits** des Wanderweges rinnt das Bächlein im Wiesengrund.
- 167 talwärts : Der Hauptort liegt 34 km südlich von Bologna, am rechten Flussufer vom Idice und **talwärts** des Ortsteils Monterezenzio, von dem der Gemeindepnamen stammt. Ein Stück weiter **talwärts** vom Lodge liegt das Camp.
- 172 über ... hin : Bei guter Fernsicht ist der Glärnisch bis weit **über** den Zürichsee **hin** zu sehen.
- 174 über ... weg : Dazu der Blick **über** den Fluss **weg** nach Süden ins Kaisergebirge Als er fast bei mir ist, macht der Leopard einen gewaltigen Satz **über** den Fluß **weg** an das Ufer, wo ich stand, landete im hohen Grase und war verschwunden.
- 175 um ... her : ... daß sie sammeln sollten von den Äckern **um** die Städte **her**, ... Dazu nun scheinen die Bergnamen **um** Calmbach **her** mitzugehören. Wie auf der Alb der Roßberg und Farrenberg, so stehen hier der Hengstberg und Kälbling neben ...
- 177 unfern : Die Stadt Chomutov liegt am Fuße des Erzgebirges, **unfern** der Grenze.
- 181 unter ... her : **Unter** der Brücke **her**, weitere 500m geradeaus, ... Auf dem Deutzer Ring immer auf der Vorfahrtstraße bleiben, **unter** der Bahnunterführung **her** und links einordnen.
- 182 unter ... hindurch : Nördlich des Stadtteils Cheriton, im Westen von Folkestone, liegt das nördliche Portal des Eurotunnels, der **unter** dem Ärmelkanal **hindurch** verläuft. In Angelroda **unter** dem Viadukt **hindurch** auf der Hauptstraße, vorbei am Alten Zollhaus, dann links in die "Große Gasse" und nach ca. ...
- 183 unterkant : In zwei bis **unterkant** des Fundamentes und im Leitungsraben in der Strasse erstellten Schlitzsondierungen wurde zum Teil brauner, torfig-sandiger Untergrund festgestellt. Weiter müsse bei Versickerungsanlagen des Typs b ab **unterkant** der Versickerungsanlage bis zum höchsten Grundwasserspiegel eine natürlich gewachsene ungestörte vertikale Filterstrecke von mindestens einem Meter vorhanden sein.
- 185 unterwasserseits : Zusätzlich wird **unterwasserseits** der bestehenden Brücke ein Steg für Fussgängerinnen und Fussgänger sowie für den Zweiradverkehr Fahrrihtung Unterlunkhofen – Rottenschwil erstellt.
- 186 unweit : Klobbicke liegt mitten im Barnim, nordöstlich von Berlin, **unweit** von Bad Freienwalde.
- 187 via : Li startet ihren Flug in Dresden und fliegt **via** Frankfurt/Main nach Costa Rica.
- 188 vis-à-vis : Das Atlon liegt in Berlin am Pariser Platz, **vis a vis** des Brandenburger Tores.
- 189 vom Ufer : Das älteste Viertel der Stadt, Ribeira, zieht sich **vom Ufer** des Douro steil hügelan.
- 191 von ... an : **Von da an** steigt das Gelände gemächlich und breit zu den bewaldeten Bergkuppen hinauf.

- 193 von beiden Seiten : *Und folgerichtig hat er gemeinsam mit dem WWF nichtstaatliche Akteure **von beiden Seiten** der Oder hierher nach Frankfurt eingeladen. In Reden und Interviews führender Politiker **von beiden Seiten** des Atlantiks kommt das Gemeinsame ebenso zum Ausdruck wie das Trennende.*
- 194 von beiderseits : *Am 20. September wurde der Bauplatz in Marckolsheim von UmweltschützerInnen **von beiderseits** des Rheins besetzt.*
- 197 von hinter : *Haas lässt alle Salzburger Glocken schlagen. Es läuten: der Dom, St. Peter, die Ursulinenkirche, die Kajetanerkirche, **von hinter** dem Berg die Maxglaner Kirche und direkt vom Berg die Glocken von Kloster Nonntal und wieder von unten die Blasiuskirche, dann direkt auf dem Mönchsberg die Palotiner Glocken.*
- 202 von vor : *Wir müssen quasi die Ladung **von vor** dem Haus nach hinten bekommen. Versetzung der Bushaltestelle in der Diekircher Straße **von vor** dem Haus Nr. 55 bis Ende der Bauarbeiten vor das Haus Nr. 47, ...*
- 207 vor die Tore : *Auf seinem Feldzug in Ägypten ist Caesar mit seinem Truppen bis **vor die Tore** Alexandrias vorgerückt.*
- 208 vorwärts : *Zunächst Vernichtung des Feindes **vorwärts** des Don, dann Gewinnung der Ölgebiete im kaukasischem Raum und Übergang über den Kaukasus selbst. Antemurale heißt bis zu den Kreuzzügen jedes **vorwärts** der Ringmauer gelegene Werk Durch Erstürmung des Clabucetu Baiului wurden die bisherigen Erfolge **vorwärts** des Predealpasses vervollständigt.*
- 209 wasserseits : *Die Fläche liegt **wasserseits** der vorhandenen Hochwasserschutzanlage (Stadtdeich Wesel) im Überschwemmungsgebiet des Rheins.*
- 211 zu ... hin : *An der Marnefront hatte der Gegner die Räumung des Südufers durch die Deutschen nirgends bemerkt, er begann in den Morgenstunden die verlassenen Stellungen und das Gelände bis **zur** Marne **hin** heftig zu beschießen.*
- 212 zunächst : *Untersuchungen über die Entwicklung der materiellen Kultur des platten Landes auf Grund der Quellen **zunächst** des Mosellandes*
- 213 zuseiten : *In allen Kammern finden sich **zuseiten** der Fenster steinerne Sitzbänke. Hufeisenförmig **zuseiten** der Turnhalle gruppierte Zwillingengebäude einer Doppelschule von 1906 nach Entwurf von G. Schaumann und R. Restle in Architekturformen aus Neobarock und Jugendstil*
- 214 zu : *In der Nähe von Alès, im Süden, erstreckt sich die Landschaft hin **zum** Mittelmeer, wo auf Terrassen Weinreben, Olivenbäume und Brombeersträucher auf die ... Die Fähre fährt **zur** Insel Langeoog.*
- 215 zwischen : *Um möglichst viele Dörfer gleichmässig zu bedienen, wurden vielfach die Bahnhöfe **zwischen** die Dörfer gelegt, so auch in Freidorf, damals noch ein Weiler, ... Das Gebiet el Somontano ist eingebettet **zwischen** die Flüsse Alcantare und Guatizalema und kann auf eine lange Tradition im Weinanbau zurück blicken.*
- 216 zwischen : *Die 2,5 km lange und bis 1,2 km breite Sandbank liegt **zwischen** den Inseln Borkum und Juist, in der Nähe der Vogelschutzinsel Memmert.*
- 217 zwischen ... hindurch : *Sie müssen **zwischen** den Bäumen **hindurch** abbiegen.*

C.2 Verben

Die Verbtabelle auf den folgenden Seiten fassen jeweils mehrere Typen (Satzschemata) in einer Tabelle zusammen, um nicht für Schemata, dem nur wenige Verben (teilweise nur ein einziges Verb) folgen, eine eigene Tabelle anlegen zu müssen. Die Satz-schemata sind als binäres Merkmal (+/-), aber auch als Zwischenüberschriften in der Tabelle angegeben.

Die drei Tabellen beschreiben Bewegungs- (Tab. C.3, vgl. Kap. 4.5.1), statische Relationen Mensch-Ort (Tab. C.4, vgl. Kap. 4.5.4) und Relationen zwischen Orten (Tab. C.5, vgl. Kap. 4.5.5). Die Tabellen teilen sich eine Reihe von Merkmalen, die für alle Verben gleich beschrieben werden:

- Bildung der Perfektformen mit *haben* oder *sein*
- reflexives Verb, falls das Reflexivpronomen *sich* angegeben ist.
- Präfigierte Verben werden als Folge von Präfix und Stamm angegeben. Trennbare und nicht trennbare Präfixe werden auf zwei verschiedene Spalten verteilt, die jeweils links- bzw. rechtsbündig ausgerichtet sind.
- Restriktionen bzgl. des semantischen Typs von N_0 sind angegeben, z.B.

$N_0=Hum$ Menschen

$N_0=HumColl$ Gruppen von Menschen

$N_0=anim$ belebte Wesen, d.h. alle Tiere

$N_0=transport$ Verkehrs- und Transportmittel (nicht weiter spezifiziert)

$N_0=ship$ Schiffe

$N_0=Toponym$ Orte (Toponyme)

$N_0=Choronym$ Regionen (Choronyme)

$N_0=Hydronym$ Gewässer (Hydronyme)

$N_0=Fluss$ Fließgewässer

$N_0=Hodonym$ Verkehrswege (Hodonyme)

Falls die weiteren Aktanten (N_1 usw.) variieren können, werden Selektionsrestriktionen auch für diese in analoger Weise angegeben.

- Realisierungsmöglichkeiten der Präposition, sofern das Verb eine Präposition stark regiert (alle Schemata, die *Prep* enthalten). Als stark regierte Präpositionen wurden beobachtet: *auf_{akk}*, *an_{dat}*, *aus*, *in_{akk}*, *in_{dat}*, *über_{akk}*, sowie bei den Verben zum Ausdruck von Relationen zwischen Orten *zu* und *mit*.
- ist die Präposition nur schwach regiert, wird der semantische Typ der Präposition (*Loc_{stat}*, *Loc_{dst}*, *Loc_{pass}*, *Loc_{src}*) im Satzschema notiert (z.B. $N_0 V N_1 Loc_{dst} N_1 dest$).
- das Merkmal *Pfad*= \emptyset kennzeichnet Verben, die neben dem fokussierten keine weiteren Pfadbestandteile realisieren können, vgl. S. 159.
- Für viele Verben wurden auch abgeleitete Nomina *Vn* angegeben, aber nur dann, wenn sie nicht mit dem Infinitiv identisch sind.

Tabelle C.3: Bewegungsverben

id	N ₀ =Hum	N ₀ =anim	N ₀ =transport	N ₀ =ship	reflexiv	Prfx, trennbar	Prfx, fest	Lemma	Perfekt <i>haben</i>	Perfekt <i>sein</i>	N ₀ V N ₁ _{akk}	N ₀ V N ₁ _{dat}	N ₀ V Prep N ₁	N ₀ V Loc N ₁	N ₀ V Loc _{stat} N ₁	Prep= <i>auf</i> _{akk}	Prep= <i>an</i> _{dat}	Prep= <i>aus</i>	Prep= <i>in</i> _{akk}	Prep= <i>in</i> _{dat}	Prep= <i>über</i> _{akk}	N ₁ =source	N ₁ =passage	N ₁ =destinat.	Pfad=∅	Kommentar	V _{n1}	V _{n2}	V _{n3}		
N₀ VN₁src_{akk}																															
1	+	+	+	+			ver lassen	+	-	+	-	-	-	-								+	-	-		<i>Peter verlässt das Zimmer</i>					
2	+	+	+	+		zurück	lassen	+	-	+	-	-	-	-								+	-	-							
N₀ VN₁dest_{akk}																															
1	+	-	+	-		an	fahren	+	-	+	-	-	-	-								-	-	+						Anfahrt	
2	-	-	+	-		an	fliegen	+	-	+	-	-	-	-								-	-	+						Anflug	
3	-	-	-	+		an	laufen	+	-	+	-	-	-	-								-	-	+		<i>Die Yacht läuft den Inselhafen an</i>					
4	+	-	+	+		an	steuern	+	-	+	-	-	-	-								-	-	+							
5	+	-	-	-			besteigen	+	-	+	-	-	-	-								-	-	+						Besteigung	
6	+	-	-	-			betreten	+	-	+	-	-	-	-								-	-	+							
7	+	+	-	-			erklettern	+	-	+	-	-	-	-								-	-	+							
8	+	+	-	-			erklimmen	+	-	+	-	-	-	-								-	-	+							
9	+	+	+	+			erreichen	+	-	+	-	-	-	-								-	-	+							
10	+	+	-	-			ersteigen	+	-	+	-	-	-	-								-	-	+							
11	+	-	-	-			erstürmen	+	-	+	-	-	-	-								-	-	+							
12	+	-	-	-			stürmen	+	-	+	-	-	-	-								-	-	+		<i>vgl. N₀ VLoc_{dst} N₁ dest</i>					
N₀ VN₁dest_{dat}																															
1	+	+	+	+	sich		nähern	+	-	-	+	-	-	-								-	-	+							
N₀ VN₁pass_{akk}																															
1	+	-	+	+		ab	fahren	+	+	+	-	-	-	-								-	+	-						Abfahren	
2	+	-	-	-		ab	laufen	+	+	+	-	-	-	-								-	+	-							
3	+	-	-	-		ab	schreiten	+	+	+	-	-	-	-								-	+	-							

Tabelle C.3: Bewegungsverben (Fortsetzung)

id	N ₀ =Hum	N ₀ =anim	N ₀ =transport	N ₀ =ship	reflexiv	Prefix, trennbar	Prefix, fest	Lemma	Perfekt <i>haben</i>	Perfekt <i>sein</i>	N ₀ V N ₁ akk	N ₀ V N ₁ dat	N ₀ V Prep N ₁	N ₀ V Loc N ₁	N ₀ V Loc _{stat} N ₁	Prep= <i>auf</i> _{akk}	Prep= <i>an</i> _{dat}	Prep= <i>aus</i>	Prep= <i>in</i> _{akk}	Prep= <i>in</i> _{dat}	Prep= <i>über</i> _{akk}	N ₁ =source	N ₁ =passage	N ₁ =destinat.	Pfad=∅	Kommentar	V _n	V _{n₂}	V _{n₃}	
4	+	-	+	+				be fahren	+	-	+	-	-	-	-							-	+	-						
5	+	-	+	-				be fliegen	+	-	+	-	-	-	-								-	+	-					
6	+	-	-	-				be gehen	+	-	+	-	-	-	-								-	+	-					
7	+	-	-	-				be reisen	+	-	+	-	-	-	-								-	+	-					
8	+	-	+	-				durch brausen	+	-	+	-	-	-	-								-	+	-					
9	+	-	+	-				durch eilen	+	-	+	-	-	-	-								-	+	-					
10	+	-	+	+				durch fahren	+	-	+	-	-	-	-								-	+	-					
11	+	-	+	-				durch fliegen	+	-	+	-	-	-	-								-	+	-					
12	+	-	-	+				durch gleiten	+	-	+	-	-	-	-								-	+	-					
13	+	+	+	-				durch jagen	+	-	+	-	-	-	-								-	+	-					
14	+	-	-	-				durch laufen	+	-	+	-	-	-	-								-	+	-					
15	+	+	+	+				durch queren	+	-	+	-	-	-	-								-	+	-		Durchquerung			
16	+	-	+	-				durch rasen	+	-	+	-	-	-	-								-	+	-					
17	+	-	+	-				durch reisen	+	-	+	-	-	-	-								-	+	-					
18	+	-	-	-				durch reiten	+	-	+	-	-	-	-								-	+	-					
19	+	-	-	-				durch schreiten	+	-	+	-	-	-	-								-	+	-					
20	+	+	-	-				durch schwimmen	+	-	+	-	-	-	-								-	+	-					
21	+	-	-	+				durch segeln	+	-	+	-	-	-	-								-	+	-		Durchsegelung			
22	+	-	-	-				durch steigen	+	-	+	-	-	-	-								-	+	-					
23	+	+	-	-				durch streifen	+	-	+	-	-	-	-								-	+	-					
24	+	+	-	-				durch wandern	+	-	+	-	-	-	-								-	+	-		Durchwanderung			
25	+	+	-	-				durch waten	+	-	+	-	-	-	-								-	+	-					
26	+	-	+	-			entlang	brausen	-	+	+	-	-	-	-								-	+	-					
27	+	-	-	-			entlang	eilen	-	+	+	-	-	-	-								-	+	-					
28	+	-	+	+			entlang	fahren	-	+	+	-	-	-	-								-	+	-					
29	+	-	-	-			entlang	flanieren	-	+	+	-	-	-	-								-	+	-					
30	+	+	+	-			entlang	fliegen	-	+	+	-	-	-	-								-	+	-					
31	+	-	-	-			entlang	gehen	-	+	+	-	-	-	-								-	+	-					
32	+	-	-	-			entlang	humpeln	-	+	+	-	-	-	-								-	+	-					

Tabelle C.3: Bewegungsverben (Fortsetzung)

id	N ₀ =Hum	N ₀ =anim	N ₀ =transport	N ₀ =ship	reflexiv	Prfx, trennbar	Prfx, fest	Lemma	Perfekt <i>haben</i>	Perfekt <i>sein</i>	N ₀ V N ₁ akk	N ₀ V N ₁ dat	N ₀ V Prep N ₁	N ₀ V Loc N ₁	N ₀ V Loc _{stat} N ₁	Prep= <i>auf</i> _{akk}	Prep= <i>an</i> _{dat}	Prep= <i>aus</i>	Prep= <i>in</i> _{akk}	Prep= <i>in</i> _{dat}	Prep= <i>über</i> _{akk}	N ₁ =source	N ₁ =passage	N ₁ =destinat.	Pfad=∅	Kommentar	V _n	V _{n2}	V _{n3}		
33	+	-	-	-			entlang	kommen	-	+	+	-	-	-	-																
34	+	+	-	-			entlang	kriechen	-	+	+	-	-	-	-																
35	+	-	+	-			entlang	kurven	-	+	+	-	-	-	-																
36	+	+	-	-			entlang	laufen	-	+	+	-	-	-	-																
37	+	-	-	-			entlang	marschieren	-	+	+	-	-	-	-																
38	+	-	+	-			entlang	rasen	-	+	+	-	-	-	-																
39	+	-	-	-			entlang	rudern	-	+	+	-	-	-	-																
40	+	-	+	-			entlang	sausen	-	+	+	-	-	-	-																
41	+	-	-	-	sich		entlang	schlängeln	+	-	+	-	-	-	-																
42	+	+	-	-			entlang	schleichen	-	+	+	-	-	-	-																
43	+	-	-	-			entlang	schreiten	-	+	+	-	-	-	-																
44	+	+	-	-			entlang	schwimmen	-	+	+	-	-	-	-																
45	+	-	-	-			entlang	segeln	-	+	+	-	-	-	-																
46	+	-	-	-			entlang	spazieren	-	+	+	-	-	-	-																
47	+	+	-	-			entlang	streichen	-	+	+	-	-	-	-																
48	+	-	-	-			entlang	streifen	-	+	+	-	-	-	-																
49	+	-	-	-			entlang	stürmen	-	+	+	-	-	-	-																
50	+	-	-	-			entlang	tappen	-	+	+	-	-	-	-																
51	+	-	-	-			entlang	traben	-	+	+	-	-	-	-																
52	+	-	-	-			entlang	ziehen	-	+	+	-	-	-	-																
53	+	+	+	+				passieren	+	-	+	-	-	-	-																Passage
54	+	+	+	+				queren	+	-	+	-	-	-	-																Querung
55	+	+	+	-			über	fliegen	+	-	+	-	-	-	-																Überflug
56	+	+	+	+			über	queren	+	-	+	-	-	-	-																Überquerung
57	+	-	-	-			über	schreiten	+	-	+	-	-	-	-																Überschreitung
58	+	+	-	-			über	springen	+	-	+	-	-	-	-																
59	+	-	-	-			über	treten	+	-	+	-	-	-	-																Übertritt
60	+	-	+	-			über	rollen	+	-	+	-	-	-	-																
61	+	-	+	+			um	fahren	+	-	+	-	-	-	-																Umfahrung

Tabelle C.3: Bewegungsverben (Fortsetzung)

id	N ₀ =Hum	N ₀ =anim	N ₀ =transport	N ₀ =ship	reflexiv	Prfx, trennbar	Prfx, fest	Lemma	Perfekt <i>haben</i>	Perfekt <i>sein</i>	N ₀ V N ₁ akk	N ₀ V N ₁ dat	N ₀ V Prep N ₁	N ₀ V Loc N ₁	N ₀ V Locstat N ₁	Prep= <i>auf</i> _{akk}	Prep= <i>an</i> _{dat}	Prep= <i>aus</i>	Prep= <i>in</i> _{akk}	Prep= <i>in</i> _{dat}	Prep= <i>über</i> _{akk}	N ₁ =source	N ₁ =passage	N ₁ =destinat.	Pfad=∅	Kommentar	Vn	Vn ₂	Vn ₃		
62	+	-	-	-				um gehen	+	-	+	-	-	-	-								-	+	-					Umgehung	
63	+	-	-	-				um rudern	+	-	+	-	-	-	-								-	+	-						
64	+	-	-	+				um schiffen	+	-	+	-	-	-	-								-	+	-					Umschiffung	
65	+	-	-	+				um segeln	+	-	+	-	-	-	-								-	+	-					Umsegelung	
66	+	-	-	-				um wandern	+	-	+	-	-	-	-								-	+	-						
N₀ VN₁pass_{dat}																															
1	+	+	+	+				folgen	-	+	-	+	-	-	-								-	+	-						
N₀ VPrepN₁src																															
1	-	-	+	-		aus		fahren	-	+	-	-	+	-	-	-	-	+	-	-	-	+	-	-							<i>Ein IC fährt aus dem Augsburger Hbf aus</i>
N₀ VPrepN₁dest																															
1	+	+	+	-		ein		dringen	-	+	-	-	+	-	-	-	-	-	+	+	-	-	-	-	+						
2	+	-	+	-		ein		fahren	-	+	-	-	+	-	-	-	-	-	+	+	-	-	-	-	+						Einfahrt
3	+	+	-	-		ein		fallen	-	+	-	-	+	-	-	-	-	-	+	+	-	-	-	-	+						Einfall
4	+	-	-	-		ein		kehren	-	+	-	-	+	-	-	-	-	-	+	+	-	-	-	-	+						Einkehr
5	+	-	-	+		ein		laufen	-	+	-	-	+	-	-	-	-	-	+	+	-	-	-	-	+						
6	+	-	-	-		ein		marschieren	-	+	-	-	+	-	-	-	-	-	+	+	-	-	-	-	+						Einmarsch
7	+	-	-	-		ein		reisen	-	+	-	-	+	-	-	-	-	-	+	+	-	-	-	-	+						Einreise
8	+	-	-	-		ein		reiten	-	+	-	-	+	-	-	-	-	-	+	+	-	-	-	-	+						
9	+	-	-	-		ein		rücken	-	+	-	-	+	-	-	-	-	-	+	+	-	-	-	-	+						
10	+	-	-	-		ein		steigen	-	+	-	-	+	-	-	-	-	-	+	-	-	-	-	-	+						Einstieg
11	+	-	-	-		ein		treten	-	+	-	-	+	-	-	-	-	-	+	-	-	-	-	-	+						
12	+	-	-	-		ein		wandern	-	+	-	-	+	-	-	-	-	-	+	+	-	-	-	-	+						Einwanderung
13	+	-	-	-		ein		ziehen	-	+	-	-	+	-	-	-	-	-	+	+	-	-	-	-	+						Einzug
14	+	-	-	-	sich	heran		pirschen	+	-	-	-	+	-	-	-	-	-	+	-	-	-	-	-	+						
15	+	+	+	+	sich	zu		bewegen	+	-	-	-	+	-	-	+	-	-	+	-	-	-	-	-	+						

Tabelle C.3: Bewegungsverben (Fortsetzung)

id	N ₀ =Hum	N ₀ =anim	N ₀ =transport	N ₀ =ship	reflexiv	Prfx, trennbar	Prfx, fest	Lemma	Perfekt <i>haben</i>	Perfekt <i>sein</i>	N ₀ V N ₁ _{akk}	N ₀ V N ₁ _{dat}	N ₀ V Prep N ₁	N ₀ V Loc N ₁	N ₀ V Loc _{stat} N ₁	Prep= <i>auf</i> _{akk}	Prep= <i>an</i> _{dat}	Prep= <i>aus</i>	Prep= <i>in</i> _{akk}	Prep= <i>in</i> _{dat}	Prep= <i>über</i> _{akk}	N ₁ =source	N ₁ =passage	N ₁ =destinat.	Pfad=∅	Kommentar	V _n	V _{n2}	V _{n3}	
16	+	-	+	+		zu	fahren	-	+	-	-	+	-	-	+	-	-	-	-	-	-	-	-	+						
17	+	+	+	-		zu	fliegen	-	+	-	-	-	+	-	-	+	-	-	-	-	-	-	-	-	+					
18	+	-	+	+		zu	halten	-	+	-	-	-	+	-	-	+	-	-	-	-	-	-	-	-	+					
19	+	+	+	-		zu	jagen	-	+	-	-	-	+	-	-	+	-	-	-	-	-	-	-	-	+					
20	+	+	-	-		zu	kriechen	-	+	-	-	-	+	-	-	+	-	-	-	-	-	-	-	-	+					
21	+	+	-	-		zu	laufen	-	+	-	-	-	+	-	-	+	-	-	-	-	-	-	-	-	+					
22	+	-	-	-		zu	marschieren	-	+	-	-	-	+	-	-	+	-	-	-	-	-	-	-	-	+					
23	+	-	+	-		zu	schießen	-	+	-	-	-	+	-	-	+	-	-	-	-	-	-	-	-	+					
24	+	+	-	-		zu	schwimmen	-	+	-	-	-	+	-	-	+	-	-	-	-	-	-	-	-	+					
25	+	+	+	+		zu	steuern	-	+	-	-	-	+	-	-	+	-	-	-	-	-	-	-	-	+					
26	+	+	-	-		zu	stürzen	-	+	-	-	-	+	-	-	+	-	-	-	-	-	-	-	-	+					
N₀ V Prep N₁ pass																														
1	+	-	-	+		über	setzen	-	+	-	-	+	-	-	-	-	-	-	-	-	+	-	+	-						
2	+	-	-	+			setzen	-	+	-	-	+	-	-	-	-	-	-	-	-	+	-	+	-						
3	+	-	+	+		vorbei	fahren	-	+	-	-	+	-	-	-	-	+	-	-	-	-	-	-	+	-					
4	+	+	+	-		vorbei	fliegen	-	+	-	-	+	-	-	-	-	+	-	-	-	-	-	-	+	-					
5	+	+	-	-		vorbei	galoppieren	-	+	-	-	+	-	-	-	-	+	-	-	-	-	-	-	+	-					
6	+	-	-	-		vorbei	gehen	-	+	-	-	+	-	-	-	-	+	-	-	-	-	-	-	+	-					
7	+	-	-	-		vorbei	hasten	-	+	-	-	+	-	-	-	-	+	-	-	-	-	-	-	+	-					
8	+	-	-	-		vorbei	hetzen	-	+	-	-	+	-	-	-	-	+	-	-	-	-	-	-	+	-					
9	+	-	+	-		vorbei	kommen	-	+	-	-	+	-	-	-	-	+	-	-	-	-	-	-	+	-					
10	+	+	-	-		vorbei	laufen	-	+	-	-	+	-	-	-	-	+	-	-	-	-	-	-	+	-					
11	+	-	-	-		vorbei	rasen	-	+	-	-	+	-	-	-	-	+	-	-	-	-	-	-	+	-					
12	+	-	-	-		vorbei	reiten	-	+	-	-	+	-	-	-	-	+	-	-	-	-	-	-	+	-					
13	+	+	-	-		vorbei	rennen	-	+	-	-	+	-	-	-	-	+	-	-	-	-	-	-	+	-					
14	+	-	+	-		vorbei	rollen	-	+	-	-	+	-	-	-	-	+	-	-	-	-	-	-	+	-					
15	+	+	+	-		vorbei	sausen	-	+	-	-	+	-	-	-	-	+	-	-	-	-	-	-	+	-					
16	-	-	-	+		vorbei	segeln	-	+	-	-	+	-	-	-	-	+	-	-	-	-	-	-	+	-					

Tabelle C.3: Bewegungsverben (Fortsetzung)

id	N ₀ =Hum	N ₀ =anim	N ₀ =transport	N ₀ =ship	reflexiv	Prfx, trennbar	Prfx, fest	Lemma	Perfekt <i>haben</i>	Perfekt <i>sein</i>	N ₀ V N ₁ akk	N ₀ V N ₁ dat	N ₀ V Prep N ₁	N ₀ V Loc N ₁	N ₀ V Loc _{stat} N ₁	Prep= <i>auf</i> _{akk}	Prep= <i>an</i> _{dat}	Prep= <i>aus</i>	Prep= <i>in</i> _{akk}	Prep= <i>in</i> _{dat}	Prep= <i>über</i> _{akk}	N ₁ =source	N ₁ =passage	N ₁ =destinat.	Pfad= \emptyset	Kommentar	Vn	Vn ₂	Vn ₃		
17	+	-	-	-		vorbei	wandern	-	+	-	-	+	-	-	-	-	+	-	-	-	-	-	+	-							
18	+	-	-	+		vorbei	ziehen	-	+	-	-	+	-	-	-	-	+	-	-	-	-	-	+	-							
19	-	-	+	-		vorüber	brausen	-	+	-	-	+	-	-	-	-	+	-	-	-	-	-	+	-							
20	+	-	-	-		vorüber	eilen	-	+	-	-	+	-	-	-	-	+	-	-	-	-	-	+	-							
21	+	-	+	-		vorüber	fahren	-	+	-	-	+	-	-	-	-	+	-	-	-	-	-	+	-		Vorüberfahrt					
22	+	-	+	-		vorüber	flitzen	-	+	-	-	+	-	-	-	-	+	-	-	-	-	-	+	-							
23	+	-	-	+		vorüber	laufen	-	+	-	-	+	-	-	-	-	+	-	-	-	-	-	+	-							
N₀ V Loc_{src} N₁src																															
1	+	-	-	-		ab	rücken	-	+	-	-	-	+	-								+	-	-							
2	+	-	-	-	sich	ab	setzen	+	-	-	-	-	+	-								+	-	-							
3	+	+	-	-		ab	wandern	-	+	-	-	-	+	-								+	-	-		Abwanderung					
4	+	-	-	-		ab	ziehen	-	+	-	-	-	+	-								+	-	-		Abzug					
5	-	-	-	+		aus	laufen	-	+	-	-	-	+	-								+	-	-		<i>Die Fähre läuft aus</i>					
6	+	-	-	-		aus	reisen	-	+	-	-	-	+	-								+	-	-							
7	-	-	-	+		aus	segeln	-	+	-	-	-	+	-								+	-	-							
8	+	-	-	-		aus	wandern	-	+	-	-	-	+	-								+	-	-							
9	+	-	-	-		aus	ziehen	-	+	-	-	-	+	-								+	-	-		Auszug					
10	+	-	+	-		davon	brausen	-	+	-	-	-	+	-								+	-	-							
11	+	-	+	+		davon	eilen	-	+	-	-	-	+	-								+	-	-							
12	+	+	-	-		davon	galoppieren	-	+	-	-	-	+	-								+	-	-							
13	+	+	+	+		davon	rasen	-	+	-	-	-	+	-								+	-	-							
14	+	-	-	-		davon	reiten	-	+	-	-	-	+	-								+	-	-							
15	+	+	-	-		davon	rennen	-	+	-	-	-	+	-								+	-	-							
16	+	-	-	-	sich	davon	schleichen	+	-	-	-	-	+	-								+	-	-							
17	+	-	-	-		davon	schleichen	-	+	-	-	-	+	-								+	-	-							
18	+	-	-	-	sich	davon	stehlen	+	-	-	-	-	+	-								+	-	-							
19	+	-	-	-		davon	stürmen	-	+	-	-	-	+	-								+	-	-							
20	+	-	-	-			emigrieren	-	+	-	-	-	+	-								+	-	-		Emigration					

Tabelle C.3: Bewegungsverben (Fortsetzung)

id	N ₀ =Hum	N ₀ =anim	N ₀ =transport	N ₀ =ship	reflexiv	Prfx, trennbar	Prfx, fest	Lemma	Perfekt <i>haben</i>	Perfekt <i>sein</i>	N ₀ V N ₁ akk	N ₀ V N ₁ dat	N ₀ V Prep N ₁	N ₀ V Loc N ₁	N ₀ V Loc _{stat} N ₁	Prep= <i>auf</i> _{akk}	Prep= <i>an</i> _{dat}	Prep= <i>aus</i>	Prep= <i>in</i> _{akk}	Prep= <i>in</i> _{dat}	Prep= <i>über</i> _{akk}	N ₁ =source	N ₁ =passage	N ₁ =destinat.	Pfad=∅	Kommentar	V _n	V _{n2}	V _{n3}
21	+	+	+	+	sich		ent	fernen	+	-	-	-	-	+	-							+	-	-					
22	+	+	-	-				fliehen	-	+	-	-	-	+	-							+	-	-		Flucht			
23	+	-	+	-		fort		brausen	-	+	-	-	-	+	-							+	-	-					
24	+	-	+	-		fort		dampfen	-	+	-	-	-	+	-							+	-	-					
25	+	-	+	-		fort		fahren	-	+	-	-	-	+	-							+	-	-					
26	+	+	-	-		fort		flattern	-	+	-	-	-	+	-							+	-	-					
27	+	-	-	-		fort		gehen	-	+	-	-	-	+	-							+	-	-		Fortgang			
28	+	+	-	-		fort		laufen	-	+	-	-	-	+	-							+	-	-					
29	+	-	-	-	sich	fort		machen	+	-	-	-	-	+	-							+	-	-					
30	+	+	-	-		fort		reiten	-	+	-	-	-	+	-							+	-	-					
31	+	+	-	-		fort		rennen	-	+	-	-	-	+	-							+	-	-					
32	+	+	-	-	sich	fort		schleichen	+	-	-	-	-	+	-							+	-	-					
33	+	+	-	-		fort		schleichen	-	+	-	-	-	+	-							+	-	-					
34	+	+	-	-	sich	fort		schleppen	+	-	-	-	-	+	-							+	-	-					
35	+	+	-	+		fort		segeln	-	+	-	-	-	+	-							+	-	-					
36	+	-	-	-		fort		stiefeln	-	+	-	-	-	+	-							+	-	-					
37	+	+	-	-		fort		traben	-	+	-	-	-	+	-							+	-	-					
38	+	+	-	-		fort		ziehen	-	+	-	-	-	+	-							+	-	-					
39	+	-	-	-		herab		eilen	-	+	-	-	-	+	-							+	-	-					
40	+	-	+	-		herab		rollen	-	+	-	-	-	+	-							+	-	-					
41	+	-	-	-		herab		steigen	-	+	-	-	-	+	-							+	-	-					
42	+	+	-	-		herab		stürzen	-	+	-	-	-	+	-							+	-	-					
43	+	-	-	-				türmen	-	+	-	-	-	+	-							+	-	-					
44	+	-	-	-		weg		fahren	-	+	-	-	-	+	-							+	-	-					
45	+	+	-	-		weg		fliegen	-	+	-	-	-	+	-							+	-	-					
46	+	-	-	-		weg		gehen	-	+	-	-	-	+	-							+	-	-					
47	+	+	-	-		weg		laufen	-	+	-	-	-	+	-							+	-	-					
48	+	-	-	-	sich	weg		machen	+	-	-	-	-	+	-							+	-	-					
49	+	-	-	+		weg		segeln	-	+	-	-	-	+	-							+	-	-					

Tabelle C.3: Bewegungsverben (Fortsetzung)

id	N ₀ =Hum	N ₀ =anim	N ₀ =transport	N ₀ =ship	reflexiv	Prfx, trennbar	Prfx, fest	Lemma	Perfekt <i>haben</i>	Perfekt <i>sein</i>	N ₀ V N ₁ akk	N ₀ V N ₁ dat	N ₀ V Prep N ₁	N ₀ V Loc N ₁	N ₀ V Loc _{stat} N ₁	Prep= <i>auf</i> _{akk}	Prep= <i>an</i> _{dat}	Prep= <i>aus</i>	Prep= <i>in</i> _{akk}	Prep= <i>in</i> _{dat}	Prep= <i>über</i> _{akk}	N ₁ =source	N ₁ =passage	N ₁ =destinat.	Pfad=∅	Kommentar	Vn	Vn ₂	Vn ₃	
27	+	+	+	+				gelangen	-	+	-	-	-	+	-															
28	+	-	-	-				hechten	-	+	-	-	-	+	-															
29	+	-	-	-	sich	heim		begeben	+	-	-	-	-	+	-															
30	+	-	-	-		heim		eilen	-	+	-	-	-	+	-															
31	+	-	-	-		heim		fahren	-	+	-	-	-	+	-															Heimfahrt
32	+	-	-	-		heim		gehen	-	+	-	-	-	+	-															
33	+	-	-	-		heim		kehren	-	+	-	-	-	+	-															Heimkehr
34	+	-	-	-		heim		kommen	-	+	-	-	-	+	-															
35	+	-	-	-		heim		laufen	-	+	-	-	-	+	-															
36	+	-	-	-		heim		reisen	-	+	-	-	-	+	-															Heimreise
37	+	-	-	-		heim		rennen	-	+	-	-	-	+	-															
38	+	-	-	-		heim		ziehen	-	+	-	-	-	+	-															
39	+	-	-	-				hetzen	-	+	-	-	-	+	-															
40	+	-	+	+		hin		fahren	-	+	-	-	-	+	-															Hinfahrt
41	+	-	+	-		hin		fliegen	-	+	-	-	-	+	-															Hinflug
42	+	-	-	-		hin		gehen	-	+	-	-	-	+	-															
43	+	+	-	-				hoppeln	-	+	-	-	-	+	-															
44	+	+	-	-				hopsen	-	+	-	-	-	+	-															
45	+	+	-	-				humpeln	-	+	-	-	-	+	-															
46	+	+	-	-				hupfen	-	+	-	-	-	+	-															
47	+	+	-	-				hüpfen	-	+	-	-	-	+	-															
48	+	+	-	-				huschen	-	+	-	-	-	+	-															
49	+	-	-	-				jetten	-	+	-	-	-	+	-															
50	+	-	-	-				jumpen	-	+	-	-	-	+	-															
51	+	+	-	-				klettern	-	+	-	-	-	+	-															Kletterei Klettertour
52	+	+	-	-				kommen	-	+	-	-	-	+	-															
53	+	+	-	-				krabbeln	-	+	-	-	-	+	-															
54	+	-	-	-				kraulen	-	+	-	-	-	+	-															
55	+	-	-	-				krawlen	-	+	-	-	-	+	-															

Tabelle C.3: Bewegungsverben (Fortsetzung)

id	N ₀ =Hum	N ₀ =anim	N ₀ =transport	N ₀ =ship	reflexiv	Prfx, trennbar	Prfx, fest	Lemma	Perfekt <i>haben</i>	Perfekt <i>sein</i>	N ₀ V N ₁ akk	N ₀ V N ₁ dat	N ₀ V Prep N ₁	N ₀ V Loc N ₁	N ₀ V Loc _{stat} N ₁	Prep= <i>auf</i> _{akk}	Prep= <i>an</i> _{dat}	Prep= <i>aus</i>	Prep= <i>in</i> _{akk}	Prep= <i>in</i> _{dat}	Prep= <i>über</i> _{akk}	N ₁ =source	N ₁ =passage	N ₁ =destinat.	Pfad= \emptyset	Kommentar	Vn	Vn ₂	Vn ₃	
56	+	+	+	-				kriechen	-	+	-	-	-	+	-							-	-	+						
57	+	+	-	-				kullern	-	+	-	-	-	+	-							-	-	+						
58	+	-	-	-				latschen	-	+	-	-	-	+	-							-	-	+		Latscherei				
59	+	+	-	-				laufen	-	+	-	-	-	+	-							-	-	+		Lauf				
60	+	-	-	-				marschieren	-	+	-	-	-	+	-							-	-	+		Marsch				
61	+	+	-	-				paddeln	-	+	-	-	-	+	-							-	-	+						
62	+	-	+	-				pesen	-	+	-	-	-	+	-							-	-	+						
63	+	-	-	-				pilgern	-	+	-	-	-	+	-							-	-	+						
64	+	-	-	-				radeln	-	+	-	-	-	+	-							-	-	+		Radltour	Radl-Tour			
65	+	-	-	-		rad		fahren	-	+	-	-	-	+	-							-	-	+		Radfahrt	Radtour			
66	+	+	+	-				rasen	-	+	-	-	-	+	-							-	-	+		Raserei				
67	+	-	-	-				reisen	-	+	-	-	-	+	-							-	-	+		Reise				
68	+	-	-	-				reiten	-	+	-	-	-	+	-							-	-	+		Reiten	Ritt			
69	+	+	-	-				rennen	-	+	-	-	-	+	-							-	-	+		Rennen				
70	+	+	-	+	sich			retten	+	-	-	-	-	+	-							-	-	+						
71	+	+	-	-				robben	-	+	-	-	-	+	-							-	-	+						
72	+	+	+	-				rollen	-	+	-	-	-	+	-							-	-	+						
73	+	-	-	-				rudern	-	+	-	-	-	+	-							-	-	+		Rudern				
74	+	+	+	-				sausen	-	+	-	-	-	+	-							-	-	+						
75	+	+	-	-				schleichen	-	+	-	-	-	+	-							-	-	+						
76	+	-	-	-				schlendern	-	+	-	-	-	+	-							-	-	+		Schlendern				
77	+	+	-	-	sich			schleppen	+	-	-	-	-	+	-							-	-	+						
78	+	-	-	-	sich			schmuggeln	+	-	-	-	-	+	-							-	-	+						
79	+	+	-	-				schwimmen	-	+	-	-	-	+	-							-	-	+		Schwimmen				
80	+	-	-	+				segeln	-	+	-	-	-	+	-							-	-	+		Segeln	Segeltörn	Törn		
81	+	-	-	-				skaten	-	+	-	-	-	+	-							-	-	+						
82	+	-	-	-		spazieren	fahren		-	+	-	-	-	+	-							-	-	+		Spazierfahrt				
83	+	-	-	-		spazieren	gehen		-	+	-	-	-	+	-							-	-	+		Spaziergang				
84	+	-	-	-		spazieren	reiten		-	+	-	-	-	+	-							-	-	+		Spazierritt				

Tabelle C.3: Bewegungsverben (Fortsetzung)

id	N ₀ =Hum	N ₀ =anim	N ₀ =transport	N ₀ =ship	reflexiv	Prfx, trennbar	Prfx, fest	Lemma	Perfekt <i>haben</i>	Perfekt <i>sein</i>	N ₀ V N ₁ akk	N ₀ V N ₁ dat	N ₀ V Prep N ₁	N ₀ V Loc N ₁	N ₀ V Loc _{stat} N ₁	Prep= <i>auf</i> _{akk}	Prep= <i>an</i> _{dat}	Prep= <i>aus</i>	Prep= <i>in</i> _{akk}	Prep= <i>in</i> _{dat}	Prep= <i>über</i> _{akk}	N ₁ =source	N ₁ =passage	N ₁ =destinat.	Pfad=∅	Kommentar	V _n	V _{n2}	V _{n3}								
85	+	-	-	-				spazieren	-	+	-	-	-	+	-																						
86	+	+	-	-				springen	-	+	-	-	-	+	-																			Sprung			
87	+	+	-	-				sprinten	-	+	-	-	-	+	-																						
88	+	-	-	-				stapfen	-	+	-	-	-	+	-																						
89	+	-	-	-				steigen	-	+	-	-	-	+	-																						
90	+	+	-	-				stolzieren	-	+	-	-	-	+	-																						
91	+	-	-	-				stürmen	-	+	-	-	-	+	-																				≠ erobern (nur Bewegung)		
92	+	-	-	-				treten	-	+	-	-	-	+	-																						
93	+	-	-	-			über	siedeln	-	+	-	-	-	+	-																				Übersiedelung		
94	+	-	-	-			über	siedeln	-	+	-	-	-	+	-																				Übersiedelung		
95	+	-	-	-			vor	rücken	-	+	-	-	-	+	-																						
96	+	-	-	-				wallfahren	-	+	-	-	-	+	-																					Wallfahrt	
97	+	+	-	-				wandern	-	+	-	-	-	+	-																					Wanderung	
98	+	-	+	+			weiter	fahren	-	+	-	-	-	+	-																					Weiterfahrt	
99	+	+	+	-			weiter	fliegen	-	+	-	-	-	+	-																					Weiterflug	
100	+	-	-	-			weiter	gehen	-	+	-	-	-	+	-																						
101	+	-	-	-			weiter	laufen	-	+	-	-	-	+	-																						
102	+	-	-	-			weiter	marschieren	-	+	-	-	-	+	-																						Weitermarsch
103	+	-	-	-			weiter	reisen	-	+	-	-	-	+	-																						Weiterreise
104	+	-	-	-			weiter	rennen	-	+	-	-	-	+	-																						
105	+	+	-	-			weiter	ziehen	-	+	-	-	-	+	-																						
106	+	-	-	-			wieder	kommen	-	+	-	-	-	+	-																						
107	+	-	+	-				zuckeln	-	+	-	-	-	+	-																						
108	+	-	-	-	sich		zurück	begeben	+	-	-	-	-	+	-																						
109	+	-	-	-	sich		zurück	bemühen	+	-	-	-	-	+	-																						
110	+	-	+	+			zurück	fahren	-	+	-	-	-	+	-																						Rückfahrt
111	+	+	+	-			zurück	fliegen	-	+	-	-	-	+	-																						
112	+	-	-	-			zurück	gehen	-	+	-	-	-	+	-																						

Tabelle C.3: Bewegungsverben (Fortsetzung)

id	$N_0=Hum$	$N_0=anim$	$N_0=transport$	$N_0=ship$	reflexiv	Prfx, trennbar	Prfx, fest	Lemma	Perfekt <i>haben</i>	Perfekt <i>sein</i>	$N_0 V N_1_{akk}$	$N_0 V N_1_{dat}$	$N_0 V Prep N_1$	$N_0 V Loc N_1$	$N_0 V Loc_{stat} N_1$	Prep= <i>auf</i> _{akk}	Prep= <i>an</i> _{dat}	Prep= <i>aus</i>	Prep= <i>in</i> _{akk}	Prep= <i>in</i> _{dat}	Prep= <i>über</i> _{akk}	$N_1=source$	$N_1=passage$	$N_1=destinat.$	Pfad= \emptyset	Kommentar	Vn	Vn ₂	Vn ₃			
113	+	+	+	+			zurück	kehren	-	+	-	-	-	+	-							-	-	+								
114	+	+	+	+			zurück	kommen	-	+	-	-	-	+	-								-	-	+							
115	+	+	-	+			zurück	laufen	-	+	-	-	-	+	-								-	-	+							
116	+	-	-	-			zurück	marschieren	-	+	-	-	-	+	-								-	-	+		Rückmarsch					
117	+	+	+	-			zurück	rasen	-	+	-	-	-	+	-								-	-	+							
118	+	-	-	-			zurück	reisen	-	+	-	-	-	+	-								-	-	+		Rückreise					
119	+	-	-	-			zurück	reiten	-	+	-	-	-	+	-								-	-	+							
120	+	+	-	-			zurück	rennen	-	+	-	-	-	+	-								-	-	+							
121	+	-	+	-			zurück	rollen	-	+	-	-	-	+	-								-	-	+							
122	+	-	-	-			zurück	rudern	-	+	-	-	-	+	-								-	-	+							
123	+	+	-	-			zurück	schwimmen	-	+	-	-	-	+	-								-	-	+							
124	+	+	-	-			zurück	wandern	-	+	-	-	-	+	-								-	-	+		Rückwanderung					
125	+	-	-	-	sich		zurück	ziehen	+	-	-	-	-	+	-								-	-	+		Rückzug					
$N_0 V Loc_{pass} N_1 pass$																																
1	+	-	-	-				bummeln	-	+	-	-	-	+	-								-	+	-		Bummelei					
2	+	-	+	-		durch		brausen	-	+	-	-	-	+	-								-	+	-							
3	+	-	-	-	sich	durch		drängeln	+	-	-	-	-	+	-								-	+	-							
4	+	-	-	-	sich	durch		drängen	+	-	-	-	-	+	-								-	+	-							
5	+	-	+	+		durch		fahren	-	+	-	-	-	+	-								-	+	-		Durchfahrt					
6	+	-	+	-		durch		flitzen	+	-	-	-	-	+	-								-	+	-							
7	+	-	-	-		durch		gehen	-	+	-	-	-	+	-								-	+	-							
8	+	+	+	-		durch		jagen	-	+	-	-	-	+	-								-	+	-							
9	+	+	-	-		durch		klettern	-	+	-	-	-	+	-								-	+	-							
10	+	+	+	-		durch		kommen	-	+	-	-	-	+	-								-	+	-							
11	+	+	-	-		durch		kriechen	-	+	-	-	-	+	-								-	+	-							
12	+	+	-	-		durch		laufen	-	+	-	-	-	+	-								-	+	-							
13	+	-	-	-		durch		marschieren	-	+	-	-	-	+	-								-	+	-							
14	+	-	-	-		durch		plumpsen	-	+	-	-	-	+	-								-	+	-							

Tabelle C.3: Bewegungsverben (Fortsetzung)

id	N ₀ =Hum	N ₀ =anim	N ₀ =transport	N ₀ =ship	reflexiv	Prfx, trennbar	Prfx, fest	Lemma	Perfekt <i>haben</i>	Perfekt <i>sein</i>	N ₀ V N ₁ akk	N ₀ V N ₁ dat	N ₀ V Prep N ₁	N ₀ V Loc N ₁	N ₀ V Loc _{stat} N ₁	Prep= <i>auf</i> _{akk}	Prep= <i>an</i> _{dat}	Prep= <i>aus</i>	Prep= <i>in</i> _{akk}	Prep= <i>in</i> _{dat}	Prep= <i>über</i> _{akk}	N ₁ =source	N ₁ =passage	N ₁ =destinat.	Pfad=∅	Kommentar	V _n	V _{n2}	V _{n3}				
15	+	-	+	-		durch		rasen	-	+	-	-	-	+	-																		
16	+	-	-	-		durch		reisen	-	+	-	-	-	+	-																		Durchreise
17	+	-	-	-		durch		reiten	-	+	-	-	-	+	-																		
18	+	-	+	-		durch		rollen	-	+	-	-	-	+	-																		
19	+	-	-	-		durch		rücken	-	+	-	-	-	+	-																		
20	+	-	+	-		durch		sausen	-	+	-	-	-	+	-																		
21	+	+	+	-	sich	durch		schlängeln	+	-	-	-	-	+	-																		
22	+	+	+	-	sich	durch		schleichen	+	-	-	-	-	+	-																		
23	+	-	-	-	sich	durch		schmuggeln	+	-	-	-	-	+	-																		
24	+	+	-	-		durch		schwimmen	-	+	-	-	-	+	-																		
25	+	-	-	-		durch		steigen	-	+	-	-	-	+	-																		
26	+	+	-	-		durch		tappen	-	+	-	-	-	+	-																		
27	+	-	-	-	sich	durch		tasten	+	-	-	-	-	+	-																		
28	+	-	-	-		durch		taumeln	-	+	-	-	-	+	-																		
29	+	-	-	-		durch		treten	-	+	-	-	-	+	-																		
30	+	+	-	-		durch		waten	-	+	-	-	-	+	-																		
31	+	+	-	-		durch		ziehen	-	+	-	-	-	+	-																		
32	+	-	-	-				flanieren	-	+	-	-	-	+	-																		
33	+	-	-	-				geistern	-	+	-	-	-	+	-																		
34	+	-	-	-				gondeln	-	+	-	-	-	+	-																		
35	+	-	+	-				hasten	-	+	-	-	-	+	-																		
36	+	-	+	-				heizen	-	+	-	-	-	+	-																		
37	+	-	+	-				holpern	-	+	-	-	-	+	-																		
38	+	-	-	-				irren	-	+	-	-	-	+	-																		
39	+	+	+	-				jagen	-	+	-	-	-	+	-																		
40	+	-	-	-				joggen	-	+	-	-	-	+	-																		
41	+	-	+	-				klappern	-	+	-	-	-	+	-																		
42	+	-	+	-				kurven	-	+	-	-	-	+	-																		
43	+	-	-	-				nomadisieren	-	+	-	-	-	+	-																		

Tabelle C.3: Bewegungsverben (Fortsetzung)

id	$N_0=Hum$	$N_0=anim$	$N_0=transport$	$N_0=ship$	reflexiv	Prfx, trennbar	Prfx, fest	Lemma	Perfekt <i>haben</i>	Perfekt <i>sein</i>	$N_0 V N_1_{akk}$	$N_0 V N_1_{dat}$	$N_0 V Prep N_1$	$N_0 V Loc N_1$	$N_0 V Loc_{stat} N_1$	Prep= <i>auf</i> _{akk}	Prep= <i>an</i> _{dat}	Prep= <i>aus</i>	Prep= <i>in</i> _{akk}	Prep= <i>in</i> _{dat}	Prep= <i>über</i> _{akk}	$N_1=source$	$N_1=passage$	$N_1=destinat.$	Pfad= \emptyset	Kommentar	Vn	Vn ₂	Vn ₃		
44	+	-	+	+				patrouillieren	+	+	-	-	-	+	-							-	+	-							
45	+	+	-	-				streuen	-	+	-	-	-	+	-							-	+	-							
46	+	+	-	-				stromern	-	+	-	-	-	+	-							-	+	-							
47	+	-	-	-				tingeln	-	+	-	-	-	+	-							-	+	-							
48	+	-	-	-				tippeln	-	+	-	-	-	+	-							-	+	-							
49	+	-	-	-				toben	-	+	-	-	-	+	-							-	+	-							
$N_0 V Loc_{stat} N_1 dest$																															
1	+	+	+	+		an		kommen	-	+	-	-	-	-	+							-	-	+							Ankunft
2	+	-	-	-		an		langen	-	+	-	-	-	-	+							-	-	+							
3	-	-	-	+		an		legen	-	+	-	-	-	-	+							-	-	+							
4	+	-	-	-		an		tanzen	-	+	-	-	-	-	+							-	-	+							
5	+	-	-	-	sich	ein		finden	+	-	-	-	-	-	+							-	-	+							
6	+	-	+	-		ein		treffen	-	+	-	-	-	-	+							-	-	+							
7	+	+	+	+				landen	-	+	-	-	-	-	+							-	-	+							Landung
8	+	-	-	+				stranden	-	+	-	-	-	-	+							-	-	+	+						
$N_0 V N_1 Loc_{dst} N_1 dest$																															
1	+	-	-	-			ver	reisen	-	+	-	-	-	+	-							-	-	+							
$N_0 V Loc_{src/stat} N_1 src$																															
1	+	-	+	+		ab		fahren	-	+	-	-	-	+	+							+	-	-							Abfahrt
2	+	+	+	-		ab		fliegen	-	+	-	-	-	+	+							+	-	-							Abflug
3	-	-	+	-		ab		heben	-	+	-	-	-	+	+							+	-	-							
4	-	-	-	+		ab		legen	-	+	-	-	-	+	+							+	-	-							
5	+	-	-	-		ab		marschieren	-	+	-	-	-	+	+							+	-	-							Abmarsch
6	+	-	-	-		ab		reisen	-	+	-	-	-	+	+							+	-	-							Abreise
7	-	-	-	+		ab		segeln	-	+	-	-	-	+	+							+	-	-							

Tabelle C.4: Statische lokative Verben „Mensch – Ort“

id	$N_0=Hum$	$N_0=HumColl$	reflexiv	Prfx, trennbar	Prfx, fest	Lemma	Perfekt <i>haben</i>	Perfekt <i>sein</i>	$N_0 V N_1 loc_{akk}$	$N_0 V Loc_{src} N_1 loc$	$N_0 V Loc_{stat} N_1 loc$	Beispiel	Kommentar	Vn_1	Vn_2	Vn_3
$N_0 VN_1 loc_{akk}$																
1	-	+				besiedeln	+	-	+	-	-	Über 20 Indianerstämme besiedeln die Region.	meist inchoativ			
2	-	+				bevölkern	+	-	+	-	-	Kleine Handwerker und vor allem Bauern und Hirten bevölkerten die Region.				
3	-	+				bewohnen	+	-	+	-	-	Die Ye-Maek bewohnten das Gebiet um den Oberlauf des Yalu-Stromes.				
4	+	+				bewohnen	+	-	+	-	-	Wir bewohnen ein kleines Häuschen mit großem Garten.				
$N_0 VLoc_{src} N_1 loc$																
1	+					kommen	-	-	-	+	-	Burns kommt aus Porto.				
2	+					stammen	-	-	-	+	-	Der 23-jährige Mittelfeldspieler stammt aus der Region um Nyon, in der ...				
$N_0 VLoc_{stat} N_1 loc$																
1	+					sein	-	+	-	-	+	Max ist in München.	Vsup			
2	+		sich auf			halten	+	-	-	-	+	Sebald hat sich in England aufgehalten.	Vsup			
3	+		sich			befinden	+	-	-	-	+	Angela Merkel befindet sich derzeit in den USA.	Vsup			
4	+					bleiben	-	+	-	-	+	Vier Tagen blieben wir in Italien am Lago Maggiore.	(Vsup)			
5	+					biwakieren	+	-	-	-	+	Wir biwakierten auf einer Moräne auf knapp 3000m Höhe.		Biwak		
6	+					campen	+	-	-	-	+	Wir campen auf dem Potuhisa Campground, gleich am Parkeingang.		Camp	Camping	
7	+					campieren	+	-	-	-	+	Die Kameraden vom DRK Campierten an der Grillhütte.				
8	+					gastieren	+	-	-	-	+	David Byrne gastierte in Wien.				
9	+					hausen	+	-	-	-	+	Wir hausten in Wellblechbuden.				
10	+					kampieren	+	-	-	-	+	Trent, der in einem Wohnwagen im Wald kumpiert, ...				
11	+					lagern	+	-	-	-	+	Zu Ostern 1525 lagerte der Neckartaler Haufen bei Weinsberg ...		Lagern	Lager	
12	+					leben	+	-	-	-	+	Stefan Müller lebt in der Nähe von Hanau am Main.		Leben	Lebensmittelpunkt	
13	+					logieren	+	-	-	-	+	James Last logierte im Maritim-Badehotel.				
14	+					nächtigen	+	-	-	-	+	Wir nächtigten in der Biopension Alpenheim.				
15	+	+				residieren	+	-	-	-	+	Lady Cochran residiert auf Schloß Surssock im christlichen Ostteil der Stadt.		Residieren	Residenz	

Tabelle C.4: Statische lokative Verben „Mensch – Ort“ (Fortsetzung)

id	N ₀ =Hum	N ₀ =HumColl	reflexiv	Prfx, trennbar	Prfx, fest	Lemma	Perfekt <i>haben</i>	Perfekt <i>sein</i>	N ₀ V N ₁ loc _{akk}	N ₀ V Loc _{src} N ₁ loc	N ₀ V Loc _{stat} N ₁ loc	Beispiel	Kommentar	V _n	V _{n2}	V _{n3}
16	+	+				siedeln	+	-	-	-	+	<i>Lange bevor . . . , siedelten in der Gegend der heutigen Hansestadt bereits Germanen.</i>		Siedeln	Siedlung	
17	+	+				sitzen	-	+	-	-	+	<i>Das Ingenieürüro für Organisation und Technik (IBOT) sitzt in Remscheid.</i>		Sitz		
18	+				über	nachten	+	-	-	-	+	<i>Wir übernachteten im Hotel.</i>		Übernachtung		
19	+					weilen	+	+	-	-	+	<i>Gegenüber der Polizei sagte er, dass er privat in Italien geweilt habe.</i>				
20	+					wohnen	+	-	-	-	+	<i>Mr. White wohnt auf Helgoland.</i>		Wohnung	Wohnsitz	Wohnort
21	+					zelten	+	-	-	-	+	<i>Wir zelten an der Ostsee.</i>				

Tabelle C.5: Verben zum Ausdruck von Relationen zwischen geographischen Entitäten

id	N_0 =Choronym N_0 =Hydronym N_0 =Fluss N_0 =Hodonym reflexiv Prfx, trennbar Prfx, fest Lemma	Perfekt <i>haben</i> Perfekt <i>sein</i> N_0 V N_1 _{akk} N_0 V N_1 _{dat} N_0 V Prep N_1 N_0 V Loc N_1 N_0 V Loc _{stat} N_1 P rep= <i>aus</i> P rep= <i>in</i> _{akk} P rep= <i>an</i> _{akk} P rep= <i>zu</i> P rep= <i>mit</i> N_1 =Toponym N_1 =Choronym N_1 =Hydronym N_1 =Fluss N_1 =Choron+admin N_1 =Hodonym P rep2= <i>zu</i> N_2 =Hydronym N_2 =Fluss Beispiel
N_0locVN$_1$loc_{akk}		
1	- - + - durchfließen	+ - + - - - - - - + - - - - - <i>Der Fluss Suzhou durchfließt Shanghai auf einer Länge von 54 km.</i>
2	- - + - durchfließen	+ - + - - - - - - - - + - - - - <i>Weiter nördlich durchfließt die Wolga den Rybinsker Stausee. . .</i>
3	- - + - durchlaufen	+ - + - - - - - - - - - + - - - - <i>Der Kleine Regen durchläuft die Trinkwassertalsperre Frauenau.</i>
4	- - + + durchlaufen	+ - + - - - - - - - - + - - - - - <i>Die von Turin kommende Autobahn A5 durchläuft das gesamte Aostatal. Der Randelbach durchläuft die Bauernschaft Bentlage bei Rheine.</i>
5	- - + + durchqueren	+ - + - - - - - - - - + - - - - - <i>Die Moskwa durchquert das Stadtgebiet von Moskau in Mäandern von Nordwest nach Südost.</i>
6	- - + - durchströmen	+ - + - - - - - - - - - - + - - - - <i>Der Shire durchströmt den Malombesee.</i>
7	- - + - durchströmen	+ - + - - - - - - - - - + - - - - - <i>Der Fluss Öxará durchströmt den Nationalpark.</i>
8	- - + - queren	+ - + - - - - - - - - - - - - - + <i>Am westlichen Ende quert die Beeke die Bundesstrasse 70.</i>
9	- - + + queren	+ - + - - - - - - - - - + - - - - - <i>Der Pfad quert den Hang des Raukopfs auf halber Höhe.</i>
10	- - - + queren	+ - + - - - - - - - - - - - + - - - <i>Die Karlsbrücke quert die Moldau seit dem 14 Jhd.</i>
11	- - + - speisen	+ - + - - - - - - - - - - + - - - - <i>Der Gabelbach speist am Unterlauf den Ritzebühler Teich. . .</i>
12	- - - + überbrücken	+ - + - - - - - - - - - - - + - - - <i>Die Ponte Romano überbrückt den Fluss Ofanto.</i>
13	+ - - - umfassen	+ - + - - - - - - - - - + - - - - - <i>Der Bezirk Berlin Steglitz-Zehlendorf umfasst die Ortsteile Steglitz, Lichterfelde, . . .</i>
14	- - + + unterqueren	+ - + - - - - - - - - - - - + - + <i>Die Südaue unterquert den Mittellandkanal bei Stromkilometer 140.</i>
N_0locVN$_1$loc_{dat}		
1	- - + - zu fließen	- + - + - - - - - - - - - + - - <i>Hier fließt die Samara der Wolga von links zu.</i>
N_0locVPrepN$_1$loc		

Tabelle C.5: Verben zum Ausdruck von Relationen zwischen geographischen Entitäten (Fortsetzung)

id	N_0 =Choronym	N_0 =Hydronym	N_0 =Fluss	N_0 =Hodonym	reflexiv	Prfx, trennbar	Prfx, fest	Lemma	Perfekt <i>haben</i>	Perfekt <i>sein</i>	N_0 V N_1 _{akk}	N_0 V N_1 _{dat}	N_0 V Prep N_1	N_0 V Loc N_1	N_0 V Loc _{stat} N_1	Prep= <i>aus</i>	Prep= <i>in</i> _{akk}	Prep= <i>an</i> _{akk}	Prep= <i>zu</i>	Prep= <i>mit</i>	N_1 =Toponym	N_1 =Choronym	N_1 =Hydronym	N_1 =Fluss	N_1 =Choron+admin	N_1 =Hodonym	Prep2= <i>zu</i>	N_2 =Hydronym	N_2 =Fluss	Beispiel				
1	-	-	+	-		ab		fließen	-	+	-	-	+	-	-	-	+	-	-	-	-	-	+	-	-	-	-	-	-	-	-	<i>Die Binnenalster fließt im Süden über die Kleine Alster ab.</i>		
2	+	-	-	-		an		grenzen	+	-	-	-	+	-	-	-	-	+	-	-	+	-	-	-	-	-	-	-	-	-	-	<i>... liegt im Norden des Bundesstaates Minas Gerais, der an den Nachbarstaat Bahia angrenzt.</i>		
3	-	-	+	-		aus		treten	-	+	-	-	+	-	-	+	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	<i>Bei Savona tritt der Thompson River aus dem rund 30 Kilometer langen Kamloops Lake aus.</i>		
4	-	-	+	-		ein		münden	-	+	-	-	+	-	-	-	+	-	-	-	-	-	+	-	-	-	-	-	-	-	-	<i>der Strymonas, welcher in den Nordosten des Kerkini Sees einmündet</i>		
5	-	-	-	+		ein		münden	-	+	-	-	+	-	-	-	+	-	-	-	-	-	-	-	-	-	-	+	-	-	-	<i>An der Osterstraße, die in den östlichen Marktplatz einmündet, befindet sich ...</i>		
6	-	-	+	-				entspringen	-	+	-	-	+	-	-	+	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	<i>Der Schwarze Drin entspringt aus dem Ohridsee.</i>	
7	-	-	+	-				entstehen	-	+	-	-	+	-	-	+	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	<i>+ Die Elbe selbst entsteht aus den beiden Gewässern Weißwasser und der Kleinen Elbe.</i>	
8	-	-	+	-	sich			ergießen	+	-	-	-	+	-	-	+	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	<i>... des Odiel, der sich hier in den Atlantischen Ozean ergießt.</i>	
9	-	-	+	-				fließen	-	+	-	-	+	-	-	+	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	<i>Der Mystic River fließt in den Long Island Sound.</i>	
10	-	-	-	-				gehören	+	-	-	-	+	-	-	-	-	-	+	-	-	-	-	-	-	-	-	+	-	-	-	-	<i>Der Gonsenheimer Wald gehört zum Naturschutzgebiet "Mainzer Sand".</i>	
11	+	-	-	-				grenzen	+	-	-	-	+	-	-	-	+	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	<i>das an den Bayerischen Wald grenzende Mühlviertel</i>	
12	-	-	+	-		heraus		treten	-	+	-	-	+	-	-	+	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	<i>Der Malefinkbach tritt nun aus dem Tal heraus und fließt ...</i>
13	-	-	+	-				münden	-	+	-	-	+	-	-	+	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	<i>Der Río de la Plata mündet in den Atlantischen Ozean.</i>
14	-	+	-	-	sich			speisen	+	-	-	-	+	-	-	+	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	<i>Die Weser speist sich aus den Flüssen Werra und Fulda.</i>
15	-	-	+	-				treten	-	+	-	-	+	-	-	+	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	<i>Der Kamp tritt hier aus dem engen Kampthal in das Tullnerfeld.</i>
16	-	-	+	-	sich			vereinigen	+	-	-	-	+	-	-	-	-	-	-	+	-	-	-	-	+	-	-	-	-	-	-	-	-	<i>Die Linth vereinigt sich in Schwanden mit dem Sernft.</i>
17	-	-	+	-				zusammenfließen	-	+	-	-	+	-	-	-	-	-	-	+	-	-	-	-	+	-	-	-	-	-	-	-	-	<i>Dort fließt der Tallapoosa River mit dem Coosa River zusammen, um den Alabama River zu bilden.</i>

 N_0 locVLoc_{dst} N_1 loc

Tabelle C.5: Verben zum Ausdruck von Relationen zwischen geographischen Entitäten (Fortsetzung)

id	N_0 =Choronym	N_0 =Hydronym	N_0 =Fluss	N_0 =Hodonym	reflexiv	Prfx, trennbar	Prfx, fest	Lemma	Perfekt <i>haben</i>	Perfekt <i>sein</i>	N_0 V N_1 _{akk}	N_0 V N_1 _{dat}	N_0 V Prep N_1	N_0 V Loc N_1	N_0 V Loc _{stat} N_1	Prep= <i>aus</i>	Prep= <i>in</i> _{akk}	Prep= <i>an</i> _{akk}	Prep= <i>zu</i>	Prep= <i>mit</i>	N_1 =Toponym	N_1 =Choronym	N_1 =Hydronym	N_1 =Fluss	N_1 =Choron+admin	N_1 =Hodonym	Prep2= <i>zu</i>	N_2 =Hydronym	N_2 =Fluss	Beispiel
1	+	-	-	-	sich		erstrecken	+	-	-	-	-	+	-							+	-	-	-	-	-				<i>Der Naturpark "Hohe Mark" erstreckt sich von Dülmen bis zum Niederrhein.</i>
2	+	-	-	-			reichen	+	-	-	-	-	+	-							+	-	-	-	-	-				<i>Der Wald reicht bis zur Stadtgrenze.</i>
N_0locVLoc_{pass}N_1loc																														
1	-	-	+	-			fließen	-	+	-	-	-	+	-							-	+	-	-	-	-				<i>Die Nau fließt durch Langenau.</i>
2	-	-	+	-			führen	+	-	-	-	-	+	-							-	+	-	-	-	-				<i>Der Fluss führt durch intensiv landwirtschaftlich genutztes Hügelland.</i>
3	-	-	+	-	sich		schlängeln	+	-	-	-	-	+	-							-	+	-	-	-	-				<i>Die Spree schlängelt sich durch Berlin.</i>
4	-	-	-	+	sich		schlängeln	+	-	-	-	-	+	-							-	+	-	-	-	-				<i>Über die Höhen schlängeln sich Paßstraßen.</i>
5	-	-	+	-			verlaufen	-	+	-	-	-	+	-							-	+	-	-	-	-				<i>Ein kleiner Bach verläuft durch die Mitte des Tales.</i>
6	-	-	-	+			verlaufen	-	+	-	-	-	+	-							-	+	-	-	-	-				<i>Die Straße verläuft durch das Gebiet der früheren Kämme-reiheide.</i>
N_0locVLoc_{stat}N_1loc																														
1	-	-	+	-			entspringen	-	+	-	-	-	-	+							-	+	-	-	-	-				<i>Der Neckar entspringt am Rande des Schwarzwaldes bei Villingen-Schwenningen.</i>
N_0locVN₁locPrepN_2loc																														
1	-	-	+	-			entwässern	+	-	-	-	+	-	-	-	+	-	-	-	-	-	-	+	-	-	-	+			<i>Die kurze Buna entwässert den Shkodrasee in die Adria.</i>
N_0locVPrepN_1locPrepN_2loc																														
1	-	-	+	-	sich		vereinigen	+	-	-	-	+	-	-	-	-	-	-	+	-	-	-	+	-	-	+	-	+		<i>Nördlich von Bremen-Blockland, bei Wasserhorst, vereinigt sich die Wümme mit der Hamme zur Lesum.</i>

C.3 Nomina

Die Tabelle auf den folgenden Seiten beschreibt die Eigenschaften (quasi)prädikativer Nomina des Typs „Hauptstadt“, die eine Relation zwischen zwei Orten beschreiben und dem Satzschema $N_0loc V_{sup} Nloc N_1loc$ folgen (siehe Kap. 4.6.3). Die Spalten bzw. Symbole in der Tabelle beschreiben dabei folgende Eigenschaften :

- Restriktionen beim semantischen Typs von N_0 , N_1 bzw. N_2 :
 - $N_0=Oikonym$ Siedlung (Oikonyme)
 - $N_0=Choronym$ Regionen (Choronyme)
 - $N_0=admin. Choronym$ administrative Regionen
 - $N_0=Hydronym$ Gewässer (Hydronyme)
 - $N_0=Fluss$ Fließgewässer
 - $N_0=Hodonym$ Verkehrswege (Hodonyme)
- Anschluss von N_0 über Stützverb/Kopula, und zwar *sein*, *bilden*, bzw. *darstellen*
- Anschluss von N_0 in Apposition (§ N_0appos): *Frankreichs Hauptstadt Paris* mit *Frankreich* als N_1 und *Paris* als N_0 .
- Modifikation des Nomens durch *Ex-*: *Deutschlands Ex-Hauptstadt Bonn*
- syntaktische Anbindung von N_1 (§ N_{1gen} , usw.), siehe Kap. 4.6 mit Beispielen.
- gleichzeitige Anbindung von N_0 und N_1 durch:
 - N_0-N_1 -§ *Rhein-Main-Zusammenfluss*
 - § von N_0 und N_1 *Zusammenfluss von Rhein und Main*
- für Nomina, die eine ternäre Relation ausdrücken (z.B. *Enklave*), ist die Präposition, durch die N_2 angeschlossen wird, angegeben.

Tabelle C.6: Relationen zwischen zwei geographischen Entitäten: Nomina Typ «Hauptstadt»

id	N_0 =Oikonym	N_0 =admin.Region	N_0 =Choronym	N_0 =Hodonym	N_0 =Fluss	Copul= <i>sein</i>	Copul= <i>bilden</i>	Copul= <i>darstellen</i>	§ N_0 appos	Ex-	Lemma	N_1 -§	N_1 #§	N_1 adj §	§ N_1 gen	§ von N_1	§ in N_1 dat	§ zu N_1	N_0 - N_1 -§	§ von N_0 und N_1	N_1 =admin.Region	N_1 =Hodonym	N_1 =Hydronym	ternäre Relation	in N_2 dat	in N_2 akk	Beispiel
1	-	+	-	-	-	+		+			Bezirksgemeinschaft	-	-	+	+	+				§ von N_0 und N_1	+	-	-	-			<i>Hamburg-Mitte ist ein Bezirk der Freien und Hansestadt Hamburg.</i>
2	-	+	-	-	-	+		+			Bundesland	-	-	+	+	+	+			§ von N_0 und N_1	+	-	-	-			<i>Baden-Württemberg ist ein Bundesland der Bundesrepublik Deutschland.</i>
3	-	+	-	-	-	+		+			Bundesstaat	-	-	+	+	+	+			§ von N_0 und N_1	+	-	-	-			<i>Hawaiï ist der 50. Bundesstaat der USA.</i>
4	-	+	-	-	-	+		+			Département	-	-	+	+	+	+			§ von N_0 und N_1	+	-	-	-			<i>das französische Département Alpes-Maritimes</i>
5	-	+	-	-	-	+		+			Gliedstaat	-	-	+	+	+	+			§ von N_0 und N_1	+	-	-	-			<i>Von 1871 bis 1945 war Preußen ein Gliedstaat des Deutschen Reiches.</i>
6	-	+	-	-	-	+		+			Gouvernement	-	-	+	+	+	+			§ von N_0 und N_1	+	-	-	-			<i>Witebsk war ein Gouvernement des Russischen Reiches.</i>
7	-	+	-	-	-	+		+			Gouvernorat	-	-	+	+	+	+			§ von N_0 und N_1	+	-	-	-			<i>Musandam ist ein Gouvernorat im Sultanat Oman.</i>
8	-	+	-	-	-	+		+			Kanton	-	-	+	+	+	+			§ von N_0 und N_1	+	-	-	-			<i>Waadt ist ein Schweizer Kanton.</i>
9	-	+	-	-	-	+		+			Kolonie	-	-	+	+	+	+			§ von N_0 und N_1	+	-	-	-			<i>St. Vincent blieb bis 1969 britische Kolonie.</i>
10	-	+	-	-	-	+		+			Kronkolonie	-	-	+	+	+	+			§ von N_0 und N_1	+	-	-	-			<i>Gibraltar ist seit 1730 britische Kronkolonie.</i>
11	-	+	-	-	-	+		+			Provinz	-	-	+	+	+	+			§ von N_0 und N_1	+	-	-	-			<i>Livland und Kurland waren Provinz des zaristischen Russlands.</i>
12	-	+	-	-	-	+		+			Teilgebiet	-	-	+	+	+	+			§ von N_0 und N_1	+	-	-	-			<i>auf der Laichinger Alb, einem Teilgebiet der Schwäbischen Alb</i>
13	-	+	-	-	-	+		+			Teilrepublik	-	-	+	+	+	+			§ von N_0 und N_1	+	-	-	-			<i>die jugoslawische Teilrepublik Mazedonien</i>
14	-	+	-	-	-	+		+			Überseegebiet	-	-	+	+	+	+			§ von N_0 und N_1	+	-	-	-			<i>Bermuda ist ein Überseegebiet des Vereinigten Königreichs.</i>
15	-	+	-	-	-	+		+			Überseeterritorium	-	-	+	+	+	+			§ von N_0 und N_1	+	-	-	-			<i>1946 wurde Kongo französisches Überseeterritorium.</i>
16	-	+	-	-	-	+		+			Unionsterritorium	-	-	+	+	+	+			§ von N_0 und N_1	+	-	-	-			<i>Delhi ist ein indisches Unionsterritorium mit Sonderrechten.</i>
17	-	+	-	-	-	+		+			Verwaltungseinheit	-	-	+	+	+	+			§ von N_0 und N_1	+	-	-	-			<i>Der Bezirk Waldenburg ist eine Verwaltungseinheit im Kanton Basel-Landschaft.</i>
18	+	-	-	-	-	+		+	+		Bezirkshauptstadt	+	-	+	+	+	-			§ von N_0 und N_1	+	-	-	-			<i>Hallein ist Bezirkshauptstadt des Tennengaus.</i>
19	+	-	-	-	-	+		+	+		Bundeshauptstadt	-	-	+	+	+	-			§ von N_0 und N_1	+	-	-	-			<i>die Schweizer Bundeshauptstadt Bern</i>
20	+	-	-	-	-	+		+	+		Hauptstadt	+	-	+	+	+	-			§ von N_0 und N_1	+	-	-	-			<i>Bern ist die Hauptstadt der Schweiz.</i>
21	+	-	-	-	-	+		+	+		Inselhauptstadt	+	-	+	+	+	-			§ von N_0 und N_1	+	-	-	-			<i>Die Inselhauptstadt von Pemba ist Chake Chake.</i>

Tabelle C.6: Relationen zwischen zwei geographischen Entitäten: Nomina Typ «Hauptstadt» (Fortsetzung)

id	N_0 =Oikonym	N_0 =admin.Region	N_0 =Choronym	N_0 =Hodonym	N_0 =Fluss	Copul= <i>sein</i>	Copul= <i>bilden</i>	Copul= <i>darstellen</i>	§ N_0 appos	Ex-	Lemma	N_1 -§	N_1 #§	N_1 adj §	§ N_1 gen	§ von N_1	§ in N_1 dat	§ zu N_1	N_0 - N_1 -§	§ von N_0 und N_1	N_1 =admin.Region	N_1 =Hodonym	N_1 =Hydronym	ternäre Relation	in N_2 dat	in N_2 akk	Beispiel	
22	+	-	-	-	-	+			+	+	Kantonshauptstadt	+	-	+	+	+	-			+	-	-	-					die Baselbieter Kantonshauptstadt Liestal
23	+	-	-	-	-	+			+	+	Kreisstadt	+	-	+	+	+	-			+	-	-	-					Bad Homburg vor der Höhe ist die Kreisstadt des Hochtaunuskreises.
24	+	-	-	-	-	+			+	+	Landeshauptstadt	+	-	+	+	+	-			+	-	-	-					Düsseldorf ist Landeshauptstadt von Nordrhein-Westfalen.
25	+	-	-	-	-	+			+	+	Provinzhauptstadt	+	-	+	+	+	-			+	-	-	-					Die Provinzhauptstadt von Kandal heißt Ta Khmau.
26	+	-	-	-	-	+			+	+	Regierungssitz	+	-	+	+	+	-			+	-	-	-					Bonn war bis 1999 Regierungssitz der Bundesrepublik Deutschland.
27	+	-	-	-	-	+			+	+	Regionalhauptstadt	+	-	+	+	+	-			+	-	-	-					Dijon ist Regionalhauptstadt der französischen Region Burgund.
28	+	-	-	-	-	+			+	+	Reichshauptstadt	+	-	+	+	+	-			+	-	-	-					Berlin war 1871–1945 Reichshauptstadt des Deutschen Reichs.
29	+	-	-	-	-	+			+	+	Verwaltungssitz	+	-	+	+	+	-			+	-	-	-					Friedrichshafen ist Verwaltungssitz des Bodenseekreises.
30	+	+	+	-	-	+			+		Zentrum	-	-	+	+	+	-			+	-	-	-					Schöllkrippen ist das Zentrum des oberen Kahlgrunds.
31	-	-	-	-	+	+			+		Altarm	+	+	-	+	+	-			-	-	+	-					Die Alte Donau in Wien ist ein Altarm der Donau.
32	-	-	-	-	+	+			+		Arm	+	+	-	+	+	-			-	-	+	-					ein Arm des Orinoco, der Casiquiare, fließt ...
33	-	-	-	-	+	-			-		Konfluenz	-	-	-	-	-	-	-	+	-	-	+	-					Konfluenz von Weißem und Blauem Nil
34	-	-	-	-	+	+	+	+	+	+	Hauptzufluss	+	+	-	+	+	+	+		-	-	+	-					Der Hauptzufluss des Zürichsees ist die Linth.
35	-	-	-	-	+	+	+	+	+	+	Nebenarm	+	+	-	+	+	+	+		-	-	+	-					Der Kittelbach ist ein Nebenarm der Nördlichen Düssel.
36	-	-	-	-	+	+	+	+	+	+	Nebenfluss	+	+	-	+	+	+	+		-	-	+	-					Die Prims ist ein Nebenfluss der Saar.
37	-	-	-	-	+	+	+	+	+	+	Nebenflüsschen	+	+	-	+	+	+	+		-	-	+	-					der Jacobigraben, ein kleines Nebenflüsschen der Erse
38	-	-	-	-	+	+	+	+	+	+	Nebengewässer	+	+	-	+	+	+	+		-	-	+	-					das Klosterwasser, ein Nebengewässer der Schwarzen Elster
39	-	-	-	-	+	+	+	+	+	+	Quellbach	+	+	-	+	+	+	+		-	-	+	-					Das Großschweidnitzer Wasser ist der linke Quellbach des Löbauer Wassers.
40	-	-	-	-	+	+	+	+	+	+	Quellfluss	+	+	-	+	+	+	+		-	-	+	-					Die Hache ist ein Quellfluss der Ochtum.
41	-	-	-	-	+	+	+	+	+	+	Quellflüsschen	+	+	-	+	+	+	+		-	-	+	-					die Doyca, ein Quellflüsschen der Obra
42	-	-	-	-	+	+	+	+	+	+	Seitenbach	+	+	-	+	+	+	+		-	-	+	-					der Ruisseau de Motélon, ein südlicher Seitenbach des Jaunbachs, ...

Tabelle C.6: Relationen zwischen zwei geographischen Entitäten: Nomina Typ «Hauptstadt» (Fortsetzung)

id	N_0 =Oikonym	N_0 =admin.Region	N_0 =Choronym	N_0 =Hodonym	N_0 =Fluss	Copul= <i>sein</i>	Copul= <i>bilden</i>	Copul= <i>darstellen</i>	§ N_0 appos	Ex-	Lemma	N_1 -§	N_1 #§	N_1 adj §	§ N_1 gen	§ von N_1	§ in N_1 dat	§ zu N_1	N_0 - N_1 -§	§ von N_0 und N_1	N_1 =admin.Region	N_1 =Hodonym	N_1 =Hydronym	ternäre Relation	in N_2 dat	in N_2 akk	Beispiel	
43	-	-	-	-	+	+	+	+	+		Seitenfluss	+	+	-	+	+	-	+			-	-	+	-				<i>Die Hron (deutsch Gran) ist ein 298 km langer, linker Seitenfluss der Donau.</i>
44	-	-	-	-	+	+	+	+	+		Zufluss	+	+	-	+	+	-	+			-	-	+	-				<i>Die Glinze ist ein rechter Zufluss der Dosse.</i>
45	-	-	-	-	+	+			+		Zusammenfluss	+	+	-	+	+	-	+			-	-	+	-				<i>Zusammenfluss der Aare mit dem Rhein</i>
46	-	-	-	-	+	+			-		Zusammenfluss	-	-	-	-	-	-	-	+	+	-	-	+	-				<i>Isar-Amper-Zusammenfluss</i>
47	-	-	-	-	+	+	+	+	+		Abfluss	+	+	-	+	+	-		+	+	-	-	+	+	-	+		<i>Die Narva bildet den etwa 78 Kilometer langen Abfluss des Peipussees zur Ostsee.</i>
48	-	-	+	-	-	+			-		Einmündung	+	+	-	+	+	-				-	-	+	+	-	+		<i>Einmündung der Rotach in den Bodensee</i>
49	-	-	+	-	-	+	+	+	-		Mündung	+	+	-	+	+	-				-	-	+	+	-	+		<i>die Mündung der Drau in die Donau</i>
50	-	-	+	-	-	+	+	+	-		Mündungsgebiet	+	+	-	+	+	-				-	-	+	+	-	+		<i>Das Donaudelta ist das Mündungsgebiet der Donau ins Schwarze Meer.</i>
51	-	-	+	-	-	+			+		Insel	+	+	+	+	+					-	-	+	-				<i>Guadalcanal ist eine pazifische Insel.</i>
52	-	+	-	-	-	+	+	+	+		Anrainer	+	+	-	+	+	-	-			+	-	+	-				<i>Iran, selbst Anrainer des Kaspischen Meeres, ...</i>
53	-	+	-	-	-	+	+	+	+		Anrainerstaat	+	+	-	+	+	-	-			+	-	+	-				<i>Griechenland ist ein Mittelmeer-Anrainerstaat.</i>
54	+	-	-	-	-	+			+		Stadtteil	-	-	-	+	+	-	-			+	-	-	-				<i>im Berliner Stadtteil Charlottenburg</i>
55	-	-	+	-	-	+	+	+	+		Hauptinsel	+	+	+	+	+					+	-	-	-				<i>die japanische Hauptinsel Honshū</i>
56	-	+	-	-	-	+			+		Enklave	-	-	+	+	+	-	-			+	-	-	+	+	-		<i>Büdingen ist eine deutsche Enklave in der Schweiz.</i>
57	-	+	-	-	-	+			+		Exklave	-	-	+	+	+	-	-			+	-	-	+	+	-		<i>Das Kaliningrader Gebiet ist eine russische Exklave.</i>

C.4 Adverbien

Die Tabelle der Adverbien enthält Adverbien dreier Satz schemata, die bereits in Kap. 4.7 diskutiert wurden. In den Spalten der Tabelle werden folgende Eigenschaften beschrieben (von links nach rechts):

- der semantische Typ von N_0 :

$N_0=Hum$ Menschen

$N_0=HumColl$ Gruppen von Menschen

$N_0=anim$ belebte Wesen, d.h. alle Tiere

$N_0=veget$ Pflanzen

$N_0=conc$ Konkreta

$N_0=transport$ Verkehrsmittel

- die Kopula kann immer durch *sein* realisiert werden, auf eine besondere Kennzeichnung wurde deshalb verzichtet. Für einzelne Adverbien sind weiter möglich: *liegen* und *sich befinden*.
- der Beginn des durch das Adverb beschriebenen Zustandes kann durch ein als ‚IncepFunc₁‘ bezeichnetes (Stütz)Verb ausgedrückt werden, und zwar *werden*, oder auch *gehen* bzw. *sich machen*. Kann das Adverb mit einem solchen Verb kombiniert werden, wird dies als Merkmal kodiert.
- ein vorangestellter Modifikator: *polizeilich gemeldet*
- die Verwendung als attributives Adjektiv: *Die Linde ist in ganz Europa heimisch* → *die in ganz Europa heimische Linde*. Dies ist nicht für alle Adverbien möglich.
- die Satz schemata, in denen das Adverb realisiert werden kann. Neben den drei klassifizierenden Schemata sind dies auch die Varianten:

$N_0 V_{sup} \S$ *Ein US-Lazarettschiff ist unterwegs (in die Krisenregion).*

$N_0 V_{sup} Path \S$ *Die verunglückte Boeing war auf dem Weg von Manaus in die Hauptstadt Brasilia.*

Tabelle C.7: Adverbien und Adjektive mit lokativen Komplementen (N_0 V_{sup} Loc N_1loc Adv)

id	$N_0=Hum$	$N_0=anim$	$N_0=veget$	$N_0=loc$	$N_0=conc$	$N_0=transport$	$N_0=HumColl$	$Copul=liegen$	$Copul=sich\ befinden$	$IncepFunc_1=werden$	$IncepFunc_1=\dots$	vorangest. Modif.	Lemma	ADJ	N_0 V_{sup} Loc N_1loc §	N_0 V_{sup} Loc N_1dest §	N_0 V_{sup} $Path$ §	N_0 V_{sup} §	N_0 V_{sup} von N_1loc N_2mess §	Beispiel
N_0 V_{sup} Loc_{stat} N_1loc Adv																				
1	+	+	-			-	+	-	-	+			ansässig	+	+	-	-	-	-	<i>Das Unternehmen ist in Trångsviken ansässig.</i>
2	+	-	-	-		-	-	-	-	+			wohnhaft	+	+	-	-	-	-	<i>Sie ist in Ennetbürgen wohnhaft.</i>
3	+	+	-	+	+	-	-	-	-	-			befindlich	+	+	-	-	-	-	<i>ihre bereits in Israel befindlichen Angehörigen</i>
4	+	-	-			-	+	-	+	-			daheim	-	+	-	-	-	-	<i>Frau Müller ist in Gelsenkirchen daheim.</i>
5	+	+	+			-	-	-	+	-			zuhause	-	+	-	-	-	-	<i>Jeder fünfte Österreicher ist in Wien zuhause.</i>
6	-	+	+			-	-	-	-	+			heimisch	+	+	-	-	-	-	<i>Die Linde ist in ganz Europa heimisch.</i>
7	+	-	-			-	-	-	-	-		<i>polizeilich</i>	gemeldet	+	+	-	-	-	-	<i>Der Antragsteller ist in Potsdam gemeldet.</i>
8	-	-	-			-	+	-	-	-			geschäftsansässig gemeldet	+	+	-	-	-	-	<i>Das Unternehmen ist in Epfach geschäftsansässig gemeldet.</i>
9	-	-	-			-	+	-	-	+			geschäftsansässig	+	+	-	-	-	-	<i>Die Detektei ist in Cottbus geschäftsansässig.</i>
10	-	-	-			-	+	-	-	-			registriert	+	+	-	-	+	-	<i>Die GmbH ist in Oberhausen registriert.</i>
11	-	+	+			-	+	-	-	-			beheimatet	+	+	-	-	-	-	<i>Die Küstentanne ist in Nordamerika beheimatet.</i>
12	+	-	-			-	+	-	-	+			ortsansässig	+	+	-	-	-	-	<i>Das Ingenieurbüro ist in Halle (Saale) ortsansässig.</i>
13	+	-	-			-	+	-	-	+			ortsfremd	+	+	-	-	-	-	<i>Leute, die in Frankfurt ortsfremd sind, ...</i>
14	+	-	-			-	+	-	-	+			gebietsansässig	+	+	-	-	-	-	<i>eine natürliche oder juristische Person, die in Österreich gebietsansässig ist, ...</i>
15	+	-	+		+	-	+	-	-	+			gebietsfremd	+	+	-	-	-	-	<i>... ob der Steuerpflichtige gebietsfremd ist, ...</i>
16	+	-	-			-	-	+	-	-			begraben	+	+	-	-	-	-	<i>Beethoven ist in Wien begraben.</i>
N_0 V_{sup} $Prep$ N_1loc N_2mess Adv																				
17	+	+	+	+	+	-	+	+	-				entfernt	+	-	-	-	-	+	<i>Das Hotel ist nur 50 Meter vom Strand entfernt.</i>
18	+	+	+	+	+	-	+	+	-				weg	-	-	-	-	-	+	<i>Das Haus ist keine 100m weit weg.</i>
19	-	-	-	+	-	-	+	+	-				abgelegen	+	-	-	-	-	+	<i>Tysnes ist ungefähr 1,5 Autostunden von Bergen abgelegen.</i>

Tabelle C.7: Adverbien und Adjektive mit lokativen Komplementen (N_0 V_{sup} Loc N_1loc Adv) (Fortsetzung)

id	$N_0=Hum$	$N_0=anim$	$N_0=veget$	$N_0=loc$	$N_0=conc$	$N_0=transport$	$N_0=HumColl$	Copul= <i>liegen</i>	Copul= <i>sich befinden</i>	IncepFunc ₁ = <i>werden</i>	IncepFunc ₁ =...	vorangest. Modif.	Lemma	ADJ	N_0 V_{sup} Loc N_1loc §	N_0 V_{sup} Loc N_1dest §	N_0 V_{sup} Path §	N_0 V_{sup} §	N_0 V_{sup} von N_1loc N_2mess §	Beispiel
N_0	V_{sup}	Loc_{dst}	N_1dest	Adv																
20	+		+	-	-	-							unterwegs	-	+	+	+	+	-	<i>Die zwei Kasachen sind unterwegs nach Almaty.</i>
21	+		+	-	+	-							auf dem Weg	-	-	+	+	-	-	<i>Sie ist auf dem Weg zur Schule.</i>
22	+			-	+	-							auf der Durchreise	-	-	+	+	+	-	<i>Der junge Deutsche ist auf der Durchreise nach Nicaragua.</i>
23	+			-	-	-		<i>sich machen</i>					auf und davon	-	-	+	+	+	-	<i>Peter ist auf und davon (nach Rom).</i>
24	+		-	-	+	-		<i>gehen</i>					auf Reisen	-	+	-	-	+	-	<i>Sie war auf Reisen nach Tunesien und in die Türkei.</i>

D Lokativa im Erklärend-Kombinatorischen Wörterbuch

Lokativa im weiteren Sinn, d.h. Bezeichnungen für Lokationen und Situationen, die Lokationen als Aktanten haben, sind in den vorhandenen Erklärend-Kombinatorischen Wörterbüchern unsystematisch, aber doch exemplarisch zu finden. Das russische TKSR enthält *peč²* (‘Ofen’), *passažir* (‘Passagier’), *sudno* (‘Boot’), *parochod* (‘Dampfer’), *lyži* (‘Ski’), *kinno/kinoteatr* (‘Kino’); Das französische DEC die Adverbiale *à pied* ‘zu Fuß’ (DEC: II, 99), das Nomen *vitesse* ‘Geschwindigkeit’ (II, 321–4), und zahlreiche Bezeichnungen für Gewässer (étendues d’eau): *anse²* ‘kleine Bucht’ (III, 150), *baie¹* ‘Bai, Bucht’ (III, 159), *cataracte²* 1. ‘Katarakt, Wasserfall’ (IV, 175), *étang* ‘Teich’ (III, 202), *flaque* ‘Pfütze, Lache’ (III, 216), *golfe* ‘Golf, Meerbusen’ (III, 221), *lac* ‘See’ (III, 230), *marais* ‘Sumpf, Moor(see), Marsch’ (III, 234), *mare* ‘Tümpel’ (III, 234), *marécage* ‘Moor(gebiet), Sumpfgebiet’ (III, 235), *mer* ‘Meer’ (III, 236–241), *mer intérieure* ‘sehr großer Salzsee (z.B. Aralsee, Kaspisches Meer)’ (III, 242), *océan* ‘Ozean’ (III, 260–1), *voie d’eau* II ‘Wasserstraße, Wasserweg’ (DEC: III, 317, vgl. Mel’čuk 1995: 224) Einige der Gewässerbezeichnungen sind unten exemplarisch abgedruckt. In Kap. D.1 folgen Skizzen von Einträgen für einige deutsche Lexeme, die wie z.B. *wohnen* (Kap. 2.5.3.1) in dieser Arbeit detaillierter besprochen wurden.

VOIE D’EAU, nominal, fem., lit. ‘way/path of water’

I. $\lceil \text{Voie d’eau} \rceil \text{ dans } X [\approx \text{‘a leak’}] = \text{an accidental hole in a submerged part } X^1 \text{ of a vessel } X^2 \text{ through which water enters } X^2 \text{ such that it represents a danger for } X^2 \text{ and its occupants because it can cause that } X^2 \text{ sinks.}$

[...]

II. $\lceil \text{Voie d’eau} \rceil \text{ entre } X \text{ et } Y \text{ [pour } Z] [\approx \text{‘a waterway’}] = \lceil \text{Body of water} \rceil \text{ whose elongated form implies an axe “X–Y” and that serves or can serve as a waterway for navigation of } Z \text{ between } X \text{ and } Y.$

Government Pattern

X = I	Y = II	Z = III
1. <i>entre</i> ‘between’ N et N		—

$C_1 + C_2$: *une voie d’eau entre Montréal et Québec*

Lexical Functions

Syn_U : canal ‘canal’; détroit ‘strait’; passage ‘passage’
 Real₃ : emprunter ‘[to] borrow’ [ART ~]; exploiter ‘[to] exploit’, utiliser ‘[to] utilize’ [ART ~]; naviguer ‘[to] navigate [on]’ [sur ART ~]

Examples

Le Saint-Laurent est une voie d’eau importante en Amérique. Les Grands Lacs sont une voie d’eau très exploitée entre l’est et l’ouest du pays.

ANSE², nom, fém.
Anse = Petite BAIE¹.

Lexical Functions

Les FL des sections «Une anse et son environnement», «Comportement d'une anse», «Déplacement dans une anse», «Exploitation d'une anse»: ↑ EAU 2*

Syn_C : BAIE¹
 Syn_U : crique
 Loc_{in/ad} : dans [ART ~]

Notes

EAU 2* \simeq Masse plus ou moins considérable d'eau 1 qui est un élément géographique [*marche au bord de l'eau; la surface des eaux*]; cf. EAU 1 \simeq Liquide transparent nécessaire à la vie [*boire un verre d'eau*]

BAIE¹, nom, fém.

Baie = Partie d'une MER I.3 (, d'un lac ou d'une rivière), délimitée par une échancrure de la côte (ou de la rive) et ayant la grandeur d'une «étendue d'eau» qui est entre «grand» et «petit».

N.B. 1 : La composante 'd'un lac ou d'une rivière' est considérée optionnelle à cause d'un flottement d'usage : pour certains locuteurs, une baie ne peut être située que dans une mer, alors que pour d'autres, une baie peut se trouver autant dans un lac ou dans une rivière que dans une mer.

N.B. 2 : N.B. dans ↑ MER I.3

Lexical Functions

Syn_C : anse²
 Syn_U : golfe; fjord/fjord; crique
 Contr : pointe, presqu'île
 Loc_{in/ad} : dans [ART ~]
 Loc_{in}² → le bord de la B. : au ⟨sur le⟩ bord de [ART ~]; sur [ART ~] | G désigne (la localisation d')une agglomération [*New-Richmond est situé sur la Baie des Chaleurs*]
 Loc_{in}² → région de la B. : à [la ~] | B. est accompagné de la dénomination [*New-Richmond est situé sur la Baie des Chaleurs*]

CATARACTE², nom, fém.

I. *Cataracte* = Très grande masse d'eau d'un grand «cours d'eau» à gros débit qui tombe de très haut à cause d'un changement_{I.1a} de pente très brusque.

Lexical Functions

Syn_C : «chute d'eau», chute(s)
 Syn_U : cascade
 Son : gronder, murir
 S₀Son : fracas, grondement, mugissement [de ART ~]

Exemples

Les cataractes du Niagara sont célèbres dans le monde entier. Au spectacle de ces cataractes, il croit assister à l'une de ces scènes constitutives de l'univers. Le fracas de l'eau qui heurte les parois rocheuses, le bruit caverneux de la cataracte l'ont inspiré pour sa composition.

ÉTANG, nom, masc.

Étang = 'Étendue d'eau' (douce) à l'intérieur des terres qui est entre «assez grand» et «petit», et dont la majeure partie est libre de végétation.

Lexical Functions

Les FL des sections «Un étang et son environnement», «Comportement d'un étang», «Déplacement dans un étang», «Exploitation d'un étang»: ↑ EAU 2*

F₂, F₃, Loc_{ad} : ↑ MER I.1
Syn_U : lac, mare 1
Gener : 'étendue d'eau', spéc 'nappe d'eau'
Loc_{in} : dans [ART ~]
étant très calme, Func₀ : dormir

Notes

EAU 2* : ↑ ANSE²

LAC, nom, masc.

Lac = (Assez) grande 'étendue d'eau' à l'intérieur des terres dont une grande partie est libre de végétation.

N.B. : N.B. dans ↑ MER I.3

Lexical Functions

Les FL des sections «Un lac et son environnement», «Comportement d'un lac», «Déplacement dans un lac», «Exploitation d'un lac»: ↑ EAU 2*

F₂, F₃, Loc_{ad} : ↑ MER I.1
Syn_U : étang
Anti : marais
Contr : île
Gener : 'étendue d'eau', spéc 'nappe d'eau'
A₀ : lacustre
Loc_{in} : dans [ART ~]
Loc_{in}² → région du L. : au [~] | L. est accompagné de la dénomination [*New-Richmond est situé sur la Baie des Chaleurs*]
étant plutôt grand, Func₀ : s'étaler, s'étendre | a un dépendant
étant calme, Func₀ : dormir

Notes

EAU 2* : ↑ ANSE²

MER, nom, fem.

I.1 L'immense 'étendue d'eau' salée qui couvre une grande partie de la Terre ... [*La mer s'étendait à perte de vue*]

I.2a Immense 'étendue d'eau' salée – partie de la MER I.1 [*une mer tropicale*]

I.2b L'ensemble des mers MER I.2A [*parcourir les mers*]

I.3 MER I.2A délimitée par une ou des terre(s) ... [*la mer Méditerranée*]

II. Immense étendue de X ... [*une mer de manifestants*]

MER INTÉRIEURE, loc. nom, fem.

'Mer intérieure' = Très grand lac salé.

N.B. 1 : Une mer intérieure s'appelle soit «lac α» soit «mer α».

N.B. 2 : N.B. dans ↑ MER I.3

Lexical Functions

F₂, F₃, Loc_{ad} : ↑ MER I.1
Syn_▷ : lac

Examples

La plupart des mers intérieures actuelles sont des vestiges d'anciennes mers qui étaient reliées aux océans.

D.1 Skizzenhafte Einträge deutscher Lokativa

Die folgenden skizzenhaften Einträge einiger lokativer Lexeme beziehen sich auf im Textteil der Arbeit angesprochene Wörter oder Konzepte z.B. *wohnen*, *Straße* und die Präposition *hinter*. Nur die wenigsten Einträge sind vollständige Einträge im Sinne des Erklärend-Kombinatorischen Wörterbuchs:

- weitere, nicht-lokative Bedeutungen/Lesarten der Stichwörter sind meist nur angedeutet oder fehlen ganz. Das gilt insbesondere für Fälle regulärer Polysemie (*die ganze Stadt war auf den Beinen*).
- nur die wichtigsten lexikalischen Funktionen sind angegeben.

HINTER¹, Präp., mit Dativ

I. *X hinter Y* = X befindet sich in einer Position relativ zu Y, so dass Ys Rücken oder Rückseite X zugekehrt ist.

Max steht hinter Peter.

Hinter dem Haus liegt der Garten.

Hinter der Absperrung untersucht die Polizei den Tatort.

Hinter dem Schalter sitzt der Postbeamte und gähnt, die Schlange vor dem Schalter wird immer länger.

[Präposition, räumlich, statisch, intrinsisch]

[Rücken oder Rückseite von Y sind konventionalisiert, bei Lebewesen ist es die dem Wahrnehmungsapparat entgegengesetzte Seite, bei Fahrzeugen die Seite, die entgegen der üblichen, kanonischen Bewegungsrichtung weist. Bei Häusern ist es die, der Straße abgewandte Seite, bei Objekten wie Altar, Thresen, Schalter ist es die Seite, auf der das Publikum nicht ist.]

II. *X hinter Y (von Z aus gesehen)* = X befindet sich auf der dem Beobachter Z abgewandten Seite von Y.

Lexical Functions

Syn_⊂ : *jenseits*
Conv₂₁ : *vor*

Von dir aus gesehen liegt der Ball vor dem Baum.

Nimm die Kugel, die von dir aus gesehen hinter dem Würfel liegt.

Hinter dem Hügel liegt ein See.

[Präposition, räumlich, statisch, relativ; transitive Relation: *hinter dem Berg fließt ein Fluss* und *hinter dem Fluss liegt ein Wald*, dann befindet sich *der Wald hinter dem Berg*]

III.

[(für Z auf der Route W kommt) X hinter Y] = das Objekt oder die Person Z bewegt sich auf der Route W und passiert von den zwei Orten X und Y, die sich beide auf oder neben W befinden, Y zuerst, und später X.

Lexical Functions

Syn : *nach*
Conv₂₁ : *vor*

15 km hinter Nürnberg auf der A3 liegt die Raststätte Greding.

IV.

X *hinter* Y = X und Y bewegen sich in der gleichen Richtung und auf der gleichen Route, dabei passiert Y die Wegpunkte der Route zeitlich früher als X.

Lexical Functions

A₀ : *folgend, nachfolgend*
V₀ : *folgen*
Conv₂₁ : *vor*

*Hinter dem Lastwagen fahren zwei Polizeiautos.
Häkkinen fährt weit abgeschlagen hinter Schuhmacher.*

V. X (*ist/fährt/liegt*) Z Zeiteinheiten *hinter* Y = wie HINTER¹ IV, aber X und Y sind Teilnehmer eines Rennens, Laufs oder einem ähnlichen (sportlichen) Wettkampfes bei dem die Zeit, die zur Überwindung einer Distanz oder zum Erfüllen einer Aufgabe benötigt wird, darüber entscheidet, wer den Wettkampf gewinnt. X hat dabei Z Zeiteinheiten weniger als Y benötigt, um die (oder einen Teil der) Distanz zu überwinden oder die (oder einen Teil der) Aufgabe zu lösen.

Lexical Functions

Conv₂₁ : *vor*

Häkkinen ist nur 0,5 Sekunden hinter Vilneuve.

VI. X *hinter* Y = X und Y sind Teil einer linear angeordneten Reihe oder Liste von Personen oder Objekten (z.B. Platzierte in einem Wettkampf, Wörter im Satz, Buchstaben im Alphabet, Stichwörter im Lexikon, usw.); werden die Elemente der Liste in ihrer linearen Reihenfolge aufgezählt, wird X zuerst, später Y genannt.

Lexical Functions

Syn : *nach*
Conv₂₁ : *vor*

B kommt hinter A.

Präpositionen stehen hinter dem Nomen.

Schumacher muss sich als Dritter erstmals dem Silberpfeil-Duo geschlagen geben. Häkkinen belegt beim McLaren-Mercedes-Doppelsieg Platz zwei hinter David Coulthard. [D.h. Coulthard ist erster, Häkkinen zweiter, Schumacher dritter der Liste]

HINTER², Pröp., mit Akkusativ

[direktionale Entsprechungen zu HINTER¹ I und HINTER¹ II, ev. auch HINTER¹ IV]

I. *X (bewegt sich) hinter Y = X* bewegt sich oder wird bewegt, so dass X sich am Ende der Bewegung in einer Position relativ zu Y (und in Abhängigkeit zu dessen Ausrichtung) befindet, die durch HINTER¹ I beschrieben wird.

Der Ball rollt hinter den Ofen.

Peter rollt den Ball hinter den Ofen.

Peter geht hinters Haus.

II. *(von Z aus gesehen bewegt sich) X hinter Y = X* bewegt sich oder wird bewegt, so dass X sich am Ende der Bewegung in einer Position relativ zu Y und Z befindet, die durch HINTER¹ I beschrieben wird.

Die rote Kugel rollt hinter die grüne.

Peter stellt sich hinter den Baum.



[für *hinter* mit Dat. und Akk.]

hinter jmdm. liegen

etw. hinter sich haben; etw. hinter sich bringen; etw. hinter sich lassen

hinter/nach jmdm. kommen

hinter jmdm. stehen; sich hinter jmdn. stellen

hinter jmdm./etw. her sein

hinter jmdm. in etw. zurückbleiben ↔ jmdn. in etw. hinter sich lassen

etw. hinter jmds. Rücken tun

es faustdick hinter den Ohren haben

mit seiner Meinung nicht hinter dem Berge halten

sich etw. hinter die Ohren schreiben

hinter die Wahrheit kommen

hinter den Busch gehen/müssen

hinter dem Steuer (sitzen)

HINTER³, Adj., nur attr.

[*das hintere Zimmer*]

HINTER⁴, Adv.

= nach hinten [*Max schickte Peter hinter in den Garten*]

D.1.1 Wortfeld *wohnen*

BAUWERK, N, neutr.

I. eine von Menschen errichtete Konstruktion, die fest mit dem Untergrund verbunden und unbeweglich ist.

II. großes, eindrucksvolles GEBÄUDE.

GEBÄUDE, N, neutr. ein oberirdisches BAUWERK I, das einen oder mehrere Räume umschließt und diese von der Umgebung trennt. Gebäude schützen Menschen, Tieren oder Sachen vor Witterungseinflüssen wie Hitze, Kälte, Niederschläge, usw.

Lexical Functions

Gener	:	<i>Bauwerk</i>
⁴²⁶ AntiBon	:	<i>baufällig // Gemäuer</i>
CausFunc ₀	:	<i>bauen, errichten</i>
LiquFunc ₀	:	<i>abreißen, zerstören</i>

HAUS, N, neutr.

I.1 *das Haus von X* = GEBÄUDE, das dazu bestimmt ist, dass eine Person X (oder eine Gruppe von Personen) es zum Wohnen (siehe WOHNEN I.1) nutzt. (vgl. auch WOHNUNG)

Government Pattern

X = 1 (siehe WOHNUNG)
1. N _{gen} 2. von N _{dat} 3. A _{poss}

Lexical Functions

S ₁	:	<i>Bewohner, Hausbewohner</i>
MultS ₁	:	<i>Hausgemeinschaft</i>
Gener	:	<i>Gebäude, Behausung</i>
Bon	:	<i>stilvoll, luxuriös // Villa, Palast</i>
AntiBon	:	<i>armselig, heruntergekommen, // Baracke, Hütte, Schuppen, Loch, Bruchbude</i>
Contr	:	<i>Garten, Hof</i>
Loc _{in}	:	<i>in</i>
CausFunc ₀ , LiquFunc ₀	:	= ↑ GEBÄUDE
Fact ₀ , Real ₁ (+Komb.)	:	= ↑ WOHNUNG

I.2 = Gebäude

Bürohaus

I.3 Die in einem HAUS I.1 wohnenden Personen. [= S₁Real₁]

Das ganze Haus war auf den Beinen

Lexical Functions

Syn : *Hausgemeinschaft, Bewohner, Hausbewohner*

I.4 = Herrscherhaus, Adelsgeschlecht, Familie

I.5 in der Astrologie ...

I.6 Gewerbebetrieb oder Geschäft, ...

[...]



altes Haus

frei Haus

Hohes Haus

das Weiße Haus

der Herr im Hause sein

mit der Tür ins Haus fallen

...

WOHNUNG, N, fem.

[=S₂(WOHNEN I.1)] *die Wohnung von X* = ein abgeschlossener Teil eines Hauses, bestehend aus meist mehreren Räumen, der dazu bestimmt ist, dass eine Person X (oder eine Gruppe von Personen) ihn zum Wohnen (WOHNEN I.1) nutzt. Eine Wohnung verfügt über die notwendigen Einrichtungen zur Erfüllung der Grundbedürfnisse von X: Schlafen, Ausruhen, Körperpflege, Hygiene, Kochen, usw.

Government Pattern

X = I
1. N _{gen} 2. von N _{dat} 3. A _{poss}

- D_{I.1} : *Peters Wohnung*
 D_{I.2} : *die Wohnung von Peter*
 D_{I.3} : *seine Wohnung*

Lexical Functions

- Syn : *「die eigenen vier Wände」*
 Syn_C : *Apartment, Penthouse, Maisonette*
 Syn_D : *Bude, Heim, Bleibe, Unterkunft*
 Gener : *Behausung*
 Mult : *Haus, Wohnhaus, Wohnkomplex, Wohngebäude, Wohnanlage, Wohnblock*
 S₁ : *Bewohner*
 Real₁ : *bewohnen* [_{akk}], *wohnen in* [_{dat}]
 IncepReal₁ : *beziehen, einziehen in* [_{dat}/_{akk}],
 FinReal₁ : *ausziehen aus* [_{dat}]; *räumen, aufgeben, auflösen* [_{akk}]
 AntiFact₀ : *leer stehen*
 A₀AntiFact₀ : *leerstehend, unbewohnt; unvermietet, frei*
 S₀AntiFact₀ : *Leerstand*
 Loc_{in} : *in*
 CausFunc₀, LiquFunc₀ : = ↑ GEBÄUDE
 Magn : *groß, großzügig*
 AntiMagn : *klein // Kleinstwohnung*
 Ver : *schön, gemütlich*
 Bon : *gut, komfortabel, luxuriös // Luxuswohnung*
 AntiVer : *hässlich // Loch*

DIE EIGENEN VIER WÄNDE

Lexical Functions

- Syn : *Wohnung*
 Loc_{in} : *in, innerhalb*

WOHNEN V

I.1 *X wohnt in Y* = eine Person X (oder eine Gruppe von Personen) nutzt ein Haus, eine Wohnung Y (oder einen Teil davon) zur ständigen Unterkunft: X hält sich regelmäßig dort auf, erfüllt dort

Grundbedürfnisse wie Schlafen, Essen, Körperpflege und verbringt einen wesentlichen Teil seiner Zeit dort. [=Real₁(Haus), Real₁(Wohnung)]

Government Pattern

X = I	Y = II
1. N _{nom}	1. Loc _{in} N

D_{I.1}+D_{II.1} : Peter wohnt in einer Maisonettewohnung
D_{I.1}+D_{II.1} : Peter wohnt auf dem Dachboden

Lexical Functions

Syn : bewohnen
Anti : \lceil unter der Brücke wohnen \rceil , \lceil auf der Straße leben \rceil
S₁ : Bewohner
S₂ : Wohnung, Haus

I.2 X wohnt in Y = die Wohnung oder das Haus, das X zum Wohnen (WOHNEN I.1) nutzt, befindet sich in der Gemeinde Y, und X hat seinen Lebensmittelpunkt in Y, d.h. X hält sich überwiegend in Y auf, schläft/übernachtet regelmäßig in der Wohnung/Haus in Y, verbringt einen Teil seiner Freizeit dort (er muss nicht in Y arbeiten), unterhält persönliche Beziehungen zu anderen Menschen, die in Y wohnen, wenn X eine Familie hat, lebt sie in Y; und X hat Verpflichtungen gegenüber der Gemeinde Y (z.B. die Pflicht, Steuern zu zahlen), aber auch Rechte (z.B. das aktive Wahlrecht).

Government Pattern

X = I	Y = II
1. N _{nom}	1. Loc _{in} N

D_{I.1}+D_{II.1} : Peter wohnt in München.

Lexical Functions

Syn : leben 2
A₁ : ansässig, wohnhaft, meldeamtlich erfasst
S₀ : ständiger Aufenthalt
S₁ : Einwohner
S₂ : Wohnort, Wohnsitz
Incep : sich niederlassen [Loc_{in} N], ziehen [Loc_{ad} N]
Fin : wegziehen, fortziehen [Loc_{ab} N]
Fin+Incep : ziehen/umziehen [Loc_{ab} N] [Loc_{ad} N] (von Berlin nach Hamburg ziehen)

I.3 X wohnt Loc_{stat} Y = die Wohnung oder das Haus, das X zum Wohnen (WOHNEN I.1) nutzt, befindet sich in einer durch Loc_{stat} spezifizierten Beziehung zu Y.

Government Pattern

X = I	Y = II
1. N _{nom}	1. Loc _{stat}

D_{I.1}+D_{II.1} : Max wohnt bei Hamburg.
 : Max wohnt an der Nordsee.

I.4 *X wohnt in Y* = wie WOHNEN I.1, aber X nutzt Y nur vorübergehend als Unterkunft.

GP ↑ WOHNEN I.1

Wir wohnen für 6 Tage im Hotel Marbell.

Lexical Functions

Syn : *übernachten, absteigen* [Loc_{in} N]
 S₂ : *Quartier, Nachtquartier*
 S₂usual : *Hotel, Pension, Herberge*
 Incep : *sich einquartieren, absteigen* [Loc_{in} N]

II. *X wohnt in Y* = Eine Gruppe von Menschen X (ein Volk, eine ethnische Gemeinschaft, usw.) lebt ständig in einer Region Y, hat dort Siedlungen (SIEDLUNG I.1) errichtet und nutzt die in der Region Y vorhandenen Ressourcen (Land, Pflanzen, Tiere, Bodenschätze) für ihren Lebensunterhalt.

Lexical Functions

Syn : *siedeln, bewohnen*
 S₀ : ↑ SIEDLUNG II
 S₂ : *Siedlungsgebiet*
 Incep : *besiedeln, sich niederlassen*

Die Römer wohnten rund um das Mittelmeer.

III. Func₁: *In ihm wohnt unbändiger Ehrgeiz* = *Er ist unbändig ehrgeizig; Zwei Seelen wohnen, ach, in meiner Brust.*

WOHNSITZ *der Wohnsitz von X* = Gemeinde, in der X meldeamtlich erfasst ist

Government Pattern

X = 1
1. N _{gen} 2. von N _{dat} 3. A _{poss}

Lexical Functions

Oper₁ : *haben*
 FinOper₁+IncepOper₁ : *verlegen* [Loc_{ab} N] [Loc_{ad} N]

ANWESEN

Anwesen = größeres Grundstück, das mit Gebäuden bebaut ist, in der Regel mit einem (zum Wohnen genutzten) Haus.

HOTEL *X' Hotel* = GEBÄUDE, das einzelne Zimmer zeitweise zum Wohnen (WOHNEN I.4) an

eine Person X gibt, und von X dafür bezahlt wird. Ein Hotel bietet X außerdem die Möglichkeit, dort Mahlzeiten einzunehmen (wie ein Restaurant).

Government Pattern

X = I
1. N _{gen} 2. von N _{dat} 3. A _{poss}

- D_{I,1} : Peters Hotel
- D_{I,2} : das Hotel von Herrn Maier
- D_{I,3} : unser Hotel

Lexical Functions

- S₁ : *Gast, Hotelgast*
- Bon : *erstklassiges, teures // Luxushotel, Fünf-Sterne-Hotel*
- AntiBon : *billiges, schäbiges*
- Cap : *Direktor, Hoteldirektor*
- Equip : *Personal*
- IncepReal₁ : *absteigen in [~_{akk}]*
- PreparReal₁ : *buchen*
- Real₁ : *wohnen in [~_{dat}]*

D.1.2 Wortfeld *Siedlung*

SIEDLUNG, N, fem.

I.1 Ein mit Häusern zusammenhängend bebauter Ort, wo eine Gruppe von Menschen X zusammen lebt und als Gemeinschaft organisiert ist [=S_{res}(*siedeln*)]

Government Pattern

X = I
1. N _{gen} 2. A 3. N§

- D_{I,1} : *eine Siedlung der Römer*
- D_{I,2} : *eine merowingische Siedlung, eine keltische Siedlung*
- D_{I,3} : *Römersiedlung, Keltensiedlung*

Lexical Functions

- Syn : *Ort II, Ortschaft*
- SingS₁ : *Bewohner, Einwohner*
- Real₁ : *bewohnen, wohnen*
- FinReal₁ : *verlassen*

LiquFact	:	<i>auflassen</i>
CausFunc	:	<i>gründen</i>
LiquFunc	:	<i>zerstören, auslöschen</i>
S ₀ FinReal, S ₀ LiquFact	:	<i>Wüstung</i>

I.2 Eine zusammenhängender, abgeschlossener Teil einer SIEDLUNG I.1, der aus gleichartigen Häusern besteht.

Lexical Functions

Syn	:	<i>Kolonie</i>
-----	---	----------------

Arbeitersiedlung; Weißenhofsiedlung; Maikäfersiedlung

II. Prozess des Siedelns [= S₀(*siedeln*), S₀(*besiedeln*)]

Kultur, Mobilität, Migration und Siedlung von Juden im mittelalterlichen Europa; die Siedlung der Einwanderer im Westen Nordamerikas in der Mitte des vorigen Jahrhunderts

III. *Siedlung von X* = Gruppe von Tieren der Art X, die an einem Ort zusammenleben

Unsere Fotos zeigen, wie Ameisen gerade beginnen, eine Siedlung von Blattläusen anzulegen.

STADT, N, fem. größere SIEDLUNG I.1 mit eigener Verwaltung und Einrichtungen, die zur Versorgung der dort lebenden Menschen dienen

Lexical Functions

A ₀	:	<i>städtisch</i>
Gener	:	<i>Siedlung, Ort</i>
Dimin	:	<i>Städtchen</i>
Magn	:	<i>// Großstadt, Metropole</i>
Antimagn	:	<i>// Kleinstadt</i>
Cap	:	<i>Bürgermeister</i>
S ₁	:	<i>Bürgerschaft, Einwohnerschaft, Bevölkerung</i>
SingS ₁	:	<i>Bürger, Bewohner, Einwohner, Städter</i>
CausFunc	:	<i>gründen, erbauen, errichten</i>
Real ₁	:	<i>= ↑ SIEDLUNG I.1</i>
Loc _{in}	:	<i>in</i>
S ₀ FinReal, S ₀ LiquFact	:	<i>Geisterstadt</i>

KÖLN, EN, neutr., ohne Artikel STADT am Rhein in Nordrhein-Westfalen.

Lexical Functions

A ₀	:	<i>Kölner, kölnisch</i>
Bewohner von ~	:	<i>Kölner</i>
Dialekt, der in ~ gesprochen wird	:	<i>kölsch</i>

DORF, N, neutr. kleinere SIEDLUNG I.1, deren Bewohner (teilweise) Landwirtschaft betreiben.

Lexical Functions

siehe auch SIEDLUNG I.1

AntiMagn	: // Weiler
Cap	: Dorfältester, Dorfvorstand
SingS ₁	: Bewohner, Einwohner, Dorfbewohner, Dorfleute (Pl.), Dörfler
Loc _{in}	: in, auf



Potemkinsches Dorf
 Olympisches Dorf
 die Kirche im Dorf lassen

D.1.3 Wortfeld *Straße, Weg*

WEG

I.1 *der Weg für X von Y über W nach Z* = ein schmaler Geländestreifen, ohne Hindernisse passierbar, meist ausgebaut und befestigt, der dazu bestimmt ist, dass eine Person X vom Ort Y über W nach Z gelangen oder Objekte befördern kann.

Lexical Functions

Loc _{in}	: auf [\sim_{dat}]
Func ₂₃	: verbinden N _{akk} und N _{akk} , verbinden N _{akk} mit N _{dat}
Func ₂₄₃	: führen/gehen von N _{dat} über N _{akk} Loc _{ad} N
Func ₄	: passieren N _{akk} , führen/gehen durch N _{akk} , durchqueren N _{akk} , verlaufen über N _{akk}
Func ₂	: beginnen in N _{dat}
Func ₃	: enden in N _{dat}
A ₂	: erreichbar auf [\sim_{dat}], über [\sim_{akk}]
Real ₁	: nehmen N _{akk}
CausFunc	: anlegen, bauen
LiquFunc	: auflassen
LiquFunc _[zeitweise]	: sperren
A ₀ LiquFunc _[zeitweise]	: gesperrt

I.2 WEG I.1, unbefestigt oder nur gering befestigt, zur Benutzung durch Fußgänger oder geländegängige Fahrzeuge.

Lexical Functions

Magn	: breit
AntiMagn	: schmal // Pfad, Trampelpfad, Steig
Bon	: eben
AntiBon	: holprig, steil

II. *X' Weg von Y über W nach Z* = Route, auf der X von Y über W nach Z gelangt

Government Pattern

X = 1	Y = 2	Z = 3	W = 4
1. N _{gen} 2. von N _{dat} 3. A _{poss}	1. von N _{dat} 2. aus N _{dat}	1. über N _{dat} 2. via N _{dat} 3. [weitere ‚viale‘ direkt. Präp.]	1. in N _{akk} 2. nach N _{dat} 3. [weitere ‚finale‘ direkt. Präp.]

Lexical Functions

Syn	:	<i>Route</i>
Syn _U	:	<i>Reise, Zug¹</i>
Func ₂₄₃	:	<i>führen/gehen von N_{dat} über N_{akk} Loc_{ad} N</i>
IncepOper ₁	:	<i>den ~ einschlagen, sich auf den ~ machen</i>
Oper ₁	:	<i>auf dem ~ sein</i>
Bon	:	<i>direkt, kürzer</i>
AntiBon	:	<i>// Umweg</i>

der Weg der Störche von Europa nach Zentralafrika.

Wir waren auf dem Weg von Hannover über Fulda nach Nürnberg.

Sein Weg führte ihn von Paderborn nach Wien und von dort per Schiff durch das Schwarze Meer nach Palästina. Es war eine abenteuerliche Reise!

III. ... [weitere Bedeutungen von Weg]

STRASSE

I.1 wie WEG I.1, dazu geeignet und bestimmt, dass eine Person X ihn mit einem Fahrzeug benutzt; breiter als ein WEG I.2 und gut befestigt, meist asphaltiert.

Lexical Functions

wie ↑ WEG I.1 und zusätzlich:

Magn	:	<i>// Autobahn</i>
CausReal ₀	:	<i>dem Verkehr übergeben</i>

I.2 eine STRASSE I.1 innerhalb einer Ortschaft, einschließlich der Gehwege und der an ihr gelegenen Häuser, die dazu bestimmt ist, dass die Bewohner X einer Gruppe von Häusern oder Grundstücken diese erreichen können.

Lexical Functions

S ₁	:	<i>Anwohner, Anrainer</i>
Loc _{in}	:	<i>in</i>

I.3 eine STRASSE I.1, die dazu dient eine Region X zu erschließen, d.h. den An- und Abtransport von Gütern in X zu ermöglichen.

Lexical Functions

Syn	:	<i>Erschließungsstraße, Wirtschaftsstraße</i>
Func ₁	:	<i>erschließen N_{akk}</i>

II. Meerenge, als Wasserweg benutzbar.

Lexical Functions

Syn_▷ : *Meerenge*
Syn_▷ : *Wasserstraße*

III. unmittelbar aufeinanderfolgende aufsteigende Kartenwerte oder Nummern in einem Karten- oder Würfelspiel.

Große Straße, Kleine Straße



auf der Straße leben

auf die Straße setzen, auf die Straße werfen

auf der Straße liegen

auf offener Straße, auf hellichter Straße

auf die Straße gehen

E Korpora

E.1 Named Entities

E.1.1 Bewertung des CoNLL-2003-Shared-Task-Korpus

Die nachfolgende Liste enthält die ersten 200 Lokationen aus dem Testkorpus A des CoNLL-2003-SHARED-TASK. Durch 'fehlt' wurden fehlende, durch 'err' falsche Annotationen markiert. '??' kennzeichnet Annotationen, an deren Korrektheit Zweifel angebracht sind, die aber einer Diskussion bedürfen.

Von den 200 Lokationen sind 161 sicher korrekt annotiert. 13 sind falsch annotiert und 3 fehlen. 23 Annotationen sind fragwürdig.

Nr.	Toponym mit Kontext (Konkordanz)	Fehler/Kommentar
1	e Nordendler ein<S> NORDEND.<S> Einen Fotowettbew	
2	Leben und feiern im Nordend und mit den Nordendle	
3	Nordend und mit den Nordendlern" hat der Vorstand	err: Bewohner
4	2 Uhr), im Saal des Gehörlosenzentrums (Rothschil	err: Klassenbez.
5	Gehörlosenzentrums (Rothschildallee 16a) bekanntg	
6	e der "Nordendler", Lenaustraße 70 (Glauburgbunke	
7	r", Lenaustraße 70 (Glauburgbunker), geschickt we	
8	Stöffche kommt aus Hochstadt Bundesäppelwoifest	fehlt
9	muß auf Sauren aus Steinheim verzichten<S> HANAU	
10	nheim verzichten<S> HANAU.<S> Seit 1740 wurde im	
11	:<S> Die einzige in Steinheim verbliebene Keltere	
12	e Kelterei Höhl aus Hochstadt zurückgreifen, wenn	
13	Besucher/ innen im Schloßhof erwartet werden.<S>	err: Klassenbez.
14	Kirchengemeinde<S> BAD VILBEL.<S> Die Evangelisc	
15	n dauern, werden im Grünen Weg 4-6 abgehalten.<S>	
16	für die Bürger der Niederlande, daß sie ihr Depu	
17	cherbewegung in den USA hat ökologische Maßstäbe	
18	-Umwelt-Netzwerk in Großbritannien, geht einen Sc	
19	alden.<S> Allein in Großbritannien belasten tägli	
20	m Frauen-Seminar<S> FRIEDBERG.<S> Zu einem Wochen	
21	meldungen nimmt das Frauenzentrum unter Tel.<S> 0	
22	Sonntagsblattes" in Hamburg.<S> Schultz werde sei	
23	ietrich Sattler, in Hamburg mit.<S> Schultz war v	
24	n Pressedienstes in Genf, bevor er als Redakteur	
25	Fairneßabkommen<S> WETTERAUKREIS.<S> Die Wettera	
26	sprechungstermin in Friedberg vorschlagen, kündig	
27	<S> Im benachbarten Main-Kinzig-Kreis habe der SP	
28	entsprüche, und in Friedberg hätten die Sozialde	
29	s Sportabzeichen<S> KARBEN.<S> Aufgrund der große	
30	KSV Klein-Karben am Günter-Reutzel-Sportplatz um	
31	meterStrecke an der Nidda entlang zur Burg Bad Vi	
32	r Nidda entlang zur Burg Bad Vilbel fällt um 9 Uh	??: Klassenbez. zu Ort?
33	können samstags im Hallenbad Karben von 9 bis 10	??: Klassenbez. zu Ort?

Nr.	Toponym mit Kontext (Konkordanz)	Fehler/Kommentar
34	Die Riesenlampe aus Weißbleich ruht auf einem Hau	err ¹⁶⁵
35	Federohrringe aus Kenia - weg vom Massen-Image.	
36	kt stellt die "Casa Barcelona" dar.<S> Die besten	err ¹⁶⁶
37	lt.<S> Das Haus aus Barcelona wandert nach der Me	
38	r Messe weiter nach New York in das Museum of Mod	
39	ach New York in das Museum of Modern Art.<S> Eher	
40	die Nachbildung des Schlosses Neuschwanstein aus	?: Klassenbez. zu Ort?
41	sack (ebenfalls aus Sachsen) halten.<S> Und die S	
42	s Aserbaidtschanern "Türken" werden<S> "Wir haben	err: Bewohner
43	mpften Krisenregion Berg-Karabach.<S> Die Zahl de	
44	Unabhängigkeit von Aserbaidtschan und einer Verei	
45	d einer Vereinigung BergKarabachs mit Armenien be	fehlt ¹⁶⁷
46	g BergKarabachs mit Armenien begann ein blutiger	
47	orium der Ortschaft Saluga, etwa eine Stunde von	
48	anischen Hauptstadt Baku entfernt.<S> Etwa 1500 F	
49	charows Heimatstadt Schuscha.<S> Nur sechs Kilome	
50	von der Hauptstadt BergKarabachs, Stepanakert, e	
51	tadt BergKarabachs, Stepanakert, entfernt, war Sc	
52	kert, entfernt, war Schuscha für die aserbaidtscha	
53	das tiefer gelegene Stepanakert bis Anfang Mai un	
54	r und Soldaten nach Suscha einrückten.<S> Es habe	
55	che junge Frau, ist Ukrainerin.<S> Als Krankensch	err: Bewohner
56	in einer Klinik in Schuscha lernte sie ihren ase	
57	nen waren zu Anfang Armenierinnen, weil sie besse	err: Bewohner
58	ien aufgetaucht.<S> Armenische Nachbarn hätten da	err: Adjektiv
59	pucken und sie als "Türken" zu beschimpfen.<S> We	err: Bewohner
60	ck nach Hause, nach Schuscha."<S> Mein Großvater,	
61	aner miteinander in Berg-Karabach leben, wagt kei	
62	en Fernseheteams aus Schuscha.<S> Es scheinen älte	
63	r noch preiswert<S> WETTERAUKREIS.<S> Prall gefül	
64	sinteresse über die Wetterau hinausgeht, können s	
65	tag der "Entdeckung Amerikas" beschäftigen.<S> Mi	
66	svolkshochschule in Friedberg und Büdingen sowie	
67	le in Friedberg und Büdingen sowie bei allen Auße	
68	en kann man sich in Friedberg, Büdingen und in al	
69	sich in Friedberg, Büdingen und in allen anderen	
70	n.<S> -DOCSTART-<S> Japans Notprogramm Hauptsache	
71	wieder Geld<S> Als Nippons Finanzminister Tsutom	
72	en darauf will sich Tokio offenkundig nicht mehr	
73	örsenschock bewegte Nippons nervöse Regierung zu	
74	Steuerverzichte.<S> Tokios Logik scheint zu sein:	
75	imer Grundschule<S> KELKHEIM.<S> Heftige Angriffe	
76	rstag abend sind in Frankfurt mit 43,7 Litern Nie	
77	geplattert.<S> Über Seligenstadt waren die Wolken	
78	ehrheit wankt nicht Israels Regierung steht trotz	
79	lassen, versicherte Israels Innenminister Aryeh D	
80	verstummen, wonach Israels Ministerpräsident Yit	
81	on Von Armin Wertz (Jerusalem) in der vorigen Reg	
82	erreicht", feierte Israels Presse den Schritt. S	
83	chungen".<S> Und in Jerusalem pfeifen die Spatzen	
84	togenes Training<S> BAD VILBEL.<S> Autogenes Trai	

165 Orthographischer Fehler im Ortsnamen, richtig: *aus Weißblech*.

166 Die *Casa Barcelona* ist ein Kunstwerk.

167 *Vereinigung BergKarabachs* als ORG markiert.

Nr.	Toponym mit Kontext (Konkordanz)	Fehler/Kommentar
85	um 16.30 Uhr in der Christuskirchengemeinde, Grün	?: Organisation?
86	tuskirchengemeinde, Grüner Weg 4-6.<S> Anmeldunge	
87	enbildungsstätte in Friedberg entgegen.<S> -DOCST	
88	on Sportplätzen?<S> WETTERAUKREIS.<S> Was mit den	
89	er Sportflächen der Karl-Weigand-Schule in Florst	
90	l-Weigand-Schule in Florstadt und der Brüder-Grim	
91	n Florstadt und der Brüder-Grimm-Schule in Dorhei	
92	der-Grimm-Schule in Dorheim anfallen, daran schei	
93	einem Salzstock in Baden-Württemberg.<S> Die Grü	
94	essischen Salzstock Herfa-Neurode lagern, bis gee	?: Klassenbez. zu Ort?
95	nausflug führt nach Lahnstein<S> ROSBACH.<S> Die	
96	t nach Lahnstein<S> ROSBACH.<S> Die Stadt lädt al	
97	ner Tagesfahrt nach Lahnstein ein.<S> Um 9.30 Uhr	
98	ssen vom Marktplatz Ober-Rosbach, Marktplatz Rodh	
99	Rosbach, Marktplatz Rodheim und der Gaststätte We	
100	atz Rodheim und der Gaststätte Wehrheim in Nieder	?: Klassenbez. zu Ort?
101	tstätte Wehrheim in Nieder-Rosbach aus starten. S	
102	chiffsfahrt auf dem Rhein vorgesehen.<S> Anmeldun	
103	CE-Stellungnahme<S> HOFHEIM.<S> Wenn der Wallauer	
104	t, ab 19 Uhr in der Ländcheshalle tagt, gibt es n	
105	lungnahme der Stadt Hofheim zur ICE-Trasse Köln-R	
106	heim zur ICE-Trasse Köln-Rhein/Main.<S> Zuvor tre	??
107	zur ICE-Trasse Köln-Rhein/Main.<S> Zuvor treffen	??
108	E-Trasse Köln-Rhein/Main.<S> Zuvor treffen sich P	??
109	in um 18 Uhr in der Weiherstraße.<S> -DOCSTART- S	
110	uf dem Bauernhof<S> BAD VILBEL.<S> Für eine Herbs	
111	r- und Bauernhof in LichtenfelsNeukirchen sind no	
112	ich Ausflüge an den Edersee an.<S> Sauna und ein	
113	hletik-Nachwuchs<S> KARBEN.<S> Olympia ruft - unt	
114	Ost nach West, von Fechenheim bis Höchst, eröffn	
115	von Fechenheim bis Höchst, eröffnet ein neuer De	
116	einer kennt.<S> Auf Frankfurts Industriebauten fä	
117	eß es am Freitag im Technischen Rathaus, stehen u	
118	ßen vor:<S> Das ist Frankfurt am Main, im wahrste	
119	tatut für die Stadt Frankfurt und Sachsenhausen"	
120	Stadt Frankfurt und Sachsenhausen" aus dem Jahre	
121	n zu jener Zeit:<S> Griesheim, Bockenheim, Fechen	
122	Zeit:<S> Griesheim, Bockenheim, Fechenheim, Höchs	
123	esheim, Bockenheim, Fechenheim, Höchst am Main. S	
124	enheim, Fechenheim, Höchst am Main.<S> Innerhalb	?: komplex. Ortsname ¹⁶⁸
125	chenheim, Höchst am Main.<S> Innerhalb der Stadt	?: komplex. Ortsname
126	n von Cassella, die Hafenmühle im Osthafen	?:fehlt
127	, die Hafenmühle im Osthafen, die Bauer'sche Gieß	?:err
128	le im Osthafen, die Bauer'sche Gießerei, das hera	?: Organisation?
129	hungsmanöver vor<S> MAIN-KINZIG-KREIS.<S> Einem T	
130	von Schulleitern im Lahn-Dill-Kreis für das Staat	
131	chen Schulamtes des Main-Kinzig-Kreises in Hanau	
132	n-Kinzig-Kreises in Hanau ist, keine Rechtskraft,	
133	ahr 1991/92 auch im Main-Kinzig-Kreis weiter.<S>	
134	ndentafel begründet Wiesbaden mit der Einführung	
135	ie Teilspernung des Oeder Wegs zusammengekommen,	
136	bahnrichtung in der Finkenhofstraße - vorwiegend	
137	ße - vorwiegend aus Heinestraße, Bornwiesenweg, L	

168 Höchst am Main, vgl. in Zeile 118: Frankfurt am Main.

Nr.	Toponym mit Kontext (Konkordanz)	Fehler/Kommentar
138	nd aus Heinestraße, Bornwiesenweg, Lersnerstraße.	
139	aße, Bornwiesenweg, Lersnerstraße.<S> Dort würden	
140	Autofahrer von der Eschersheimer Landstraße ihre	
141	en Schleichpfad zum Oeder Weg suchen, die vom Esc	
142	Weg suchen, die vom Eschenheimer Tor her wegen de	
143	unsten der Orgel<S> BAD VILBEL.<S> Der Fördervere	
144	m-Klavierduo in die Alte Mühle ein.<S> Der Erlös	
145	kretär will weg aus Wiesbaden<S> WIESBADEN.<S> So	
146	eg aus Wiesbaden<S> WIESBADEN.<S> So kompliziert	
147	te ja "nur" weg aus Wiesbaden, sein Landtagsmanda	
148	"Dualen System" in Bonn arbeiten, wo auch sein f	
149	rde).<S> Weil er in Bonn schon arbeitet, ist das	
150	Sozialdezernent in Frankfurt.<S> Dort war er 198	
151	s Mitteln der Stadt Frankfurt, weshalb er entspre	
152	rdneten Sutter nach Bonn zu folgen.<S> Und selbst	
153	rr auf und über der Zeil, dessen vortrefflichste	fehlt
154	, die 80 Radler von Frankfurt nach Erfurt schickt	
155	von Frankfurt nach Erfurt schickte, um Geld für	
156	te, um Geld für die Erfurter Kinderklinik zu erst	?: Adj. + Klassbez.
157	nplotz kommt in die Stadthalle<S> HATTERSHEIM.<S>	err: Klassenbez.
158	n die Stadthalle<S> HATTERSHEIM.<S> "Räuber Hotze	
159	. September, in die Hattersheimer Stadthalle, Kar	?: Adj. + Klassbez.
160	sheimer Stadthalle, Karl-Ekkel-Weg.<S> Das Frankf	
161	es unter anderem im Kulturbüro, Tel.<S> 0 61 90/	err: Klassenbez.
162	räuber, der sich in Frankfurt als Polizist ausgeg	
163	ein 22jähriger aus Montenegro zu drei Jahren Haf	
164	Juli 1991 nachts im Gallusviertel einen 35 Jahre	
165	die S-Bahn zwischen Karben und Frankfurt, wo aben	
166	zwischen Karben und Frankfurt, wo abends ein 17 J	
167	Nationalpark<S> sp HANNOVER, 23. August.<S> Drei	
168	andesregierungen in Potsdam, Schwerin, Magdeburg	
169	erungen in Potsdam, Schwerin, Magdeburg und Hanno	
170	Potsdam, Schwerin, Magdeburg und Hannover appell	
171	erin, Magdeburg und Hannover appelliert, die Plän	
172	die Pläne für einen Nationalpark Elbtal-Auen nich	?: Klassenbez. zu Ort?
173	Naturerbe jetzt in Hannover.<S> Die Kernzone sol	
174	fassen.<S> Die fürs Elbtal typischen Grünflächen	
175	al am Rednerpult<S> WETTERAUHKREIS.<S> Männer woll	
176	nen unter den 50 im Niddaer Bürgerhaus versammelt	?: Adj. + Klassbez.
177	fstellen:<S> in der Herrnstraße 16 in Offenbach,	
178	r Herrnstraße 16 in Offenbach, Telefon 069/ 81181	
179	811811, und in der Hauptstraße 32-36 in Dreieich	
180	auptstraße 32-36 in Dreieich, Telefon 06103/ 6494	
181	st auch Träger von "Bella Vista" in der Berliner	
182	Bella Vista" in der Berliner Straße 118, Telefon	
183	Drogenabhängige aus Stadt und Kreis Offenbach auf	err: Klassenbez.
184	aus Stadt und Kreis Offenbach auf, ein Drittel da	
185	es Projekt:<S> der "Bauhof" in Offenbach, wo Süch	?: Organisation?
186	<S> der "Bauhof" in Offenbach, wo Süchtige in ein	
187	inrichtung ist die "Stiftung Waldmühle" in Mühlhe	?: Organisation?
188	ftung Waldmühle" in Mühlheim.<S> Dort warten Drog	
189	ittherapie.<S> Die "Waldmühle" nimmt Leute aus ga	?: Organisation?
190	immt Leute aus ganz Hessen auf und zählt daher ni	
191	htige aus dem Kreis Offenbach.<S> -DOCSTART-<S> N	
192	nd im Bürgerhaus<S> NEU-ANSPACH."<S> Neue Grundsc	

Nr.	Toponym mit Kontext (Konkordanz)	Fehler/Kommentar
193	ugust, um 20 Uhr im Bürgerhaus Neu-Anspach in der	?: Klassenbez. zu Ort?
194	Neu-Anspach in der Gustav-Heinemann-Straße 3 mit	
195	ter der Grundschule Neu-Anspach Horst Eufinger, d	?: Klassenbez. zu Ort?
196	s Schulangebotes in NeuAnspach durch eine betreut	
197	nes ANC-Informanten Südafrikas Präsident droht mi	
198	hannes Dieterich<S> JOHANNESBURG, 23. August.<S>	
199	g in der Hafenstadt Port Elisabeth die Fahndung n	
200	aatsanwaltschaft in Port Elisabeth inzwischen wie	

E.2 Allgemeine Korpora

Die untenstehende Tabelle gibt einen Überblick über die für verschiedene Experimente und lexikographische Arbeiten verwendeten deutschsprachigen Korpora. Auch viele der Beispiele im Text entstammen diesen Quellen.

Korpus	Größe (Mio Wortformen)
Süddeutsche Zeitung	800
<i>andere Tageszeitungen</i>	430
WWW (gecrawlte Seiten)	2 850
WIKIPEDIA.DE und Schwesterprojekte	400
dewac (WACKY)	1 360
<i>andere</i>	410
	6 250

Tabelle E.2: Größe der verwendeten Korpora

Lebenslauf

ZUR PERSON

Name Sebastian Johannes Nagel
geboren 15. September 1970 in München
Familienstand mit Freundin, zwei Kinder

AUSBILDUNG

Sept. 1977 – Juli 1981 Volksschule an der Stielerstraße
Sept. 1981 – Juli 1990 Gymnasium: Theresiengymnasium (humanistisch) in München
Abschluss Abitur
Sept. 1990 – Sept. 1992 Berufsausbildung zum Zimmerer – Firma Heinrich GmbH
Abschluss Gesellenprüfung
Okt. 1992 – Dez. 1993 Zivildienst: Werkstatt Pronova, Einrichtung (Schreinerei) zur Wiedereingliederung psychisch Kranker ins Arbeitsleben

STUDIUM

April 1994 – März 2001 LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN (LMU)
Hauptfach Slavische Philologie (Linguistik)
Schwerpunkte Semantik, Lexikologie, kognitive Linguistik, Aspekt
Nebenfächer Vor- und Frühgeschichte, Völkerkunde (Ethnologie)
Magisterarbeit „Zur Semantik der Grundfarbadjektive im Russischen und Tschechischen“
Abschluss M.A.
Sept. – Okt. 1995 MOSKAUER LINGUISTISCHE UNIVERSITÄT (MGLU)
Sprach- und Dolmetschkurs
Sept. 1997 – Jan. 1998 UNIVERSITÄT KAZAN' (KGU)
Schwerpunkte Sprachkurse Russisch und Tatarisch, Aspekt im Russischen, Ethnographie und Archäologie des mittleren Wolgagebiets
Okt. 1999 – Jan. 2000 KARLSUNIVERSITÄT PRAG (Univerzita Karlova)
Schwerpunkte Sprachkurse, Korpuslinguistik, lexikalische Semantik, Archäologie der Slaven
seit April 2001 LMU, Centrum für Informations- und Sprachverarbeitung (CIS)
April 2001 – Sept. 2002 Aufbaustudium Computerlinguistik
Abschlussarbeit „Formenbildung im Russischen: Formale Beschreibung und Automatisierung für das CISLEX-Wörterbuchsystem“
seit Okt. 2002 Promotion in Computerlinguistik bei Prof. Dr. Franz Guenther
Thema: „Lokale Grammatiken zur Beschreibung von lokativen Sätzen“
Prüfer: Prof. Dr. Franz Guenther, Prof. Dr. Klaus U. Schulz, Prof. Dr. Ulrich Schweier (Nebenfach Slavistik)

PRAKTIKA

Sommer 1996, 1999 Novgorod, Teilnahme an archäologischen Ausgrabungen
Sept. 1998 – April 1999 Osteuropa-Institut, Digitalisierung

BERUFLICHE LAUFBAHN

April 2001 – Sept. 2003 wissenschaftliche Hilfskraft am CIS, LMU München
Dez. 2004 – Febr. 2008 wiss. Angestellter am CIS
seit April 2008 Softwareentwickler, exorbyte GmbH, Konstanz

München, den 22. Dezember 2009