



universität
wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„Automatische Gewinnung österreichischer Terminologie
aus offiziellen Antragsformularen der Einwanderungs-
und Niederlassungsbehörden in Österreich“

verfasst von / submitted by

Anchalin Ruckwatin

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of
Master of Arts (MA)

Wien, 2021 / Vienna 2021

Studienkennzahl lt. Studienblatt /
degree programme code as it appears on
the student record sheet:

UA 070 342 331

Studienrichtung lt. Studienblatt /
degree programme as it appears on
the student record sheet:

Masterstudium Translation UG2002
Englisch
Deutsch

Betreut von / Supervisor:

Ass. -Prof. Mag. Dr. Dagmar Gromann, BSc

Inhaltsverzeichnis

Abbildungsverzeichnis	iii
Tabellenverzeichnis	iv
0 Einleitung	1
1 Die deutsche Sprache und ihre österreichische Varietät	6
1.1 Deutsch als plurizentrische Sprache	6
1.2 Die österreichische Nationalvarietät	9
1.2.1 Sprachgeografische, sprachhistorische und sprachsoziologische Voraussetzungen der Sprachsituation in Österreich	10
1.2.2 Austriazismen – Besonderheiten der österreichischen Varietät	13
1.2.3 Linguistische Eigenheiten der österreichischen Varietät	14
2 Verwaltung und Sprache	19
2.1 Die öffentliche Verwaltung	19
2.2 Das sprachliche Handeln im Verwaltungsverfahren	20
2.2.1 Verwaltungssprache in der Gegenwart	21
2.2.2 Österreichische Verwaltungssprache	23
2.2.3 Texte im Verwaltungsbereich	28
3 Formulare als Verwaltung-Bürger-Schnittstelle	31
3.1 Historischer und juristischer Hintergrund von Formularen	31
3.2 Zwecksetzung und Struktur von Formularen	33
3.3 Formular und Verwaltungssprache in der Diskussion	34
4 Migration und behördliches Verfahren	37
4.1 Motivation und Muster von Wanderungsbewegungen	37
4.2 Einwanderung und Aufenthalt in Österreich	39
4.3 Verständigung im Niederlassungs- und Aufenthaltsverfahren	45
5 Maschinengestützte Terminologiegewinnung als Teil der praktischen Terminologearbeit	47
5.1 Grundlagen der Terminologie	47
5.2 Grundelemente der Terminologielehre	50
5.3 Terminologiegewinnung	53
5.4 Maschinengestützte Terminologieextraktion	55
5.5 Terminologieextraktionssysteme	57
5.6 Verfahren der maschinengestützten Terminologieextraktion	59
5.6.1 Statistische Extraktionsverfahren	60
5.6.2 Linguistische Extraktionsverfahren	62
5.6.3 Hybride Extraktionsverfahren	64
5.6.4 Grenzen der maschinengestützten Terminologieextraktion	66
5.7 Evaluierung der Terminologieextraktion	69
5.7.1 Grundlagen der Termextraktionsevaluierung	69
5.7.2 Quantitative Evaluierung der Terminologieextraktion	71

6	Relevante Forschungsarbeiten	76
7	Praktische Evaluierung der Terminologieextraktionsysteme.....	80
7.1	Untersuchungsmethodologie.....	81
7.1.1	Beschaffung des Dokumentationsmaterials	82
7.1.2	Erstellung der Referenztermliste	83
7.1.3	Praktische maschinengestützte Terminologieextraktion	88
7.1.3.1	AntConc	88
7.1.3.2	Tilde Wrapper System for CollTerm (TWSC)	90
7.1.3.3	Termsuite	92
7.1.4	Quantitative Evaluierung der Extraktionsergebnisse.....	94
7.1.4.1	Term Evaluator	94
7.1.4.2	Evaluierungsprozess	96
7.2	Terminologieextraktionsergebnisse	99
7.3	Ergebnisse der Leistungsevaluierung	102
7.4	Diskussion.....	106
7.4.1	Herausforderungen bei der RTL-Erstellung	107
7.4.2	Evaluierungsergebnisse.....	109
7.4.3	Schlussfolgerung.....	112
8	Fazit	115
9	Quellenverzeichnis	118
	Anhang I.....	138
	Anhang II	139
	Abstract (Deutsch).....	155
	Abstract (English)	156

Abbildungsverzeichnis

Abbildung 1: Varietäten der deutschen Standardsprache	7
Abbildung 2: Österreichische und bundesdeutsche Varietäten	13
Abbildung 3: Semiotisches Dreieck	50
Abbildung 4: Ablauf der hybriden Extraktionsverfahren	65
Abbildung 5: Maße zur quantitativen Evaluierung der Termextraktion	72
Abbildung 6: Typisches und ideales Verhältnis von Recall und Precision.....	73
Abbildung 7: Ermittlung des F1-Maßes	74
Abbildung 8: Zusammenfassung von Daten anhand einer Pivot-Tabelle.....	85
Abbildung 9: Verifizierung der manuell extrahierten Termkandidaten.....	87
Abbildung 10: Stoppwortliste in Werkzeug-Einstellungen – AntConc 3.5.9	89
Abbildung 11: Extraktionsablauf von TWSC (vgl. Pinnis et al. 2012: 195)	91
Abbildung 12: Voreinstellung der Terminologieextraktion – TWSC (TaaS).....	92
Abbildung 13: Konfiguration der linguistischen Anreicherung (TermSuite)	93
Abbildung 14: Pipeline-Einstellungen (TermSuite).....	94
Abbildung 15: Die Auswertungsoberfläche (Term Evaluator)	95
Abbildung 16: Annotation-Listenansicht (Term Evaluator)	96
Abbildung 17: Datei importieren und Voreinstellungen (Term Evaluator)	98
Abbildung 18: Annotationsspeicher (Term Evaluator).....	99
Abbildung 19: Wortliste-Ergebnis – AntConc 3.5.9.....	99
Abbildung 20: N-Gramm-Ergebnis – AntConc 3.5.9	100
Abbildung 21: N-Gramm-Ergebnis sortiert nach Ende des Worts	101
Abbildung 22: Termkandidatenliste von TWSC (TaaS).....	101
Abbildung 23: Termkandidaten-Ansichtsbereich (TermSuite).....	102
Abbildung 24: Extraktionsleistung bei Einwortbenennungen	103
Abbildung 25: Extraktionsleistung bei Zweiwortbenennungen	104
Abbildung 26: Extraktionsleistung bei Dreiwortbenennungen	105
Abbildung 27: Extraktionsleistung bei Vierwortbenennungen	105
Abbildung 28: Extraktionsleistung bei Fünfwortbenennungen	106

Tabellenverzeichnis

Tabelle 1: Österreichische und deutsche syntaktische Varianten	16
Tabelle 2: Prozentuale Gesamtbewertung der automatischen Terminologieextraktionswerkzeuge	106

0 Einleitung

Im Gegensatz zu anderen Fachgebieten ist die Verwaltungsterminologie nicht auf die ausschließliche Kommunikation oder Verwendung einer bestimmten Gruppe von Fachleuten beschränkt, sondern erstreckt sich über diese hinaus auf alle BürgerInnen im Staat, da eine der Funktionen der öffentlichen Verwaltung darin besteht, deren Rechte und Pflichten zu regeln (vgl. Chiocchetti et al. 2013a: 11). Da Recht und Verwaltung somit nahezu alle Lebensbereiche von BürgerInnen durchdrungen haben, wird die Beschäftigung mit Rechts- und Verwaltungstexten unumgänglich, oder wie Öhlinger es formuliert: „Von der Wiege bis zur Bahre: Formulare, Formulare – das ist eine Erfahrung, der heute in der Tat niemand entgeht“ (Öhlinger 1986: 26). Folglich ist jeder und jede in Österreich dazu verpflichtet, sich zumindest einmal im Leben direkt oder indirekt mit einem oder mehreren Verwaltungsorganen zu beschäftigen.

Insbesondere die Einwanderungsbehörde jedes österreichischen Bundeslands hat regelmäßig mit Personen zu tun, die nach Österreich eingewandert sind und aus verschiedenen Gründen, z. B. Studium, Auslandsbeschäftigung, Forschungsprojekten und vielem mehr, eine Aufenthaltsgenehmigung erhalten möchten. In erster Linie handelt es sich bei diesem Verwaltungsbereich um AntragstellerInnen, die aus anderen Herkunftsländern stammen und deshalb zumeist andere Muttersprachen besitzen. So variiert bei diesen Zugewanderten das Niveau der deutschen Sprache, wobei einige gering bis sogar keinerlei Deutschkenntnisse beherrschen. Laut dem jährlichen statistischen Bericht über Zuwanderung und Integration der Statistik Austria (2019), dessen Erstellung durch das Bundesministerium für Europa, Integration und Äußeres und den Asyl-, Migrations- und Integrationsfonds kofinanziert wurde, weisen von allen ZuwanderInnen mit einer anderen Muttersprache als Deutsch 55 % fortgeschrittene Deutschkenntnisse auf, während die restlichen 45 % über entweder durchschnittliche oder geringe bis keine Deutschkenntnisse verfügen.

Ungeachtet dieser Tatsachen sind die Antragsformulare, die eine der entscheidenden Dokumentenkategorien in diesem Prozess der Erteilung von Aufenthaltstiteln darstellen, ausschließlich in deutscher Sprache verfügbar und zeichnen sich zudem überwiegend durch die österreichspezifische Verwaltungsterminologie aus. Spezifisch haben sich die antragstellenden Personen im Laufe des behördlichen Verfahrens mit Formularen und weiteren Verwaltungsschreiben im sogenannten österreichischen Amtsdeutsch auseinanderzusetzen, das spezifische Begriffe und Phraseologie umfasst, die wiederum abstrakte Konzepte bezeichnen und durch einzigartige syntaktische, stilistische und vor allem lexikalische Präferenzen

gekennzeichnet sind. Insgesamt handelt es sich um Sonderausdrücke, die in der alltäglichen Sprache unüblich sind. Dies hat zur Folge, dass diese domänenspezifische Terminologie zu Sprachbarrieren und Schwierigkeiten für LaienanwenderInnen führen kann.

Angesichts der Aufforderungen zum Engagement der Behörden in der bürgernahen Konzipierung der Formulare ist anzumerken, dass nicht gleichermaßen um Verständlichkeit für nichtdeutschsprachige RezipientInnen bemüht ist. Für diese Laiengruppe wird zum Problem, dass die Verwaltungssprache als Fachsprache „eine Sprache für Menschen ist, die bereits verstanden haben, und deshalb oft darauf verzichtet, einen Verstehenshorizont aufzubauen“ (Ebert 2011: 15). In Anbetracht dieses Status quo soll in der vorliegenden Masterarbeit ein Beitrag zur Verständigungsoptimierung geleistet werden, indem sprachliche und terminologische Fragen erörtert werden. Untersucht wird, wie verschiedene Methoden der maschinengestützten Terminologieextraktion mit Texten in österreichischer Verwaltungssprache umgehen, die sich auf ein spezifisches administratives Gebiet beziehen. Konkret lässt sich die Forschungsfrage wie folgt formulieren:

Welches automatische Terminologieextraktionswerkzeug schneidet bei der Verarbeitung von fachspezifischer österreichischer Verwaltungsterminologie in Bezug auf Einwanderung und Aufenthalt vergleichsweise am besten ab?

Zu diesem Zweck werden drei Computerwerkzeuge mit unterschiedlichen Terminologieextraktionsverfahren, nämlich statistikbasierten, linguistischen und hybriden Ansätzen, auf ein Korpus angewendet, das aus sämtlichen Formularen für die Beantragung von Aufenthaltsberechtigungen besteht, die auf den Websites österreichischer Einwanderungsbehörden gesammelt wurden. Dementsprechend besteht die Zielsetzung in der Leistungsevaluierung der unterschiedlichen Terminologieextraktionswerkzeuge, um einen möglichst effektiven Ansatz zur (semi-)automatisierten Terminologiegewinnung für Texte in der österreichischen Verwaltungssprache zum Niederlassungs- und Aufenthaltsrecht in Österreich und verwandten Teilbereichen zu ermitteln. Dabei gilt es in erster Linie die folgende Hypothese, die sich anhand der theoretischen Auseinandersetzung im Rahmen der vorliegenden Arbeit begründen ließ, zu verifizieren:

Ein hybrider Ansatz zur Terminologieextraktion, der statistische und linguistische Methoden kombiniert, liefert im Vergleich zu rein statistischen und linguistischen Ansätzen die beste Leistung bei der Verarbeitung domänenspezifischer österreichischer Verwaltungsterminologie bezüglich Einwanderung und Aufenthalt.

Die Erkenntnisse, die durch die praktische Untersuchung in der Arbeit erlangt werden, könnten für künftige Beteiligte bei der Ausarbeitung potenzieller Terminologiarbeit zu diesem fachspezifischen Thema von Nutzen sein. Damit soll eine Grundlage geschaffen werden, wodurch sich sowohl menschliche als auch maschinengestützte Übersetzungen der bestehenden Antragsformulare sowie anderer relevanter Verwaltungsschreiben in weitere Sprachen realisieren lässt. Dadurch wird es den RezipientInnen ermöglicht, Sprachbarrieren zu überwinden, vollen Zugang zu Informationen zu erhalten und diese effizient nutzen zu können.

Um den wesentlichen theoretischen Rahmen abzustecken, wird im ersten Kapitel das Konzept der sprachlichen Plurizentrität erläutert, wobei Deutsch, das in mehreren Ländern als Amtssprache oder als eine der Amtssprachen verwendet wird und sich je nach Region (bzw. Sprachzentren) eigene Spracheigenheiten entwickelt haben, als plurizentrische Sprache zu betrachten ist. So wird die Existenz der Standard- bzw. Nationalvarietät Österreichs begründet, auf deren Entstehung und Entwicklung unter geschichtlichen, regionalen und gesellschaftlichen Einflüssen in Unterkapitel 1.2 eingegangen wird. Außer den einzigartigen Hintergründen weist die österreichische Standardsprache eigene linguistische Besonderheiten auf, die sogenannten Austriazismen, die ausführlich im Abschnitt 1.2.3 geschildert werden.

Im zweiten Kapitel steht die österreichische Nationalvarietät im Zusammenhang mit der öffentlichen Verwaltung im Vordergrund. Um die Bedeutung der Verwaltungssprache als Instrument im Verständigungsprozess nachvollziehbar darzustellen, wird zunächst ein Überblick der Staatsverwaltung gegeben, bevor die Austauschbeziehung zwischen Behörden und BürgerInnen in Verwaltungsverfahren aus kommunikativer Sicht betrachtet wird, wobei sich die Verwaltungssprache aufgrund ihrer funktionsbedingten Eigenschaften von der Fachsprache anderer Gebiete unterscheiden lässt. Diese werden anschließend hinsichtlich einzelner linguistischer Merkmale betrachtet, die kennzeichnend für die österreichische Verwaltungssprache sind. Schließlich erfolgt die Charakterisierung der Verwaltungstexte, die als Basis einer eingehenden Betrachtung der Textsorte Formulare dient, der häufigsten Erscheinungsform der Verwaltungssprache.

Das Letztgenannte ist Gegenstand des dritten Kapitels. Hier werden der Ursprung dieser Textsorte sowie ihre primäre Funktion erörtert, auf die sich die Gestaltung der gegenwärtigen Formulare zurückführen lässt. Dabei werden die grundlegenden textsortenspezifischen Elemente aufgeführt, die durch die Zwecksetzung von Formularen bestimmt werden. Schließlich werden Formulare und Verwaltungssprache in den Kontext der bisherigen

Forschungsarbeiten eingebettet, mit denen das gemeinsame Hauptziel der Verständlichkeitsoptimierung und Bürgernähe angestrebt wird.

Das vierte Kapitel umfasst das Thema Migration und die damit einhergehenden Verwaltungsverfahren bei zuständigen Behörden. Zunächst wird ein Überblick über das Phänomen im Allgemeinen vermittelt, indem Faktoren bzw. Ursachen dargestellt werden, die Menschen zur Aus- und Einwanderung motivieren und dabei verschiedene Typen von Migration veranlassen. Österreich ist aus historischer Sicht auch vom Phänomen der Zuwanderung betroffen, besonders nach dem zweiten Weltkrieg. Immigration erfordert den Erwerb der entsprechenden Kategorie des Niederlassungs- und Aufenthaltsrechts, was den Kontakt mit zuständigen Einwanderungsbehörden nach sich zieht. Die damit verbundene Austauschbeziehung der Beteiligten des Verwaltungsprozesses wird nach der Darstellung der wesentlichen Aufenthaltsberechtigungen erläutert.

Im fünften Kapitel werden die wesentlichen Grundlagen der Terminologielehre und der Terminologearbeit beleuchtet, wobei der Fokus aufgrund der vorgesehenen Zielsetzung primär auf der Terminologieextraktion und der Evaluierung der -Extraktionssysteme liegt. Dieses Kapitel umfasst somit die Erläuterung der maschinengestützten Terminologiegewinnung und die unterschiedlichen Verfahren zur Durchführung dieses Prozesses. Hinsichtlich der Bewertung der Extraktionsleistung wird die quantitative Methode zur Evaluierung der Extraktionsergebnisse dargestellt, bei der in erster Linie auf die Berechnung von Recall-, Precision- und F1-Maßen zurückgegriffen wird. Daraufhin wird im sechsten Kapitel Überblick über die relevanten Forschungsarbeiten zum Thema Evaluierung von Terminologieextraktionssystemen gegeben, an denen sich die Methodologie im Rahmen der praktischen Untersuchung dieser Arbeit ausrichtet.

Das siebte und letzte Kapitel ist der praktischen Untersuchung gewidmet. Im Rahmen dessen wird das Vorgehen bei der automatisierten Terminologieextraktion mithilfe von drei Computerwerkzeugen und der anschließenden quantitativen Evaluierung beschrieben. Bei der Terminologieextraktion wurden drei Systeme eingesetzt: das korpusanalytische Programm AntConc, die Online-Plattform zur linguistisch basierten Identifizierung von Terminologie Tilde Wrapper System for CollTerm (TWSC) und das hybride Tool TermSuite. Diese sollen jeweils unterschiedliche Methoden zur automatisierten Terminologiegewinnung repräsentieren. Die Qualität der extrahierten Termkandidaten wird anhand ihrer Korrektheit und terminologischen Relevanz durch menschliche RaterInnen mithilfe des Annotationstools Term Evaluator bewertet. Um die Effizienz des jeweiligen Terminologieextraktionswerkzeugs zu beurteilen, erfolgt die Berechnung von Recall- und Precision-Werten gegenüber der

Referenztermliste, die manuell im Vorfeld der automatisierten Terminologiegewinnung erstellt wurde und die als ein Goldstandard für die Evaluierung dienen soll. Eine anschließende Berechnung der durchschnittlichen Recall- und Präzisionswerte gibt Aufschluss über die Gesamtleistung der einzelnen Tools. Dadurch wird wiederum jenes mit der höchsten Effizienzbewertung ermittelt.

1 Die deutsche Sprache und ihre österreichische Varietät

Im ersten Kapitel wird sich mit der deutschen Sprache und ihren verschiedenen nationalen Ausprägungen innerhalb des deutschsprachigen Raums der Gegenwart auseinandergesetzt. Diese Grundlage soll einer eingehenden Betrachtung der Sprachsituation in Österreich dienen, die für die vorliegende Arbeit von Bedeutung ist. Daher wird zunächst eine Übersicht über den Aufbau der deutschen Schrift- und Standardsprache erstellt, die sich aus Subsystemen (in Form von Varietäten) zusammensetzt. Dies wird anhand des Begriffs Plurizentralismus erläutert, durch den die Existenz der in Österreich verbindlichen Varietät der deutschen Sprache untermauert wird. Um die Besonderheiten des österreichischen Deutsch nachzuvollziehen, wird anschließend auf grundlegende Aspekte wie den geschichtlichen Hintergrund eingegangen. Zudem sollen die Spracheigenheiten der österreichischen Standardvarietät beleuchtet werden, allerdings mit dem Fokus auf der österreichischen Schriftsprache, da diese einen Teil des Untersuchungsgegenstands im empirischen Teil dieser Arbeit bildet.

1.1 Deutsch als plurizentrische Sprache

In der Gegenwart umfasst die Region mit Deutsch als Amtssprache insgesamt sieben Staaten, d. h. Deutschland, Österreich, die Schweiz, Luxemburg, Liechtenstein, Belgien und Italien. Allerdings besitzt sie nur in den ersten fünf genannten Ländern nationalen Status (vgl. Clyne 1992: 133). Sie gilt als alleinige Staats- bzw. Amtssprache auf dem gesamten Gebiet in Deutschland, Österreich und Liechtenstein, während sie in der Schweiz sowie in Luxemburg als kooffizielle Amtssprache neben anderen Sprachen wie Französisch, Italienisch und Rätoromanisch bzw. Französisch und Letzeburgisch verwendet wird (vgl. Ammon et al. 2011: XXXI). Hingegen ist Deutsch lediglich eine regionale Amtssprache in Belgien und Italien und weist dadurch eine deutlich eingeschränkte Geltung auf (vgl. Ammon 1995a: 12). Da sich in diesen Nationen eigene länderspezifische Spracheigenheiten entwickelt haben, zählen sie zu den Zentren der deutschen Sprache, und somit wird das Deutsche als eine sogenannte plurizentrische Sprache betrachtet.

Dementsprechend werden Sprachen als plurizentrisch bezeichnet, wenn deren Gebrauch sich über mehr als ein Land verbreitet und sie aus geschichtlichen bzw. geografischen Gründen mehrere interagierende Zentren aufweisen, in denen sich jeweils eine nationale Standardvarietät mit eigener Kodifizierung oder eigenen Sprachbesonderheiten herausgebildet hat (vgl. de Cilia 1995: 6; Ammon et al. 2011: XXXI). Der Begriff des Plurizentralismus wurde

erstmals von Kloss (1978) vorgestellt und von Clyne (1984, 1992, 1995) als Grundlage für die Klassifizierung der nationalen Varietäten des Deutsch im deutschsprachigen Raum angewandt.

Dem plurizentrischen Modell zufolge ist eine Sprache „[e]ine Gesamtsprache, die mindestens eine Standardvarietät als Subsprache (Subsystem) enthält. Zwischen Standardsprache und Standardvarietät besteht also eine Menge-Element-Beziehung“ (Ammon 1995b: 112), das heißt, dass die Varietäten der einzelnen deutschsprachigen Länder, vor allem in Deutschland, in Österreich und im größten Teil der Schweiz, die Bestandteile darstellen, welche die deutsche Standardsprache der Gegenwart bilden (siehe Abb. 1). Da die Standardvarietäten üblicherweise in Verbindung mit spezifischen Nationen stehen bzw. mit diesen identifiziert werden, sind sie mit dem Begriff nationale Standardvarietäten oder Nationalvarietäten aufzufassen (vgl. Clyne 1988: 359; Ammon 2005: 31).

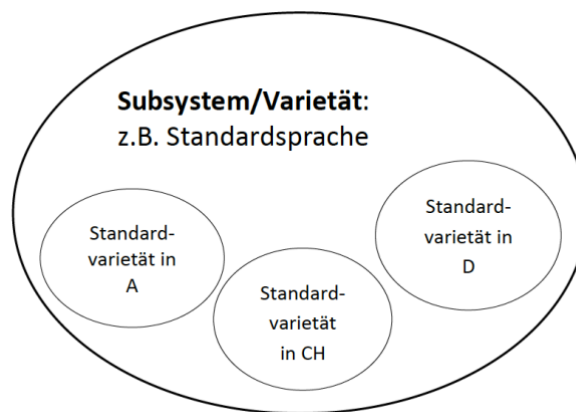


Abbildung 1: Varietäten der deutschen Standardsprache (adaptiert von Ammon 1995b: 112)

Allerdings werden nicht alle Ausprägungen in einzelnen Zentren der deutschen Sprache trotz ihrer besonderen länderspezifischen Sprachvarianten als Standardvarietäten anerkannt. In dieser Hinsicht unterscheidet Ammon (2004: XXXI) zwischen Voll- und Halbzentren plurizentrischer Sprachen. Dabei besteht das wesentliche Kriterium darin, dass in einem Vollzentrum die Kodifizierung und teilweise die amtliche Institutionalisierung der standardsprachlichen Merkmale seiner Varietät, z. B. die Schreibweise (Orthografie), durchgeführt werden. Demgemäß sind Liechtenstein, Luxemburg, die autonome Provinz Bozen in Italien und die deutschsprachige Gemeinde in Ostbelgien als nationale Halbzentren zu bewerten, in denen es an offiziellen Instanzen mangelt, die über die Festlegung der Standardvarietät in Form von beispielsweise Regelwerken erst entscheiden (vgl. Ammon 1995b: 113). Hingegen verfügen Deutschland, Österreich und die deutschsprachige Schweiz über mehrere vom Staat approbierte Kodifizierungen, deren Standardvarietäten, z. B. das

„Wörterbuch der Standardsprache in der deutschen Schweiz“ von Bickel und Landolt (2012) und das „Österreichische Wörterbuch“ (erstmalig erschienen 1951). So gelten sie als die einzigen nationalen Vollzentren der deutschen Sprache. Infolgedessen beruht die Kategorisierung der gegenwärtigen deutschen Standardsprache im Wesentlichen auf der Aufteilung in drei nationale Varietäten: das deutsche (oder deutschländisches) Standarddeutsch, das österreichische Standarddeutsch und das Schweizerdeutsch (oder Schweizer Hochdeutsch), ebenso kollektiv als die DACH-Varietäten bekannt. (vgl. Ammon 1995a: 229ff; Dollinger 2019: 3).

Das ursprüngliche Paradigma des Plurizentralismus wurde im Laufe des 20. Jahrhunderts als realitätsfern und revisionsbedürftig empfunden. Ursprünglich ging es von der Grenzziehung zwischen den deutschsprachigen Nationen aus, obwohl in der Sprachwirklichkeit sowohl Überschneidungen des Sprachgebrauchs über die Staatsterritorien hinweg als auch die Uneinheitlichkeit der jeweiligen deutschen Sprachvarietäten bestehen (Vgl. Muhr 1997: 42; Wiesinger 2008a: 6). So wurde der alternative Begriff des Pluriarealismus von Wolf (1994: 74) vorgeschlagen, der den „zugrundeliegende[n] areale[n] Gliederung[en] des Sprachgebietes“ entspricht (Pohl 1997: 69). Auch Scheuringer (1996) äußert sich grundsätzlich gegen die Anwendung des Begriffs plurizentrisch für die Beschreibung der deutschen Sprache und zieht die Bezeichnung pluriareale Sprache vor, denn plurizentrisch weist auf die Existenz von „nationalen oder staatlich einheitlichen Varietäten des Deutschen in relativ strikter Abgrenzung voneinander“ hin, „die es so nicht gibt“ (Scheuringer 1996: 152).

Die Hauptunterschiede zwischen den zwei Paradigmen bestehen in den Sprachbeschreibungsvorgängen, die unterschiedliche Einstellungen hinsichtlich des Sprachgebrauchs und der Sprachvariation als Ausgangspunkt haben (vgl. Scheuringer 1996: 150). Einerseits gilt die deutsche Standardsprache beim plurizentrischen Paradigma, wie erwähnt, als die „Schnittmenge der drei Vollvarietäten“ (DACH), wobei die Varietäten des Deutschen durch das staatliche Territorium bestimmt werden. Zudem wird dabei der Sprachgebrauch innerhalb der einzelnen Zentren festgelegt, um einen Vergleich zwischen den anderen nationalen Varietäten zu ziehen. Der Fokus liegt primär auf dem kommunikativ-pragmatischen Aspekt: Vertreten wird, dass das bestimmte Sprachverhalten und die Anwendung der Sprachformen für Kommunikation durch sprachexterne Faktoren wie Länder, Regionen und soziale Gemeinschaften gesteuert werden (vgl. Muhr 1997: 43 ff.). Andererseits wird die deutsche Sprache aus der pluriarealen Sicht als Gesamterscheinung betrachtet und damit wird die grenzüberschreitende Gültigkeit der deutschen Standard- bzw. Schriftsprache in diesem Zusammenhang hervorgehoben. Das heißt, innerhalb des von einer Standardsprache

überdachten deutschsprachigen Raums treten mehrere überregionale Varietäten auf, deren Verwendung und Verbreitung aufgrund der ausgeprägten historischen und politischen Verflechtungen nicht mit Staatsgrenzen übereinstimmen. Infolgedessen wird hier nicht von nationalen, sondern von arealen Varietäten und damit auch von Varianten gesprochen (vgl. Wiesinger 1996: 221).

Schließlich hat die Entscheidung darüber, welcher der beiden Ansätze vertreten wird, keinen Einfluss auf das Wesen der deutschen Sprache selbst und diese bleibt somit unverändert. In erster Linie basieren die Argumente auf differenzierten sprachpolitischen Perspektiven, die den tatsächlichen Streitpunkt darstellen (vgl. Schrodt 1997: 14 ff.). Dennoch erscheint es sinnvoll, im Rahmen der vorliegenden Masterarbeit den plurizentrischen Ansatz zu verfolgen. Der Schwerpunkt liegt hierbei auf der Verwaltungssprache und -terminologie, die mit dem Rechtssystem des jeweiligen Staats verbunden ist und deshalb keinesfalls grenzübergreifende Gemeinsamkeiten aufweist (vgl. Muhr 1993: 111 f.). Aufgrund dieser engen Zugehörigkeit zu einem Staat erweist sich der Begriff Nationalvarietät im Sinne der vorgesehenen Untersuchung als notwendig und somit als relevant. Dementsprechend sind die österreichischen rechts- und verwaltungsbezogenen Varianten ausschließlich im Kontext der österreichischen nationalen Standardvarietät zu bewerten.

1.2 Die österreichische Nationalvarietät

Die breite Anerkennung des österreichischen Standarddeutsch als eigene Nationalvarietät wird zum einen durch die langjährige Souveränität Österreichs und zum anderen durch die normierten linguistischen Besonderheiten begründet (vgl. Clyne 1992: 120). Konkret führen de Cilia (1995) sowie Clyne (1993) die autonome Existenz der österreichischen Sprachvarietät auf die historische, kulturelle und soziopolitische Sonderentwicklung zurück, die bereits während des Zeitalters des aufgeklärten Absolutismus im späten 18. Jahrhundert einsetzte. Vor allem seitdem die staatliche Unabhängigkeit des Jahres 1955 nach der siebenjährigen Herrschaft des Nationalsozialismus von 1938 bis 1945 und den anschließenden zehn Jahren Besatzungszeit unter den Alliierten wiederhergestellt wurde, hat sich in Österreich das verstärkte Nationalbewusstsein deutlich in der Sprache manifestiert (vgl. Wiesinger 1995a: 59). Dies hängt laut Muhr (1993: 111) damit zusammen, dass „die Sprache für die Identitätsbildung und den Ausdruck von Identität eine zentrale Rolle spielt.“

Als konkrete Maßnahme gilt die im Jahr 1951 veröffentlichte lexikografische Kodifizierung des österreichischen Wörterbuchs, dessen Erarbeitung vom Bundesministerium für Unterricht beauftragt wurde und das aktuell in fortgeführter 38. Auflage des Jahres 1997

zur Verfügung steht (vgl. Wiesinger 2000: 41 ff.). Das Nachschlagewerk soll einerseits anhand der Hervorhebung der österreichischen Wortschatzeigenheiten, somit der lexikalischen Nationalvarianten, eine einigende Funktion aufweisen. Zugleich werden aber die Ausdrücke besonders mittel- und norddeutscher Herkunft als „ungebräuchlich in Österreich“ oder als „bundesdeutsch“ (Wiesinger 1995a: 59) vermerkt, was laut Dressler und Wodak (1983: 258) einer separatistischen Wirkung entspricht. Zu dieser Handlungstendenz wurde von Muhr (1993: 111) wie folgt formuliert: „Sprache harmonisiert nach innen und differenziert nach außen“. Damit zeigt sich der Versuch, die deutsche Sprache in Österreich als eine eigene und selbstständige Varietät zu etablieren (vgl. Wiesinger 1995a: 60; Wiesinger 2008a: 7).

Dennoch ist das gegenwärtige österreichische Standarddeutsch selbst nicht als einheitliche Varietät der deutschen Sprache zu verstehen. Vielmehr setzt sich dieses aus der Gesamtheit der in Österreich bestehenden sprachlichen Varianten zusammen (vgl. Ebner 1980: 215). Spezifisch zur Begründung der Diversität der österreichischen Varianten schreibt Wiesinger (2008a):

Die Ursachen dafür liegen in der Diachronie und gehen damit auf die jahrhundertealte Geschichte der deutschen Sprache mit verschiedenartigen stammessprachlichen Grundlagen, wechselnden kommunikativen und kulturellen Beziehungen und sich unterschiedlich entwickelnden territorialen und sprachräumlichen Verhältnissen zurück. (Wiesinger 2008a: 6)

In diesem Zusammenhang werden im folgenden Abschnitt Faktoren wie geografische Lage, historischer Hintergrund sowie soziale und pragmatische Situationen erläutert, die zur Herausbildung der gegenwärtigen österreichischen Spracheigenheiten beigetragen haben.

1.2.1 Sprachgeografische, sprachhistorische und sprachsoziologische Voraussetzungen der Sprachsituation in Österreich

In sprachgeografischer Hinsicht gilt Österreich neben Süddeutschland und der Schweiz als Teil der oberdeutschen Dialektregion, die sich im Süden des deutschsprachigen Raums befindet (vgl. Mentrup 1980a: 527). Diese Großdialektgruppe lassen sich weiter in das ostoberdeutsche Bairisch und das westoberdeutsche Alemannisch unterteilen. Ersteres bildet das größte Dialektgebiet in Österreich. Dieses erstreckt sich von Tirol im Westen bis Niederösterreich und Burgenland im Osten, umfasst damit beinahe das gesamte Staatsgebiet, bis auf das Bundesland Vorarlberg gemeinsam mit dem kleinen benachbarten Areal von Westtirol, das dem alemannischen Dialektraum zugeordnet ist. So weist das österreichische Deutsch oberdeutsche Gemeinsamkeiten einerseits im bairischen Bereich mit Ober-, Niederbayern sowie der Oberpfalz und andererseits im alemannischen Gebiet mit der Schweiz, Liechtenstein und dem

süddeutschen Allgäu auf. Die Unterschiede zwischen der bairischen und der alemannischen Dialektregion bestehen nicht nur in der geografischen Verteilung, sondern allen voran im Bereich Pragmatik: Während im Sprachgebrauch des bairischen Raums keine klare Trennung zwischen Dialekt und Standardsprache vorliegt, werden die beiden in der Region des Alemannischen deutlich voneinander abgegrenzt (vgl. Ammon 1995a: 197 f.). Die Einflüsse der zwei unterschiedlichen Dialekte des Oberdeutschen zeigen sich vor allem in den Aussprachegewohnheiten und in der Alltagssprache, wirken aber auch in die österreichische Standardsprache hinein, z. B. in Bezug auf den Wortschatz (siehe Abschnitt 1.2.3). Als weitere Quelle der Spracheinflüsse gelten die Sprachen der Österreich umgebenden Länder, z. B. Italienisch, Alpenromanisch, Slowenisch, Ungarisch, Slowakisch und Tschechisch, die in Form von Fremdwörtern und Entlehnungen im österreichischen Deutsch erscheinen (vgl. Wiesinger 2008a: 8).

Des Weiteren herrscht in Österreich ein breites Spektrum sprachlicher Variationen; dabei wird die gegenwärtige österreichische Sprachsituation grundsätzlich mit dem Dreiermodell Mundart – Umgangssprache – Hochsprache charakterisiert (vgl. Mentrup 1980a: 528). In diesem Zusammenhang wird unter der Umgangssprache die Sprachform verstanden, die „zwischen Hochsprache und Mundart liegt und von beiden angezogen hat“ (Hornung & Roitinger 1950: 8). Eine moderne Äquivalenz dieses Modells wurde von Löffler (2005: 11) vorgestellt: Dialekt – Substandardsprache – Standardsprache. So verfügen die ÖsterreicherInnen bei der (mündlichen) Sprachverwendung über eine Auswahl an Varianten, die sich anhand ihres „unterschiedlichen Gehalts an dialektalen, umgangssprachlichen und hochsprachlichen Formen voneinander unterscheiden“ (Reiffenstein 1977: 176), wobei die Präferenz bzw. die Entscheidung über deren Nutzung im Wesentlichen auf soziale und pragmatische Faktoren zurückzuführen ist. Beispielsweise dient die österreichische Standardsprache (im Sinne von Löfflers Darstellung) primär dem Gebrauch in wenigen öffentlichen Bereichen, darunter Justiz und Verwaltung, Medien sowie Bildungs- und religiösen Einrichtungen, aber auch als üblich verwendete Sprachform exklusiv innerhalb einer kleinen städtischen Bildungsschicht, während im Alltag in erster Linie auf die Substandardsprache – die Umgangssprache bzw. den Dialekt – zurückgegriffen wird. Insoweit werden die sprachliche Variation sowie der Grad der Abweichung und der Approximation an die Standardsprache bei verschiedenen NutzerInnen und SprecherInnen durch den gesellschaftlichen Status mit Bildungshintergrund und beruflicher Tätigkeit sowie demografische Faktoren, wie Generationszugehörigkeit und Geschlechtsgruppe, bedingt (vgl. Wiesinger 2008: 9; Mentrup 1980a: 530).

Betrachtet aus der sprachhistorischen Perspektive reicht die Genese des österreichischen Deutsch bis ins 16. Jahrhundert zurück. Vor allem leiteten sich das gegenwärtige österreichische Amtsdeutsch und seine Terminologien von der bürokratischen Verwaltung und Regelung des damaligen Territorialgebildes ab, das sich über mehrere Jahrhunderte zum gegenwärtigen österreichischen Staat entwickelte (vgl. Wiesinger 1990: 501 f.). Beispielsweise wurden die Charakterzüge der Schriftsprache im österreichischen Bildungs- und Verwaltungswesen bis um Mitte des 18. Jahrhunderts von der Kanzleisprache¹ Kaiser Maximilians I beeinflusst (vgl. Polenz 2009: 86). Durch die Ausdehnung der habsburgischen Besitzungen in Ungarn, Böhmen und Siebenbürgen u. a. zwischen dem 16. und 17. Jahrhundert verwandelte sich Österreich mit Wien als Hauptsitz in ein multikulturelles Zentrum und somit eine multilinguale Gesellschaft. Zusammen mit den diplomatischen Beziehungen zwischen dem Wiener Hof und mehreren europäischen Großmächten wie Frankreich, Italien und Spanien trug dies zu Entlehnungen aus verschiedenen Fremdsprachen im österreichischen Deutsch bei. Besonders als Österreich im Jahr 1804 unabhängiges Kaisertum wurde, wurden mit der Erweiterung des Staats- und Militärwesens eine Reihe neuer bereichsspezifischer österreichischer Bezeichnungen in den behördlichen Domänen eingeführt (vgl. Mentrup 1980a: 528).

Durch den sprachlichen Einfluss und die allmähliche Assimilation der sozialen Oberschichten Wiens an die hochdeutsche Aussprache kam es allerdings zum Bemühen auf nationaler Ebene, die deutsche Schriftsprache und die Orthografie zu vereinheitlichen und die spezifisch österreichische Sprachtradition zugunsten der Einheitlichkeit des Sprachgebrauchs innerhalb des gesamten deutschsprachigen Raums zu unterdrücken (vgl. Mentrup 1980b: 327; Wiesinger 1995b: 336). Erst mit dem Austritt Österreichs aus dem Deutschen Bund, folglich der kleindeutschen Lösung um die Mitte des 19. Jahrhunderts und der konsequenten Gründung der österreichisch-ungarischen Monarchie im Jahr 1866, wuchs zunehmend die sprachliche Verschiedenheit zwischen dem österreichischen Deutsch und dem Binnendeutsch. Dabei blieben z. B. in Österreich viele alte, oft von Fremdsprachen übernommene amtssprachliche Ausdrücke erhalten, während es in Deutschland zur Verdeutschung vor allem der Verwaltungssprache tendierte (vgl. Polenz 2009: 161). Dieses nationale Selbstbewusstsein ließ sich, wie eingangs in Unterkapitel 1.2 erklärt, noch verstärkt durch die Wiedererlangung der Souveränität Österreichs nach seinem Anschluss ans nationalsozialistische Deutsche Reich

¹ Form der deutschen Sprache im geschäftlichen Schriftverkehr seit der ersten Hälfte des 13. Jahrhunderts, besonders in Urkunden, Akten und Rechtsvorschriften (Duden).

wiederbeleben und führte zur Etablierung des österreichischen Deutsch als national verbindliche Sprachvarietät.

1.2.2 Austriazismen – Besonderheiten der österreichischen Varietät

Die sprachlichen und pragmatischen Unterschiede im Gebrauch einer plurizentrischen Sprache wie Deutsch in einzelnen Regionen bzw. Ländern werden in der Variationslinguistik als Varianten bezeichnet (vgl. Wissik 2012: 56). Als Beispiel einer Variante nennt Ammon (1995a) Karfiol und Blumenkohl, welche die Benennungen desselben Objekts sind und somit die gleiche Bedeutung haben. Jedoch lassen sich die beiden durch ihren regional-sprachlichen Gebrauch unterscheiden, denn die Benennung Karfiol findet ihre Verwendung lediglich in Österreich, wohingegen Blumenkohl eine bundesdeutsche Variante ist. Bei überwiegenden Gemeinsamkeiten innerhalb des Verwendungsbereichs einer plurizentrischen Sprache fungieren solche sprachlichen Varianten als *differentia specifica*, die in Summe eine bestimmte (nationale) Standardvarietät ausmachen und diese von den anderen abgrenzen (vgl. Ammon 1995a: 64 ff.). Anders ausgedrückt stellen Varianten spezifische Einheiten bzw. Bestandteile dar, durch die ein gesamtes sprachliches System, d. h. eine Sprachvarietät, konstruiert und konkretisiert wird (siehe Abb. 2).

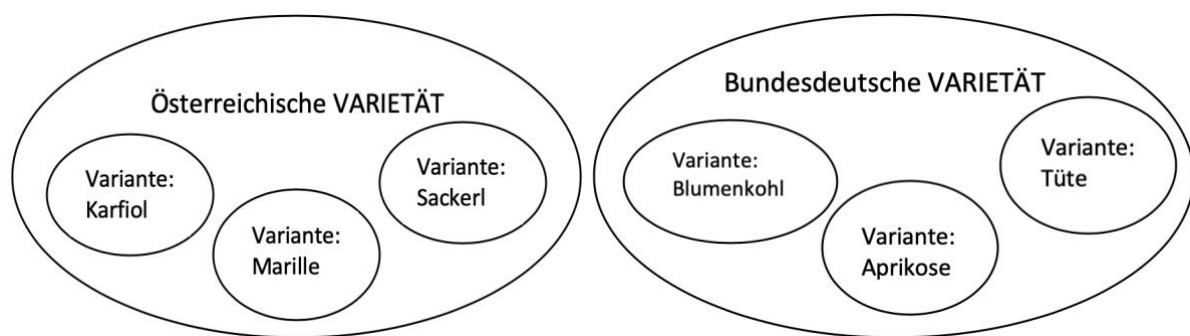


Abbildung 2: Österreichische und bundesdeutsche Varietäten (adaptiert von Ammon 1995a: 65)

Die Varianten gelten in unterschiedlichem Ausmaß auf allen linguistischen Ebenen wie der phonetisch-phonologischen, der morphologischen, der syntaktischen, der pragmatischen und der lexikalisch-semantischen Ebene einschließlich der Phraseologie (vgl. Wiesinger 2008a: 5). In Bezug auf die österreichische Varietät werden die markierenden Besonderheiten Austriazismen bezeichnet; hingegen wird für die nationalen Ausprägungen der deutschsprachigen Schweiz der Terminus Helvetismen verwendet. Hinsichtlich der nur in Deutschland geltende Varianten sind mehrere Bezeichnungen vertreten, wobei „Teutonismen“ (Ammon 1994: 51), „Deutschlandismen“ (Polenz 1996: 210; Muhr 2000: 30) und

„Germanismen“ bzw. „Germanizismen“ (Markhardt 2005: 21) zu den weit bekannten Bezeichnungen gehören.

Auch wenn die Austriazismen, wie erläutert, auf allen sprachlichen und pragmatischen Ebenen angesiedelt sind, betreffen diese die Lexik am meisten, die im Vergleich zu den anderen Charakteristiken den auffälligsten Aspekt darstellt (vgl. Wermke 1995: 201; Schrodtt 1997: 13). Allerdings bilden die zu Austriazismen gehörenden Bezeichnungen, die zum einen die amtssprachliche Terminologie der staatlichen Verwaltung und Einrichtungen und zum anderen den gemeinsprachlichen Wortschatz umfassen, quantitativ gesehen einen Anteil von lediglich drei Prozent der gesamten deutschen Standard- und Schriftsprache (vgl. Krumm 2010: 360). Das heißt, in hundert Wörtern in einem deutschsprachigen Text werden im Durchschnitt drei Austriazismen festgestellt, während die restlichen als außerhalb von Österreich allgemein verständlich beurteilt werden. Jedoch variiert dieses Verhältnis dabei selbstständig auch je nach Inhalt und Sachgebiet. Trotz dieser geringen Anzahl sind sämtliche bestehende Varianten der österreichischen Standardvarietät laut Moser (1989: 25) aufgrund der weitgehenden Akzeptanz in einzelnen Gebieten „hinsichtlich ihrer normativen Gültigkeit als gleichwertig und gleichberechtigt anzusehen“ (vgl. auch Wiesinger 2008a: 6). Darüber hinaus weist Muhr (1995) auf das häufig übersehene Merkmal nationaler Varianten hin, wonach es sich vielmehr um „Präferenzunterschiede“ handelt, „die durch die Wahl unterschiedlicher lexikalischer Mittel innerhalb desselben Ausdrucksfeldes entstehen“ (Muhr 1995: 213).

1.2.3 Linguistische Eigenheiten der österreichischen Varietät

In diesem Abschnitt werden charakteristische Erscheinungen der österreichischen Schriftsprache auf verschiedenen sprachlichen Ebenen dargelegt. Dies soll einen Überblick über die Merkmale dieser Sprachvarietät im Allgemeinen vermitteln, um die Grundlage für die Vertiefung in die linguistischen Besonderheiten der österreichischen Verwaltungssprache im folgenden Kapitel zu ermöglichen, die das Kernstück dieser Arbeit darstellt. Allerdings wird auf die Beschreibung folgender Aspekte verzichtet: umgangssprachliche sowie dialektale Austriazismen; phonetisch-phonologische Eigenschaften, z. B. Lautbildung, Sprechmelodie und Akzentuierung; und pragmatische Merkmale, bei denen es sich überwiegend um mündliche Sprachvarianten handelt. Diese werden, wie erklärt, nicht in die empirische Untersuchung einbezogen und haben somit keine Relevanz.

Morphologische Besonderheiten

Die österreichischen Varianten zeigen sich in dieser Hinsicht deutlich im Genus- und Formengebrauch bei Substantiven. Anhand der Auflistung von Muhr (1995: 215 ff.) heißt es beispielsweise in Österreich der Akt (im Sinne von Gerichtsakt), wohingegen die bundesdeutsche Variante dieses Begriffs die Akte ist. Hier liegen sowohl ein Genus- als auch ein Formunterschied zwischen den beiden Varianten vor. Nach Wiesinger (2008a: 13) soll die Genusänderung durch die Formabweichung ausgelöst werden, wie bei der Karren (Ö) und die Karre (D), der Spalt (Ö) und die Spalte (D), das Offert (Ö) und die Offerte (D) usw. Es gibt aber auch Austriazismen, die trotz unveränderter Form einen unterschiedlichen Genus besitzen, z. B. das Gehalt (Ö) und der Gehalt (D), das Service (Ö) und der Service (D) – weitere Beispiele lassen sich Ebner (2009: 462 ff.) entnehmen.

Auch gilt die morphologische Differenzierung in manchen Fällen der Pluralbildung des österreichischen Deutsch: Hierbei wird zur Umlautung von umlautfähigen Vokalen bei der Pluralform geneigt, so in die Erlässe, die Bögen, die Generäle, während diese im Bundesdeutschen in der Regel nicht umgelauteet sind. Darüber hinaus wird hinsichtlich der Wortkomposition im österreichischen Deutsch die Hinzufügung von Fugenzeichen zwischen Substantiven bevorzugt. So tritt das Fugenmorphem -s beispielsweise bei den österreichischen Varianten wie Fabriksdirektor, Personsbeschreibung, Aufnahmsprüfung und Schweinsbraten gegenüber Fabrikdirektor, Personenbeschreibung, Aufnahmeprüfung und Schweinebraten im Bundesdeutschen auf. Andererseits lässt sich mitunter auch das Gegenteil bemerken, indem auf das Fugenmorphem bei österreichischen Varianten verzichtet wird, z. B. bei Taglohn (Ö) und Tagelohn (D) (vgl. Ammon 1995a: 173).

Doch betrifft die besondere Wortbildungspräferenz nicht nur Substantive, sondern auch Verben. Häufig wird in der Schriftsprache auf die Verbalableitung mit -ieren zurückgegriffen, sodass es delogieren (hinauswerfen), exekutieren (Auftrag ausführen), transferieren (versetzen), refundieren (ersetzen) und vidieren (beglaubigen) heißt (vgl. Wiesinger 2008a: 14). Außerdem bestehen relevante Unterschiede bei zahlreichen Präfixverben, denn ihre Verwendung in der jeweiligen Varietät greift zum Teil tief ins Sprachsystem ein. Nachfolgend findet sich eine Übersicht zu österreichischen Verbalvarianten in Anlehnung an Muhr (1995: 217 ff.), der diesbezüglich drei Kategorien von Unterschieden ermittelt. Dabei werden die österreichischen Varianten jeweils immer vor den bundesdeutschen angeführt:

1. Dasselbe Stammverb in Kombination mit unterschiedlichen Präfixen; die Varianten haben dieselbe Grundbedeutung und drücken im Prinzip denselben Inhalt aus, z. B. aufscheinen

<> erscheinen, beheben <> abheben, beistellen <> bereitstellen, beiziehen <> zurate ziehen, entfallen <> ausfallen.

2. Dasselbe Präfix in Kombination mit unterschiedlichen Verben; die Varianten haben dieselbe Grundbedeutung und drücken im Prinzip denselben Inhalt aus, z. B. absperren <> abschließen, ausständig <> ausstehend, beischließen <> beilegen, einbekennen <> eingestehen, einlangen <> eintreffen.
3. Sowohl Stammverben als auch Präfixe unterscheiden sich. Die Varianten haben allerdings annähernd dieselbe Bedeutung, z. B. aufnehmen <> anstellen/einstellen, zurücklegen <> abgeben, erstrecken <> verlängern.

Syntaktische Besonderheiten

Im Hinblick auf die Grammatik ist der Bereich Satzbau die sprachliche Ebene, auf der sich die Merkmale der österreichischen Varietät am geringsten manifestieren und somit am wenigsten vom Bundesdeutschen divergieren (vgl. Reiffenstein 1983: 20). Zu den typischen syntaktischen Eigenheiten gehört die Bildung des Perfekts mit sein statt des in Deutschland überwiegend verwendeten Hilfsverbs haben bei den folgenden Zustands- und Bewegungsverben: stehen, sitzen, liegen, hängen, knien, lehnen, stecken und schweben. Dementsprechend heißt es in Österreich beispielsweise ich bin gesessen. Auch wenn eine solche Perfektbildung teilweise im Süddeutschen und im Schweizerdeutschen gängig ist, wird diese in Deutschland oft als umgangssprachlich bewertet (vgl. Ammon et al. 2011: LXIX).

Ferner besteht ein Unterschied zwischen dem österreichischen Deutsch und dem Bundesdeutschen in der Abfolge der verbalen Elemente von Nebensätzen im Perfekt oder Konjunktiv II mit dreiteiligem Prädikat. Nach dem bundesdeutschen Standard steht im Schlussfeld eines Nebensatzes an erster Stelle das konjugierte Hilfsverb haben, gefolgt vom Infinitiv des Hauptverbs und dem Modalverb, während die Reihenfolge in der österreichischen Variante genau umgekehrt erfolgt (vgl. Wiesinger 2008a: 14). Zur Übersicht lässt sich wie in Tabelle 1 darlegen:

Österreich	dass er den Anspruch der Klägerin im Einzelnen	prüfen	<u>hätte</u>	müssen.
Deutschland	dass er den Anspruch der Klägerin im Einzelnen	<u>hätte</u>	prüfen	müssen.

Tabelle 1: Österreichische und deutsche syntaktische Varianten

Laut Stubkjær (1993:48) wird gemäß der österreichischen Reihenfolgennorm besonderer Wert darauf gelegt, dass das Hauptverb und seine nominalen Glieder möglichst nah aneinander und vor den anderen Verbformen innerhalb des Nebensatzes angeordnet werden, denn diese stellen grundlegende und damit hervorzuhebende Informationen dar.

Neben den in der Fachliteratur weitbekannten Eigenheiten, wie den oben genannten, hat Muhr (1995) weitere syntaktische Merkmale entdeckt, die hauptsächlich in der abweichenden oder zusätzlichen Verwendung von Präpositionalgruppen zu lokalen und temporalen Bestimmungen festzustellen sind. Dabei beruht die Abweichung hinsichtlich des Ausdrückens von Lokalität und zeitlicher Punktualität im österreichischen Deutsch auf „systematischen Unterschieden in der semantischen Basis“ (Muhr 1995: 225). Die Differenzierung liegt somit an der Auslegung und der Wahrnehmung der Begriffe zur Orts- und Zeitangabe in der jeweiligen Variante. Zum Beispiel wird in Österreich der Begriff *Fahrbahnmitte* als Raum betrachtet und tritt daher in Kombination mit der Präposition *in* auf. Hingegen gilt dieser in der bundesdeutschen Variante als Punkt und wird meistens mit *auf* ausgedrückt. Schließlich zeigt Muhr (1995: 227 f.) anhand des Korpus aus der österreichischen Mediensprache, dass die österreichischen Varianten einiger Verben zusätzliche Präpositionen oder sogar Reflexivpronomen mit sich bringen, was teilweise auf einen tschechischen bzw. slowakischen Spracheinfluss zurückgeführt werden kann, z. B. auf etwas vergessen oder sich etwas erwarten.

Lexikalische Besonderheiten

Wie bereits in Unterkapitel 1.2.2 erwähnt, machen die *Austriazismen* auf dem lexikalischen Gebiet den auffälligsten Anteil an der österreichischen Standardvarietät aus und zählen zu den kennzeichnenden Merkmalen, an denen sich österreichische SprecherInnen erkennen lassen. Laut Ammon (1995a: 154) „spielen die lexikalischen *Austriazismen* auch mit Abstand die größte Rolle in populären oder öffentlichkeitswirksamen Darstellungen des österreichischen Deutsch.“ Dabei lässt sich in diesem Bereich in zwei Kategorien unterteilen: Einerseits gibt es neutrale *Austriazismen*, die in der Standardsprache ohne jegliche Wertung gebraucht werden, wie *Kommerzialrat*, *Gebarungsjahr*, *sich verkühlen* (gegenüber den bundesdeutschen Varianten *Kommerzienrat*, *Geschäftsjahr* bzw. *sich erkälten*). Andererseits gehören zum österreichischen Wortschatz auch sprachsoziologisch gebundene Ausdrücke, die unter anderem als umgangssprachlich oder salopp aufzufassen sind (vgl. Wiesinger 2008a: 14). Zwar findet sich österreichischer Wortschatz in allen Sachgebieten, sein Umfang fällt aber jeweils unterschiedlich aus. In diesem Bezug hat Ammon (1995a: 157 ff.) die Erscheinungshäufigkeit von *Austriazismen* anhand einer charakteristischen Auswahl von 418 Belegen statistisch

untersucht. Ausgehend von der daraus resultierenden Rangliste verzeichnet der Bereich Speisen und Mahlzeit die meisten lexikalischen Besonderheiten, gefolgt von Verwaltung, Justiz, Gesundheitswesen, Schule und Militär und Geschäftsleben, Handwerk, Landwirtschaft und Verkehr als der zweit- bzw. drittgrößten Quelle des österreichischen Wortschatzes. Zusammen machen Austriazismen aus diesen drei Sachgebieten mehr als die Hälfte der österreichspezifischen Terminologie aus.

Obwohl der Gebrauch der meisten Austriazismen innerhalb des Staatsgebiets erfolgt, gehen einige der zum österreichischen Deutsch gehörenden Begriffe durch Mobilität und Medien über die Grenzen Österreichs hinaus. Deshalb lässt sich im Sinne der räumlichen Verbreitung lexikalischer Austriazismen in „grenzüberschreitenden uneigentlichen österreichischen Wortschatz“ und „gesamt- oder teilösterreichischen eigentlichen österreichischen Wortschatz“ gliedern (Wiesinger 1988a: 25 ff.). Dabei geht die erste Bezeichnungsgruppe auf sprachgeografische Einflüsse zurück und besteht einerseits aus dem oberdeutschen Wortschatz, der Österreich und Süddeutschland sowie der Schweiz gilt, und andererseits dem bairisch-österreichischen Wortschatz, der Österreich mit (Alt-)Bayern „aufgrund der gemeinsamen Stammesgrundlage bzw. späterer Kulturbeziehungen“ verbindet (Wiesinger 2008a: 15). Die zweite Gruppe, die sich somit zum Großteil mit den österreichischen Staatsgrenzen deckt, wird wiederum in einen ost- und westösterreichischen, regionalen und gesamtösterreichischen Wortschatz eingeteilt. Bei den zwei Erstgenannten beschränken sich die sprachlichen Eigenheiten zwar – wie der Name schon sagt – zumeist auf die inländischen Regionen, werden aber zum Teil auch mit den angrenzenden Randgebieten außerhalb von Österreich geteilt, wie beispielsweise Vorarlberg mit der Ostschweiz (vgl. Wiesinger 1988b: 230). Dementsprechend bleibt der gesamtösterreichische Wortschatz die alleinige Kategorie, deren Gültigkeit ganz Österreich ohne Differenzierung betrifft und nicht über Landesgrenzen hinaus reicht. Er umfasst sowohl den verkehrssprachlichen Wortschatz, somit die neutralen Austriazismen (siehe oben), als auch eine Reihe amtssprachlicher Terminologien vor allem im Bereich Politik, Verwaltung und Recht. Diese österreichischen bereichsspezifischen Varianten sind in der staatlichen Souveränität verankert und werden damit ausschließlich in Österreich gebraucht, sodass sie laut Scheuringer (1997: 342) als die einzigen standardsprachlichen Austriazismen gelten sollten – alle anderen würden dann zu Regionalismen gezählt.

2 Verwaltung und Sprache

Da es sich in der vorliegenden Arbeit um die Sondersprache im administrativen Bereich handelt, soll an dieser Stelle zunächst die öffentliche Verwaltung im Allgemeinen erläutert werden. Hierauf aufbauend wird die Interaktion in der Verwaltung aus kommunikativer Sicht betrachtet, und zwar als sprachliches Handeln zwischen InterakteurInnen. Im Anschluss werden die funktionsbedingten Eigenschaften der gegenwärtigen Verwaltungssprache vorgestellt, einhergehend mit den einzelnen linguistischen Merkmalen der österreichischen Varianten der Verwaltungssprache. Schließlich wird auf die Charakterisierung der Verwaltungstexte eingegangen, die dann als Basis für eine Betrachtung der Textsorte Formulare dient. Um Missverständnisse zu vermeiden, gilt es allerdings vorab anzumerken, dass es ebenfalls mündliche Formen der Verwaltungssprache gibt. Nur werden diese im Rahmen der vorliegenden Arbeit nicht berücksichtigt, denn es sind hier lediglich schriftlich niedergelegte Texte bezüglich der Linguistik der Verwaltungssprache zu untersuchen.

2.1 Die öffentliche Verwaltung

Die öffentliche Verwaltung, auch als Administrative bezeichnet, gehört im Rahmen des österreichischen Rechtsstaats neben der Regierung zur Gesamtheit der Organe zur Ausübung der vollziehenden Gewalt, d. h. der Exekutive (vgl. Seifert 2003: 5). So hat die Verwaltung in erster Linie die Aufgabe, die Umsetzung der von der Legislative beschlossenen Gesetze zu gewährleisten, wobei die Befugnisse und Durchführung jeglicher administrativer Tätigkeiten von der Verfassung vorgeschrieben sind (vgl. Efinger 2020: 288). Dementsprechend ist das Verwaltungshandeln eng an die gesetzlichen Bestimmungen gebunden und darf dem Grundsatz der Gesetzmäßigkeit zufolge nicht ohne Ermächtigung bzw. im Widerspruch zu diesen geschehen. Dennoch stellt die öffentliche Verwaltung keine einheitliche Einrichtung dar, sondern gehört als Bestandteil zu jeweils unterschiedlichen Stellen auf Bundes-, Länder- bzw. kommunaler Ebene (vgl. Becker-Mrotzek & Scherner 2000: 632). Zu Handlungsträgern in der österreichischen Verwaltung werden alle staatlichen Institutionen wie Behörden und Ämter gezählt, bis hin zu Bundesregierung und BundespräsidentIn.

Allerdings beschränkt sich der Umfang der Verwaltung nicht nur auf die Aktivitäten innerhalb staatlicher Institutionen, sondern betrifft unter anderem auch private Haushalte sowie betriebswirtschaftliches Gebiet (vgl. Fuchs-Heinritz et al. 1994: 722). Das wesentlichste Kennzeichen der Industriegesellschaft des 20. Jahrhunderts besteht darin, dass die öffentliche Verwaltung beinahe alle Lebensbereiche einzelner BürgerInnen durchdrungen hat (vgl.

Albrecht 1980: 83). Die abstrakte Konstruktion der Gesetzgebung, die der gesellschaftlichen Handlungsorientierung und -steuerung in Form von beispielsweise Verordnungen und Verboten dient, wird erst mit dem konkreten Verwaltungsakt für die Privatpersonen realisiert. Folglich trägt dies sowohl zur Regelung des öffentlichen (Zusammen-)Lebens im Staat als auch zur Einhaltung der Vorschriften bei (vgl. Fotheringham 1981: 105). Dementsprechend kommen alle in Österreich ansässigen BürgerInnen ständig oder mindestens einmal mit behördlichen Stellen in Berührung. In diesem Rahmen liegt ein Verhältnis zwischen zwei unterschiedlichen Interaktionsgruppen vor, die amungsverfahren beteiligt sind: Einerseits wird den BürgerInnen im außenstehenden Umfeld, die jeweils einen bestimmten Sachverhalt vorbringen bzw. verfolgen, die Rolle von KlientInnen zugewiesen. Andererseits fungieren die zuständigen VerwaltungsbeamtInnen als ihre AgentInnen (vgl. Rehbein 1999: 662). Fluck (2008: 121) unterstreicht die Bedeutung der Sprache im Zusammenhang mit der Aufgabe der Verwaltung, geltende Gesetze in Anwendung zu bringen und zu vollstrecken: „Durch diese Aufgabenstellung ist die Sprache der Verwaltung (zusammen mit der Sprache des Rechts) ‚unmittelbar gesellschaftsprägend und Zugangsinstrumentarium zum politischen Gemeinleben‘ (Lambertz 1999: 145)“. Damit fungieren die Verwaltungsbehörden als eine Vermittlungsstelle zwischen Justiz und BürgerInnen, wobei die Verwaltungssprache der Kommunikation zwischen den AgentInnen und KlientInnen dient (vgl. Benrath 2011: 553). In den folgenden Unterkapiteln wird auf den sprachlichen Aspekt des-verfahrens und damit die Verwaltungssprache eingegangen.

2.2 Das sprachliche Handeln im-verfahren

Die Abwicklung dieser Austauschbeziehung zwischen den beiden Parteien ist kommunikativer Natur. Dabei empfangen die AgentInnen der Verwaltung Informationen bezüglich des Anliegens von KlientInnen bzw. fordern diese gegebenenfalls auf, (Zusatz-) Informationen zu vermitteln. Die Eingaben werden darauffolgend einer Bearbeitung und Überprüfung innerhalb der zuständigen Verwaltungsdienststellen unterzogen, woraus sich förmliche Verwaltungsentscheidungen ergeben, die schließlich den einreichenden KlientInnen mitgeteilt werden (vgl. Oberndörfer 2001: 2). Der gesamte Prozess – vom Anbringen des Sachverhalts bis hin zur Verkündung der Entscheidungen – erfolgt in den meisten Fällen über schriftlichen Verkehr in Form von Antragsformularen, Anzeigen, Bekanntmachungen, Bescheiden, amtlichen Entschliefungen und sonstigen Mitteilungen (vgl. §13 Abs. 1 Allgemeines-verfahrensgesetz). In solchen offiziellen Texten zeigt sich eine Sondersprache des administrativen Gebiets, die als Verwaltungssprache bezeichnet wird. Diese gilt als

bedeutsamstes „Mittel der Verständigung und Grundlage gemeinsamen Handelns einer Gemeinschaft“ (Kirchhoff 1987: 6) und ist somit zu einem wesentlichen Element des sprachlichen Lebens einzelner Individuen geworden (vgl. Wagner 1984: 7).

2.2.1 Verwaltungssprache in der Gegenwart

Die Durchführung jeder Verwaltungstätigkeit eines Staats steht gemäß seiner Verfassung nicht nur in enger Verbindung mit dem Gesetz, sondern unterliegt diesem auch. Als Folge wird die Verwaltungssprache durch das entsprechende Rechtssystem und die Rechtssprache geprägt, was sich auch im Ineinandergreifen der beiden sprachlichen Systeme äußert (vgl. Becker-Mrotzek 1999: 1392). Zusammen bilden sie ein vielschichtiges Gefüge sondersprachlicher Art, das sich aufgrund der verschiedenen bestehenden Rechts- und Verwaltungsbereiche aus mehreren sprachlichen Teilgebieten zusammensetzt. Für den charakteristischen Zusammenhang zwischen der jeweiligen Rechtsordnung und der dazugehörigen Verwaltungs- bzw. Rechtssprache wird die Bezeichnung Systemgebundenheit eingeführt (vgl. de Groot 1999: 206). Darunter ist zu verstehen, dass die Rechts- und Verwaltungssprache so im bestimmten Rechtssystem verankert ist, dass die genaue Definition des jeweiligen Begriffs nur ausschließlich innerhalb dessen nachvollzogen und ausgelegt werden kann. Dies ist auf den jeweils unterschiedlichen kulturellen, historischen, gesellschaftlichen und wirtschaftlichen Kontext einzelner Staaten zurückzuführen (vgl. Chiocchetti et al. 2013a: 16; Chiocchetti et al. 2013b: 11).

Hier lässt sich eine Gemeinsamkeit der Konzepte Systemgebundenheit und Plurizentrismus bemerken: Es liegt hinsichtlich der länderspezifischen Gültigkeit der Rechts- und Verwaltungssprache wie bei der nationalen Sprachvarietäten eine klare und sogar noch strengere Grenzziehung vor, sodass ein grenzüberschreitender Gebrauch grundsätzlich ausgeschlossen wird. Ein deutlicher Unterschied liegt allerdings in der terminologischen Bedeutungsbeziehung zwischen Benennungen vor, die demselben Begriff zugewiesen werden. Zum einen werden die gemeinsprachlichen Nationalvarianten einer plurizentrischen Sprache, z. B. Marille (Ö) und Aprikose (D) als Synonyme betrachtet. Zum anderen erfordert diese semantische Relation im Zusammenhang mit der Rechts- und Verwaltungssprache, dass die Benennungen nicht nur aus derselben natürlichen Sprache stammen, sondern einem gleichen Rechtssystem entsprechen müssen (vgl. Sandrini 1996: 138). Demnach können zwei Benennungen, beispielsweise (Eltern-)Karenz (Ö) und Elternzeit (D), die aber jeweils einem unterschiedlichen Rechtssystem angehören, trotz derselben natürlichen Sprache und ihrer ähnlichen Bedeutung nicht als Synonyme bezeichnet werden. Aufgrund der systemgebundenen

Charakteristik der Rechts- und Verwaltungssprache lassen sich diese Varianten im rechtlichen Zusammenhang keinesfalls miteinander auswechseln und sind daher vielmehr als Äquivalenz zu bewerten (vgl. Chiochetti et al. 2013a: 16).

Infolgedessen findet häufig die Paarbezeichnung der „Rechts- und Verwaltungssprache“ und der „Rechts- und Verwaltungstexte[n]“ Anwendung in der Literatur (Efing 2020: 289). Jedoch weist Sellmann (2014: 178) trotz dieser engen Verbindung auf die notwendige Differenzierung zwischen beiden hin, wenn es sich um den praktischen Einsatzbereich handelt. Die Rechtssprache der Gesetze, die beinahe ausschließlich in der internen Fachkommunikation unter juristischen Fachleuten verwendet wird, ist als juristische Fachsprache zu betrachten. Hingegen handelt sich bei der Verwaltungssprache um eine „spezifische Form kommunikativen Handelns, die Bestandteil eines Kommunikationsprozesses in oder mit einer Verwaltungsinstitution ist“ (Gühlich 1981: 326). Aus dieser Beschreibung lässt sich ein zweiseitiger Verwendungszweck ableiten: Einerseits soll die Verwaltungssprache aufgrund ihrer fachsprachlichen Eigenschaft als sprachliches Instrument zur Verständigung innerhalb einer bestimmten Profession oder unter Fachleuten dienen, die in diesem Fall Verwaltungsangehörigen im Rahmen ihrer beruflichen Ausübung entsprechen (vgl. Fisch & Margies 2014: 9). Im Gegensatz zu anderen Fachsprachen, darunter die Rechtssprache, die sich größtenteils auf einen „fachlich begrenzbaren Kommunikationsbereich“ (Hoffmann 1985: 53) beschränken, richtet sich die Verwaltungssprache andererseits aber auch an AdressatInnen außerhalb des Fachgebiets. Dies betrifft alle StaatsbürgerInnen als Laien ohne Sachkenntnis, die häufig am Verwaltungshandeln und somit an der kommunikativen Austauschbeziehung mit Behörden teilnehmen (siehe Unterkapitel 2.1).

Allerdings wird der Adressatenkreis bei der Verwaltungssprache nicht berücksichtigt, sondern der Verwaltungsakt selbst steht im Vordergrund (vgl. Günther 1898: 19). So ist aus juridischer Sicht das Ziel, Sachverhalte mit möglichster Objektivität und Präzision auszudrücken, was mithilfe eines umfangreichen Bestands an festgelegten Bezeichnungen sowie gegenstandadäquater Textstrukturen ermöglicht wird. Außerdem führt das Bemühen um Eindeutigkeit zu häufig formelhaften und unpersönlich wirkenden Wendungen (vgl. Daum 1981: 83). Deshalb stellt sich die Begegnung von KlientInnen, etwa den BürgerInnen, und der Verwaltungssprache nicht unproblematisch dar. Ihre abstrakte und alltagsfremde Ausdrucksweise mag zwar den fachspezifischen Zweck erfüllen, kann aber für die Öffentlichkeit Sprachbarrieren bedeuten. Dies wird als Paradoxon der Verwaltungssprache bezeichnet: Dabei lässt sich diese gerade aufgrund ihrer Fachsprachlichkeit – Sach- und Normbezogenheit sowie Rechtssicherheit – mit dem Anspruch auf Informationskompetenz und

Allgemeinverständlichkeit der betroffenen KlientInnen, d. h. RezipientInnen außerhalb des Fachgebiets, schwer vereinbaren (vgl. Schendera 2004: 322; Klein 2014: 19).

In Österreich führt diese Problematik sowohl zur anhaltenden Kritik an der mangelnden Verständlichkeit der Verwaltungssprache als auch zur Prägung einer pejorativen Bezeichnung als „Amtsdeutsch“ (Heinrich 2014: 48), das „für jegliche schwülstige, schwer verständliche und alltagsfremde Ausdrucksweise steht“ (Wiesinger 2008b: 106). Konkret zeigte das Ergebnis einer im Jahr 1982 von einem niederösterreichischen Meinungsforschungsinstitut durchgeführten repräsentativen Umfrage, dass beinahe drei Viertel der Befragten eine kritische Haltung zum Sprachgebrauch im Verwaltungsgebiet bekundeten und deshalb trotz ihrer Rechte bewusst auf Verwaltungsleistungen verzichteten, um den Umgang mit Amtstexten und im Falle der Unklarheit eine unangenehme Interaktion mit Verwaltungsangestellten zu meiden (vgl. Ott 1983: 64; Lerch 2014: 94). Als Folge wurden mehrere Verwaltungseinheiten zum Handeln angeregt: Im Laufe der 1980er-Jahre haben Bundeskanzleramt, Landesregierungen sowie Stadtverwaltungen zahlreiche interdisziplinäre Projekte auf nationaler Ebene zur Optimierung der Verwaltungskommunikation eingeleitet. Dabei wurden ExpertInnen aus verschiedenen relevanten Fachgebieten, darunter JuristInnen, SprachwissenschaftlerInnen und technische RedakteurInnen, miteinbezogen. Ihre Aufgaben bestanden darin, Gesetzes- und Verwaltungstexte auf ihre Verständlichkeit zu untersuchen und zu verbessern, wie auch eventuell Anweisungen für deren Gestaltungen auszuarbeiten (vgl. Otto 1981: 44). Schließlich haben sich diese ambitionierten Reformversuche aufgrund der Menge an bestehenden und aufkommenden Fassungen jedoch als eine zeit- und kostenaufwändige Aufgabe für den Staat erwiesen (vgl. Wiesinger 2008b: 107). In der Folge wurde das Ziel des sprachlich einwandfreien Umgangs der Öffentlichkeit mit Rechts- und Verwaltungstexten bisher noch nicht erreicht.

2.2.2 Österreichische Verwaltungssprache

Die Schwerverständlichkeit der gegenwärtigen Rechts- und Verwaltungssprache wird häufig mit ihren linguistischen Eigenheiten in Verbindung gebracht. Die Studie von Wagner (1984) gilt als frühestes und grundlegendes Werk im Hinblick auf deutschsprachige Amtstexte zur linguistischen Untersuchung des spezifischen Sprachgebrauchs in der Verwaltung. Ihre Analyse beschränkt sich allerdings auf die Merkmale der bundesdeutschen Verwaltungssprache und wurde von Wiesinger (2008b) in Rahmen seiner Beschäftigung mit den sprachlichen Besonderheiten der österreichischen Rechts- und Verwaltungssprache zum Vergleich herangezogen. Da das Hauptaugenmerk der vorliegenden Arbeit auf der österreichischen

Verwaltungssprache liegt, sind primär die repräsentativen Untersuchungsergebnisse von Wiesinger (2008b: 127 ff.) hier von Bedeutung. Dabei handelt es sich um eine Auswahl von unterschiedlichen Textstücken aus verschiedenen Teilbereichen der Verwaltungsinstitutionen in mehreren österreichischen Bundesländern. Diese reicht von Gesetzestexten zur Rechtsstellung der Bundesbediensteten über Antragsformulare und Ansuchen der Finanzdirektion bis hin zu Bescheiden, Kundmachungen sowie Merkblättern des Bundesheeres. Im folgenden Abschnitt werden die festgestellten syntaktischen und lexikalischen Eigenschaften sowie Stilmerkmale der österreichischen Verwaltungssprache, zum Teil mit Gegenüberstellung zu bundesdeutschen Varianten, dargestellt.

Syntaktische Besonderheiten

Im Bereich Syntax werden bei österreichischen Verwaltungsschreibern tendenziell lange Sätze mit hoher Wortanzahl gegenüber kürzeren Formulierungen bevorzugt, wobei Sätze mit 10 bis 19 Wörtern am häufigsten erscheinen und damit den größten Teil ausmachen, gefolgt von denjenigen, die 20 bis 29 Wörter umfassen (vgl. Wiesinger 2008b: 110 f.). Dies hängt auch damit zusammen, dass die Substantivsatzglieder der Subjekte und Objekte in vielen Fällen eine Häufung mehrerer Attribute aufweisen, z. B. Einkünfte aus nicht selbstständiger Arbeit oder die Zustimmung des Einigungsamtes zur Kündigung. Darüber hinaus lässt sich anhand der untersuchten Schriftstücke darauf schließen, dass die österreichische Verwaltungssprache zur Verwendung von Satzgefügen (komplexen Sätzen) neigt und einfach aufgebauten Sätzen vergleichsweise deutlich weniger Raum zugesteht. Dabei herrschen Satzgefüge, die meistens einen Nebensatz beinhalten, in mehr als der Hälfte aller Sätze in den von Wiesinger (2008b) untersuchten Amtstexten vor. Innerhalb dieser Nebensätze wird zusätzlich auf aufschwemmende Attributsätze hingewiesen (vgl. Wiesinger 2008b: 115). Anhand des folgenden Beispiels lassen sich diese Eigenheiten veranschaulichen:

Mitglieder der Prüfungskommission, die auf Grund des Gehaltsüberleitungsgesetzes oder des BDG bestellt wurden und zum Zeitpunkt des Inkrafttretens dieses Bundesgesetzes dieses Amt noch bekleiden, bleiben bis zum Ablauf ihrer Bestellungsperiode oder bis zum Eintritt eines der im § 29 dieses Bundesgesetzes angeführten Endigungsgründe, längstens aber bis zur Neuregelung der für die betreffende Verwendung in der Anlage 2 angeführten Ausbildungs- bzw. Prüfungsvorschrift gemäß § 24 Abs. 5 dieses Bundesgesetzes im Amt. (I: Beamten- Dienstrechtsgesetz, § 186 Abs. 5)

Das Vorfeld dieses Satzgefüges (fett gedruckt) allein setzt sich bereits aus einem Attributsatz in Form eines zweiteiligen Nebensatzes mit insgesamt 21 Wörtern zusammen. Hinzu kommt im Nachfeld eine unverhältnismäßig lange Satzklammer, die von bleiben bis Amt sogar noch

weitere Mehrfachattribute von Temporaladverbialen umschließt. Daraus geht hervor, dass diese syntaktischen Faktoren der Verständnisschwierigkeiten der Verwaltungssprache zu einem erheblichen Teil zugrunde liegen. Die Präferenz für lange und kompliziert gebaute Sätze sorgt für Verständnishemmungen. Vor allem die Nebensätze oder sogar Einfachsätze können die Verständlichkeit erschweren, denn die miteinander in Beziehung stehenden Satzglieder sind unterbrochen durch mehrere Attribute und damit unübersichtlich aufgestellt. Dies zeigt das fehlende Bemühen der österreichischen Verwaltungstexte um eine einfache und leicht verständliche Gestaltung hinsichtlich der Syntax im Gegensatz zu den bundesdeutschen, bei denen Einfachsätze gegenüber den Satzgefügen dominieren (vgl. Wagner 1984: 47 f.).

Stilistische Besonderheiten

In Anbetracht der Stilmerkmale legt die Verwaltungssprache Wert darauf, dass Entindividualisierung und Sachlichkeit zum Ausdruck kommen. Der Verwaltungsakt steht hierbei im Vordergrund und nicht die davon betroffenen Individuen. Dies wird sichergestellt durch die hohe Prägung des Passivgebrauchs, wobei die Bevorzugung hinsichtlich der passiven Bildung je nach Teilbereich variiert. Zum Beispiel tritt bei österreichischen Rechtstexten häufig das Vorgangs- und Zustandspassiv auf, während Passivumschreibungen mit einem imperativischen Hinweis (etwas ist zu tun) ihren Niederschlag in Schreiben des militärischen Bereichs finden (vgl. Wiesinger 2008b: 118 f.). Als weiteres linguistisches Instrument zur Aufrechterhaltung der Objektivität gilt die eingeschränkte verbale Variabilität. Nur eine begrenzte Anzahl bestimmter Verben wird wiederkehrend gebraucht, um die Eintönigkeit und damit Einheitlichkeit des Inhalts zu sichern. Der geringe Bestand an zulässigen Verben wird durch die nominale Realisierung ausgeglichen, die in der österreichischen Verwaltungssprache ausgeprägt ist. Auch hier äußert sich die unpersönlich-sachliche Charakteristik in einer Dominanz abstrakter Substantive, z. B. Verhältnis, Steuerpflicht oder Gehaltsüberleitung gegenüber solchen, die sich auf Personen oder konkrete Gegenstände beziehen, z. B. Beamter, Finanzamt oder Wehrpflichtiger (vgl. Wiesinger 2008b: 120).

Derartig abstrakte Substantivierungen betreffen auch verbale Wortbildungen mithilfe des Ableitungssuffixes -ung, um abstrakte Verbalsubstantive wie Verfügung, Endigung oder Ermächtigung zu realisieren. Ebenfalls werden Infinitive bereichs- und textspezifisch nominalisiert, wie das Inkrafttreten, das Entstehen oder das Anbringen. Die Präferenz für Abstrakta scheint eine varietätsspezifische Besonderheit der österreichischen Verwaltungssprache zu sein, denn es treten abstrakte und konkrete Substantive bei bundesdeutschen Amtstexten mit ähnlichem Häufigkeitsgrad auf (vgl. Wagner 1984: 28).

Außerdem zeigt sich das Übergewicht an Nomen im Verhältnis zu Pronomen darin, dass dieselben Substantive überwiegend wiederholt werden, anstatt durch entsprechende Pronomen ersetzt zu werden. Dies hängt mit dem Bemühen um „begriffliche Klarheit und Eindeutigkeit“ zusammen und trägt erheblich zum Nominalstil der Verwaltungssprache bei (Wiesinger 2008b: 122). Schließlich zählt der hohe Gebrauch an Mehrfachattributen bei Substantiven ebenfalls zu den stilistischen Eigenheiten, besonders in Gesetzestexten, wobei die Attribute einer ausführlichen Erläuterung und somit der Präzision der Verwaltungssprache dienen.

Lexikalische Besonderheiten

Die österreichische Verwaltungssprache zeichnet sich wie alle anderen Fach- und Sondersprachen durch eine domänenspezifische Terminologie aus. Mithilfe dieser begrifflich definierten Bezeichnungen und Phraseologie lassen sich vielfältige Sachverhalte innerhalb der Verwaltung deutlich und genau artikulieren. Vorwiegend finden die lexikalischen Eigenheiten ihren Ausdruck in Substantiven. Diese können in Form von Simplicia, die jeweils aus einem Wort bzw. einer bedeutungstragenden Einheit bestehen, oder Komposita, somit einer Zusammensetzung zweier oder mehrerer Wörter, vorkommen. Weniger häufig handelt es sich um Verben, Adjektive und Adverbien, die dennoch meistens als Teil von Mehrwortbenennungen, z. B. Wortgruppierungen, Wendungen oder Phrasen, auftreten. Auf die Bildung von sprachlichen Fachbezeichnungen wird in Unterkapitel 5.2 eingegangen.

Aufgrund ihrer ausschließlichen Geltung in Österreich werden die fachspezifischen Wörter und Wendungen zu den amts- und verwaltungssprachlichen Austriaismen gerechnet. Zur Ermittlung dieses Wortschatzes der österreichischen Verwaltungssprache hat Wiesinger (2008b: 124 ff.) das Verhältnis der Fachausdrücke zur Gemeinsprache zum Ausgangspunkt genommen, woraus sich drei Kategorien fachspezifischer Ausdrücke ergeben:

1. Fachliche Ausdrücke anstelle ihrer existierenden gemeinsprachlichen Äquivalente

Die Fachausdrücke in dieser Gruppe weisen einen altmodischen Charakter auf, denn sie beruhen vielfach auf der altherkömmlichen österreichischen Hof- und Kanzleisprache. Trotz des Bestands an modernen Äquivalenten werden diese im Rahmen der Fachkommunikation des juristischen und administrativen Bereichs zugunsten von Eindeutigkeit und Präzision bevorzugt. So werden z. B. Ruhegenuß statt Pension, Verbücherung anstelle von Eintragung in die Amtsbücher bzw. Quästor und nicht Schatzmeister in Amts- und Verwaltungstexten verwendet. Auch wenn die Substantive den größten Anteil ausmachen, finden sich auch in dieser Gruppe einige andere Wortarten, wie

einberichten (mitteilen), unbeachtlich (nicht zu berücksichtigen) und anher (hierher). Außer Wörtern gilt diese Präferenz auch Phraseologien (Fachwendungen): unbeschadet des §... (ohne Beeinträchtigung durch den §...), außer Ansatz bleiben (unberücksichtigt bleiben) und ein anhängiges Verfahren (ein laufendes Verfahren bei Gericht). Die Austriazismen in dieser Gruppe verfügen zum Teil auch über bundesdeutsche Äquivalente, wie Verlassenschaft (Ö) und Erbschaft (D), Kundmachung (Ö) und Bekanntmachung (D), Drucksorte (Ö) und Formular (D) etc. (Wiesinger 2008b: 125)

2. Gemeinsprachliche Ausdrücke mit fachsprachlicher Kontexteinbettung

Wie die erste Gruppe werden die Wörter und Wendungen dieser Kategorie ebenfalls durch die Kanzleikonventionen beeinflusst. Doch besteht der Unterschied zwischen den beiden darin, dass die Wörter und Wendungen der zweiten Gruppe der Alltagssprache entnommen werden. Jedoch repräsentieren sie im Kontext der Verwaltungssprache spezifische Begriffe und haben damit eine Bedeutung, die von jener im gemeinsprachlichen Gebrauch abweicht. Hierzu gehören beispielsweise Ausfluß einer Tätigkeit (= Ergebnis eines Akts), Abfuhr (= Bezahlung) und Anfall des Ruhegehuß (= Eintritt in die Pension).

3. Spezifische verwaltungssprachliche Fachausdrücke

Die spezifischen Fachausdrücke ergeben den größten Teil der österreichischen verwaltungssprachlichen Terminologie und umfassen in erster Linie Substantive, während andere Wortarten vergleichsweise in den Hintergrund treten. Dabei finden sich die Substantive in Form von Simplicia sowie in einer Vielzahl von Komposita. Solche spezifischen Bezeichnungen, insbesondere die Simplicia, sind im jeweiligen entsprechenden Teilbereich der Verwaltung entstanden und weisen zwar keine gemeinsprachlichen Äquivalente auf, können aber teilweise durch das Außenverhältnis der Behörden zu BürgerInnen in das Öffentlichkeitsbewusstsein und damit in die gemeinsprachliche Verwendung eindringen (vgl. Wiesinger 2008b: 126). Infolgedessen kommen beispielsweise die ursprünglich fachspezifischen Begriffe im militärischen Bereich wie Musterung, Einheit und Offizier ebenfalls in der Alltagssprache zur Anwendung. Allerdings ist die Mehrheit der spezifischen Fachausdrücke als ausschließliche Verwaltungstermini zu verstehen. Bei Komposita lässt sich weiter in Einfach- und Mehrfachkomposita gliedern, wobei Letzteres als eine Zusammensetzung von mehr als zwei lexikalischen Einheiten zu verstehen ist. Solche Zusammensetzungen mehrerer Nomen bereiten AdressatInnen wenig Verständnisschwierigkeiten, wenn sich die

Bedeutung des neuen Worts aus der Grundbedeutung der einzelner Glieder überwiegend ableiten lässt. Weist das Kompositum aber eine Eigenbedeutung auf, die nicht mehr direkt mit der Grundbedeutung des Grund- und Bestimmungsworts zusammenhängt, erschwert dies die Verständlichkeit. Ähnliches gilt für fachspezifische Wendungen bzw. Mehrwortbenennungen, die sich ohne Erläuterungen als kaum nachvollziehbar erweisen, z. B. körperliche Wirtschaftsgüter (= Gegenstände, Geräte, Maschinen und Fahrzeuge, die der Abnutzung unterliegen) (vgl. Wiesinger 2008b: 128). Schließlich ist zu bemerken, dass die Bevorzugung der fachspezifischen Adjektivkomposita vom jeweiligen Teilbereich der Verwaltung abhängt. Diese treten beispielsweise häufiger bei Amtstexten im Militär- und Finanzwesen auf, insbesondere jene mit dem Adjektiv -berechtigt wie zeichnungsberechtigt, anspruchsberechtigt bzw. mit -pflichtig und -frei wie steuerpflichtig, wehrpflichtig oder beitragsfrei. Diese Adjektivbildung soll der Satzkürzung sowie der in der Rechts- und Verwaltungssprache angestrebten Exaktheit dienen. Im Gegensatz zu den nominalen Komposita lösen diese adjektivischen Zusammensetzungen kaum bis keine Verständlichkeitsprobleme aus (vgl. Wiesinger 2008b: 127).

2.2.3 Texte im Verwaltungsbereich

Eine weitere Herausforderung bei den Optimierungsbemühungen besteht in der Vielzahl schriftlicher Textsorten, deren Sprachstil sich je nach Funktion und Verwaltungszweck voneinander unterscheidet und sich damit nicht einheitlich gestalten lässt. Zur Kategorisierung der administrativen Textsorten wird auf deren zu erfüllende Aufgaben im Verwaltungshandeln zurückgegriffen. Im Rahmen ihrer linguistischen Untersuchung von Verwaltungsäußerungen hat beispielsweise Wagner (1984) zur Untergliederung administrativer Textsorten anhand deren eingebetteten funktionalen Bedeutung in Institutionen beigetragen. Ausgehend von dieser Zweckbestimmtheit von Verwaltung wurde von Becker-Mrotzek und Scherner (1999; 2000) eine detaillierte Systematisierung erarbeitet. Festgestellt wurde, dass sämtliche Textsorten in der Verwaltung den gemeinsamen Hauptzweck verfolgen, der „Bearbeitung von Wissen bzw. Informationen nach vorgegebenen, ja institutionsspezifischen Bearbeitungsverfahren“ zu dienen (Becker-Mrotzek & Scherner 2000: 634). Dementsprechend stehen der jeweiligen Phase der Informationsbearbeitung einschlägige Textsorten zur Verfügung. In jeder Textsorte sind wiederum, abhängig von deren Teilzwecken im Bearbeitungsprozess, eigene textuelle Strukturen sowie sprachliche Mittel enthalten (vgl. Becker-Mrotzek 2001:1508). Vor diesem Hintergrund werden im Interesse dieser Arbeit Verwaltungstexte ebenfalls nach ihrem jeweils

dominanten funktionalen Zweck im kommunikativen Verwaltungshandeln wie folgt typologisiert:

Texte zur Festlegung des Zuständigkeitsbereichs von Verwaltung

Diese Art Texte weist regelnde und steuernde Funktionen auf, indem die Befugnisse und Handlungsrahmen verschiedener Verwaltungsbehörden sowie -angehörigen hier gesetzlich vorgeschrieben sind (vgl. Wagner 1984: 11; Becker-Mrotzek 1999: 1396). So fallen alle normsetzenden Texte, wie Verwaltungsvorschriften, Dienstanweisungen, Gesetze sowie Durchführungsbestimmungen von Verwaltungstätigkeiten und -abläufen, unter diese Kategorie. Aus textlinguistischer und pragmatischer Sicht sind sie von zentraler Bedeutung, denn sie gelten als primäre Verweisquellen für rechtliche Vorgaben, die dann in anderen Texten „zitiert, paraphrasiert oder genannt“ werden (Becker-Mrotzek & Scherner 2000: 634).

Texte zur Informationsvermittlung zwischen KlientInnen und AgentInnen

Anhand dieser Texte informieren die KlientInnen im Rahmen der kommunikativen Schnittstellen die AgentInnen der Verwaltung schriftlich über die Einzelheiten ihres Anliegens. Hierzu gehören demnach sämtliche Anbringen von KlientInnen, insbesondere Anträge, Beschwerden, Anfragen und Auskunftersuche, die an die Verwaltungseinrichtungen übermittelt werden. Nach Wagner (1984: 11) werden diese schriftlichen Äußerungen der Kategorie des „allgemeinen behördlichen Schriftverkehrs“ zugeordnet. Zur systematischen Informationserhebung werden zumeist Texte in Form von Formularen durch die Verwaltung bereitgestellt (vgl. Becker-Mrotzek & Scherner 2000: 634).

Texte für die verwaltungsinterne Wissensbearbeitung

Texte in dieser Kategorie lassen sich in erster Linie als Verwaltungsakte bezeichnen. Darunter sind alle Schriftstücke zu fassen, die über den Bearbeitungsverlauf in der Verwaltung entstehen und der institutionsinternen Bearbeitung unterzogen werden. Bei der Bearbeitung von Verwaltungsakten ziehen SachbearbeiterInnen relevante Wissensbestände bezüglich der Verwaltungshandlungen heran, um eine Verwaltungsentscheidung über den Sachverhalt der KlientInnen anhand der erhobenen Informationen herbeizuführen (vgl. Seibert 1981: 23). Angesichts dieser Zusammenführung von Wissen beider AkteurInnen spielen laut Becker-Mrotzek (1999: 1396) Formulare deshalb eine wesentliche Rolle, welche die „Schnittstelle in der Bürger-Verwaltungskommunikation“ darstellen.

Texte zur Bekanntgabe der Verwaltungsentscheidung

Hierunter fallen jegliche schriftliche Äußerungen, welche die Ergebnisse des Bearbeitungsverfahrens mitteilen und damit die finale Verwaltungsentscheidung verkünden. Am häufigsten handelt es sich bei dieser „Willensäußerung im Gebiet der öffentlichen Verwaltung“ um Bescheide „als hoheitliche Verwaltungsakte, die Leistungen fordern oder gewähren“ (Wagner 1984: 11; Becker-Mrotzek & Scherner 2000: 634). Dazu zählen beispielsweise Steuerbescheid, Sozialhilfebescheid sowie Erteilung der Aufenthaltsbewilligung und Beschäftigungsgenehmigung.

Wie erwähnt, geht die angeführte Typologisierung von der vorherrschenden Funktion eines jeweiligen Texts aus. Das heißt, unterschiedliche Schreiben in der Verwaltung lassen sich teilweise nicht unbedingt diesen Textsorten einheitlich zuordnen. Dies liegt an der charakteristischen multifunktionalen und mehrfachadressierten Gestaltung von Verwaltungstexten: Ein Verwaltungsschreiben kann sich aus mehreren Textsorten mit differenzierten Funktionen zusammensetzen und zugleich an unterschiedliche RezipientInnen adressiert sein (vgl. Kühn 1995: 4). So können z. B. in einem Antragsformular gleichzeitig Informationen der Verwaltung, Mitteilungen der antragstellenden Person, institutionsinterne Bearbeitungsvermerke und zuletzt der Erteilungsbescheid enthalten sein. Dieses Kennzeichen muss demzufolge bei der Erstellung von Verwaltungstexten berücksichtigt werden, insbesondere wenn kommunikativen Anforderungen einzelner Adressatengruppen mit unterschiedlichem handlungsbezogenem Wissen nachzukommen ist (vgl. Ehlich & Rehbein 1980: 342). Darüber hinaus ist darauf zu achten, bearbeitungsfähige Texte bereitzustellen, sodass die sukzessive Bearbeitung mehrerer AgentInnen aufgrund der hierarchischen Struktur von Verwaltungsprozessen erfolgen kann (vgl. Becker-Mrotzek & Scherner 2000: 634). Als Beispiel dafür gelten Formulare, die sich nicht nur an mehrfache KlientInnen richten, sondern auch für die interne Bearbeitung durch verschiedene zuständige Verwaltungsstellen vorgesehen sind.

3 Formulare als Verwaltung-Bürger-Schnittstelle

Im Verwaltungshandeln werden Formulare häufig als eine der wesentlichsten Textsorten bezeichnet, die als Instrument zum Austauschverhältnis der Behörden zu BürgerInnen dienen (vgl. Grosse 1981: 271). Anhand der obigen Textsortendifferenzierung lässt sich des Weiteren bemerken, dass Formulare nicht nur zur klientenseitigen Informationsübermittlung, sondern auch zur verwaltungsinternen Datenerhebung und -verarbeitung dienen (siehe hierzu Abschnitt 2.2.3). So verfügt diese Textsorte laut Becker-Mrotzek & Scherner (2000: 635) über eine „Doppelfunktion als Kommunikations- und Arbeitsmittel“, was eines ihrer relevantesten Merkmale ausmacht. Terminologisch werden die Begriffe Formular und Vordruck in der Literatur austauschbar für Texte verwendet, die schablonenhaft ganz oder teils vorgefertigt sind und schriftlicher Ergänzungen durch AdressatInnen bedürfen (vgl. Diederich 1980: 96; Fuchs-Khakhar 1987: 105; Toebe-Albrecht 1989: 23). Die besondere Komplexität von Formularen zeigt sich in der fachspezifischen Formulierung und Gestaltung durch sachkundige Verwaltungsangehörige der jeweiligen Institution, obwohl sich diese Lückentexte auch an laienhafte BenutzerInnen mit geringem oder keinem entsprechenden Handlungswissen richten. Infolgedessen gehören Formulare zusammen mit ihrer verbreiteten Nutzung in der Gegenwart zu den „typischsten und schwierigsten Erscheinungsformen der Verwaltungssprache“ (Gülich 1981: 328). In den folgenden Unterkapiteln wird diese spezifische Textsorte samt ihrer Besonderheiten eingehend erläutert.

3.1 Historischer und juristischer Hintergrund von Formularen

Der Ursprung dieser Textsorte geht auf deren Nutzung im rechtlichen Rahmen zu historischen Zeiten zurück. Im Ägypten der Spätzeit waren z. B. Textbausteine in bestimmten Reihenfolgen geordnet und festgeschrieben, um die Authentizität der Texte wie Urkunden oder Urteile zu gewährleisten (vgl. Seidl 1962: 41). Etymologisch stammt der Begriff Formular aus der Bezeichnung eines juristischen Verfahrens, dem Formularprozess (lignite per formulas) im antiken Rom. Diese Bezeichnung hängt mit den damals gebräuchlichen Sammlungen von Textformeln zusammen, die von einem höchsten Justizbeamten formuliert wurden und dem Richter im Rahmen der Rechtsprechungen des Zivilrechtprozesses dienten. Die formulas stellten sich somit als von der obersten Instanz erstellte schriftliche Äußerungen dar, die normgebende Anweisungen für Gerichtsentscheidungen in bestimmten Rechtsstreiten erteilten (vgl. Fotheringham 1980: 25). Sogenannte Formelsammlungen im Rechtsbereich fanden ebenfalls in den westlichen Germanenreichen Anwendung. Dabei verstanden sich diese als von

Juristen festgelegtes beispielhaftes Muster von Textelementen, das die Gerichtsbeamten zur Aufzeichnung von Ergebnissen der Beilegung in konkreten Rechtsfällen verwendeten. Neben dem schriftlichen Inhalt bezüglich entsprechender Rechtshandlungen enthielten diese Formeln auch Leerstellen, die mit Namen von betroffenen Personen, Beteiligten sowie Orts- und Zeitangaben auszufüllen waren (vgl. Fotheringham 1980: 26 f.). Der ursprüngliche Zweck dieser Textform hat sich zu der gegenwärtigen Verwendung von Formularen für eine einheitliche Datenerfassung weiterentwickelt. Darüber hinaus nahmen die damaligen Textelemente der Formeln aufgrund ihrer Entstehung in den Verwaltungskanzleien erheblichen Einfluss auf die Verwaltungssprache und untermauerten zum Großteil die Entwicklung der neuhochdeutschen Schriftsprache (vgl. Grosse 1981: 268).

Die enge Verknüpfung von Recht und Verwaltung sowie der dazugehörigen Schriftstücke trifft, wie zuvor dargestellt, auch in der Gegenwart zu. Ordnungsgemäß müssen sprachliche und textuelle Elemente innerhalb jeglicher Verwaltungstexte und damit Vordrucke, die als Teil des jeweiligen förmlichen Verwaltungsakts eingesetzt werden, mit den entsprechenden Rechtsgrundlagen in Einklang stehen. Dies ist auf die Neubestimmung des Verhältnisses zwischen Staat und BürgerInnen ab den 1970er-Jahren zurückzuführen, wodurch die Verwaltung ihre institutionell-hierarchische Obrigkeit gegenüber den BürgerInnen ablösen ließ und eine zugänglichere serviceorientierte Position einnahm (vgl. Fluck 2008: 117). In Bezug auf Österreich zeigt sich dies in der zunehmenden Demokratisierung des bundesverwaltungsgerichtlichen Handlungsrahmens, die im Jahr 1987 zur Gesetzgebung bezüglich der Auskunftspflicht der Verwaltung und der Klagemöglichkeit gegen Verwaltungsentscheidungen beigetragen hat (vgl. Ringel & Reischauer 2019: 229). Seitdem sind die österreichischen Bundesorgane dazu verpflichtet, jeden Verwaltungsakt samt entsprechenden Prozessen der Entscheidungsfindung juristisch nachvollziehbar und transparent zu gestalten: Alle Vorgänge im Rahmen des Verwaltungshandelns sind mit Verweis auf gesetzliche Grundlagen schriftlich anzugeben, um die rechtliche Überprüfung durch das Verwaltungsgericht und die Öffentlichkeit zu ermöglichen (vgl. Becker Mrotzek & Scherner 2000: 636 f.). Als Folge wirkt sich diese neue Entwicklung des Verwaltungsrechts auf die Durchführung von Verwaltungstätigkeiten und somit auf die Formulierung der damit einhergehenden Texte einschließlich Formulare aus. Die Verwaltungstexte dienen somit der „Reverbalisierung“ (Klein 2014: 18) in Rechtsbestimmungen und -vorschriften festgelegter normativer Konzepte. Konkret handelt es sich um die schriftliche Bezugnahme auf rechtliche Bestimmungen wie Verordnungen und Gesetzestexte, was aber zum Teil zur Schwerverständlichkeit dieser Textform führt. Vor allem werden die fachsprachlichen

Spezifika wie Termini und Idiomatik der Rechts- und Verwaltungssprache häufig vollständig in den Text der Formulare übernommen, weil die Vereinfachung oder Umschreibung in die Gemeinsprache aus juristischer Sicht zum Bedeutungsverlust bzw. zur Mehrdeutigkeit führen könnte (vgl. Fotheringham 1980: 29). Darüber hinaus werden die möglichen Kommunikationsbarrieren, denen die KlientInnen begegnen, kaum berücksichtigt. Die amtsinternen VordrucksverfasserInnen gehen bei der Textgestaltung tendenziell von ihrer eigenen „berufsbedingten fehlenden Distanz zum Sach- und Wortverhalt“ aus, sodass die Bedeutung der Fachbegriffe für sie selbstverständlich ist (Grosse 1980: 17). Die übliche Übergangslösung scheint laut der Untersuchung von Helbig (1980) die Hinzufügung einer Erklärungslegende für fachliche Ausdrücke zu sein, was wiederum zusätzliche Textmengen für die KlientInnen bedeutet und als überfordernd empfunden wird (vgl. Helbig 1980: 49).

3.2 Zwecksetzung und Struktur von Formularen

Die zweifache Funktion von Formularen äußert sich in ihrer Rolle innerhalb des Verwaltungshandelns, das aus handlungstheoretischer Sicht nach Lüdenbach und Herrlitz (1981: 307) als „konditionales Entscheidungsprogramm“ betrachtet wird: Eine bestimmte Handlung erfolgt erst, wenn die Voraussetzungen erfüllt werden. Zum Beispiel muss eine Genehmigung zur Niederlassung in Österreich erteilt werden, wenn der Antrag allen Auflagen genügt. So hat die Verwaltung in diesem Zusammenhang die Aufgabe, die erforderlichen Bedingungen zu überprüfen und hier kommen die Formulare zum Einsatz. Mit ihrer Hilfe kann „eine geregelte Interaktion“ zwischen den AgentInnen und KlientInnen stattfinden, indem die für die Entscheidungsfindung benötigten Informationen strukturiert und einheitlich übermittelt und erhoben werden (Brinckmann et al. 1986: 122). Damit wird der erste Teilzweck der Formulare als Erhebungsinstrument abgeschlossen. Daraufhin kann die zweite Teilaufgabe der Formulare beginnen, die in der verwaltungsinternen Bearbeitung besteht. Hier werden sie nach der arbeitsteiligen Organisationsstruktur der Verwaltung schematisch bearbeitet und es wird ermittelt, ob die erfassten Informationen der individuellen Anliegen den relevanten Rechtsbestimmungen zugeordnet werden können. Daraus ergeben sich Entscheidungen, deren Bekanntgabe in Form von Beschlüssen oder Bescheiden erfolgt. Demzufolge lässt sich feststellen, dass Formulare „ein wesentlicher Bestandteil der Akte, der verwaltungstypischen Form der Vorgangsbearbeitung“ bilden (Helbig 1980: 46). Angesichts ihrer Stellung im Verwaltungshandeln müssen Formulare vielseitigen Erfordernissen gerecht werden. Einerseits sollte bei ihrer Gestaltung als Kommunikationsmittel das laienhafte Handlungswissen der

KlientInnen und damit die Verständlichkeit berücksichtigt werden (vgl. Becker-Mrotzek 1999: 1398 f.). Andererseits müssen sie als Arbeitsmittel im Rahmen der Verwaltungsakte nicht nur den rechtlichen Vorschriften nachkommen, sondern auch die technischen Voraussetzungen im Sinne einer Bearbeitungsfähigkeit erfüllen.

Die spezifische Textstruktur von Formularen wird ebenfalls durch ihre doppelte Funktion bestimmt. In der Regel enthält die Makrostruktur eines Formulars die folgende Komponente: Name der ausstellenden Einrichtung, Bezeichnung des Formulars und der damit verbundenen Verwaltungshandlung, Anleitung zum Ausfüllen, Erhebungsteil, Ermächtigungs- bzw. Einverständniserklärung, Bestätigung über die Richtigkeit der Angaben und Feld für Datum sowie Unterzeichnung der den Antrag stellenden KlientInnen (vgl. Becker Mrotzek & Scherner 2000: 637). Der Erhebungsteil stellt im Hinblick auf die Zwecksetzung dieser Textsorte den wesentlichsten Bestandteil dar, denn die KlientInnen haben hier die Möglichkeit, erforderliche Informationen für die Verwaltungshandlung schriftlich mitzuteilen. Dies erfolgt anhand der vorkonzipierten Erhebungsformen, die sich laut Helbig (1980: 50 ff.) wie folgt untergliedern lassen: Vollständige Fragen mit Leerräumen für Antworten, stichwortartige Fragen mit Leerstellen, vorformulierte Äußerungen mit Textlücken zum Ergänzen und vorformulierte Optionen mit Kästchen zum Ankreuzen. In Anbetracht dieser textsortenspezifischen Elemente lässt sich bemerken, dass es sich bei Formularen in erster Linie um formalisierte Sprechhandlungen handelt. Dieser Informationsaustausch wird daher von Gülich (1981: 341) als „standardisierter schriftlicher Dialog zwischen einer Institution und einem Klienten“ beschrieben. Trotz der Bemühung um dialogische Konzipierung fehlen Formularen in der Praxis Kommunikationsmittel, die sonst für mündliche Interaktion typisch sind, wie unmittelbare Nach- und Rückfragen (vgl. Stickel 1981: 297). Aus diesem Grund werden Formulare zumeist mit zusätzlichen Texten zur Verfügung gestellt, insbesondere Erklärungen bezüglich der Verwaltungshandlung sowie Hinweisen für AusfüllerInnen, um das Verstehen der antragstellenden Personen zu sichern (vgl. Brinckmann et al. 1986: 162 f.).

3.3 Formular und Verwaltungssprache in der Diskussion

Aus den obigen Erläuterungen geht hervor, dass sich die gegenwärtigen Formulare als „hochstandardisiertes Kommunikationsmittel“ (Albrecht 1980: 76) darstellen und überwiegend auf verwaltungsinterne Vorlagen wie Dienstvorschriften und behördliche Anforderungen ausrichten. Darüber hinaus erfordert der Anspruch auf Verwaltungstransparenz die schriftliche Einbindung von Gesetzestexten innerhalb der Formulare, was die fachsprachliche Komplexität der Rechts- und Verwaltungssprache wiederum nach sich zieht. Dabei tritt die Anforderung an

Informationsfähigkeit seitens der KlientInnen aber in den Hintergrund. Dies hat eine Asymmetrie der Interaktion zur Folge, wobei sich die AgentInnen und die meistens laienhaften KlientInnen im Rahmen des Verwaltungshandelns auf deutlich unterschiedlichen Ebenen befinden, insbesondere jene des Fachwissens und Informationshorizonts (vgl. Stickel 1981: 296). Trotz des unzureichenden Handlungswissens haben die KlientInnen zumeist zwangsläufig mit unterschiedlich vorgegebenen Handlungsformen und Rahmenbedingungen umzugehen, die je nach Verwaltungsakten variieren und durch die Schwer- bzw. Unverständlichkeit aufgrund der sprachlichen und juristisch-technischen Faktoren verkompliziert werden (vgl. Efing 2020: 289). Die sich daraus ergebende Handlungsbeschränkung veranlasst eine allgemein negative Haltung der Öffentlichkeit gegenüber Formularen sowie den Unwillen von BürgerInnen, sich mit diesen auseinanderzusetzen (vgl. Grosse & Mentrup 1980: 123; Fisch & Margies 2014: 9). Das mangelnde Bewusstsein für die Bedeutung der Formulare als amtsexterne Texte und deren entsprechende Funktion lässt die „von Regierung und Ämtern gewünschte Überbrückung der Distanz“ schwer realisieren und trägt vielmehr zur „Staats- und Verwaltungsverdrossenheit“ bei (Grosse 1980: 12).

So bestand das Ziel zweier wissenschaftlicher Arbeitstagungen des IDS in Mannheim (Grosse & Mentrup Hg. 1980) sowie der Akademie der Sprache und Dichtung in Darmstadt (Radtke Hg. 1981) darin, das allgemeine Problembewusstsein in Bezug auf Formulare als wesentliches Medium des Amt-Bürger-Kommunikationskreises zu erhöhen und Verbesserungsmöglichkeiten zu fördern. Dabei wurde hervorgehoben, dass die Verwaltung sich aufgrund des neubestimmten, zunehmend demokratisierten Verhältnisses von Staat und BürgerInnen als Dienstleister sehen sollte, der nach transparenter Nachverfolgbarkeit und Kommunikation strebt (vgl. Becker Mrotzek & Scherner 2000: 633). Demzufolge waren die beiden Fachtagungen der Untersuchung der Gestaltung und der damit verbundenen Verständlichkeit von Vordrucken im terminologischen und syntaktischen Bereich der Verwaltungssprache sowie jenen des Textaufbaus gewidmet. Konkrete Vorschläge zur bürgernahen und leicht verständlichen Gestaltung von Formularen finden sich unter anderem in den empirischen Studien von Helbig (1980) sowie von Toebe-Albrecht (1989), die nicht nur Ansätze zum inhaltlichen, sondern auch zum äußeren Vordruckentwurf wie Typografie, Farben und Format beinhalten. Außerdem spricht Mentrup (1980: 122 f.) z. B. von Vermeidung der veralteten Rechts- und Verwaltungsbegriffe, Hervorhebung der relevantesten Gegenstände, übersichtlicher und klar visualisierter Textgliederung sowie der Möglichkeit der persönlichen Beratung durch die zuständigen SachbearbeiterInnen beim Ausfüllen. Hinsichtlich der

österreichischen Verwaltungssprache, mit der sich ebenfalls intensiv befasst wird, steht nach wie vor laut Wiesinger (2008b: 130) eine „regelmäßige, in Österreich bislang fehlende interdisziplinäre Zusammenarbeit von Juristen und Ämtern mit Sprachwissenschaftlern“ zur bürgernahen Gestaltung von Rechts- und Verwaltungstexten aus. Außer den linguistischen Untersuchungen von Wiesinger (2008b) und Wagner (1984) enthalten sonstige Literaturquellen zur Verwaltungssprache vorwiegend belehrende Absichten, ohne dabei die Textfunktionen zu beachten, und bieten lediglich allgemeine und vereinfachte Hinweise, welche die Zweckbestimmtheit dieser Sondersprache neglieren, wie z. B.: „[D]ienstliche Anweisungen sollen klar, verständlich und so kurz wie möglich abgefasst werden“. (Otto 1981: 44)

Die bestehenden Diskurse dienen, wie erwähnt, der Sensibilisierung für die Problematik der Verständlichkeit, die Amtsformulare sowie die Rechts- und Verwaltungssprache im Allgemeinen mit sich bringen. Becker Mrotzek und Scherner (2000: 635) zufolge fehlen weiterhin intensive linguistische Untersuchungen von Formularen aus verschiedenen Verwaltungsbereichen, einschließlich einer umfangreichen systematischen Behandlung der fachsprachlichen Problematik in Theorie und Praxis, in Forschung und Ausbildung. Mit der vorliegenden Masterarbeit soll daher mindestens eine Forschungslücke bezüglich der österreichischen Formularsprache im folgenden praktischen Teil erkundet und geschlossen werden. Der Ausgangspunkt der Auseinandersetzung mit dem Korpus offizieller Antragsformulare der österreichischen Einwanderungsbehörden leitet sich von Vorschlägen aus dem Resümee von Mentrup (1980: 124), Fluck (2008: 134) und Lambertz (1999: 157) ab, die vor allem die Notwendigkeit für terminografische Lösungsansätze, wie die Ausarbeitung ein- und mehrsprachiger Fachwortsammlungen betonen, die den Umgang der RezipientInnen mit der Amtssprache erleichtern sollen. Ebenfalls heben Arntz et al. (2004: 24) hinsichtlich der Kommunikationsoptimierung mithilfe der sprachlichen und terminologischen Untersuchung hervor:

Die Verständigung zwischen Fachleuten und Laien und zwischen Fachleuten verschiedener Gebiete lässt sich nur dann sichern bzw. erleichtern, wenn immer mehr am Kommunikationsprozess Beteiligte sich aktiv um diese Verständigung bemühen und sich in diesem Zusammenhang mit sprachlichen Fragen auseinandersetzen.

4 Migration und behördliches Verfahren

In diesem Kapitel wird zunächst dargelegt, was unter Migration grundsätzlich zu verstehen ist. In diesem Zusammenhang werden Faktoren bzw. Ursachen erläutert, die Menschen zu dieser motivieren und dabei verschiedene Typen von Migration veranlassen. Vor diesem Hintergrund sind die weiteren Unterkapitel dem Thema Einwanderung und Aufenthalt in Österreich gewidmet, wobei ein besonderer Fokus auf das behördliche Antragsverfahren und die damit verbundene Verständigung gelegt wird.

4.1 Motivation und Muster von Wanderungsbewegungen

Bei Migration handelt es sich um eine allgemeinmenschliche Handlung, bei der sich ein Individuum oder eine Gruppe über Grenzen hinweg bewegt, um seinen bzw. ihren gewöhnlichen Aufenthaltsort einmal, wiederholt oder sogar permanent räumlich zu wechseln (vgl. Faist et al. 2013: 1 ff.; Mecheril et al. 2010: 7). Grenzüberschreitende Bevölkerungsbewegungen haben im Laufe der Jahrhunderte beinahe überall auf dem Globus stattgefunden und sind deshalb tief in die menschliche Geschichte eingebettet (vgl. Mecheril 2007: 470). So wird im Hinblick auf dieses anthropologische Grundmuster von *homo migrans* gesprochen (vgl. Burmeister 2012: 31). Dass bestimmte Diasporagemeinden in einigen Einwanderungsländern mittlerweile größer als die Bevölkerung in den Herkunftsländern selbst sind, bestätigt den bereits herrschenden Trend der Migrationsgeschichte (vgl. Lucassen et al. 2010: 3 f.). Als signifikanter Beweggrund der historischen Wanderungen verschiedener Völker gilt laut Nuscheler (2004: 29) die Disproportionalität zwischen verfügbarem Ressourcenbestand und steigender Nachfrage aufgrund des Bevölkerungswachstums, die zu Rohstoffknappheit und Nahrungsmittelkrisen sowie weiteren „existenzbedrohenden Notlagen“ führt.

Zu den bekanntesten theoretischen Grundlagen im Rahmen der Migrationsforschung gehört das sogenannte Push-Pull-Modell, das die Ursachen grenzüberschreitender Menschenbewegungen darstellt. Im engeren Sinne handelt es sich bei diesem Konzept um das Zusammenspiel zwischen verschiedenen abstoßenden (Push-)Faktoren des meistens ärmeren Herkunftslands und unterschiedlichen anziehenden (Pull-)Faktoren des üblicherweise reicheren Ziellands (vgl. Schoorl et al. 2000: 3). Zu den häufigsten Push-Faktoren, welche die Abwanderung aus dem ursprünglichen Gebiet verursachen, zählen unter anderem Krisensituationen, beispielsweise (bürger-)kriegerische Auseinandersetzungen, rassistische oder religiöse Unterdrückung und Verfolgung gegen bestimmte Bevölkerungsgruppen,

Naturkatastrophen, aber auch wirtschaftliche Not wie Armut und Hunger. Hingegen umfassen die Pull-Faktoren Vorteile im Hinblick auf politische, gesellschaftliche und ökonomische sowie ökologische Aspekte, die eine reizende Migrationswirkung auf Individuen ausüben (vgl. Neumayer 2004: 9). Die Push- und Pull-Faktoren interagieren miteinander und beeinflussen je nach Kombination das Ausmaß und die Richtung der Migrationsströme (vgl. Portes und Böröcz 1989: 607).

Vor allem die zunehmende Globalisierung und Digitalisierung, welche die Erde durch moderne Technologieentwicklungen räumlich und zeitlich schrumpfen lassen, spielen eine bedeutende Rolle in der Migrationsentscheidung und somit -bereitschaft von Menschen. Während die neuen Schöpfungen im Bereich Verkehrstechnik zur höheren Mobilität beitragen, lässt sich das Ideal des reizvollen westlichen Lebensstils mit gegenwärtigen Kommunikationsmöglichkeiten sowie elektronischen Massenmedien und sozialen Netzwerken weltweit verbreiten (vgl. Luft 2016: 16). Aus den genannten und weiteren individuellen Gründen ergibt sich insbesondere angesichts der zunehmenden wirtschaftlichen und sozialen Ungleichheit ein rasanter Anstieg internationaler Migration, deren verschiedene Muster sich von Nuscheler (2004: 52 f.) grundsätzlich wie folgt unterteilen lassen: „[F]reiwillige Auswanderung oder Emigration“; „[D]ie aus dem Schutz der Familie abgeleitete, aber z. B. beim Zuzugsalter unterschiedliche geregelte Familienzusammenführung“; „[D]ie illegale, nach der Sprachregelung der Vereinten Nationen ‚irreguläre‘ oder nach angelsächsischer Diktion ‚nichtdokumentierte Migration‘ (weil die Zuwanderer ohne gültige Papiere einreisen)“; „[Z]eitlich begrenzte Auslandsaufenthalte, z. B. von Studierenden“; und „[D]ie durch Kriege, politische Verfolgung, existenzielle Notlagen oder Umweltkatastrophen erzwungene Flucht“.

Trotz der Untergliederung tritt die Migration nicht zwangsweise in einem deutlich erkennbaren Muster auf, sondern kann häufig durch unterschiedliche Motive bedingt werden, sodass es zur Überschneidung mehrerer Wanderungsformen kommt (vgl. Nuscheler 2004: 53). Allerdings besteht ein klarer Unterschied zwischen den Wanderungsmustern Flucht und Migration. Das Flüchtlingskommissariat der Vereinten Nationen (UNHCR) bezeichnet Flüchtlinge als Personen, die vor individueller Verfolgung aufgrund ihrer politischen Meinung oder Zugehörigkeit zu einer ethnischen Minderheit oder vor bewaffneten Konflikten in ihrer Heimat fliehen (vgl. UNHCR 2018). Dagegen verlassen MigrantInnen meistens freiwillig ihr Herkunftsland für bessere Lebensbedingungen oder aus familiären Gründen und dürfen sich gegebenenfalls dauerhaft im Einwanderungsland niederlassen oder ohne Gefahr für Leib und Leben in ihre Heimat zurückkehren (vgl. ebd.). Es gilt somit lediglich Letzteres als relevant für

die vorliegende Masterarbeit, denn der Schwerpunkt liegt auf den sprachlichen Aspekten des Verwaltungsverfahrens bezüglich der Einwanderung und des Aufenthalts in Österreich.

4.2 Einwanderung und Aufenthalt in Österreich

Seit dem Jahr 1945 nach dem Zweiten Weltkrieg sind Österreich sowie mehrere westeuropäische Staaten zum Einwanderungsland geworden und blicken somit auf eine lange Zuwanderungsgeschichte zurück. Im Laufe der Nachkriegszeit war die Migration hauptsächlich, wenn nicht ausschließlich, innerhalb Europas eingeschränkt: Vor allem ab Mitte der 1950er-Jahre stieg in West- und Zentraleuropa folglich des Wirtschaftsaufschwungs der Bedarf nach Arbeitskräften kontinuierlich an, sodass die Länder in dieser Region bis Anfang der 1970er-Jahre rund 15 Millionen GastarbeiterInnen überwiegend aus ihren südeuropäischen Nachbarstaaten und dem Mittelmeerraum anwarben (vgl. Perchinig 2009: 231).

Im Hinblick auf Österreich stammten die ausländischen Arbeitskräfte in erster Linie aus Südosteuropa, insbesondere aus Jugoslawien und Ostthrakien (Europäische Türkei). Als Folge wuchs die Zahl an GastarbeiterInnen über die 1960er-Jahre bis zum Beginn der 70er-Jahre um etwa 20 000 bis 40 000 Personen. So erlebte Österreich im Jahr 1973 den ersten Höhepunkt der Arbeitsmigration mit beinahe 230 000 EinwanderInnen (vgl. Blasge 2014: 122; Fassmann & Münz 1995: 41). Die europäische Binnenmigration erweiterte sich ab den 1980er-Jahren um EinwanderInnen aus den Ländern außerhalb des Kontinents, somit aus den sogenannten Drittstaaten. Seitdem handelt es sich nicht nur um die GastarbeiterInnen und deren Familienangehörige aus den ärmeren strukturschwächeren Ländern, sondern auch um beruflich hochqualifizierte internationale ZuwanderInnen, die am österreichischen Arbeitsmarkt teilnehmen dürfen (vgl. Nuscheler 2004: 33).

So gibt es aktuell in Österreich laut Statistik Austria (2020a) von insgesamt 8,9 Millionen EinwohnerInnen fast 1,5 Millionen Personen, die über eine nichtösterreichische Staatsbürgerschaft verfügen. Diese bilden knapp über 16,5 Prozent der gesamten Bevölkerung ab und umfassen Angehörige aus der Schweiz, verschiedenen Drittstaaten und den Mitgliedstaaten des Europäischen Wirtschaftsraums (EWR). Hierzu gehören alle Mitgliedstaaten der Europäischen Union zusammen mit Island, Liechtenstein und Norwegen. Spezifisch machen die Staatsangehörigen der Europäischen Union (EU) mehr als die Hälfte der MigrantInnen aus, wobei die meisten aus Deutschland stammen, gefolgt von Rumänien und Ungarn. Unter den ZuwanderInnen aus den Drittstaaten stellen die serbischen, türkischen und bosnischen Staatangehörigen die größten Gruppen dar (vgl. Statistik Austria 2020b). Solange ihr Aufenthalt in Österreich länger als drei bzw. sechs Monate beträgt, bedürfen alle

ausländischen ZuwanderInnen, auch als Fremde bezeichnet, einer Aufenthaltsgenehmigung und unterliegen somit dem österreichischen Niederlassungs- und Aufenthaltsgesetz (NAG). Im Rahmen des NAG wird das Aufenthaltsrecht sowohl für die EWR-BürgerInnen und SchweizerInnen als auch für die Drittstaatsangehörigen geregelt, indem die Voraussetzungen zur Erteilung, Versagung und Entziehung der Aufenthaltsgenehmigung vorgesehen sind (vgl. MA 35 Fachbereich Einwanderung).

Um einen Überblick über das behördliche Verfahren und dessen Bedeutung im Bereich Einwanderung zu schaffen, werden in den folgenden Abschnitten allgemeine Informationen zu verschiedenen Aufenthaltszwecken und -nachweisen je nach Staatsangehörigkeit dargestellt. Sie gehen einher mit detaillierten Voraussetzungen und Verfahrensabläufen für Aufenthalt und Niederlassung in Österreich, was hauptsächlich den Websites <https://www.migration.gv.at>, <https://bmi.gv.at/> und <https://www.oesterreich.gv.at> zu entnehmen ist, die vom Bundeskanzleramt und dem Bundesministerium für Inneres in Zusammenarbeit mit dem Bundesministerium für Arbeit, Soziales, Gesundheit und Konsumentenschutz betrieben werden.

4.2.1 Aufenthalt der EWR-BürgerInnen und SchweizerInnen

In der Regel dürfen sich alle EWR-BürgerInnen und SchweizerInnen einheitlich gemäß der unionsrechtlichen Bestimmungen in der Europäischen Union frei bewegen, in jeden Mitgliedsstaat einreisen und sich dort für einen bestimmten Zeitraum aufhalten oder sich dauerhaft niederlassen. Hinsichtlich des Aufenthalts im österreichischen Bundesgebiet gilt die Visumfreiheit für EWR-BürgerInnen und SchweizerInnen in den ersten drei Monaten. Wird die Aufenthaltsdauer nach drei Monaten überschritten, besteht das Aufenthaltsrecht ausschließlich für Personen mit selbständigem oder unselbständigem Beschäftigungsverhältnis im Bundesgebiet, aber auch für Nichterwerbstätige mit nachweislich ausreichenden eigenen Existenzmitteln und einem umfassenden Krankenversicherungsschutz, z. B. PensionistInnen, Studierende oder Auszubildende. Unabhängig von ihrer Staatsangehörigkeit können Familienangehörige dieser Personengruppe ebenfalls dasselbe unionsrechtliche Aufenthaltsrecht in Anspruch nehmen, solange die entsprechenden Voraussetzungen gegeben sind. Zur Sicherung des Aufenthaltsrechts sind EWR-BürgerInnen und SchweizerInnen sowie ihre Angehörigen dazu verpflichtet, einen Nachweis des Aufenthaltsrechts in Form sogenannter Dokumentation zu beantragen. Die verschiedenen Typen der Dokumentation lassen sich je nach

Aufenthaltsdauer und -bedingungen und im Falle von Familienangehörigen je nach Staatsangehörigkeit wie folgt unterteilen:

Anmeldebescheinigung und Bescheinigung des Daueraufenthalts

EWB-BürgerInnen und SchweizerInnen sowie deren Familienangehörige mit einer EWB- bzw. schweizerischen Staatsangehörigkeit haben bei einer beabsichtigten Aufenthaltsdauer von länger als drei Monaten binnen vier Monaten nach der Einreise ins Bundesgebiet diese Art Dokumentation zu beantragen. Die Gültigkeitsdauer ist unbefristet, solange Aufenthalt in Österreich besteht. Personen, die bereits über eine Anmeldebescheinigung verfügen und sich ununterbrochen fünf Jahre rechtmäßig in Österreich aufhalten, erwerben das Recht auf Daueraufenthalt. Auf Antrag wird eine Bescheinigung des Daueraufenthalts erteilt. Zusätzlich können sie bei Bedarf einen Lichtbildausweis für EWB-BürgerInnen ausstellen lassen, der gleichzeitig als Identitätsdokument fungiert.

Aufenthaltskarte und Daueraufenthaltskarte

Diesen Nachweis des Aufenthaltsrechts erhalten auf Antrag Familienangehörige, die selbst Drittstaatsangehörige sind. Die Karte ist generell fünf Jahre gültig, kann aber bei geplanter kürzerer Aufenthaltsdauer auch kürzer gelten. Die Erteilung der Daueraufenthaltskarte unterliegt denselben Aufenthaltsbedingungen für die Bescheinigung des Daueraufenthalts, gilt aber den Familienangehörigen von EWB-BürgerInnen oder SchweizerInnen aus Drittstaaten. Die Gültigkeitsdauer beträgt zehn Jahre.

Der Antrag auf Ausstellung oder die Verlängerung einer Dokumentation ist nach erfolgter Einreise bzw. erneut nach dem Ablauf der Gültigkeitsdauer persönlich bei der örtlich zuständigen österreichischen Einwanderungs- und Niederlassungsbehörde per Formular zu stellen. Die Entscheidung über die Erteilung und Aushändigung der Dokumentation erfolgt, sobald die SachbearbeiterInnen die Vorlage der Voraussetzungen anhand der vorgelegten erforderlichen Unterlagen und befüllten Antragsformulare überprüft haben.

4.2.2 Aufenthalt der Drittstaatsangehörigen

Drittstaatsangehörige sind Personen, die weder über EWB- noch schweizerische Staatsbürgerschaft verfügen. Für Aufenthalte unter sechs Monaten sind sie zwar bei der Einreise nach Österreich visumpflichtig, benötigen jedoch keine weitere Aufenthaltsgenehmigung. Sofern es sich um eine Aufenthaltsdauer von über sechs Monaten

oder um eine dauerhafte Niederlassung in Österreich handelt, müssen sie gemäß dem NAG einen Antrag auf einen entsprechenden Aufenthaltstitel, somit ein Dokument der Aufenthaltsberechtigung, stellen. Je nach persönlichem Zweck des Aufenthalts in Österreich kommen für Drittstaatsangehörige die folgenden Aufenthaltstitel infrage:

Aufenthaltstitel zur Ausübung einer beruflichen Tätigkeit

Der Aufenthaltstitel in dieser Kategorie berechtigt Drittstaatsangehörige sowohl zur Niederlassung als auch zur beruflichen Beschäftigung in Österreich. Allerdings handelt es sich nicht um eine einheitliche Berechtigung, sondern es bestehen verschiedene Arten des Aufenthaltstitels. Für die jeweilige Art des Aufenthaltstitels gelten unterschiedliche Kriterien, z. B. Tätigkeitsbereich, Qualifikationen und Beschäftigungsbedingungen. Dementsprechend werden unterschiedliche Rechte gewährt, darunter der Umfang des Zugangs zum österreichischen Arbeitsmarkt und mögliche Ableitung des Aufenthaltsrechts für weitere Familienangehörige der antragstellenden Personen. Beispielsweise ermöglichen die Rot-Weiß-Rot-Karte und Blaue Karte EU zwar eine Zuwanderung qualifizierter Arbeitskräfte aus Drittstaaten nach Österreich, sehen aber unter anderem differenzierte Kriterien bezüglich der Beschäftigungsbedingungen vor: Während der erstgenannte Aufenthaltstitel außer der Arbeitsaufnahme bei einem bestimmten Arbeitgeber ebenfalls eine selbstständige Erwerbstätigkeit zulässt, setzt letzterer ausnahmslos ein verbindliches Arbeitsplatzangebot voraus.

Aufenthaltstitel zum Familiennachzug

Die Möglichkeit und die Art des gültigen Aufenthaltsrechts sowie der Umfang des Arbeitsmarktzugangs für nachziehende Familienangehörige hängen vor allem vom Aufenthaltstitel der zusammenführenden Ankerpersonen in Österreich ab. Zum Beispiel erfolgt die Gewährung von Aufenthaltstiteln an Familienangehörige von österreichischen StaatsbürgerInnen, von drittstaatsangehörigen BesitzerInnen einer Rot-Weiß-Rot-Karte, einer Blauen Karte EU sowie von aufenthaltsberechtigten ForscherInnen aus Drittstaaten ohne Begrenzung. In allen anderen Fällen ist hingegen ein vorhandener Quotenplatz erforderlich. Die Quote wird jedes Jahr unterschiedlich durch die Bundesregierung im Einvernehmen mit dem Nationalrat beschlossen und bestimmt die Anzahl der zu erteilenden Aufenthaltstitel zum Zweck der Familienzusammenführung. Im Sinne des NAG zählen grundsätzlich EhegattInnen, eingetragene PartnerInnen bzw. minderjährige (leibliche oder adoptierte) Kinder unter 18 Jahren zu Familienangehörigen. Ein Aufenthaltstitel darf allerdings nur an einen erweiterten

Angehörigenkreis erteilt werden, solange die zusammenführende Person die österreichische Staatsbürgerschaft hat und das jährliche Kontingent noch nicht ausgeschöpft ist.

Aufenthaltstitel zu Ausbildungszwecken

Diese Art Aufenthaltstitel umfasst die Aufenthaltsbewilligung für Studierende und SchülerInnen. Damit sind die InhaberInnen des Aufenthaltstitels zum Studium an einer tertiären Hochbildungsinstitution bzw. zum Schulbesuch berechtigt. Ihnen ist es ebenfalls erlaubt, eine Erwerbstätigkeit auszuüben, solange diese sich nicht beeinträchtigend auf die Ausbildung als primärer Aufenthaltszweck auswirkt und eine Beschäftigungsbewilligung für eine Arbeitsaufnahme bei einem Arbeitgeber nachgewiesen wird. Da der Aufenthaltstitel dieser Kategorie in der Regel für zwölf Monate ausgestellt wird, ist eine Verlängerung jedes Jahr bis zum Studiums- bzw. Schulabschluss zu beantragen. Um die Erneuerung der Aufenthaltsbewilligung zu sichern, muss ein Studien- bzw. Schulerfolg im vorhergehenden Studienjahr bei der Antragstellung nachweislich vorliegen. Für HochschulabsolventInnen besteht dazu die Möglichkeit, die Aufenthaltsbewilligung einmalig um weitere zwölf Monate zum Zweck der Arbeitssuche oder der Unternehmensgründung verlängern zu lassen. Sollte dieses eine oder andere Unterfangen ihnen gelingen, können sie einen Umstieg auf einen entsprechenden Aufenthaltstitel zur Erwerbstätigkeit beantragen.

Aufenthaltstitel zu sonstigen Zwecken

Hierbei handelt es sich um Drittstaatsangehörige, deren Aufenthalt in Österreich nicht mit Erwerbszwecken zusammenhängt bzw. zusammenhängen darf. Dazu gehören einerseits Personen, die nicht erwerbstätig sind und somit ihre finanzielle Unabhängigkeit anhand einer bestimmten Höhe der Existenzmittel nachweisen müssen. Die Erteilung des Aufenthaltstitels für diese Personengruppe wird durch das Quotenprinzip bedingt. Andererseits ist diese Art der Aufenthaltsbewilligung Fremden aus Drittstaaten vorbehalten, die einen Dienst bei einer überparteilichen und gemeinnützigen Organisation ausüben oder eine Tätigkeit im Rahmen des Europäischen Freiwilligendienstes ableisten.

Aufenthaltstitel zur langfristigen Niederlassung

Mit dem Aufenthaltstitel Daueraufenthalt EU dürfen sich Drittstaatsangehörige unbefristet in Österreich niederlassen und verfügen über freien Zugang zum österreichischen Arbeitsmarkt. Die physische Aufenthaltskarte ist trotzdem mit einer fünfjährigen Gültigkeit versehen und muss daher nach dem Ablauf aktualisiert werden. Der diesbezügliche Antrag kann erst gestellt

werden, wenn sich die Personen bereits fünf Jahre rechtmäßig in Österreich aufgehalten und die Voraussetzungen im Rahmen der Integrationsvereinbarung (z. B. Kenntnisse der deutschen Sprache und der grundlegenden Rechts- und Gesellschaftsordnung) erfüllt haben. Jedoch wird die Aufenthaltsdauer von fünf Jahren als anspruchsberechtigt gezählt, nur wenn die antragstellenden Personen während dieses Zeitraums eine bestimmte Art des Aufenthaltstitels, meistens diejenige aus der ersten oder der zweiten Kategorie, innehatten. Bei manchen anderen Aufenthaltstiteln kann ein bestimmter vorangehender Zeitraum zum Teil bzw. unter Umständen sogar zur Gänze je nach Ausnahmeregelung oder Voraussetzung angerechnet werden. Die Hauptsache ist, dass die AntragstellerInnen zum Zeitpunkt der Beantragung zur langfristigen Niederlassung berechtigt sind.

In der Regel unterliegen alle Aufenthaltstitel für Drittstaatsangehörige den einheitlich geltenden allgemeinen Erteilungsvoraussetzungen. Dazu gehören erforderliche Existenzmittel, Krankenversicherungsschutz und ortsübliche Unterkunft, die bei der Antragstellung immer vorgelegt werden müssen. Außerdem sind die besonderen Voraussetzungen strengstens zu berücksichtigen, die je nach individuellem Aufenthaltstitel variieren, wie die Verfügbarkeit eines Quotenplatzes, der Nachweis von Deutschkenntnissen und die schriftliche Aufnahmevereinbarung im Falle der Erwerbstätigkeit. Ausführliche Informationen bezüglich der zweckspezifischen Voraussetzungen, Aufenthaltsrechte und -bedingungen einschließlich der Liste der notwendigen vorzulegenden Unterlagen finden sich in den Antragsformularen zur jeweiligen Art des Aufenthaltstitels.

Im Gegensatz zu den Dokumentationen für ZuwanderInnen aus EWR-Staaten oder der Schweiz sind Erstanträge auf Aufenthaltstitel für Drittstaatsangehörige bereits vor der Einreise ins Bundesgebiet persönlich bei der örtlich zuständigen österreichischen Vertretungsbehörde wie der Botschaft oder dem Konsulat im Herkunftsstaat einzuholen. Erst bei Verlängerungsanträgen muss das Verfahren in Österreich bei der zuständigen Niederlassungsbehörde eingeleitet und durchgeführt werden. Die Erteilung des Aufenthaltstitels erfolgt nur, wenn der eingelangte Antrag sämtlichen Erteilungsvoraussetzungen entspricht und keine Erteilungshindernisse gemäß § 53 und § 67 des Fremdenpolizeigesetzes gegeben sind. Grundsätzlich ist jeder gewährte Aufenthaltstitel nur mit einem bestimmten vorgegebenen Zweck verbunden. Das heißt: Das Stellen eines Antrags, der auf mehrere verschiedene Aufenthaltszwecke hinweist, führt zur Abweisung und damit zur Untersagung des Aufenthalts. Es besteht jedoch eine Möglichkeit, während des rechtmäßigen Aufenthalts gegebenenfalls eine Zweckänderung zu beantragen. Der Wechsel des Aufenthaltstitels kann sich entweder systemimmanent ergeben, z. B. als Umstieg von der

Aufenthaltsbewilligung für Studierende nach dem Studienabschluss auf die Rot-Weiß-Rot-Karte für die Niederlassung mit Erwerbstätigkeit, oder aufgrund der Veränderung der Umstände, die einen neuen Aufenthaltswitz nach sich zieht (vgl. Bundesministerium für Inneres o. J.).

4.3 Verständigung im Niederlassungs- und Aufenthaltsverfahren

Aus den vorherigen Kapiteln, in denen die kommunikative Handlung der öffentlichen Verwaltung behandelt wurde, lässt sich schließen, dass ein ungleiches Machtverhältnis nach wie vor im Rahmen der Interaktion zwischen VerwaltungsagentInnen und -klientInnen herrscht. Dabei treffen die „Professionalität“ der Verwaltungsangehörigen und die „Laienhaftigkeit“ der BürgerInnen aufeinander, was ersteren häufig eine Position der Dominanz verschafft, und letzteren eine der Abhängigkeit (Becker-Mrotzek 2001: 1509). Wie festgestellt, geht die prägende Handlungsasymmetrie einerseits von dem unterschiedlichen Wissensniveau bezüglich der komplexen Sachverhalte und Verwaltungsvorgänge aus. Andererseits wird diese vor allem im Zusammenhang mit der Verständigung des Verfahrens nicht zuletzt durch die Fachsprachlichkeit und damit die (Schwer-)Verständlichkeit der Verwaltungssprache verstärkt (vgl. Fisch & Margies 2014: 15; Lerch 2014: 92 f.).

Diese Problematik trifft insbesondere auf die Austauschbeziehung im Verlauf eines Niederlassungs- und Aufenthaltsverfahrens bei der Einwanderungsbehörde angesichts der betroffenen KlientInnen zu, bei denen sich um ausländische und somit gleichzeitig um zumeist fremdsprachige Laien handelt. So haben die ZuwanderInnen bei der Antragstellung zum einen Antragsformulare und zugehörige Schreiben zu bewältigen, die ausschließlich in der österreichspezifisch deutschsprachigen Verwaltungssprache zur Verfügung stehen. Zum anderen müssen sie zunächst anhand der einsprachig vorhandenen Informationen ihr individuelles Sachanliegen beispielsweise den vorgegebenen Kategorien des Aufenthaltswitzes unterordnen, die sich in Anbetracht verschiedener differenzierter Bedingungen des jeweiligen Aufenthaltstitels als unübersichtlich und dadurch herausfordernd erweisen können. Aufgrund der möglichen Verständnisschwierigkeiten besteht generell im Verlauf des Verfahrens das Risiko, dass beim Anbringen des Sachverhalts Fehler begangen werden und die Sachbearbeitung infolgedessen je nach Schweregrad in Verzug gerät (vgl. Mentrup 1980: 122). Im Falle des Niederlassungsverfahrens bedeutet eine fehlerhafte Antragsvorlage für die zuständige Behördenstelle dementsprechend einen höheren Arbeits- und Zeitaufwand, aber auch schwerwiegendere Konsequenzen für die antragstellenden

ZuwanderInnen, denn der rechtmäßige Aufenthalt von Fremden in Österreich muss ohne Ausnahme durch einen bestimmten Zweck gemäß dem NAG gerechtfertigt werden. Angesichts dessen kann ein Versäumnis, das sich aus Missverständnissen seitens der KlientInnen aufgrund der (fach-)sprachlichen bzw. sachlichen Schwerverständlichkeit ergibt, zum ungünstigen oder schlimmstenfalls erfolglosen Aufenthaltsverfahrensausgang führen. Mögliche Konsequenzen des Formmangels reichen beispielsweise von begrenzten Aufenthaltsrechten bis hin zu einer negativen Entscheidung über Erteilungs- bzw. Verlängerungsanträgen bei Unstimmigkeit mit den zulässigen Aufenthaltszwecken, was wiederum unter anderem finanzielle, rechtliche und persönliche Nachteile für die Betroffenen mit sich bringt.

Insgesamt wirken sich diese möglichen Verständigungshürden im Niederlassungs- und Aufenthaltsverfahren beeinträchtigend auf die selbstbestimmte Teilhabe betroffener KlientInnen an Informationen und damit auf ihren Handlungsspielraum aus, während zuständige AgentInnen zugleich mit einem erschwerten Datenerhebungsprozess konfrontiert werden. Der rechtliche Anspruch auf verständliche Sprache mit einer sprachlichen Unterstützung wie Dolmetschen und Übersetzung im Rahmen eines Verwaltungsverfahrens wird bisher in Österreich nur gesichert, wenn es sich um Verfahrensbeteiligte mit mangelnden Deutschkenntnissen in Verwaltungsstrafsachen handelt (vgl. § 38a Abs. 2 ff. Verwaltungsgerichtsverfahrensgesetz). Auf Artikel 7 Zeile 3 des Staatsvertrags von Wien kann sich ebenfalls kaum berufen werden, denn die zusätzlich zum Deutschen als Amtssprache zugelassenen Minderheitssprachen gelten nur für die inländisch anerkannten Volksgruppen. In der Folge spielen die Regelungen kaum eine Rolle für andere ZuwanderInnen, insbesondere für diejenigen aus Drittstaaten, die vorerst der Verständigung in der österreich-deutschen Verwaltungssprache weiterhin uneingeschränkt unterworfen sind.

Die aktuelle Praxis der Einsprachigkeit in jenem Verwaltungsbereich, dessen Tätigkeit und Verständigung vornehmlich internationale KlientInnen mit unterschiedlichen Sprachhintergründen betreffen, stellt laut Ritter (2011: 26) insbesondere in Zeiten der grenzüberschreitenden Mobilität und Kommunikation einen „Sonderfall“ dar. Dies bekräftigt umso mehr die vorhin erwähnte Empfehlung von Mentrup (1980), Fluck (2008) und Lambertz (1999), dass ein multilinguales terminologisches Nachschlagwerk ausgearbeitet und eventuell anhand dessen Verwaltungsschreiben in mehreren Sprachen zur Verfügung gestellt werden sollte, um Verständigungsprozesse zwischen KlientInnen und AgentInnen im Aufenthaltsverfahren zu verbessern.

5 Maschinengestützte Terminologiegewinnung als Teil der praktischen Terminologearbeit

In den vorigen Kapiteln wurden bereits die Grundlagen für die Verwaltungssprache, ihre österreichische Varietät sowie ihren schriftlichen Ausdruck in behördlichen Schreiben aufbereitet. Dazu wurde auf den Zusammenhang zwischen dieser Sondersprache und der Textsorte Formular bzw. Vordruck näher eingegangen. Zu den wesentlichen Diskussionspunkten gehören deren Auswirkungen auf die Verständigung im Laufe eines Verwaltungsverfahrens, vor allem angesichts der InteraktantInnen im Verwaltungsbereich Aufenthalt und Niederlassung. Die folgende theoretische Auseinandersetzung mit der Terminologearbeit dient als Übergang zum praktischen Teil, wobei der Fokus aufgrund der vorgesehenen Untersuchung primär auf der automatisierten Terminologieextraktion und der Evaluierung der Terminologieextraktionssysteme liegt.

5.1 Grundlagen der Terminologie

Über die letzten Jahrzehnte hinweg erleben Fachbereiche des Rechts und der Politik neben jenen der Wissenschaft und Technik infolge der Globalisierung stets neue Entwicklungen, die unter anderem die Zunahme bzw. Erweiterung des Fachwissens beinhalten. Ständig wächst damit auch der Umfang der entsprechenden Wortschatzbestände des jeweiligen Fachgebiets, d. h. der Terminologie (vgl. KÜDES 2002: 95). Die Vervielfachung der gesamten Erkenntnisse erschwert allerdings die Übersicht über jegliche aktuelle Änderungen, sodass sich zunehmend Verständigungshürden zwischen Laien und sogar zwischen Fachleuten aus demselben Gebiet ergeben. So kommt der Terminologearbeit, die darin besteht, Fachwortbestände nach einschlägigen Regeln zu erfassen, sie zu vereinheitlichen und gegebenenfalls zugänglich zu machen, gegenwärtig eine entscheidende Rolle zu (vgl. RaDT 2013: 3). Die festgelegte Terminologie und deren konsistente Verwendung unterstützen das Funktionieren von Fachsprache, die wiederum erheblich möglichst zu Eindeutigkeit und Verständlichkeit in der mündlichen bzw. schriftlichen Kommunikation unter Interessengruppen beiträgt (vgl. Drewer & Schmitz 2017: 1).

Die Aufgaben der Zusammenstellung, Systematisierung und Bearbeitung der Fachwortschätze werden in der Regel von TerminologInnen durchgeführt, die das Fachwissen eines oder mehrerer Bereiche mit fachsprachlichen Kenntnissen verbinden. In sonstigen Fällen können sie auch von Fachleuten bei der Beurteilung und Erarbeitung von Terminologien, beispielsweise in der organisationsinternen Normung und Dokumentation, zur Beratung

hinzugezogen werden (vgl. Arntz et al. 2014: 1). Die positiven Effekte einer professionellen Terminologearbeit zeigen sich einerseits in der Erleichterung und Beschleunigung der gegenwärtigen Fachkommunikation, an immer höhere Anforderungen aufgrund der steigenden Komplexität und Vernetzung von Fachgebieten gestellt werden. Andererseits gilt die systematische Bereitstellung von Terminologien, insbesondere in mehrsprachiger Form, als Schlüsselkomponente für die Tätigkeit der ÜbersetzerInnen, die mit dem steigenden Bedarf an sprachübergreifenden Fachtexten sowohl quantitativ als auch qualitativ mitzuhalten haben. Die gegenwärtige Teilnahme am Wissenstransfer und -austausch auf allen Gebieten geht infolge der Globalisierung und Technisierung verstärkt über geopolitische Grenzen hinweg (vgl. KÜDES 2002: 9).

In diesem Sinne dienen festgesetzte Terminologien als sprachlicher Qualitätssicherungsfaktor für die Verständigung in bzw. zwischen verschiedenen Disziplinen, darunter auch der öffentlichen Verwaltung und Gesetzgebung: Im Zeitalter der Globalisierung sind Rechtssysteme auch Änderungen ausgesetzt, wobei staatliche Verwaltungsorgane sich mit ständiger und zahlenmäßiger Zunahme komplexer Regelungsbereiche zu beschäftigen haben (vgl. ebd.). So trägt eine standardmäßige Terminologearbeit, die beispielsweise im Zuge eines Gesetzesentwurfs geleistet wird, zur begrifflichen Übersichtlichkeit des betreffenden Regelungsbereichs und damit zur Effizienzsteigerung der Kommunikation unter den Beteiligten bei. Außerdem sorgt die terminologische Konsistenz für begrifflich kohärente und transparente Gesetzgebung, die widerspruchsfreie Rechtsanwendung durch die Behörden ermöglicht und die Rechtssicherheit sowie -überzeugung im Staat erhöht (vgl. RaDT 2013: 3).

Da internationale Beziehungen gegenwärtig einen wesentlichen Bestandteil der staatlichen Tätigkeiten bilden, leistet der Dialog zwischen Behörden und damit die entsprechende Terminologie, die effiziente Verständigung fördert, einen maßgeblichen Beitrag zur positiven Zusammenarbeit und Rechtsharmonisierung. Angesichts dessen besteht die Notwendigkeit, enge und einheitliche Kooperation hinsichtlich eines terminologischen Austauschs zwischen betreffenden Behörden zu fördern (vgl. KÜDES 2002: 10). Dies ist nur dann sinnvoll, wenn in allen beteiligten Staaten selbst standardmäßige (zumindest einsprachige) Terminologien der gesamten Verwaltung systematisch ausgearbeitet und für den Austausch zur Verfügung gestellt werden. So sind Behörden der einzelnen Staaten in der Lage, die bestehende Fachwortsammlung mit fremdsprachigen Ergänzungen auszubauen (vgl. KÜDES 2002: 67). Mit mehrsprachigen terminologischen Datenbeständen wird dann die Übersetzung von Gesetzes- und Verwaltungstexten ermöglicht, wobei sich die Qualität und Konsistenz auch dadurch sicherstellen lassen (vgl. Drewer & Schmitz 2017: 3).

Wie eingangs erwähnt, wird Terminologearbeit auf unterschiedlichen Fachgebieten geleistet, sodass die Ziele bei der Erarbeitung, Bearbeitung oder Verarbeitung des Fachwortbestands je nach Tätigkeitsbereich differieren. Jedoch wird einheitlich nach einer methodisch korrekten Durchführung in den verschiedenen Disziplinen gestrebt, damit der Austausch von Fachwissen und der damit verbundenen terminologischen Daten möglichst reibungslos erfolgt. Dementsprechend gilt es, unter Einhaltung von Grundsätzen der Terminologearbeit konform zu arbeiten (vgl. Arntz et al. 2014: 3). Solche einheitlichen Regeln entwickelten sich im Rahmen der Terminologielehre, die in der Norm DIN 2342 (2011: 14) folgendermaßen definiert wird: „die Wissenschaft von den Begriffen und ihren Bezeichnungen in den Fachsprachen“. Der Ursprung der systematischen Auseinandersetzung mit Fachwörtern geht auf die Zeit der Industrialisierung ab Mitte des 19. Jahrhunderts zurück. In den Industriestaaten wurde sich intensiv um die sachliche und begriffliche Vereinheitlichung bemüht, als die damals einsetzende Technisierung nicht nur signifikante Fortschritte, sondern auch eine Menge ungeordneter Erkenntnisse entstehen ließ. So kam der Beschäftigung mit terminologischen Fragen im Rahmen der Sachnormungsmaßnahmen zwangsläufig Bedeutung zu, denn „Gegenstände und Verfahren lassen sich nur dann verbindlich und zweifelsfrei vereinheitlichen, wenn auch über die Verwendung der hierzu verwendeten sprachliche Mittel Einigkeit besteht“ (Arntz et al. 2014: 270).

Trotz ihrer Entstehung in technischen und naturwissenschaftlichen Bereichen ist die Terminologielehre grundsätzlich interdisziplinär konzipiert. Aufgrund dieses Merkmals kommen ihre Grundsätze zur fächerübergreifenden Anwendung, sodass der Einsatz der Terminologearbeit auf allen Gebieten möglich ist (vgl. Arntz et al. 2014: 4 ff.). Da es sich bei der Terminologielehre in der Regel um die „Wissenschaft von den Fachwortschätzen“ (Arntz et al. 2014: 5) handelt, besteht eine enge Verbindung zu den lexikologischen und lexikografischen Aspekten der Sprachwissenschaft; durch die Beschäftigung mit der Festlegung von Bedeutung des jeweiligen Fachworts ist aber auch der Bezug zur Semiotik gegeben. Darüber hinaus liegt vor allem die Verbindung zur Fachsprachenforschung nahe, denn terminologische Einheiten werden dabei als integraler Bestandteil des kommunikativen Prozesses im Fachgebiet untersucht. Zu weiteren Disziplinen, in denen Terminologielehre deutlich ihren Ausdruck findet, zählen unter anderem Sachwissenschaften, Normung und Sprachplanung, Information und Dokumentation sowie Informatik (vgl. Drewer & Schmitz 2017: 6).

5.2 Grundelemente der Terminologielehre

Unabhängig von der interdisziplinären Überschneidung stehen im Mittelpunkt der Terminologielehre stets die Grundelemente Gegenstand, Begriff und Benennung, deren Begrifflichkeit anhand der folgenden Abb. 3 dargestellt wird. Auch wenn der Fokus dieser Arbeit hauptsächlich auf den Benennungen liegt, können diese nur mithilfe der Darstellung der Verbindung von allen drei Elementen zueinander sinnvoll beschrieben werden. Im Anschluss wird die Gliederung verschiedener Erscheinungsformen von Benennungen auf linguistischer Ebene erläutert.

Das Modell in Abb. 3, das üblicherweise als semiotisches Dreieck bezeichnet wird, wurde von den amerikanischen Linguisten Ogden und Richards (1974) in die sprachwissenschaftliche Diskussion zur Terminologie introduziert (vgl. Bußmann 2002: 460; Arntz et al. 2014: 41). Durch ihre einfache und klare Darstellung gilt die Grafik als eines der meistbenutzten Instrumente zur Veranschaulichung der Beziehungen zwischen den Grundelementen der Terminologielehre (vgl. Drewer & Schmitz 2017: 7). Es gilt hier vorab darauf hinzuweisen, dass dieses Dreiecksmodell teilweise als zu vereinfacht betrachtet wird und somit kein adäquates Werkzeug für die Erklärung der komplexeren Aspekte des Begriff-Benennung-Zusammenspiels in linguistischer oder kognitiver Hinsicht bietet. Jedoch eignet sich genau diese Einfachheit für die folgende Erläuterung, die einen Überblick über die terminologischen Grundelemente und ihre fundamentalen Zusammenhänge darstellt.

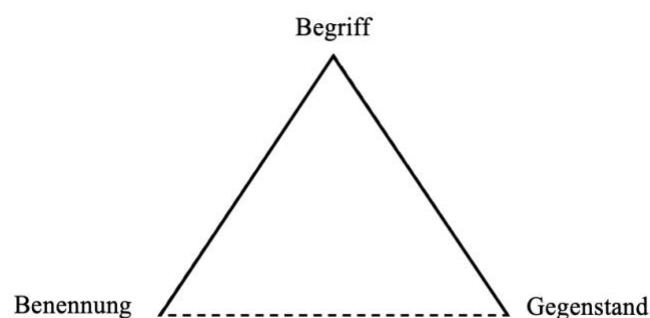


Abbildung 3: Semiotisches Dreieck (adaptiert von Arntz et al. 2014: 41)

Im Sinne des semiotischen Dreiecks stellt der Gegenstand einen „Rohstoff“ (Arntz et al. 2014: 45) aus der „wahrnehmbaren oder vorstellbaren Welt“ (DIN 2342 2011: 5) dar, von dem bei jeglicher Begriffsbildung ausgegangen wird. Dementsprechend können Gegenstände zum einen konkrete Objekte umfassen, die entweder physisch gegenwärtig oder für menschliche Sinneswahrnehmung zugänglich sind. Zum anderen gehören dazu abstrakte Gegenstände, die über keine physische Form und nur mentale Existenz verfügen (vgl. Arntz et

al. 2014: 45 ff.). Begriffe, die in der Terminologielehre eine zentrale Rolle spielen, sind wiederum die gedankliche Vorstellung bzw. die „gedanklichen Vertreter“ von Gegenständen (Felber & Budin 1989: 69). Die Begriffsbildung erfolgt, indem Menschen mehrere Gegenstände anhand deren feststellbaren gemeinsamen Eigenschaften erkennen, gedanklich ordnen und zusammenfassen (vgl. Drewer & Schmitz 2017: 8). Im Rahmen einer Terminologearbeit werden die charakteristischen Eigenschaften eines Begriffs nach der Norm DIN 2330 (2013: 6) schriftlich in Form einer Definition niedergelegt, damit dieser in ein bestimmtes Begriffssystem eingeordnet und dadurch auch gegenüber anderen Nachbarbegriffen innerhalb eines Fachgebiets abgegrenzt werden kann. Um über die Gegenstände und Begriffe mündlich oder schriftlich kommunizieren zu können, wird auf sprachliche Ausdrucksmittel zurückgegriffen. Die sprachlichen Repräsentationen des jeweiligen Begriffs werden in der Terminologielehre als Benennungen oder auch Termini (Einzahl Terminus) bezeichnet (vgl. RaDT 2013: 4). Die Zuordnung einer Benennung zu einem Begriff basiert auf den spezifischen Eigenschaften, die in der entsprechenden Definition festgelegt sind (DIN 2330: 6).

Da eine Benennung nicht nur in Form eines Worts auftritt, sondern auch aus zwei oder mehreren Wörtern bestehen kann, wird von der Unterscheidung zwischen zwei Gruppen gesprochen: Einwort- und Mehrwortbenennungen (vgl. Saß 2004: 23). Bei den Einwortbenennungen lässt sich nach der DIN 2330 (2013: 7) ferner in elementare Wörter und komplexe Wörter untergliedern. Zu den elementaren Wörtern gehören folgende Formen:

- **Simplizia** enthalten jeweils eine bedeutungstragende Einheit und dienen im Rahmen der Bildung von Benennungen als Ausgangsbasis für weitere Neubildungen von Komposita oder Mehrwortbenennungen, z. B. Magistrat, Recht, Zweck und fremd (vgl. Bußmann 2002: 686).
- **Akronyme** (auch: Initialwörter oder Buchstabenwörter) entstehen dadurch, dass die Anfangsbuchstaben einer vorhandenen Mehrwortbenennung oder eines Kompositums zusammengesetzt werden, z. B. NAG aus Niederlassungs- und Aufenthaltsgesetz, EWR aus Europäischer Wirtschaftsraum. Die Reduzierung der Buchstaben soll der Sprachökonomie dienen, wobei sich die Bedeutung des jeweiligen Akronyms aus dem Zusammenhang im entsprechenden Fachgebiet ableiten lässt, in dem es erscheint (vgl. Witschel 2004: 21; Drewer & Schmitz 2017: 79).

- **Kurzworte** entstehen durch die Auslassung eines Teils der Langform einer Benennung, wobei dies am Wortanfang, in der Wortmitte oder am Wortende erfolgen kann, z. B. Pass statt Reisepass (vgl. ebd.).

Im Gegensatz zur ersten Gruppe setzen sich die komplexen Einwortbenennungen aus zwei oder mehr bedeutungstragenden Einheiten zusammen, aber können auch mit einer Ableitung gebildet werden. Zu dieser Gruppe gehören die folgenden:

- **Komposita** werden dadurch gebildet, dass zwei oder mehr selbständige elementare Einwortbenennungen zusammengeführt werden (vgl. Bußmann 2002: 360). Die Verbindungen zwischen den zusammengesetzten Wörtern können mithilfe von Bindestrichen oder Fugenmorphemen erfolgen, um die Neubildungen übersichtlicher bzw. leichter sprechbar zu gestalten z. B. Rot-Weiß-Rot-Karte, Aufenthaltstitel oder Inlandsantragsstellung. Dieses Wortbildungsmuster spielt eine wesentliche Rolle und herrscht insbesondere in der deutschen Fachsprache vor (vgl. Drewer & Schmitz 2017: 71 ff.). Alle Wortarten können grundsätzlich für die Kompositumbildung verwendet werden, wobei Nominal-, Adjektiv- und Verbkomposita, wie an den Beispielen der österreichischen Fachausdrücke in der Verwaltung im Abschnitt 2.2.2 erkennbar ist, terminologisch am meisten von Bedeutung sind.
- **Derivate** bezeichnen die komplexen Einwortbenennungen, die durch die Kombinierung eines Stammworts mit einem oder mehreren Wortbildungsmorphemen in Form von Präfixen bzw. Suffixen entstehen (sog. Derivationsverfahren), z. B. Angehörigkeit oder Staatsbürgerschaft. Neben der Komposition ist die Ableitung von Wörtern mit Affixen im Deutschen ein bevorzugtes Wortbildungsverfahren in der Terminologearbeit. Oft ergeben sich bei der Bildung von Ableitungen Synonymvarianten, die sich nur durch morphologische Eigenschaften unterscheiden, z. B. Anmeldeformular und Anmeldungsformular (vgl. Drewer & Schmitz 2017: 74).

Mehrwortbenennungen entstehen auch durch die Zusammensetzung von mindestens zwei elementaren bzw. komplexen Einwortbenennungen. Im deutlichen Gegenteil zu Komposita werden die Wörter allerdings durch Leerstellen getrennt geschrieben, weisen aber eine syntaktische Verbindung zueinander auf (vgl. DIN 2342 2011: 5; Arntz et al 2014: 115). Die meisten terminologischen Mehrwortbenennungen in deutschen Fachsprachen bestehen aus

mehreren Substantiven oder einer Adjektiv-Substantiv-Kombination, z. B. Blaue Karte EU, aufenthaltsberechtigter Drittstaatsangehöriger, Aufenthaltserlaubnis Selbstständiger oder Sonderfälle unselbständiger Erwerbstätigkeit (vgl. Drewer & Schmitz 2017: 15). KÜDES (2002: 61 f.) unterscheidet allerdings bei den Mehrwortbenennungen ferner anhand der Wortarten innerhalb deren Kombination. Daraus ergeben sich zum einen die bereits angeführten Kombinationen von Nominalphrasen, die auch als Fachausdrücke bezeichnet werden können. Zum anderen wird von Fachwendungen und Fachphrasen gesprochen, die zusammen die Fachphraseologie bilden. Unter Fachwendungen werden pragmatische Verbindungen von Verb und Substantiv gefasst, die durch bevorzugte Verwendung oder durch Festlegung je nach Fachgebiet begründet sind (vgl. Heid 1999: 242). Dabei ist das Verb häufig selbst kein Fachwort, hat aber zusammen mit dem anderen Wort der Kombination eine fachsprachliche Bedeutung, z. B. einen Aufenthaltstitel erteilen oder einen Erstantrag einbringen. Innerhalb einer Fachphrase sind hingegen verschiedene Elemente (Verb, Substantiv, Adjektiv oder Präposition) enthalten. Die Verbindungen in Fachphrasen sind durch ihren häufigen Gebrauch „gefestigte formelhafte Satzteile“ (KÜDES 2002: 62), die in den Texten eines spezifischen Fachbereichs wie typische Textbausteine immer wieder vorkommen, z. B. Antrag auf Erteilung von Aufenthaltstiteln oder Aufenthaltstitel zur Arbeit bzw. Erwerbstätigkeit. Die Fachphraseologie dient somit überwiegend der Idiomatizität der Fachtexte, jedoch besteht im Gegensatz zu Fachausdrücken keine absolute, sondern eher eine überwiegend feste Verbindung zwischen ihren Bestandteilen. Gerade bei Verben in Fachwendungen liegt eine hohe linguistische Varianz vor, weswegen außerhalb des bestimmten fachsprachlichen Kontexts (Fachgebiet, sprachliche Landesspezifika usw.) zwischen Synonymen abgewechselt werden kann, z. B. Antrag einbringen bzw. stellen bzw. einreichen (vgl. KÜDES 2002: 63 f.; Thurmair 2003: 171).

5.3 Terminologiegewinnung

Nachdem in Unterkapitel 5.2 die wesentlichen Grundelemente der Terminologielehre geschildert wurden, wird in diesem Unterkapitel einer der relevantesten Schritte der praktischen Terminologiarbeit ausgeführt: die Gewinnung von Termini aus Textmaterial. Zunächst werden die Grundlagen der Terminologiegewinnung dargestellt, woraufhin eingehend auf das (semi-)automatisierte Verfahren des Vorgangs in den Unterkapiteln 5.4 bis 5.6 eingegangen wird. In diesem Zusammenhang steht der Einsatz von Computerwerkzeugen im Vordergrund, wobei ihre unterschiedlichen Methoden zur Identifizierung und Extraktion von Fachtermini erläutert werden.

Da Terminologearbeit in der Norm DIN 2342 (2011: 14) als die „auf der Terminologielehre aufbauende Planung, Erarbeitung, Bearbeitung oder Verarbeitung, Darstellung oder Verbreitung von Terminologie“ definiert wird, lässt sich feststellen, dass auch die Gewinnung von Terminologie aus Textquellen, die laut Arntz et al. (2014: 22) zwischen der Erarbeitung und Bearbeitung anzusiedeln ist, zur Terminologearbeit gerechnet werden soll. Bei der Terminologiegewinnung handelt es sich darum, die Termini in vorhandenen fachlichen Textdokumenten zu identifizieren und für einen weiteren terminologischen Bearbeitungsvorgang herauszufiltern (vgl. Arntz et al. 2014: 224). Bei solchen Termini, die sich in Fachtexten „versteckt“ (Drewer & Schmitz 2014: 40) befinden und die erst vor weiteren Bearbeitungsprozessen daraus extrahiert werden müssen, wird von implizit vorliegender Terminologie gesprochen. Diese Form unterscheidet sich von den sogenannten explizit bestehenden terminologischen Daten, die – wie der Name vermuten lässt – bereits zugänglich zur Verfügung stehen und deren Übernahme ins Terminologieverwaltungssystem deshalb ohne Extraktion erfolgt. Dabei handelt es sich beispielsweise um Wortlisten, Glossare, Abkürzungsverzeichnisse bestimmter Unternehmen oder Organisationen (vgl. ebd.). Eine Zusammenstellung mehrerer Textmaterialien, die der Untersuchung vor allem im Rahmen der Terminologearbeit unterzogen werden, ist als ein Korpus zu bezeichnen.

Die möglichen Ziele der Terminologiegewinnung hängen mit unterschiedlichen Gründen und deshalb auch mit unterschiedlichen Ausgangsmaterialien und Verfahren zusammen. Der erstmalige Aufbau einer Terminologiesammlung stellt zwar den häufigsten Anwendungsfall für die Terminologiegewinnung dar, dies ist jedoch nicht die einzige Zwecksetzung in der Praxis. Die Auswertung des vorhandenen Textmaterials durch die Extraktion von Termini kann außerdem auch sinnvoll durchgeführt werden, wenn ein bereits existierender Terminologiebestand um die Fachwortschätze eines neuen Fachgebiets oder Teilbereichs der Organisation aktualisiert oder erweitert werden soll (vgl. Saß 2004: 19). Darüber hinaus dient die Extraktion der ausgangssprachlichen Benennungen bei umfangreicheren Übersetzungsprojekten dazu, den Aufwand der notwendigen Terminologearbeit zu ermitteln und daraufhin die Terminologie im Vorfeld des Übersetzungsprozesses festzulegen. Nicht zuletzt werden im Rahmen der Qualitätssicherung von verfassten fachlichen Dokumenten die terminologischen Daten in den zu überprüfenden Textmaterialien durch den Extraktionsvorgang identifiziert, um die korrekte und konsistente Verwendung von Terminologie zu verifizieren (vgl. Arntz et al. 2014: 244 f.).

Die Extraktion der Terminologie aus Texten kann rein menschlich oder maschinengestützt erfolgen. Bei der menschlichen Terminologiegewinnung variiert die

Bezeichnung in der Regel je nach der Systematik der Vorgehensweise. So wird insbesondere in Bezug auf die Ad-hoc-Terminologiarbeit von Terminologierecherche gesprochen, bei der Benennungen punktuell untersucht werden, um z. B. bei einer Übersetzungsaufgabe ein gerade auftretendes terminologisches Einzelproblem spontan zu lösen oder die dazugehörigen Begriffe zu klären (vgl. Drewer & Schmitz 2017: 26). Erst gilt der Vorgang als (menschliche bzw. manuelle bzw. humane) Terminologieextraktion, auch kurz Termextraktion, wenn er systematisch auf Basis des fachlichen und terminologischen Wissens auf Benennungen in Textquellen untersucht wird (vgl. Drewer & Schmitz 2017: 41). Die menschliche Termextraktion, die laut L’Homme et al. (1996: 293) hauptsächlich auf der Expertise und Erfahrung beruht, liefert qualitativ hochwertige Ergebnisse und weist dabei weniger Fehleranfälligkeit auf. Jedoch zieht der rein manuelle Einsatz insbesondere bei größerem Quellenumfang einen hohen Zeit- und damit Kostenaufwand nach sich. Hinzu kommt, dass menschliches Versagen aufgrund von Übermüdung oder Unaufmerksamkeit der Beteiligten unvermeidlich ist (vgl. Arntz et al 2014: 245).

Erweist sich der Arbeitsaufwand insbesondere bei umfangreichen Korpora im Verhältnis zu einer menschlichen Untersuchung als unvertretbar groß, so wird auf die Unterstützung maschineller Verfahren mit bestimmten Computerwerkzeugen gesetzt (vgl. Fleischmann 2006: 122). Doch muss klargestellt werden, dass die Vorteile des automatisierten Vorgangs vielmehr in der Schnelligkeit und Produktivitätssteigerung bestehen (vgl. Zielinski & Ramírez-Safar 2005: 2). Dabei lässt sich im Hinblick auf Qualität und Tauglichkeit der Ergebnisse im Vergleich zu denen der humanen Termextraktion ein Mangel feststellen, sodass in nahezu allen Fällen eine menschliche Kontrolle und Nacharbeiten erforderlich sind, um die Brauchbarkeit der Ergebnisse zu sichern. So gebührt dieser gegenwärtigen Arbeitsweise laut Drewer und Schmitz (2017: 43 f.) eher die Bezeichnung „maschinengestützte“ oder „semiautomatisierte“ Terminologieextraktion.

5.4 Maschinengestützte Terminologieextraktion

Wie oben bereits dargelegt, sind unter maschinengestützter oder semiautomatisierter Terminologieextraktion somit sämtliche Verfahren zur Terminologiegewinnung zu verstehen, die mithilfe des Computereinsatzes durchgeführt werden. Je nach Zielsetzung, Korpusgröße und Ausgangsmaterialien variieren die sinnvollen Automatisierungsgrade des Vorgangs, die von einfachen Unterstützungsoperationen mit Standardtextverarbeitungsprogrammen bis hin zu einer hochautomatischen Terminologiegewinnung mit Computerwerkzeugen reichen, die speziell für die Erkennung und Extraktion von Benennungen in fachlichen Texten entwickelt

werden. Ungeachtet der Ausmaße der technischen Unterstützung müssen die zu verarbeitenden Texte immer in elektronisch lesbarer Form vorliegen.

Von den zur Analyse heranzuziehenden Texten ausgehend, ist grundsätzlich zwischen der monolingualen und der bi- bzw. multilingualen maschinengestützten Termextraktion zu unterscheiden (vgl. Cabré 1999: 13). Bei ersterer werden Texte oder Textkorpora, in denen nur einsprachige Daten enthalten sind, als Ausgangsmaterial für die Verarbeitung benutzt. Die Extraktionsergebnisse werden in der Praxis anschließend durch terminologische oder sachgebietsspezifische Fachleute auf fachliche Relevanz überprüft, bevor sie in die entsprechende Terminologieressource aufgenommen werden. Für die bilinguale Termextraktion gelten andererseits parallele Texte bzw. Textkorpora in zwei Sprachen, die zumeist Übersetzungen voneinander sind, als wesentliche Voraussetzung (vgl. Witschel 2005: 119). Darüber hinaus sollten diese parallelen Materialien als Alignment vorliegen. Das heißt, dass die Inhalte bereits in Phrasen-, Satz- bzw. Absatzgröße zerlegt wurden, wobei die resultierenden Segmente der einen Sprache jeweils den inhaltlichen Entsprechungen der anderen Sprache zugeordnet sind. Deshalb stellen vor allem die Inhalte von Übersetzungsspeichern eine ideale Quelle für zweisprachige Extraktion dar (vgl. Arntz et al. 2014: 245).

Sind allerdings nur die Ausgangs- und Zieltexte einer Übersetzungsarbeit getrennt verfügbar, müssen diese vor dem Extraktionsvorgang auch aligniert werden, um brauchbare Ergebnisse sicherzustellen (vgl. Drewer & Siegel 2012: 36). Die Extraktion von Termini in mehr als zwei Sprachen kann derzeit in der Regel nur dadurch erfolgen, dass eine zweisprachige Termextraktion immer mit derselben Ausgangssprache mehrfach umgesetzt wird, vorausgesetzt, dass Übersetzungen in allen gewünschten Sprachen zugänglich sind. Thurmair (2003: 173) spricht außerdem von „crosslingual retrieval“ oder einer sprachübergreifenden Terminologiegewinnung, wobei die verwendeten Ausgangsmaterialien in zwei oder mehreren Sprachen aus denselben Themengebieten stammen, aber keine Übersetzungen voneinander wiedergeben. Zum Einsatz kommt zuerst die einsprachige Extraktion von Termini aus einzelnen Texten und danach wird versucht, anhand der Ergebnisse zusammengehörige Termini zu identifizieren (vgl. Thurmair 2003: 170).

Sowohl die ein- als auch die zweisprachige maschinengestützte Termextraktion liefert eine Liste von Benennungen bzw. Benennungspaaren, die als Termkandidaten bezeichnet werden (vgl. Drewer & Schmitz 2017: 44). Solche Termkandidaten werden im Laufe des automatischen Extraktionsverfahrens durch das eingesetzte Werkzeug als potenzielle sprachliche Ausdrücke identifiziert, die ein bestimmtes Fachgebiet am ehesten repräsentieren

können und damit als mögliche Fachtermini zu bewerten sind (vgl. Lieske 2002: 110; Chiocchetti et al. 2013b: 21). Nicht alle Fachbereiche sind gleichermaßen für die maschinengestützte Termextraktion geeignet: Bei Bereichen wie Öffentlichkeitsarbeit oder der Werbebranche, deren Terminologie mehr Raum für Ausdrucksvarianz bietet, wird die Erkennung von Fachtermini insbesondere mit automatisierten Verfahren aufgrund der fehlenden Festlegung der einheitlichen Verwendung tendenziell erschwert. Im Gegenteil erweist sich die semiautomatisierte Termextraktion als sinnvoller, wenn die Disziplinen sich durch eine vergleichsweise strenge und längst bewährte Standardisierung hinsichtlich der Terminologie auszeichnen wie bei Technik, Recht und Naturwissenschaften (vgl. Love 2000: 32). Die Kennzeichen von Termkandidaten, auf die bei den maschinengestützten Extraktionsverfahren zurückgegriffen wird, und die Methoden zur Identifizierung von Terminologie werden im folgenden Unterkapitel im Zusammenhang mit den Terminologieextraktionssystemen eingehend diskutiert.

5.5 Terminologieextraktionssysteme

Die Entwicklung der Terminologieextraktionssysteme geht einerseits auf den Beweggrund zurück, dass die TerminologInnen bei der manuellen Extraktion von Termini aus immer größeren Korpora der neu entstandenen Fachgebiete dadurch entlastet werden und die hieraus folgende Reduzierung des Zeit- und Kostenaufwands zu einer höheren Arbeitseffizienz beiträgt (vgl. Strehlow 2001: 426). Andererseits handelte es sich um das Bemühen, standardisierte und somit zuverlässigere Identifizierungskriterien der sprachlichen Ausdrücke von terminologischer Relevanz zu erarbeiten, anstatt sich alleinig auf „educated guess“ (wohlbegründete Vermutung) oder menschliche Intuition zu verlassen (Love 2000: 43). Aufgrund der zunehmend verfügbaren Fachtextkorpora in maschinenlesbarer Form wurde die statistische und datenbasierte Analyse der vorhandenen Terminologie ermöglicht (vgl. Sager 2001: 764). So begannen die ersten Bestrebungen bereits in den 1980er-Jahren, Ansätze zur automatisierten Terminologiegewinnung zu erwirken (vgl. Vivaldi & Rodriguez 2006: 14). Zwar gibt es gegenwärtig mehrere kommerziell erhältliche Termextraktionsprogramme, die laut Vivaldi und Rodriguez (2007: 225) über einen gewissen Tauglichkeitsgrad verfügen, aber noch keine kann mit voller Automatisierung eine hundertprozentige Genauigkeit der Ergebnisse erbringen. Lauriston (1994: 156) betont in diesem Zusammenhang, dass existierende Programme bzw. Werkzeuge sich derzeit entweder in einem experimentellen Stadium befinden oder eine höchstens semiautomatische Ausführung bieten.

In jedem Fall besteht die wesentliche Aufgabe sämtlicher Termextraktionssysteme in der Identifizierung tokenisierter lexikalischer Einheiten bzw. Tokens, die eine entsprechende terminologische Relevanz bezüglich eines bestimmten Fachgebiets aufweisen (vgl. Frantziy et al. 2000: 115). Die Resultate werden, wie erklärt, als potenzielle Benennungen bzw. Termkandidaten bewertet und für weitere Verarbeitungsprozesse in der Terminologiearbeit bereitgestellt. Hier stellt sich die Frage, was einen Termkandidaten ausmacht, der dann bei der maschinengestützten Termextraktion berücksichtigt wird. Laut Witschel (2004: 12) sind Termini durch zwei grundlegende Merkmale gekennzeichnet, die sich für ihre Identifizierung ausnutzen lassen:

[Termini] folgen einer anderen statistischen Verteilung als gemeinsprachliche Wörter und sie werden nach bestimmten Mustern gebildet, die immer wieder – teilweise auch sprachübergreifend – auftauchen und die sich linguistisch auf bestimmte Weise charakterisieren lassen.

Aus dieser Beschreibung lässt sich herleiten, dass sich gerade die wiederholte Erscheinung und die sprachlichen Charakteristiken der Termini in den zwei weit verbreiteten Ansätzen zur maschinengestützten Termextraktion widerspiegeln: Das statistische Extraktionsverfahren (siehe Unterkapitel 5.6.1) basiert auf der Auftrittshäufigkeit bestimmter Ausdrücke, während die Analyse bei der linguistischen Methode (siehe Unterkapitel 5.6.2) mithilfe der festgelegten sprachlichen Muster auf morphosyntaktischer Ebene erfolgt (vgl. Hartley 2009: 113).

Kageura und Umino (1996) stellen eine vertiefte und abstrakte Formulierung der Eigenschaften vor, auf die sich die Extraktionswerkzeuge im Zuge der Analyse von Termkandidaten konzentrieren: Termhood und Unithood². Termhood wird in diesem Sinne als Grad der Beziehung einer linguistischen Einheit zu einem bestimmten Fachgebiet definiert (vgl. Kageura & Umino 1996: 272). Anders ausgedrückt stellt es dar, wie fachspezifisch ein Terminus ist. Da die Standardisierung von Terminologie eine formale Variation bei lexikalischen Einheiten der Termini selten zulässt, gilt nach Witschel (2004: 49) das repetitive Auftreten eines Terminus ohne Variation als Kriterium für Termhood, dessen Frequenz mithilfe z. B. der statischen Extraktionsmethode ermittelt werden kann. Andererseits kennzeichnet die Eigenschaft von Unithood das Ausmaß des Zusammenhalts bzw. der Stabilität der Kollokationen, wobei hierunter eine syntagmatische Beziehung zwischen linguistischen Einheiten innerhalb einer Benennung zu verstehen ist (vgl. Kageura & Umino 1996: 260).

² Da die englischen Bezeichnungen *Termhood* und *Unithood* aufgrund der fehlenden adäquaten deutschen Äquivalenz meistens in deutschsprachigen Veröffentlichungen wie Witschel (2004, 2005) nicht übersetzt werden, sollen die Originale auch in der vorliegenden Arbeit beibehalten werden.

Syntagmatische Beziehungen, deren Begriff von Saussure (1967: 147) geprägt wurde, beziehen sich wiederum auf die „Kombinierbarkeitsrelationen“ zweier Wörter (Witschel 2004: 23). Das heißt, zwei Wörter bilden eine syntagmatische Beziehung bzw. eine Kollokation, wenn sie kombiniert bzw. nebeneinander im Text sinnvoll erscheinen können. Dies trifft insbesondere auf Mehrwortbenennungen und Komposita zu, deren Bestandteile durch die festgelegte Verwendung von Terminologie eine hohe Zusammengehörigkeit und damit ein starkes Unithood besitzen. Extraktionswerkzeuge, die sich primär nach dieser Eigenschaft ausrichten, analysieren zuerst beispielsweise morphologische Strukturen von Mehrwortbenennungen, um die Wahrscheinlichkeit der Kombinationen der lexikalischen Einheiten festzustellen und demnach eine Art Muster für die linguistische Identifizierung zu bilden (vgl. Witschel 2004: 25; Vivaldi & Rodriguez 2007: 236). Es lässt sich abschließend zusammenfassen, dass die beiden Eigenschaften durch Extraktionsprogramme bzw. -systeme – jeweils in unterschiedlichem Umfang je nach Extraktionsstrategie – zur Untersuchung und Identifizierung von Termkandidaten herangezogen werden (vgl. Nakagawa 2001: 303 f.).

5.6 Verfahren der maschinengestützten Terminologieextraktion

In diesem Kapitel werden gängige Ansätze erläutert, die sowohl bei der ein- als auch bei der zweisprachigen Termextraktion mit maschineller Unterstützung in der Praxis eingesetzt werden können. Dabei werden die bereits im vorigen Unterkapitel 5.5 vorgestellten Extraktionsmöglichkeiten eingehend betrachtet, d. h. statistische und linguistische Verfahren, welche die inhärenten Eigenschaften der Termini Termhood und Unithood erfassen. Anschließend wird auf die hybride Nutzung im Abschnitt 5.6.3 eingegangen, bei der bestimmte Leistungsmerkmale der ersten zwei Verfahren in Kombination herangezogen werden. Zusätzlich werden mögliche Nachteile diskutiert, die das jeweilige Extraktionsverfahren mit sich bringt. Diese wirken sich auf die Qualität der Extraktionsergebnisse aus und sind deshalb bei der Evaluierung der Wirksamkeit der eingesetzten Werkzeuge zu berücksichtigen. Es gilt darauf hinzuweisen, dass aufgrund des begrenzten Umfangs dieser Arbeit nur eine Selektion bestimmter Extraktionsmethoden zum jeweiligen Verfahren erläutert werden kann. Ausführliche Informationen bezüglich spezifischer Berechnungsmodelle von unterschiedlichen Methoden, die im Zuge der entsprechenden Verfahren zur automatisierten Terminologieextraktion eingesetzt werden, finden sich beispielsweise bei Witschel (2004) und Pazienza et al. (2005). Ohnehin wird sich darum bemüht, alle wesentlichen Aspekte möglichst abzudecken, die dann einen relevanten Ausgangspunkt für weitere Diskussionen vor allem im Zusammenhang mit den Untersuchungsergebnissen im praktischen Teil bilden.

5.6.1 Statistische Extraktionsverfahren

Bei Terminologieextraktionssoftwares, die auf das statistische Verfahren zurückgreifen, wird ein probabilistisches Modell der Nutzung von Termini zugrunde gelegt (vgl. Drewer & Schmitz 2017: 49). Vor der statistischen Verarbeitung wird im Normalfall ein Segmentierungs- bzw. Tokenisierungsvorgang durchgeführt, um die eingelesenen Texte in einzelne Wörter, lexikalische Einheiten bzw. Tokens zu zerlegen, wobei Leer- und Interpunktionszeichen als Trennungsgrenzen dienen (vgl. Witschel 2004: 56). Daraufhin ermitteln die Werkzeuge grundsätzlich die statistische Verteilung bestimmter Wörter bzw. Wortkombinationen und identifizieren diejenigen Ausdrücke mit hoher Auftretenshäufigkeit als Termkandidaten. Diese Strategie beruht gleichsam auf der Annahme, dass die mehrfache Wiederholung bestimmter lexikalischer Einheiten in enger Beziehung zu ihrer terminologischen Relevanz, d. h. Termhood, steht. Diese Vermutung trifft zwar nicht immer ein, kann aber bei umfangreicheren Textkorpora eines spezifisches Fachgebiets, in denen die Schlüsseltermini logischerweise insbesondere häufig vorkommen, zu vernünftigen Ergebnissen beitragen (vgl. Love 2000: 65).

Durch die statistische Berechnung lässt sich ebenfalls der Grad von Unithood feststellen und dies ermöglicht die Erkennung von Mehrwortbenennungen. In dieser Hinsicht kann von der folgenden Annahme ausgegangen werden: Die lexikalischen Bestandteile einer fachsprachlichen Mehrwortbenennung neigen statistisch dazu, so häufig in der unmittelbaren Nähe voneinander aufzutreten, dass dies nicht als Zufall gewertet werden könnte, sondern vielmehr aufgrund ihrer ausgeprägten Zusammengehörigkeit (vgl. Hartley 2009: 114). Zur Überprüfung der Gültigkeit dieser Annahme kann einerseits eine Methode aus der Informationstheorie angewandt werden: Dabei handelt es sich um die Berechnung der sogenannten gegenseitigen Information (GI), die als eine Maßzahl zur Angabe der Stärke des statistischen Zusammenhangs zweier Zufallseinheiten dient. Im Sinne des Zusammenhalts zweier Tokens eines potenziellen Terminus wird grundsätzlich die Häufigkeit ihrer gemeinsamen Erscheinung im Verhältnis zu ihrer getrennten ermittelt (vgl. Pazienza et al. 2005: 262). Je größer der GI-Wert, desto geringer ist die Unsicherheit über den Unithood-Grad der Kollokation. Andererseits kann die heuristische Methode der sogenannten Kookkurrenz zum Einsatz gebracht werden, wobei die Häufigkeit des gemeinsamen Auftretens zweier lexikalischer Einheiten, innerhalb eines bestimmten Umfelds bzw. einer übergeordneten Einheit, z. B. in einem Satz, Text oder einem Korpus (vgl. Ahmad & Rogers 2001: 748; Arntz et al. 2014: 247) bemessen wird. Wird ein hoher Frequenzwert bei einem Wortpaar erfasst, gilt dieser als ein Indikator für einen ausgeprägten Zusammenhalt der Kollokation und

dementsprechend für einen gültigen Mehrworttermkandidaten (vgl. Love 2000: 66). Da die Wörter nicht unbedingt in einer festen Reihenfolge auftreten müssen, um identifiziert werden zu können, bietet die zweite Methode insbesondere einen vorteilhaften Umgang mit Fachphraseologie, bei der auf pragmatischer Ebene eine Anfälligkeit für leichte formale Varianz vorliegt (vgl. Witschel 2004: 102). In einigen statistisch basierten Methoden wird nicht nur die Frequenz berücksichtigt, sondern auch der Gewichtungsfaktor wird miteinbezogen, um die terminologische Relevanz einzelner Termkandidaten in den eingegebenen Dokumenten eines Korpus zu beurteilen. Bei solchen Strategien werden in der Regel einzelne Tokens unterschiedlich errechneter Gewichtungen je nach vorbestimmten Auftrittskriterien zugewiesen, wie die Stellen des Terminus im Text oder seine syntaktische Struktur (vgl. Valaski et al 2015: 479). Bei der Identifizierung von Mehrwortbenennungen kann die erfolgreich festgestellte Termhood-Eigenschaft eine entscheidende Rolle spielen: Einem Wortpaar ist eine erhöhte Erkennungswahrscheinlichkeit zugewiesen, wenn eine ihrer intrinsischen lexikalischen Einheiten bereits aufgrund ihrer hohen Auftretenshäufigkeit im gesamten Korpus als Einworttermkandidat bewertet wurde und somit „schon eine gewisse Fachspezifik (Termhood) mitbringt“ (Witschel 2004: 25).

Zur Verfeinerung und Präzisierung der Extraktionsergebnisse besteht bei einigen Termextraktionssystemen die Möglichkeit, Stoppwortlisten miteinzubeziehen (Cabr  et al. 2001: 79). Diese Listen enthalten unter anderem Funktionsw rter (Artikel, Konjunktionen, Pr positionen), aber auch gemeinsprachliche Ausdr cke, die w hrend des Extraktionsprozesses als terminologisch irrelevant eingestuft und somit ausgeschlossen werden (vgl. Ahmad & Rogers 2001: 741; Drewer & Schmitz 2017: 46). Da die extrahierten Benennungskandidaten beim statistischen Verfahren in der Regel nach ihrem Frequenzwert gereiht werden, soll zur sinnvollen Reduktion des Ergebnisumfangs ein Schwellenwert vor dem Extraktionsvorgang festgelegt werden. Bei dieser sogenannten Thresholding-Konfiguration kann eine Mindestfrequenz bestimmt werden, ab der die analysierten Benennungen als Ergebnisse auf der Termkandidatenliste angezeigt werden. Durch die Festlegung einer hohen H ufigkeitserfordernis l sst sich die Qualit t des Extraktionsergebnisses steigern, was zu weniger Nachbearbeitungsaufwand f hrt (vgl. Arntz et al 2014: 247).

Der gro e Vorteil der rein statistischen Verfahren besteht darin, dass sie bei der Identifizierung und Extraktion von Benennungskandidaten nicht auf einzelsprachspezifisches Regel- oder Muster-Repertoire angewiesen sind. Durch diese Sprachunabh ngigkeit wird eine terminologische Untersuchung von Textmaterialien in allen Sprachen gew hrleistet (vgl. Cabr  et al. 2001: 55), vor allem f r die Verarbeitung derjenigen seltenen Sprachen, f r die keine

Termextraktionssysteme mit linguistisch basierter Analyse zur Verfügung stehen (vgl. Drewer & Schmitz 2017: 49). Aufgrund der probabilistischen Grundlage stellen die statistisch basierten Methoden schließlich einen sinnvollen Ansatz für die Ermittlung großvolumiger Korpora dar: Je umfangreicher die Textmenge, desto höher ist die Auftrittswahrscheinlichkeit potenzieller Benennungen und somit besserer Extraktionsergebnisse (vgl. Love 2000: 65).

5.6.2 Linguistische Extraktionsverfahren

Linguistisch basierte Terminologieextraktionsverfahren greifen zur Identifizierung von Termkandidaten auf spezifische linguistische Daten einzelner Sprachen zurück, die bereits in die Extraktionsprogramme integriert wurden. Dieses linguistische Wissen stammt normalerweise aus eingearbeiteten morphologischen und syntaktischen Regeln sowie aus festgelegten Musterstrukturen (vgl. Drewer & Schmitz 2017: 49). Mithilfe des umfassenden einzelsprachspezifischen linguistischen Regelwerks wird den Programmen ermöglicht, vor der Extraktion die Flexion der Wörter zu neutralisieren. Hier wird von der Lemmatisierung gesprochen, bei der unterschiedliche flektierte Wortformen auf ihre Lemmata, d. h. Grundformen, zurückgeführt werden (vgl. Witschel 2004: 58). Dieser Vorgang sorgt dafür, dass sich die Trefferquote von Termkandidaten in Folge der Abgleichung der lemmatisierten lexikalischen Einheiten mit den hinterlegten Wörterbuchinformationen, in denen die Wörter auch in Grundformen vorliegen, erhöht (vgl. Love 2000: 45). Außerdem verringert sich der Nachbearbeitungsaufwand, denn die Termkandidaten werden nicht durch zahlreiche variantenreiche Flexionsformen wiederholt ausgegeben (vgl. Zielinski & Ramírez-Safar 2005: 3). Darüber hinaus stehen die lemmatisierten Termkandidaten bereits in der notwendigen Form für die weiteren Schritte in der Terminologieverwaltung zur Verfügung (vgl. Arntz et al. 2014: 248).

Gleichzeitig kommen die vorbestimmten linguistischen Musterstrukturen zum Einsatz und werden mit den auftretenden Einzelwörtern und Wortkombinationen in eingelesenen Texten verglichen. Das Ziel ist es, Übereinstimmungen festzustellen und diese als Termkandidaten zu identifizieren (vgl. Witschel 2004: 43). So handelt es sich laut Wanhoff (2007: 55 f.) bei dieser Methode zur Termextraktion um „pattern matching“ oder die Mustererkennung. Diese Strategie befähigt die Termextraktionssysteme dazu, Benennungen über die Informationen der integrierten Lexika hinaus zu erkennen (vgl. Love 2000: 46). Bei der Identifizierung von Einwortbenennungskandidaten wird sich auf ihre morphologischen Merkmale konzentriert, somit auf Wortbestandteile, die für die Terminologie des jeweiligen Fachgebiets charakteristisch sind (vgl. Witschel 2004: 40 ff.). Dazu gehören nicht nur Präfixe

und Suffixe bei Derivaten, sondern auch typische Wortbausteine (Morpheme) und selbständige Wortstämme, sodass Komposita auch als Einwortbenennungen identifiziert werden können (vgl. Zielinski & Ramírez-Safar 2005: 4). Andererseits erfolgt die Mustererkennung hinsichtlich der Mehrwortbenennungen dadurch, dass sowohl Morphologie als auch Syntax der jeweiligen Sprache analysiert werden. Diese Methode erfordert die Festlegung von morphosyntaktischen Musterstrukturen, welche die Mehrwortbenennungen eines spezifischen Fachgebiets typischerweise aufweisen (vgl. Love 2000: 46; Wanhoff 2007: 57). Anhand der dargestellten charakteristischen Mehrwortbenennungen in deutschen Fachsprachen sollen somit beispielsweise mindestens die Kombinationen von Wortarten Adjektiv + Nomen; Nomen + Verb; und Nomen + Präposition + Nomen) gebildet werden, um die grundlegenden Muster der Fachausdrücke (z. B. Nominalphrasen) sowie der Fachphraseologie abzudecken (vgl. Heid 1999: 2 f.). Entsprechen die ermittelten Ausdrücke der hinterlegten Musterauflage, so werden sie als Termkandidaten bewertet.

Die Methode zum Mustervergleich anhand der festgelegten Wortartmuster führt zu möglichst brauchbaren Extraktionsergebnissen erst dann, wenn sich die Wortarten sämtlicher lexikalischer Einheiten in den zu untersuchenden Korpora auch feststellen lassen (vgl. Witschel 2005: 42 f.). Um dies zu gewährleisten, findet im Zuge des Vorverarbeitungsprozesses der linguistischen Verfahren im Anschluss an den automatisierten Segmentierungsvorgang noch ein weiterer Schritt statt: die linguistische Anreicherung (vgl. Mahlstedt 2010: 21). Der Zweck dieses Vorgangs besteht darin, die Eingangsmaterialien mit linguistischen Informationen zu annotieren (vgl. Heid 2004: 472 f.). Zu den oft eingesetzten Methoden zur linguistischen Anreicherung zählen unter anderem Part-of-Speech-Tagging (POS-Tagging) und Parsing. Unter Ersterem wird die Erfassung und Kennzeichnung einzelner lexikalischer Einheiten bzw. Wörter mit ihren entsprechenden Wortarten (deshalb auch Wortart-Tagging) verstanden. Hingegen stellt Parsing einen komplexeren Vorgang dar: Ein Parser führt zunächst eine syntaktische Analyse durch, indem jeder Satz innerhalb der Eingangskorpora in seine grammatikalischen Bestandteile zerlegt wird, und erstellt aus diesen Daten einen Ableitungsbaum (oder auch Parse-Baum) von Satzelementen, die mit POS-Markierungen versehen sind (vgl. Mahlstedt 2010: 22). Die Syntaxanalyse dient darüber hinaus der Disambiguierung hinsichtlich derjenigen Tokens, denen mehrere Wortarten zugeordnet werden können. Die Darstellung der Relationen innerhalb einer Satzhierarchie weist die eigentliche Funktion und damit die entsprechende Wortart einzelner Satzbestandteile auf (vgl. Heid 2006: 101).

Die linguistisch basierte Beschaffenheit der beiden Methoden schließt ein, dass linguistische Ressourcen, wie grammatikalische Regelsätze und detaillierte Wörterbuchinformationen, als eine abdingbare Voraussetzung für die erfolgreiche linguistische Überprüfung gelten. Dementsprechend ist die Qualität der Vorverarbeitungs- und somit der Extraktionsergebnisse in hohem Maß abhängig von der Angemessenheit des einzelsprachlichen linguistischen Wissens, das für das jeweilige Werkzeug vorhanden ist (vgl. Heid 2006: 102 ff.). Im Gegensatz zu den statistischen Extraktionsverfahren werden beim linguistischen Ansatz morphologische sowie morphosyntaktische Musterübereinstimmungen als Indiz für Termkandidaten betrachtet, sodass auch einige potenzielle Benennungen trotz ihrer geringen Auftretenshäufigkeit in den Korpora gefunden und extrahiert werden (vgl. Mahlstedt 2010: 24).

5.6.3 Hybride Extraktionsverfahren

Von hybriden Termextraktionsverfahren wird gesprochen, wenn zusätzlich zur statistischen Ermittlung auch die morphosyntaktische Analyse dem Prozess zur automatisierten Identifizierung von Termkandidaten hinzugefügt wird (vgl. Hartley 2009: 114). Der hybride Ansatz nutzt die Vorteile der jeweiligen Verfahren, um ihre inhärenten Nachteile (siehe 5.6.4) gegenseitig ausgleichen zu lassen und damit die Qualität der Extraktionsergebnisse zu maximieren (vgl. Georgantopoulos & Piperidis 2000: 2). Grundsätzlich werden zunächst im Rahmen der linguistischen Überprüfung der Vorverarbeitungsvorgänge die morphosyntaktische Analyse und Annotation (z. B. POS-Tagging und Parsing) und Grundformenbildung (Lemmatisierung) durchgeführt. Die sich daraus ergebende Liste potenzieller Termkandidaten durchläuft dann eine statistisch basierte Untersuchung, damit die terminologische Relevanz zum Fachgebiet (Termhood) sowie die Stärke der Kollokationen (Unithood) dieser Benennungen numerisch mit einer spezifischen maschinengestützten Berechnungsmethode festgestellt werden können. Daraufhin folgen die Einordnung nach den Ausgangswerten und die Ausschließung derjenigen Tokens, die keine statistische Signifikanz aufweisen und somit nicht als relevante Termkandidaten gelten. Die linguistisch begründeten Extraktionsergebnisse aus der Vorstufe sollen die Zuverlässigkeit der nachkommenden statistisch basierten Untersuchung erhöhen (vgl. Pazienza et al. 2005: 263). Dementsprechend dient der hybride Ansatz der Aufrechterhaltung des Gleichgewichts zwischen zwei Anforderungen: der möglichst hohen Abdeckung von Termkandidaten und deren Qualität (vgl. Justeson & Katz 1995: 16).

Inzwischen wurden unterschiedliche Einsatzmöglichkeiten der hybriden maschinengestützten Termextraktion von verschiedenen AutorInnen untersucht und

vorgeschlagen. Earl (1970) stellte z. B. einen der ersten Anwendungsfälle des hybriden Ansatzes bei einer maschinengestützten Terminologieextraktion vor, wobei das System zunächst nach Nominalphrasen in einem Ausgangskorpus sucht, die mit denjenigen im hinterlegten Stichwortverzeichnis eines Buchs zu demselben Fachgebiet übereinstimmen, und diese als Termkandidaten bewertet. Die Hauptwörter bzw. nominalen Glieder einzelner Ergebnisse werden daraufhin statistisch auf deren Auftretenshäufigkeit untersucht, um die Wörter mit einem geringen Frequenzwert aus der Termkandidatenliste zu entfernen (vgl. 1970: 274 f.). Daille (1994: 3 ff.) setzt ebenfalls einen zweifachen Ansatz ein: Die durch POS-Tagging linguistisch angereicherten Korpora unterziehen sich einem morphosyntaktischen Mustervergleich, um potenzielle Termkandidaten herauszufiltern. Die Ergebnisse werden dann statistisch mithilfe der simplistischen Frequenzermittlung und der GI-Berechnung einsortiert. Justeson und Katz (1995: 17 ff.) verfolgen einen ähnlichen Ansatz, fügen allerdings zur Verfeinerung der Termkandidatenliste zusätzlich noch gemeinsprachliche Wörterbuchinformationen dem linguistischen Analysevorgang hinzu. So fassen Georgantopoulos und Piperidis (2000: 2) anhand des folgenden Flussdiagramms (Abb. 4) zusammen, wie ein üblicher Algorithmus der hybriden Extraktionsverfahren abläuft.

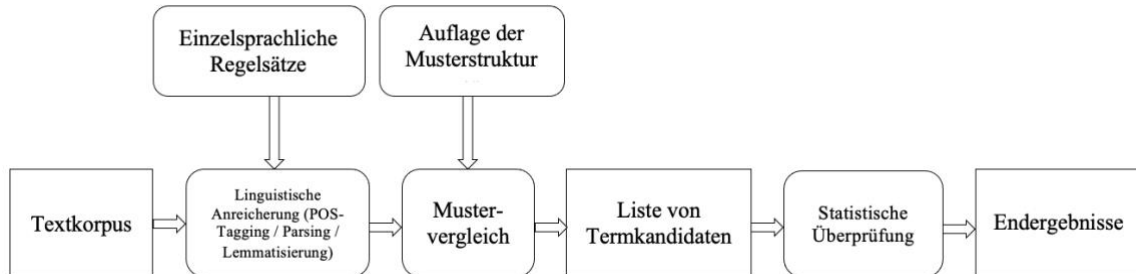


Abbildung 4: Ablauf der hybriden Extraktionsverfahren
(adaptiert von Georgantopoulos & Piperidis 2000: 2)

Die Endergebnisse des Extraktionsprozesses werden nach der Frequenz einsortiert, wobei die Termkandidaten mit den höchsten Frequenzwerten am wahrscheinlichsten als echte Termini des entsprechenden Fachgebiets zu beurteilen sind. Zur Präzisierung der Termkandidatenliste kann im Vorfeld der statistischen Überprüfung eine Mindestfrequenz des Auftretens bestimmt werden, die aber auch in einem angemessenen Verhältnis zum Umfang des untersuchten Korpus stehen soll. Mit diesem Ansatz kann eine möglichst umfangreiche Identifizierung von Benennungen erzielt werden, vorausgesetzt, dass sämtliche lexikalische Einheiten der Eingabe zweckmäßig vorverarbeitet werden, und dass das Terminologieextraktionsprogramm im Einsatz selbst über umfassende linguistische Regelsätze

und Musterstrukturen verfügt (vgl. Love 2000: 68). Valaski et al. (2015: 483) stellten im Rahmen ihrer Untersuchung zu verschiedenen Terminologieextraktionsansätzen und -methoden fest, dass hybride Verfahren die bedeutsamsten Ergebnisse produzieren. Insbesondere die Mischform, die sich zum einen auf die linguistische Mustersuche zur Identifizierung nominaler Glieder von Termkandidaten konzentriert und die zum anderen das numerische Gewichtungsschema einsetzt, erweist sich in der Studie als die effizienteste Strategie.

5.6.4 Grenzen der maschinengestützten Terminologieextraktion

Der Vorteil der maschinengestützten Terminologieextraktion liegt zwar hauptsächlich in der leichteren Verarbeitbarkeit der Daten und insbesondere in der Zeitersparnis, jedoch spielt, wie erläutert, menschliches Sprach- und Sachwissen eine entscheidende Rolle in der Qualitätssicherung der daraus resultierenden Termkandidatenlisten. Die maschinelle Unterstützung zieht je nach genutztem Werkzeug und Verfahren unterschiedliche nachteilige Aspekte nach sich. Im Zusammenhang mit der Diskussion über die möglichen Grenzen der maschinengestützten Terminologieextraktion sollen zunächst die sogenannten Silence- und Noise-Effekte erörtert werden, somit unerwünschte Reduzierungen bzw. Erweiterungen bei den Ausgabedaten. Da diese in unterschiedlichem Maß durch inhärente Beschränkungen des jeweiligen Extraktionsansatzes verursacht werden, werden sie als Faktoren in die Evaluierung der entsprechenden Extraktionsleistung einbezogen.

Als Noise-Effekte werden Wörter bezeichnet, die das Terminologieextraktionssystem auf die Termkandidatenliste aufnimmt, die aber eigentlich im Zuge der menschlichen Überprüfung aufgrund terminologischer Irrelevanz verworfen werden müssen. Anders ausgedrückt erzeugen solche fehlerhaften Vorschläge somit unnötigen Lärm. Hingegen gibt es Silence-Effekte beim Versäumnis, relevante Termini in Korpora zu identifizieren oder alle Bestandteile der Termini, insbesondere hinsichtlich der Mehrwortbenennungen, zu erkennen (vgl. Drewer & Schmitz 2017: 46 f.). In der Regel sind die automatischen Terminologieextraktionswerkzeuge so konzipiert, dass sich der Anteil von Silence und Noise durch Einstellungen des sogenannten Qualitätsfilters steuern lässt. Es kommt darauf an, ob dabei Wert auf eine möglichst hohe Abdeckung der potenziellen Benennungen oder die Relevanz der Ergebnisse gelegt wird. In der praktischen Terminologiarbeit wird sogar häufig Ersteres Letzterem als Ziel vorgezogen, indem somit möglichst viele Termkandidaten extrahiert werden sollen, auch auf die Gefahr hin, dass gemeinsprachliche oder andere nicht gewünschte Ausdrücke in die Termkandidatenliste geraten. Es ist wesentlich einfacher für die

überprüfenden Personen, die Noise-Effekte durch Aussortierung der falschen Termkandidaten zu beheben, als manuell nach ausgelassenen potenziellen Termini zu suchen, insbesondere im Zusammenhang mit umfangreichen Korpora (vgl. Love 2000: 41). Das Ausmaß der Noise- und Silence-Effekte variiert, je nachdem welches Extraktionsverfahren verwendet wird.

Im Hinblick auf die statistischen Extraktionsverfahren hängen die Noise- und Silence-Effekte mit dem vorbestimmten Schwellenwert der Auftretenshäufigkeit zusammen. Wird eine hohe Mindestfrequenz gefordert, kann dies hohe Silence-Werte zur Folge haben, da relevante Benennungen aufgrund deren seltenen Erscheinung in den Korpora bei der Extraktion nicht identifiziert werden (vgl. Mahlstedt 2010: 23). Der Silence-Effekt steigert sich insbesondere durch die wesentliche Abhängigkeit der Ergebnisqualität statistischer Verfahren von der Größe der Datenmenge, wenn es sich um kurze Texte bzw. kleine Korpora handelt. Einzelne Wörter kommen logischerweise aufgrund des beschränkten Textumfangs kaum bis gar nicht wiederholt vor (vgl. Witschel 2005: 98). Umgekehrt sorgt eine tendenziell niedrige Häufigkeit als Schwellenwert dafür, dass alle potenziellen Benennungen gemeldet werden, dies löst aber gleichzeitig auch einen hohen Anteil von Noise-Werten aus (vgl. Cabré et al. 2001: 83). Daher lässt sich ein optimaler Schwellenwert für bestimmte Textkorpora laut Arntz et al (2014: 247) lediglich durch Probieren und Anpassung ermitteln. Darüber hinaus ist der rein statistische Ansatz aufgrund der fehlenden linguistischen Ressourcen vor allem bei Sprachen mit komplexen morphosyntaktischen Eigenschaften vergleichsweise fehleranfällig und liefert somit mangelhafte Ergebnisse. Dies trifft deutlich auf Mehrwortbenennungen zu, deren Kollokationen sich nicht statistisch erschließen. Insgesamt bleibt die Qualität der statistisch abgeleiteten Ergebnisse in dieser Hinsicht hinter denen mit linguistischen Verfahren zurück (vgl. ebd.).

Programme bzw. Werkzeuge mit linguistisch basierten Extraktionsmethoden liefern zwar aufgrund des umfassenden sprachspezifischen Wissens bessere Ergebnisse, aber gerade die ausgeprägte Sprachgebundenheit stellt bei diesem Ansatz einen allgemeinen Nachteil dar (vgl. Mahlstedt 2010: 25). Linguistische Terminologieextraktionsverfahren lassen sich nur beschränkt auf diejenigen Sprachen anwenden, für die bisher die damit verbundenen linguistischen Regelwerke und Informationen erarbeitet und in die Systeme eingebaut wurden. Daher stehen die Daten, die sowohl für die morphosyntaktische Analyse als auch für die linguistische Anreicherung erforderlich sind, nicht für alle Sprachen zur Verfügung (vgl. Zielinski & Ramírez-Safar 2005: 5). Auch wenn sich solche voreingestellten morphosyntaktischen Muster insbesondere vorteilhaft für die Identifizierung von Mehrwortbenennungen eignen, können sie zu einem hohen Anteil von Noise-Effekten

beitragen, indem irrelevante Ausdrücke, die zufällig aus Tokens in den entsprechenden Musterstrukturen bestehen, ebenfalls als Termkandidaten bewertet werden (vgl. Hartley 2009: 115). Andererseits wies Cabré (2001: 84) auf ein mögliches Silence-Szenario hin: Sind bestimmte formale Varianten und unterschiedliche Wortstellungsmöglichkeiten bei der Fachphraseologie aufgrund ihrer Trennbarkeit in der Gesamtheit der Suchmusterkriterien nicht eingeschlossen, so verhindern jegliche Abweichungen die Zuordnung des entsprechenden POS-Musters während des Extraktionsvorgangs und werden vom Werkzeug nicht identifiziert. L'Homme et al. (1996: 296 f.) und Daille (1996: 49) stellten dementsprechend hinsichtlich der linguistisch basierten Termextraktionsverfahren fest, dass die Zunahme oder die Senkung des Noise-Anteils, der durch Extraktionsprogramme ausgelöst wird, in direkter Relation zur Anzahl der zu ermittelnden terminologischen Musterstrukturen steht. Hingegen wächst der Silence-Anteil umso deutlicher, je mehr lexikalische Regeln zur Ausschließung bestimmter Wörter bzw. Wortgruppen eingearbeitet werden.

Um optimale Ergebnisse zu erzielen, müssen laut Heid (2006: 104 ff.) idealerweise ein vollständiger Parsing-Prozess durchgeführt werden und die linguistisch annotierten Korpora dann vor der Eingabe in die Extraktionsprogramme einer gründlichen menschlichen Kontrolle unterzogen werden. Jedoch lässt sich diese Arbeitsmethode aufgrund des hohen Aufwands aus praktischer und finanzieller Sicht nicht realisieren. Deshalb wird, wie eingangs bemerkt, tendenziell angestrebt, den Silence-Wert auf das Minimum zu reduzieren und möglichst umfassende Vorschläge von Termkandidaten, somit Noise-Effekte, zu erlangen. Zur Optimierung der Terminologieextraktionsergebnisse wird gegenwärtig die Anwendung der beiden Verfahren in einem hybriden Ansatz in Betracht gezogen. Dabei soll der zweischichtige Qualitätsfilter genutzt werden, bei dem der Schritt mit der Analyse der (morpho-)syntaktischen sowie syntagmatischen Analyse erfolgt, deren Ausgabe anschließend durch eine Methode zur statistischen Ermittlung bestätigt wird (vgl. Vázquez & Oliver 2018: 125 f.). Obwohl die hybriden Verfahren die Beschränkungen der statistischen und linguistischen Termextraktion vor allem hinsichtlich der Noise- und Silence-Effekte zu einem gewissen Grad bewältigen können, ist auch bei ihrem Einsatz unvermeidlich der Nachteil der Sprachabhängigkeit in Kauf zu nehmen, die mit den linguistisch basierten Verfahren einhergeht (vgl. Drewer & Schmitz 2017: 50). Das heißt, neben der umfangreichen Neufestlegung der sprachspezifischen Musterkriterien für jede einzelne Sprache zieht eine Erweiterung um zusätzliche Sprachen außerdem eine vollständige Er- und Einarbeitung der linguistischen Ressourcen in Programmen nach sich (vgl. Love 2000: 47 f.).

5.7 Evaluierung der Terminologieextraktion

Nach der Erörterung des umfassenden theoretischen Rahmens bezüglich der gebräuchlichsten Verfahren und dazugehörigen Methoden zur maschinengestützten Terminologieextraktion soll die Evaluierung der Extraktionsleistung im Zusammenhang mit dem Einsatz von Terminologieextraktionsprogrammen eingehend betrachtet werden. Grundsätzlich handelt es sich bei der Evaluierung im Bereich der maschinellen Sprachverarbeitung immer um eine vergleichende Adäquatheitsanalyse: Hinsichtlich der (semi-)automatisierten Terminologiegewinnung besteht die Evaluierungsmotivation dementsprechend darin, je nach vorgegebenen Anforderungen und Zwecksetzungen eine bessere – wenn nicht gar universell optimale – Extraktionsstrategie zu ermitteln (vgl. Thurmair 2003: 169). Zunächst wird in den Abschnitten 5.7.1 und 5.7.2 auf die wesentlichen Aspekte der komparativen Termextraktionsevaluierung eingegangen, wobei der Schwerpunkt auf der quantitativen Beurteilung des Termidentifizierungsergebnisses liegt.

5.7.1 Grundlagen der Termextraktionsevaluierung

Dass jedes Werkzeug aufgrund seiner spezifischen Konstruktionselemente auf unterschiedliche Methoden zur Termextraktion zurückgreift und eigene Grenzen mit sich bringt, muss bei der Termextraktionsevaluierung in Kauf genommen werden. Deshalb lässt sich die Entwicklung einer standardisierten Evaluierungsmethodologie so schwer verwirklichen, dass es bisher keine konsensfähige Aussage bezüglich des allgemein effizientesten Extraktionsverfahrens bzw. -Programms getroffen werden kann (vgl. Zerfaß 2006: 24). Im Vorhinein gilt es, die Festlegung des Evaluierungsziels und -umfangs vorzunehmen, damit die entsprechende Evaluierungsmethodologie und die Kriterien bestimmt werden können (vgl. Zerfaß 2006: 25).

In dieser Hinsicht haben L’Homme et al. (1996: 294) eine Metrik zur umfassenden Evaluierung von Termextraktionsprogrammen erarbeitet. Diese beinhaltet zwei Untersuchungsschwerpunkte, die sich in die Vorevaluierung und die Qualitätsauswertung untergliedern. Die erste Phase findet vor der eigentlichen Verarbeitung statt: Das jeweilige Werkzeug wird anhand der bedienungsorientierten Anforderungen auf seine generelle Eignung zur geplanten Nutzung untersucht. In Betracht gezogen werden dementsprechend allgemeine Anwendungsbereiche, darunter hauptsächlich NutzerInnengruppen, Termidentifizierungsmethoden, unterstützte Sprachen und damit verbundene einzelsprachliche Ressourcen, Einarbeitungsmöglichkeiten für zusätzliche linguistische Daten, der Umfang der extrahierten Termkandidaten sowie die Ergebnisdarstellung. Zu den Überlegungen gehören

aber auch Aspekte hinsichtlich der Kompatibilität des Programms mit anderen sprachverarbeitenden Systemen, wie vorhandene Parameter und Formate der Datenübertragung (vgl. L’Homme et al. 1996: 295 f.). Die zweite Evaluierungsphase wird erst nach dem Extraktionsprozess durchgeführt, indem die Extraktionsleistung des jeweiligen Programms mithilfe der quantitativen Adäquatheitsanalyse seiner unbearbeiteten Extraktionsergebnisse bewertet wird. Grundsätzlich bezieht sich die Auswertung auf die Frage, ob das Werkzeug in der Lage ist, möglichst viele terminologische Einheiten innerhalb des Ausgangskorpus mit einem minimalen Noise-Anteil zu identifizieren (vgl. L’Homme et al. 1996: 301 f.).

Der oben angeführte Evaluierungsrahmen gilt als Grundlage für die gegenwärtig etablierte Unterscheidung zwischen der ganzheitlichen Beurteilung und der Betrachtung der internen Konfigurationen jeglicher (semi-)automatisierter sprachverarbeitender Werkzeuge und somit auch Termextraktionsprogramme (vgl. Vivaldi & Rodríguez 2007: 226). In Bezug auf die ganzheitliche Beurteilung handelt es sich um das sogenannte Black-Box-Evaluierungsmodell, bei dem der intrinsische Aufbau des zu analysierenden Programms selbst im Rahmen der Evaluierung nicht bekannt ist bzw. nicht berücksichtigt wird. Somit beschränkt sich die Untersuchung zumeist auf die äußerlichen Programmeigenschaften und dessen allgemein zugängliche Benutzerschnittstelle (vgl. Saß 2004: 66 f.). Das zweite Evaluierungsmodell wird hingegen als Glass-Box-Modell bezeichnet, da hierbei ein tiefer Einblick in die technische Funktionalität einzelner innerer Komponenten des Programms und somit bestimmter Algorithmen der Termextraktionsverfahren ermöglicht wird, wie integrierte Methoden zur Segmentierung und zur linguistischen Anreicherung, die Bearbeitung von Sonderzeichen, das Berechnungsmodell zur statistischen Ermittlung des Termhood- und Unithood-Grads von Termkandidaten usw. (vgl. Cabré et al. 2001: 79 f.). Mit diesem Wissen wird sich bei der Glass-Box-Methodologie gezielt auf die Extraktionsleistung konzentriert, die in erster Linie numerisch basiert anhand der Ausgabe von Termkandidaten ermittelt wird (vgl. Mahlstedt 2010: 10).

Aufgrund der diversen Anwendergruppen von Terminologieextraktionsprogrammen sollten keine allgemeingültigen Kriterien insbesondere bei der Vorevaluierung bzw. dem nutzungsbedingten Black-Box-Ansatz verwendet werden. Vielmehr sind die Anforderungen und die Methodologie je nach Bedürfnissen zweckmäßig möglichst detailliert auszuarbeiten, sodass aussagekräftige Urteile über die Angemessenheit des jeweiligen Programms getroffen werden können (vgl. L’Homme et al. 1996: 309; Kageura et al. 2001: 154f; Heid 2006: 98). Vivaldi und Rodríguez (2007: 227) stimmen zwar zu, dass sich die Bestimmung eines „goldenen Standards“ für die Evaluierung nicht ohne Schwierigkeit realisieren lässt. Jedoch

schlagen sie vor, wenigstens einige wesentliche Evaluierungsparameter, wie die Korpusgestaltung sowie Standardmessungen und -instrumente, festzulegen, die aber je nach Bedingungen und Zwecken angepasst werden können (vgl. Vivaldi & Rodríguez 2007: 244). In diesem Zusammenhang betonen Frantziy et al. (2000: 116), dass die Qualitätsanalyse der Extraktionsergebnisse mittels einer quantitativen Messung eine entscheidende Rolle in vergleichenden Evaluierungen unterschiedlicher Terminologieextraktionsverfahren bzw. -Programme spielt bzw. einnehmen sollte, da diese Methode minimal auf die Anwenderunabhängigkeit. Dabei können entweder die Noise- und Silence-Anteile im Verhältnis zu den gesamten Termkandidaten ermittelt oder das Wertpaar Recall und Precision kann als Maß des Qualitätsindikators verwendet werden (vgl. Vivaldi & Rodríguez 2006: 12). Die quantitative Evaluierungsmethode wird im folgenden Abschnitt eingehend erläutert.

5.7.2 Quantitative Evaluierung der Terminologieextraktion

Eine quantitative Evaluierung der Terminologieextraktion kann aus zwei Perspektiven erfolgen: Linguistik und Information-Retrieval (IR). Der Ansatz aus dem sprachwissenschaftlichen Bereich verwendet Zahlen zu Noise- und Silence-Werten, um die Qualität des Outputs automatischer Terminologie-Extraktionswerkzeuge zu beurteilen, während das IR die Bewertung auf Maße für Precision (=Genauigkeit) und Recall (=Trefferquote) stützt (vgl. Vivaldi & Rodríguez 2007: 230). Wie geschildert, sind Noise- und Silence-Fehler, die durch die Einschränkungen von Programmen bzw. Verfahren zur Termextraktion verursacht werden. Ersteres tritt auf, wenn das Werkzeug Wörter vorschlägt, die nicht als terminologische Kandidaten gelten, während Letzteres sich auf das Versagen des Programms bezieht, relevante Termini zu erkennen und diese in die Termkandidatenliste aufzunehmen (vgl. Mahlstedt 2010: 19). Andererseits basiert die Berechnung von Recall und Precision laut Sauron (2002: 8) auf der Messung der Relevanz der gefundenen Informationen: So stellt das Erstgenannte in diesem Zusammenhang einen Bruchteil der Gesamtmenge relevanter Termini dar, der tatsächlich vom Werkzeug extrahiert wird, wogegen Precision misst, wie viele der extrahierten Termkandidaten korrekt sind. Obwohl diese beiden Konzepte das Extraktionsergebnis unterschiedlich zu interpretieren scheinen, liefern sie im Wesentlichen die gleichen Informationen, und zwar die Richtigkeit und die Abdeckung der extrahierten Termkandidaten, wie dies in der von Vivaldi und Rodríguez (2007: 229) vorgestellten Berechnung in Abb. 5 zu sehen ist.

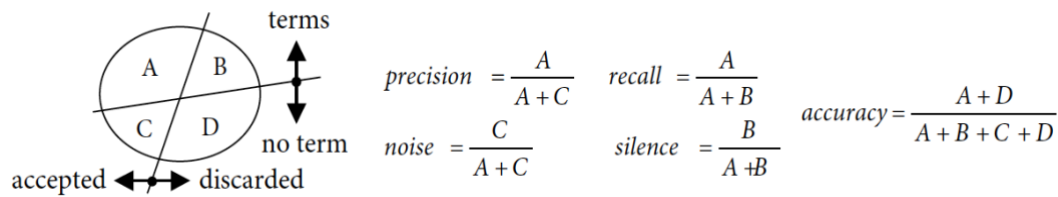


Abbildung 5: Maße zur quantitativen Evaluierung der Termextraktion
(vgl. Vivaldi & Rodríguez 2007: 229)

Der Abb. 5 zufolge spiegeln sich die beiden Wertpaare im Hinblick auf die Verhältnisberechnung gegenseitig wider: Im Rahmen der Ermittlung des Korrektheitsgrades von Extraktionsergebnissen wird die Gesamtzahl der Termvorschläge ($A + C$) als gemeinsame Basis in Betracht gezogen, wobei die Precision-Erfassung von der Anzahl der korrekt extrahierten Termini (A) ausgeht und somit auf die Richtig-Positiv-Rate abzielt, während der Noise-Wert anhand des Bezugs auf die ungültigen Termkandidaten (C) die Falsch-Positiv-Rate repräsentiert. Um andererseits den Vollständigkeitsgrad der Termkandidatenliste im Verhältnis zur gesamten Menge der im Korpus enthaltenen Termini ($A + B$) festzustellen, kann entweder Recall oder Silence ausgerechnet werden. Der Unterschied zwischen den beiden Maßen besteht darin, ob der Schwerpunkt auf der Richtig-Positiv-Rate (A) liegt oder auf der Richtig-Negativ-Rate, indem die relevanten Termini, die durch das Werkzeug im Laufe des Extraktionsvorgangs nicht erkannt werden (B), als Ausgangspunkt für die Berechnung dienen.

Die Berechnung von Recall bzw. Silence kann sich laut Vivaldi und Rodríguez (2006: 13) als anspruchsvoller als die anderen Maße herausstellen. Sie verlangt, dass alle relevanten Termini im Korpus und deren Anzahl im Vorhinein bekannt gemacht werden, damit eine genaue Bewertung der Termextraktionssysteme ermöglicht wird. Aufgrund der Schwierigkeit, den gesamten Satz von Termini in einem Korpus zu erhalten, berechnen einige Forschende alternativ nur den Noise- oder Precision-Wert und verzichten vollständig auf die Berechnung von Silence und Recall. Auf diese Weise ist es einfacher, eine spezialisierte Fachperson die Korrektheit der gegebenen Termkandidaten überprüfen zu lassen, als nach allen Termini in einem gegebenen Textkorpus zu suchen (vgl. Vivaldi & Rodríguez 2007: 231). Allerdings weist Sauron (2000: 9) darauf hin, dass der Einsatz der Metriken von Recall und Precision im Rahmen der Beurteilung der Leistung von Terminologieextraktionsprogrammen als eine effizientere Methode gilt – und dies idealerweise in Kombination, damit eine zuverlässige Schlussfolgerung erlangt werden kann. Durch die Feststellung dieser zwei Metriken kann überprüft werden, ob alle und nur die relevanten lexikalischen Einheiten extrahiert wurden.

Precision und Recall werden normalerweise mit den Werten von 0 bis 1 dargestellt. Je näher der Wert an 1 liegt, desto besser ist die Leistung des getesteten Programms. Alternativ können sie auch in Prozent berechnet werden. Van Rijsbergen (1979: 129) zufolge sollte eine Extraktion im Idealfall die maximale Punktzahl sowohl für Precision als auch für Recall ergeben, jedoch lässt sich dies in der Praxis kaum oder gar nicht erreichen. Darüber hinaus sind Precision und Recall asymmetrische Maße: Je größer der Umfang der Termkandidatenliste wird, desto höher ist die Chance, dass relevante Termini wiedererkannt werden, somit wächst der Recall-Wert. Zugleich umfasst die Termkandidatenliste auch mehr ungültige Identifizierungsergebnisse, sodass der Precision-Wert abnimmt, wie in der folgenden Abb. 6 nach Vivaldi & Rodríguez (2006: 14) dargestellt wird.

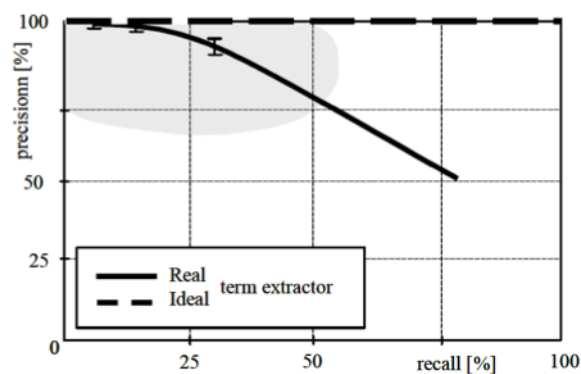


Abbildung 6: Typisches und ideales Verhältnis von Recall und Precision
(vgl. Vivaldi & Rodríguez 2006:14)

Aus diesem Grund gilt es, zur Bestimmung der Extraktionsleistung des jeweiligen Programms aus den beiden Metriken eine einzige Gesamtbewertung zu errechnen (vgl. Thurmair 2003: 174). Es existiert allerdings kein einziges Patentrezept für diesen Zweck, denn das Bewertungssystem und die damit verbundenen Faktoren hängen von der beabsichtigten Verwendung des Systems ab. Daher werden tendenziell mehrere Methoden bisher im IR bzw. in verwandten Disziplinen verwendet, um Precision und Recall zu einem Einzelmaß zusammenzufassen. Stehen lediglich die Extraktionsergebnisse im Mittelpunkt der Untersuchung, so reicht es laut Sauron (2000: 9 f.) auch, das F1-Maß im Rahmen der Ermittlung eines durchschnittlichen Genauigkeitswerts zu beziehen. Dieses ist ein bekanntes Verfahren zur Kombinierung von Recall und Precision, indem es ein gewichtetes harmonisches Mittel der beiden Metriken darstellt, die eine hohe Divergenz aufweisen (vgl. Vázquez & Oliver 2018: 134 f.). Auch hier wird das F1-Maß durch Werte von 0 bis 1 dargestellt, wobei 1 der höchsten Punktzahl für das untersuchte Werkzeug entspricht. Auf dieser Grundlage wird

üblicherweise die Akzeptanzschwelle auf 0,5 festgelegt (vgl. Sauron 2000: 10). Insgesamt lässt sich das F1-Maß mit der folgenden Formel in Abb. 7 feststellen.

$F1 = \frac{1}{\alpha \frac{1}{P} + (1 - \alpha) \frac{1}{R}}$	Wobei: P: Precision R: Recall α: Gewichtung	Wenn $\alpha = \frac{1}{2}$, dann $F1 = 2 \cdot \frac{P \cdot R}{P + R}$
--	--	--

Abbildung 7: Ermittlung des F1-Maßes (vgl. Vivaldi & Rodríguez 2007: 213)

Normalerweise werden in Abwesenheit zusätzlicher Informationen beide Komponenten, somit Precision und Recall, gleichermaßen gewichtet. Daher wird die GewichtungsvARIABLE α auf die Hälfte ($\frac{1}{2}$) gesetzt, wie in der abgebildeten Formel (Abb. 7 rechts) dargestellt. Soll aber je nach Evaluierungszweck ein größerer Wert auf den einen oder anderen Komponenten der Programmleistung gelegt werden, lassen sich Precision und Recall dementsprechend auch unterschiedlich gewichten (vgl. Witschel 2004: 78).

Eine quantitative Evaluierung mit Recall- und Precision-Werten benötigt allerdings einen Überprüfungsmaßstab, der als Bezugspunkt für die Bewertung der Extraktionsergebnisse zur Feststellung der Extraktionsleistung des jeweiligen automatischen Systems dient. Dafür werden in der Regel zwei verschiedene Methoden eingesetzt: Referenzliste und Validierung (vgl. Pazienza et al. 2005: 264 ff.). Bei dem ersten Ansatz wird zunächst ein Korpus zu einem bestimmten Gebiet von menschlichen ExpertInnen untersucht, um daraus eine a priori-Liste von Benennungen manuell zu erstellen. Nachdem aus demselben Korpus Termkandidaten automatisch extrahiert werden, soll diese Referenzliste als eine gewisse Benchmark dienen, auf deren Basis die Precision- und Recall-Werte zur Bewertung der Extraktionsleistung berechnet werden. Hingegen wird die Validierungsmethode bevorzugt, wenn sich der genaue Umfang der Terminologie zum entsprechenden Fachgebiet nicht feststellen lässt, sodass kein Maßstab im Vorhinein definiert werden kann. In diesem Fall wird die Systemleistung durch eine menschliche Fachperson bewertet, indem sie die Korrektheit der vom System extrahierten Termkandidaten validiert. Ein Precision-Wert wird so als der Anteil der extrahierten Termkandidaten errechnet, die von der überprüfenden Person als korrekt annotiert werden (vgl. Inkpen et al. 2016: 153).

Beide Methoden ziehen allerdings eigene Grenzen nach sich. Im Hinblick auf die Leistungsmessung ist die Referenzlistentechnik nicht das geeignetste Mittel zur Berechnung des Precision-Werts. Es kann passieren, dass das System wahre terminologische Ausdrücke

extrahiert, die in der Referenzliste nicht vorhanden sind: Obwohl es sich um relevante Termini handelt, werden diese Kandidaten theoretisch dann bei der Gegenprüfung mit der Referenzliste als falsch erkannt (vgl. Daille 1996: 53). Auf der anderen Seite ist die Validierungsmethode nicht in der Lage, Recall zu erfassen, da keine Benchmark für den Vergleich existiert. Außerdem ist die manuelle Validierung eine zeitaufwändige Tätigkeit und muss idealerweise von mehr als einer Fachperson durchgeführt werden, um eine möglichst zuverlässige Ressource zu haben. Weiters ist die Validierung vom persönlichen Urteil der Beteiligten abhängig, die sich im Laufe ihrer Validierungsaufgabe durch externe Faktoren und durch die Liste der bereits untersuchten Termkandidaten beeinflussen lassen. Auch wenn die Beteiligten im Vorfeld bereits hinsichtlich der Bewertungskriterien geschult worden sind, besteht die Möglichkeit, dass sie aufgrund der Abstraktheit der Benennungseigenschaften trotzdem uneinheitliche Validierungen liefern, die auf ihrer eigenen Intuition basieren (vgl. Pazienza et al. 2005: 266).

6 Relevante Forschungsarbeiten

In diesem Abschnitt sollen mehrere Forschungspublikationen zur Evaluierung von Terminologieextraktionssystemen vorgestellt werden. Aufgrund des beschränkten Umfangs der vorliegenden Masterarbeit gilt es, sich auf die wesentlichsten Werke des Bereichs zu beschränken. Da das Ziel der praktischen Untersuchung darin besteht, die Extraktionsleistung unterschiedlicher Programme und Verfahren festzustellen und zu vergleichen, ist der Schwerpunkt auf Studien zu legen, bei denen die Extraktionsergebnisse im Rahmen der Glass-Box-Methodologie quantitativ analysiert wurden. Da einige der nachfolgend aufgeführten Untersuchungsmethoden zum Teil als Grundlagen für das Forschungsdesign dieser Masterarbeit dienen sollen, wird darüber hinaus hauptsächlich auf die Evaluierungsansätze und weniger auf die tatsächlichen Evaluierungsergebnisse eingegangen.

Eine quantitative Evaluierung auf Basis der Maße Recall und Precision wurde von Inkpen et al. (2016) vorgenommen, um ein effizientes automatisiertes Terminologieextraktionswerkzeug bei der korpusbasierten Gewinnung domänenspezifischer Termini zu ermitteln. In dieser in Kanada durchgeführten Studie wurde das Werkzeug Term Evaluator entwickelt, das über ein fachspezifisches Wörterbuch, eine Reihe von Musterstrukturen und Speichern von früheren Bewertungen verfügt, um die Benutzer bei der Korrektheitsvalidierung auf die extrahierten Termkandidaten zu unterstützen. Ein statistikbasiertes Werkzeug und zwei hybride Programme mit unterschiedlichen Termidentifizierungsmustern wurden der Bewertung der Extraktionsleistung unterzogen. Die Validierung der Extraktionsausgabe dieser drei Tools wurde von zwei menschlichen AnnotatorInnen (auch als RaterInnen oder BewerterInnen bezeichnet) durchgeführt. Die als korrekt beurteilten Termkandidaten wurden dann auf ihre Relevanz hin analysiert, indem sie den vor der Extraktion als korrekt vorhergesagten Termini gegenübergestellt wurden. Die sich daraus ergebenden Verhältnisse zwischen den zwei Gruppen gaben die Recall- und Precision-Werte für das jeweilige Terminologieextraktionswerkzeug an. Das Ranking der Gesamtleistung zeigt, dass das Extraktionsprogramm TermoStat dank seiner Extraktionsstrategie, die den Fachspezifitätsgrad eines Worts vor der Extraktion gegen ein eingebautes nichttechnisches Referenzkorpus berechnet, die höchste Punktzahl erreichte (vgl. Inkpen et al. 2016: 158).

Die Frage der Evaluationsgenauigkeit wurde in einer Studie von Longinova et al. (2012) im Rahmen des FP7-EU-Projekts Terminology Extraction, Translation Tools and Comparable Corpora (TTC) mit dem Ziel der Entwicklung von Werkzeugen zur automatischen Extraktion zweisprachiger Terminologie aus domänenspezifischen Korpora angegangen. In diesem

Bestreben, den Termextraktionsprozess zu verbessern, wurde eine Referenztermliste (Reference Term List bzw. RTL) erstellt, die als Goldstandard der quantitativen und qualitativen Evaluierung der automatischen Terminologieextraktion dient (Longinova et al. 2012: 4). Zur Erstellung einer Referenztermliste wurden sowohl Ein- als auch Mehrwortbenennungen, die als spezifisch für das Fachgebiet der Windenergie im Spanischen galten, aus relevanten schriftlichen Materialien gesammelt und geordnet. Daraufgehend wurden sämtliche Termini in ihren Lemmata zusammen mit ihren flektierten Formen sowie ihren grafischen, morphologischen und syntaktischen Varianten aufgelistet. Für die Evaluierung wurden Wörter aus verschiedenen Texten des gleichen Fachgebiets mit dem im Rahmen des Projekts entwickelten Programm TermSuite einsprachig extrahiert. Auch hier wurden die Werte für Recall und Precision verwendet, um die Relevanz der extrahierten Termkandidaten in Bezug auf die Referenztermliste zu messen und die Leistung des Terminologieextraktionsprogramms zu bestimmen. Recall und Precision hängen demnach von der relativen Größe der Referenztermliste und der Liste der extrahierten Termkandidaten ab (vgl. Longinova et al. 2012: 10 f.). Umfasst letztere zahlreiche potenzielle Termini, steigt zwar der Recall aufgrund der höheren Wahrscheinlichkeit, dass die Referenztermini erkannt werden, die Precision sinkt jedoch entsprechend, da einige Termkandidaten möglicherweise nicht von der Referenztermliste abgedeckt werden.

Eine manuell erstellte Referenztermliste wurde in Maynards (2000) Studie zum hybriden Terminologieextraktionssystem TRUCKS eingesetzt, die außerdem zu den gründlichsten Werken im Bereich der Evaluierung der automatisierten Extraktionsleistung gehört. Ursprünglich wurde die biomedizinische Ontologiedatenbank im Rahmen des Projekts Unified Medical Language System als Bestand der vordefinierten Soll-Termini zum Vergleich mit den Ist-Extraktionsergebnissen verwendet. Jedoch ist der erste Versuch laut der Autorin in ihrer späteren Arbeit (Maynard et al. 2008: 108 f.) aufgrund der Unvollständigkeit solcher Ressourcen vor allem im Hinblick auf das vorhandene linguistische Wissen gescheitert. So wurde beschlossen, im zweiten Evaluierungsanlauf eine händisch zusammengestellte Referenztermliste für den Vergleich mit den Extraktionsergebnissen von TRUCKS anzuwenden, um Precision- und Recall-Werte als Bewertungsmaße zu berechnen. In diesem Fall arbeiteten zwei Fachpersonen als AnnotatorInnen zusammen, aber es wurde kein Vergleich zwischen ihnen durchgeführt bzw. gab es keine Diskussion bezüglich ihrer Entscheidungen und Begründungen bei einzelnen vergebenen Bewertungen.

Wie Inkpen et al. (2016) zielten die Studien von Vivaldi und Rodríguez (2007) ebenfalls auf die Bewertung von Extraktionsergebnissen ab, die mit einem hybriden Ansatz durch das

Terminologieextraktionsprogramm YATE automatisch erhoben wurden. Zudem wird auf die Bedeutung einer quantitativen Validierung des Grads der Bewertungsübereinstimmung zwischen mehreren menschlichen AnnotatorInnen in Untersuchungen hingewiesen, die sich auf deren qualitative Urteile verlassen (vgl. Vivaldi & Rodríguez 2007: 227). Demzufolge empfiehlt es sich, die Feststellung des Kappa-Koeffizienten, der von Cohen (1960) definiert wurde, in die vergleichenden Evaluierungsstudien mit mehreren BeurteilerInnen zu integrieren. Dabei reicht der Wert von 1, was eine perfekte Übereinstimmung bedeutet, bis -1 , was einer vollständigen Uneinigkeit entspricht. Eine Kappa-Statistik von 0,7 oder höher wird im Allgemeinen als eine gute statistische Korrelation angesehen, aber hierbei gilt: Je höher der Wert, desto besser ist die Korrelation. Diese Anwendung der Kappa-Korrelationswerte gewährleistet die Zuverlässigkeit des Ergebnisses der Bewertungsübereinstimmungen, da sowohl wahre als auch zufällige Übereinstimmungen zwischen den BeurteilerInnen berücksichtigt werden (Vivaldi & Rodríguez 2007: 238).

Zur Bewältigung der bisher unübersichtlichen Unterscheidung zwischen domänenspezifischen und gemeinsprachlichen Ausdrücken, die eine der wesentlichen Beschränkungen der automatisierten Terminologiegewinnung aus Fachkorpora darstellt, wurde von Rigouts Terryn et al. (2020) die Forschungsplattform *TermEval 2020* initiiert. Dabei erfolgte die Erstellung und Veröffentlichung einer Sammlung manuell annotierter Vergleichskorpora, *ACTER* (*Annotated Corpora for Terminology Extraction Research*), welche mit dreisprachiger Abdeckung (Englisch, Niederländisch und Französisch) und vier spezifischen Domänen (Korruption, Dressur, Windenergie und Herzinsuffizienz) als ein möglichst umfangreicher Goldstandard für die Identifizierung domänenspezifischer Termini dienen soll (vgl. Rigouts Terryn et al. 2020: 87 f.). Aktuell steht *ACTER 1.2* in Form von Termlisten zur Verfügung, deren Einträge jeweils mit entsprechenden Annotationen versehen sind, die durch ihren Grad an Domänenspezifität (Relevanz zu entsprechenden Domänen) und an Fachsprachlichkeit definiert werden. Somit unterscheiden sich die Annotationskategorien zwischen *Specific Term* (domänenbezogen und fachsprachlich), *Common Term* (domänenbezogen, jedoch nicht fachsprachlich) und *Out-of-Domain* (fachsprachlich, jedoch nicht domänenbezogen). Im Rahmen dieses Shared-Task-Projekts evaluierten fünf beteiligte ForscherInnengruppen die Leistung ihrer eigenen Terminologieextraktionssysteme, deren Verarbeitungsmethoden sich voneinander unterschieden, unter Einsatz von *ACTER*. Zunächst wurden die annotierten Korpora der drei erstgenannten Domänen als Trainings- bzw. Referenzmaterial benutzt, um das Extraktionssystem der jeweiligen Gruppe für die Verarbeitungsaufgabe zu konfigurieren. Zur Prüfung der Systemleistung wurden die Korpora

zum Thema Herzinsuffizienz anschließend der automatischen Terminologiegewinnung unterzogen, wobei alle drei Sprachen berücksichtigt wurden. Außer der komparativen Extraktionsqualität der einzelnen Systeme durch die quantitative Ermittlung der Precision-, Recall- und F1-Maße wurde zusätzlich auf die Zusammensetzung der Ausgabedaten eingegangen. Dabei wurde beispielsweise analysiert, ob sich die Trefferquote auf die Übereinstimmungen mit den *Specific Terms* zurückführen ließ, oder vielmehr wurden diejenigen Tokens erfasst, die als *Common Terms* bzw. *Out-of-Domain* zu annotieren gewesen wären (vgl. Rigouts Terry et al. 2020: 93). Darüber hinaus lässt sich anhand der resultierenden Precision-Werte, die bei allen untersuchten Systemen unter 50 % lagen, implizieren, dass die optimale Genauigkeit hinsichtlich der Identifizierung relevanter Termkandidaten noch nicht erreicht werden kann.

7 Praktische Evaluierung der Terminologieextraktionssysteme

Der Hauptschwerpunkt der vorliegenden Masterarbeit liegt in dem Bestreben, die Leistung verschiedener automatischer Terminologieextraktionssysteme bei der Verarbeitung eines domänenspezifischen Textkorpus im österreichischen Verwaltungsbereich Einwanderung und Aufenthalt zu bewerten. Dies soll ermöglichen, einen oder mehrere geeignete Ansätze zur Extraktion österreichischer Verwaltungsterminologie unter Berücksichtigung ihrer einzigartigen linguistischen Eigenschaften zu bestimmen. Auf Basis der in den vorhergehenden Kapiteln erarbeiteten theoretischen Grundlagen zu relevanten Themenbereichen lautet die Forschungsfrage, die eingangs in der Einleitung vorgestellt wurde, wie folgt:

Welches automatische Terminologieextraktionswerkzeug schneidet bei der Verarbeitung von fachspezifischer österreichischer Verwaltungsterminologie in Bezug auf Einwanderung und Aufenthalt vergleichsweise am besten ab?

Die Effizienz des jeweiligen automatischen Terminologieextraktionsverfahrens bzw. -systems hängt nicht nur von der Methode ab, die zur Extraktion der Termkandidaten angewandt wird, sondern auch von anderen Faktoren, wie Textsorten, Sprache(n) sowie vom Fachgebiet und von den Spezifika, durch die der Textkorpus gekennzeichnet ist. So kann beispielsweise der statistisch basierte Ansatz hypothetisch davon profitieren, dass bestimmte Benennungen im Korpus wiederholt vorkommen, da eine Diversifizierung des Vokabulars in der Verwaltungsterminologie nicht bevorzugt wird, um die Eindeutigkeit und Konsistenz zu erhalten (vgl. Wiesinger 2008: 119 f.). Auf der anderen Seite könnten Programme, die linguistische Ressourcen zur Identifizierung der Benennungskandidaten nutzen, gerade im deutschsprachigen Korpus ein zuverlässiges Ergebnis liefern, da sie durch die integrierten morphologischen und syntaktischen Regeln in der Lage sind, flektierte Formen eines Terminus zu erkennen und in sein Lemma umzuwandeln. Mit der Überlegung, dass sich die jeweilige Stärke beider Terminologieextraktionsansätze bei der Verarbeitung des ausgewählten Korpus effektiv gegenseitig ergänzen können, scheint deshalb die folgende Aussage vorerst die nachvollziehbarste Hypothese zu sein:

Ein hybrider Ansatz zur Terminologieextraktion, der statistische und linguistische Methoden kombiniert, liefert im Vergleich zu rein statistischen und linguistischen Ansätzen die beste Leistung bei der Verarbeitung domänenspezifischer österreichischer Verwaltungsterminologie bezüglich Einwanderung und Aufenthalt.

Dieses Kapitel stellt den Übergang zwischen Theorie und Praxis dar. Um die oben aufgeführte Annahme zu überprüfen, wird eine quantitative Evaluierung der Terminologieextraktionsergebnisse durchgeführt. Zunächst wird die Methodologie der Untersuchung erläutert und deren Zweckdienlichkeit im Sinne der vorgesehenen Untersuchung begründet. Darunter finden sich auch die ausführlichen Darstellungen der wesentlichen Komponenten wie die Referenztermliste und das Validierungstool Term Evaluator, die eine entscheidende Rolle bei der quantitativen Evaluierung der ausgegebenen Termkandidaten und damit der Extraktionsleistung des jeweiligen Programms spielen. Abschließend wird das Evaluierungsergebnis dargestellt und gegebenenfalls diskutiert.

Dabei ist vorab darauf hinzuweisen, dass der Umfang der vorliegenden Masterarbeit den Richtlinien zur Verfassung einer Masterarbeit an der Universität Wien entspricht. So ist die gesamte Leistung selbstständig erbracht worden, indem keine weiteren MitarbeiterInnen an dem Verfassen teilgenommen haben. Die Hinzuziehung externer Fachpersonen bzw. sachkundiger Personen zur Überprüfung der fachspezifischen Terminologie soll lediglich der Objektivität der Untersuchung und der maximalen Tauglichkeit der Extraktionsausgabe dienen. Ihre Beteiligung an dieser Aufgabe hat deshalb keine Auswirkung auf die Inhalte der Arbeit.

7.1 Untersuchungsmethodologie

Wie erwähnt, wurde beschlossen, den Untersuchungsschwerpunkt ausschließlich auf die quantitative Adäquatheitsanalyse der Extraktionsergebnisse im Rahmen des Glass-Box-Ansatzes zu legen und auf die Vorevaluierung im Sinne von L'Homme et al. (1996) zu verzichten. Die Entscheidung lässt sich hauptsächlich wie folgt begründen: In erster Linie handelt sich nicht um den Vergleich der extrinsischen Funktionalität der Systeme im Allgemeinen, sondern vielmehr gezielt darum, ob und wie gut ihre intrinsischen Extraktionsmethoden mit den fachspezifischen linguistischen Besonderheiten und der damit einhergehenden Komplexität der Ausgangsmaterialien umgehen können. Zweitens erfolgt die anwenderInorientierte Beurteilung des Black-Box-Modells anhand der vergleichsweise subjektiven Kriterien und hängt beinahe vollständig von den Bedürfnissen der einzelnen AnwenderInnengruppe ab, wenn nicht sogar von denen jedes Individuums. Zudem verfolgen die BenutzerInnen bei der automatischen Terminologiegewinnung in der Regel ohnehin ein gemeinsames Ziel: ein Höchstmaß an Genauigkeit bei der Identifizierung von Termini bei möglichst geringen Beschränkungen in Form von Noise- und Silence-Anteilen. Daher erscheint es angemessen zu behaupten, dass die vorliegende Nachbearbeitungsbewertung als eine

nützliche Ausgangsbasis für die Auswahl der Verfahren zur Terminologieextraktion bei Textmaterialien aus den verwandten Fachbereichen gelten kann. Allerdings wird die Untersuchung aus rein terminologischer Sicht vor dem Hintergrund einer translatorischen Anwendung und Tätigkeit durchgeführt. Somit gehören vertiefende Elemente aus der Computerlinguistik bzw. dem IR-Gebiet zur Analyse der Extraktionsleistung, wie statistische Berechnungsmodelle zur Termidentifizierung, dem priorisierten Methodologieumfang bezüglich der Evaluierung nicht an.

Die Methodologie gliedert sich in vier Teile: Zunächst erfolgt die Beschaffung und Analyse des geeigneten Dokumentationsmaterials, dessen Inhalte als Ausgangsdaten für den zweiten Schritt, somit die manuelle Erstellung einer Referenztermliste, dient. Daraufhin findet die Terminologiegewinnung mithilfe der drei unterschiedlichen automatischen Extraktionssysteme statt, wonach zum Schluss die sich ergebenden Termkandidaten einer quantitativen Evaluierung unterzogen werden. Auch wenn es sich beim gesamten Verfahren nicht um den Aufbau eines fachgebietsbezogenen Begriffssystems bzw. Terminologiebestands handelt, wird im Rahmen der vorliegenden Masterarbeit auch die entsprechende Arbeitsmethode zur systematisch terminologischen Erschließung eines Fachgebiets konsequent berücksichtigt, um die Qualität der Untersuchung und damit der Ergebnisse zu sichern.

7.1.1 Beschaffung des Dokumentationsmaterials

Zunächst wurden unterschiedliche Arten von amtlichen Formularen für den Antrag auf die Erteilung des Aufenthalts- und Niederlassungsrechts in Österreich zusammengestellt, die in erster Linie von den offiziellen Websites des Bundesministeriums für Inneres, von für Einwanderung und Niederlassung zuständigen österreichischen Behörden sowie vom Digitalen Amt Österreich stammen. Es handelte sich um insgesamt 44 Formulare, die 27 Kategorien der Aufenthaltsberechtigung betreffen. Dazu gehören ebenfalls erforderliche Zusatzformulare, die bei der Antragstellung vorzulegen sind, z. B. als Wohnorts- bzw. Beschäftigungsnachweise, insbesondere im Fall der Aufenthaltskategorien, die mit der Erwerbstätigkeit verbunden sind. Diese Textsorte wurde im Sinne der vorgesehenen Untersuchung als zweckmäßig erachtet, denn Formulare zählen, wie im Laufe des Kapitel 3 erläutert, durch ihre formellen sowie sprachlichen Besonderheiten zu den typischen Erscheinungsformen der Verwaltungssprache. Vor allem setzt die enge Verbindung dieses Verwaltungsbereichs mit dem österreichspezifischen Recht und den Bestimmungen das Auftreten von

verwaltungssprachlichen Austriaismen voraus, was wiederum die Aussagekraft der Untersuchung fördert.

Dadurch, dass diese Antragsformulare durch offizielle nationale Stellen veröffentlicht wurden, konnte im Interesse der sachlichen und sprachlichen Zuverlässigkeit bei der Auswahl des Dokumentationsmaterials festgestellt werden, dass bei der Gestaltung der Dokumente das Erstsprachenprinzip eingehalten wird und die verfassenden Personen auf dem betreffenden Gebiet Fachleute sind (vgl. Arntz et al 2014: 221). Wird darüber hinaus von dem vermeintlichen Bestreben der Verwaltung nach rechtlicher Transparenz und Nachverfolgbarkeit ausgegangen (siehe dazu Unterkapitel 3.3), kann auch angenommen werden, dass sich um terminologische Präzision bemüht wurde. Dies soll eine hohe Tauglichkeit des Ausgangsmaterials gewährleisten. Zuletzt wurde überprüft, dass bei den ausgewählten Quellen der Gesichtspunkt der Aktualität gewährleistet wurde, indem die enthaltenen Informationen den aktuellen Stand des Fachgebiets wiedergeben. Alle von den Websites heruntergeladenen Dateien, die sich im PDF-Format befinden, wurden ins Microsoft-Word-Format konvertiert, in einem Einzelkorpus aus 70 295 Wörtern bzw. Tokens verknüpft und schließlich als eine Text-Datei abgespeichert. Dieser Korpus erfüllt zwei Zwecke: Er hat als eine der Quellen für die Erstellung einer Referenztermliste gedient und wurde bei der automatisierten Terminologieextraktion durch die gegenständlichen Extraktionswerkzeuge eingesetzt.

7.1.2 Erstellung der Referenztermliste

Zur Evaluierung der Terminologieextraktionswerkzeuge wurde eine einsprachige Referenztermliste (nachfolgend als RTL bezeichnet) verwendet, die als ein Maßstab für die terminologisch relevanten Extraktionsergebnisse fungiert. Dieser Ansatz wird in der Regel üblicherweise wie im Kapitel 6 geschildert und laut Pazienza et al. (2005: 265) bei mehreren anderen Studien zur Evaluierung der natürlichen Sprachverarbeitung verfolgt: Ein manuell konstruierter Goldstandard wird aus einer Menge von Texten abgeleitet, die dann auch von den zu evaluierenden Werkzeugen verarbeitet werden. So gilt es als Grundvoraussetzung, dass alle Termini auf der RTL auch tatsächlich von dem erstellten Korpus aus Formularen abgedeckt sind (vgl. Longinova et al. 2012: 5). Um dies mit möglichst hoher Genauigkeit sicherzustellen und da es sich nicht um einen allzu umfangreichen oder komplexen Textkorpus handelte, wurden die Referenztermini durch ein manuelles Extraktionsverfahren gewonnen. Ein automatisiertes Extraktionsverfahren wäre bei einer entsprechenden Korpusgröße nicht zielführend gewesen und könnte sich zudem als fehleranfällig erweisen, sodass die Ausgabe

zusätzlicher zeitaufwändiger Nachbearbeitungen und Überprüfungen bedurft hätte, insbesondere hinsichtlich der Silence-Effekte (vgl. Arntz et al. 2014: 248).

Unter menschlicher Erstellung der RTL ist allerdings nicht zu verstehen, dass die Benennungen rein händisch aus dem Korpus erhoben und zusammengestellt wurden, vielmehr wurde das Verfahren teilweise durch den Einsatz eines vorhandenen Computerprogramms im Hinblick auf eine Zeitersparnis unterstützt, das sogenannte einfache Verfahren mit „Bordmitteln“ (Arntz et al. 2014: 246). Hierzu erfolgte die Vorverarbeitung des Korpus im Textverarbeitungsprogramm Microsoft Word, wobei der Textinhalt für die manuelle Terminologiegewinnung aufbereitet wurde. Grundsätzlich umfasste der Ansatz die folgenden Schritte: Zunächst wurden alle Interpunktionen, darunter Punkt, Beistrich, Strichpunkt, Anführungszeichen und Klammer und Zahlen entfernt, jedoch galt es Bindestriche in diesem Schritt beizubehalten, denn einige davon dienten bei Zusammensetzungen als Verbindungszeichen zwischen den lexikalischen Bestandteilen, z. B. EWR-Bürger und Rot-Weiß-Rot-Karte. Zudem wurde beim Gedankenstrich – vorerst ohne Rücksicht auf seine tatsächliche Stelle im Text – kein Leerzeichen zwischen ihm und den vor und nach ihm stehenden Wörtern gesetzt. Somit konnte verhindert werden, dass die Bestandteile einiger Bezeichnungen, insbesondere die der bestimmten Aufenthaltsberechtigung, die zum Teil durch mindestens einen Gedankenstrich verbunden sind, nicht zerlegt wurden, als der Textinhalt in (lexikalische) Einheiten mithilfe des Tabellenkalkulationsprogramms Microsoft Excel, je nach Leerzeichensetzung, segmentiert wurde. Danach wurden die Einträge zusammengefügt und in Form einer Pivot-Tabelle zusammengefasst, indem einzelne lexikalische Segmente jeweils einmalig in einer Spalte angezeigt und diese daraufhin alphabetisch sortiert wurden. Damit konnte gleichzeitig die Übersichtlichkeit gewährleistet und die Liste von zahlreichen Duplikaten befreit werden (Abb. 8). Zu diesem Zeitpunkt unterlag die Liste der manuellen Termidentifizierung, die unter Berücksichtigung der linguistischen Besonderheiten der österreichischen Verwaltungssprache im Abschnitt 2.2.2 erfolgte. Darüber hinaus wurde eine sachkundige Fachperson (siehe anonymisiertes Profil 1 im Anhang I) im Hinblick auf die Bestimmung von Termhood bzw. Unithood der Termkandidaten zurate gezogen.

Lexikalische Einheiten	Auftretenshäufigkeit
ab	1
abdeckenden	1
aber	1
abgebe	1
abgenommen	1
abgeschlossen	1
abgeschlossene	1
abgeschlossenen	2
abholen!	26
Abholung	1
Abkürzung	1
Ablauf	3
Ablaufdatum	1
abläuft	1
Ableistung	1
Abordnungsschreiben	1
Abs	21
Absatz	5
Abschluss	10
Abschlussklärung	1
absichtlich	26
Absolvierung	6
absteigender	1
adoptieren	1
adoptierten	1
Adoptionen	1
Adresse	53
ähnliches	1
akademische	52
akademischer	26
akkreditierten	1

Abbildung 8: Zusammenfassung von Daten anhand einer Pivot-Tabelle

Aufgrund des dominierenden Nominalstils der österreichischen Verwaltungssprache wurde beschlossen, die Erhebung der Substantive in Form von Simplizia und Komposita zu priorisieren. Hinsichtlich der Mehrwortbenennungen mussten zuerst die wesentlichsten terminologischen Kollokationen festgelegt werden, die dann als Kriterien der Termkandidaten gelten sollten. Ebenfalls aufgrund des Nominalstils der österreichischen Verwaltungssprache wurde beschlossen, dass sich das menschliche Verfahren hauptsächlich auf die Gewinnung der Fachausdrücke in Form von Nominalphrasen beschränkte. Auf Fachwendungen wurde im Rahmen der Untersuchung verzichtet, denn diese unterliegen in der Regel einer hohen linguistischen Varianz und weisen dementsprechend im Vergleich zu Nominalphrasen einen niedrigeren Unithood-Grad auf. Im Zusammenhang mit der Identifizierung von Mehrwortbenennungen wurde bereits im Vorhinein damit gerechnet, dass die Bestandteile der potenziellen Mehrwortbenennungen durch den Segmentierungsvorgang auseinandergerissen werden, was einen der Nachteile dieses Ansatzes darstellt. Deshalb wurden nur Einheiten von der Liste entfernt, vor allem jene, die ausschließlich zur Gemeinsprache gehörten und nicht unter die lexikalische Kategorie Gemeinsprachliche Ausdrücke mit fachsprachlicher Kontexteinbettung (siehe hierzu S. 26) fielen. Die verbleibenden alleinstehenden Adjektive, Adverbien und Genitive dienten als Ausgangspunkt für die sorgfältige Überprüfung gegenüber dem Korpus mithilfe der Suchfunktion. Waren diese Einheiten ein Bestandteil einer nominalen Mehrwortbenennung, so wurden sie in die Liste zusammen mit den restlichen zusammengehörigen Bestandteilen aufgenommen, z. B. polizeiliches Führungszeugnis,

Dokumentation des Aufenthaltsrechts oder mobiler unternehmensintern transferierter Arbeitnehmer.

Um die Anforderung bei der Erstellung der RTL zu erfüllen, dass die Einträge die Erscheinungsformen der entsprechenden verwaltungssprachlichen Terminologie widerspiegeln sollten, ist nicht einzig auf die Darstellung von Ein- und Mehrwortbenennungen zu achten, sondern die im Korpus auftretenden Flexionen der jeweiligen Benennung sowie ihre Varianten mussten auch auf der RTL vorliegen. Eine möglichst umfassende RTL sollte dazu beitragen, dass die Evaluierung ein aufschlussreiches Ergebnis liefert (vgl. Longinova et al. 2012: 5). Allerdings war hinsichtlich der Varianz an Benennungen von einer geringen Erscheinung auszugehen, da sich die verwaltungssprachliche Terminologie, vor allem im Zusammenhang mit der Textsorte Vordrucke und Formulare, erheblich durch eine strenge Standardisierung auszeichnet. Jedoch wurden im Zuge der manuellen Terminologiegewinnung potenzielle Varianten der Benennungen trotzdem berücksichtigt und in die RTL übernommen. Bei der Ermittlung der Varianten wurde auf die von Daille (2005: 183 ff.) formulierte Typologie wie folgt zurückgegriffen, wobei die herrschende Ausgangsvariante als Basisterminus (base term) bezeichnet wird:

- (Leichte) **syntaktische Varianten:** Diese liegen vor, wenn die intrinsische Struktur des entsprechenden Terminus modifiziert wird. Die Variation kann beispielsweise durch optionale Hinzufügung von einem oder mehreren Funktionswörtern (Präpositionen, Artikel, Konjunktionen) zwischen den Bestandteilen des Basisterminus erfolgen, wie Daueraufenthaltsrecht und Recht auf Daueraufenthalt;
- **morphosyntaktische Varianten:** Dabei werden die intrinsische Struktur, d. h. die Syntax der Benennungen modifiziert, sodass ihre lexikalischen Bestandteile von einer morphologischen Modifikation betroffen sind (z. B. inländische Antragstellung und Inlandsantragsstellung);
- **morphologische Varianten:** Hierunter sind Wörter zu verstehen, die sich bei derselben Bedeutung bezüglich der Schreibung durch Auslassung oder Hinzufügung des Morphems unterscheiden (z. B. Zustellbevollmächtigte/r und Zustellungsbevollmächtigte/r). Im Gegensatz zu den morphosyntaktischen Varianten geschieht die Variation nicht durch Änderungen der intrinsischen Struktur des Basisterminus.

Zudem wurden Akronyme als **anaphorische Varianten** der entsprechenden Benennungen behandelt (z. B. AusLBG für Ausländerbeschäftigungsgesetz). Paradigmatische Varianten, bei denen es sich um die Ersetzung eines lexikalischen Bestandteils der Benennung durch sein Synonym handelt, wären bei der Evaluierung einer bilingualen Terminologieextraktion mit Parallelkorpora sinnvoller gewesen. In diesem Fall ist in Betracht zu ziehen, dass ein Terminus in der Ausgangssprache durch mehr als eine Äquivalenz in der Zielsprache übersetzt werden kann (vgl. Daille 2005: 185). Aus diesem Grund ist eine Berücksichtigung dieser Art der Varianten im Rahmen der einsprachigen Untersuchung nicht angezeigt.

Nach Eingabe aller wesentlichen Daten wurde die RTL einer finalen Verifizierung durch die oben genannte sachkundige Fachperson unterzogen. Darüber hinaus bestand die Möglichkeit, die terminologische Fachgebietsrelevanz einiger Benennungskandidaten mithilfe der öffentlich verfügbaren Begriffslexika zu den verwandten Themen verdoppelt zu verifizieren. Hierbei kamen das „Integrationsglossar“ (Der Österreichische Integrationsfonds/ÖIF, 2014) und das „Glossar zu Asyl und Migration“ (Das Europäische Migrationsnetzwerk, 2018) sowie die offiziellen Websites der österreichischen behördlichen Stellen, wie das „Magistrat MA 35 Fachbereich Einwanderung der Stadt Wien“ und Österreichs „Digitales Amt“ zum Einsatz. Die Gegenprüfung erfolgte ebenfalls zwar händisch, wurde aber auch durch die Excel-Funktion S-Verweis unterstützt. Waren die Termkandidaten auch in der Kontrollliste erfasst, so wurden sie als fachgebietsbezogen bestätigt (Abb. 9). Die vollständige RTL, die sich im Anhang II der vorliegenden Masterarbeit findet, enthält 242 Referenz(-Basis-)Termini. Unter Einbezug ihrer im Korpus auftretenden Varianten und flektierten Formen führte dies zu insgesamt 321 Einträgen, wobei die Wortanzahl einer Benennung zwischen zwei und fünf lexikalischen Bestandteilen lag.

Termkandidaten aus dem Korpus	In der Kontrollliste enthalten?	Kontrollliste – offizielle Glossare
Meldeverpflichtung		Nostrifikation
Minderjähriger	JA	Nostrifizierung
mobile ICT		Notariatsordnung
MOBILER UNTERNEHMENSINTERNE		
TRANSFERIERTER ARBEITNEHMER		Obzorge
Natürliche Person	JA	Qualifikationsnachweis
NEUAUSSTELLUNG		Quotenfrei
Niederlassung	JA	Quotengröße
Niederlassungs- und Aufenthaltsgesetz		Rass-Clash-Abschmen
Niederlassungsbehörde		Rassismus
Niederlassungsbewilligung	JA	Reintegration
notarielle oder gerichtliche		
Regulierung		Resettlement
ortsübliche Unterkunft		Rot-Weiß-Rot Karte
Partnerschaftsurkunde		Rot-Weiß-Rot Karte
politisches Führungszeugnis		Rot-Weiß-Rot Karte Plus
Quote	JA	Rückkehrmigration
quotenfrei	JA	Saisonarbeiter/innen
Quotenpflicht	JA	Sanktionsmaßnahmen
Reintegration	JA	Schlüsselkraft
Reisedokument	JA	Schlüsselkraft
Rot-Weiß-Rot Karte	JA	Segregation
Rot-Weiß-Rot Karte plus		Sicherungsbescheinigung
Rückkehrentscheidung		Soziales Interesse
Schlüsselkraft	JA	Staatsbürgerschaft
Schubhaft		Staatssekretariat für Integration
selbständige Mobilität		Statutshilfe
selbständige Schlüsselkraft		Stichtagsantrag
Selbständige Schlüsselkraft		Subsidiär Schutzberechtigte
selbständige Schlüsselkräfte		Touristenrisikovermerk
SELBSTÄNDIGER	JA	Unbedenklichkeitsbescheinigung des Finanzamtes
Sicherung des Lebensunterhaltes		Unterhaltspflichtige Person
Sicherungsbescheinigung	JA	Validierung

Abbildung 9: Verifizierung der manuell extrahierten Termkandidaten

7.1.3 Praktische maschinengestützte Terminologieextraktion

Nachdem die Liste der manuell extrahierten Referenztermini fertiggestellt und möglichst perfektioniert worden war, bestand die nächste Phase darin, den bereitgestellten Korpus einer automatischen Terminologiegewinnung zu unterziehen. Für den vorbestimmten Zweck dieser korpuslinguistischen Untersuchung wurden insgesamt drei Terminologieextraktionswerkzeuge für die komparative Leistungsevaluierung ausgewählt, nämlich AntConc, Tilde Wrapper System for CollTerm (TWSC) und TermSuite, deren Einzelheiten und Verarbeitungsprozesse im Rahmen der praktischen Terminologieextraktion im Anschluss besprochen werden.

Die Auswahl ließ sich in erster Linie dadurch begründen, dass das jeweilige Werkzeug deutlich auf unterschiedliche Methoden zur Terminologieextraktion zurückgreift, damit der Vergleich möglichst aufschlussreiche Informationen liefern konnte. Darüber hinaus wurden nur diejenigen Tools in Betracht gezogen, die im Internet zum Herunterladen zur Verfügung standen bzw. die sich als cloudbasierte Plattform anboten. Eine dauerhafte Installation von maschinengestützter Termextraktionssoftware wäre kostspielig gewesen und würde sich tendenziell zur regelmäßigen Anwendung eignen. Zuletzt lag der entscheidende Faktor bei der Auswahl des hybriden bzw. linguistisch basierten Tools in der Verfügbarkeit deutschsprachiger linguistischer Ressourcen. Dies war keine einfache Aufgabe, da die meisten online zugänglichen Terminologieextraktionssysteme über eine beschränkte Sprachauswahl verfügen.

Schließlich gilt anzumerken, dass der zu verarbeitende Korpus vor dem Einlesen ins Terminologieextraktionssystem keiner Bearbeitung zur Förderung der Genauigkeit bzw. der Leistung unterzogen wurde. In der Praxis wäre dies vor allem bei umfangreichen Korpora kein zeit- und kosteneffizienter Ansatz. Außerdem durchliefen die elektronischen Texte keiner Rechtschreibprüfung in einem Textverarbeitungsprogramm, um die realen Bedingungen so nah wie möglich zu simulieren. Keine der für diese Studie verwendeten Texte waren eingescannte Kopien von Papieroriginalen, sodass eventuelle typografische Fehler auf menschliches Versagen zurückzuführen sind.

7.1.3.1 AntConc

AntConc Version 3.5.9 (Anthony, 2020) begann zwar als einfaches Konkordanz-Programm, hat sich aber seit der ersten Konzipierung im Jahr 2004 zu einem der weitverbreitetsten Freeware-Tool zur Textanalyse entwickelt (vgl. Anthony 2013: 22). Dieses Werkzeug wird vor allem in linguistischen bzw. korpuslinguistischen Untersuchungen eingesetzt, um Wörter bzw. Wortgruppen aus einem maschinenlesbaren Textkorpus oder aus

mehreren Textdateien herauszufiltern. Grundsätzlich wird im Zuge der Verarbeitung durch AntConc das eingelesene Material statistisch hinsichtlich der Auftretenshäufigkeit einzelner lexikalischer Einheiten ausgewertet, woraufhin die Ergebnisse je nach Voreinstellungen in Form eines Konkordanz-Diagramms oder einer Rangliste dargestellt werden. Dieser Ansatz mit AntConc soll als Evaluierungsbasis dienen, welche die Termidentifizierung durch reine Frequenzermittlung im Rahmen dieser Arbeit widerspiegelt.

Die Konkordanz-Funktionalität wurde dementsprechend hierbei nicht berücksichtigt, denn ein bestimmter Suchbegriff muss zunächst eingegeben werden, damit das System eine geordnete Liste des gesuchten Worts mit seinen Textbelegstellen, eine sogenannte Kontextliste, liefern kann. So wurden die ebenfalls vom Programm angebotenen Module Wortliste und N-Gramme als angemessenere Methode erachtet, die sämtliche im Korpus vorkommende Wortformen sortiert nach ihrer jeweiligen Auftretenshäufigkeit auflistet. Auch lässt sich darüber hinaus einstellen, dass die Verarbeitung des Korpus ohne Beachtung der Groß- und Kleinschreibung erfolgt, wodurch die Ergebnisse automatisch in kleingeschriebener Form generiert werden (vgl. Anthony 2020). Jedoch war diese Funktion bei der Untersuchung nicht in Anspruch zu nehmen, denn die grafische Variationen, die einen der Aspekte im Zuge des Vergleichs mit der RTL darstellten, sollten deshalb ebenfalls bei den Termkandidaten repräsentiert werden.

Nachdem der Korpus ins Werkzeug importiert wurde, wurde er einer zweifachen automatischen Korpusanalyse unterzogen. Im ersten Schritt war die Funktion Wortliste zunächst einzusetzen, um eine Liste potenzieller Einwortbenennungen herauszufiltern. Da dieses Extraktionsmodul sich ohnehin ausschließlich auf Einworttermkandidaten konzentriert, konnte zusätzlich eine Stoppwortliste³ unter Tool Preferences (Werkzeug-Einstellungen) als Textdatei ins Werkzeug importiert werden, um die Ergebnisliste möglichst zu präzisieren (Abb. 10).

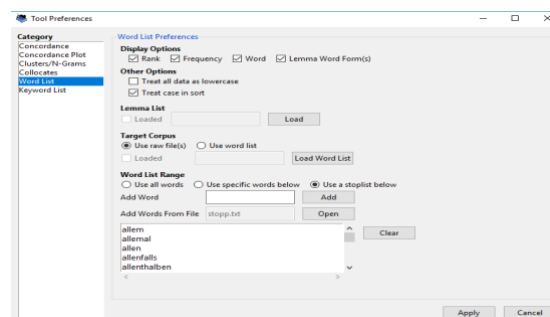


Abbildung 10: Stoppwortliste in Werkzeug-Einstellungen – AntConc 3.5.9

³ Die Stoppwortliste für die deutsche Sprache findet sich unter https://raw.githubusercontent.com/solariz/german_stopwords/master/german_stopwords_full.txt. (Stand: 25.01.2021)

Die Mehrwortbenennungen wurden wiederum mithilfe der Funktion N-Gramme behandelt, wobei ein N-Gramm aus N aufeinanderfolgenden lexikalischen Einheiten besteht. Da sich bei der RTL-Erstellung bereits feststellen ließ, dass die Wortlänge bei den Mehrwortbenennungen im Korpus zwischen zwei und fünf lexikalischen Einheiten bzw. Tokens lag, war dies bei der Einstellung der minimalen und maximalen Größe (Wortanzahl) eines jeweiligen N-Gramms demgemäß festzulegen. Zudem konnte im Zuge der Analyse mit dem Modul N-Gramm eine Mindestfrequenz bestimmt werden, wobei der Schwellenwert auf 5x gesetzt wurde, um Silence-Effekte insbesondere bei längeren Termkandidaten (ab fünf Bestandteilen) zu vermeiden. Dennoch existiert beim Ansatz N-Gramme keine Möglichkeit zur Hinterlegung einer Stoppwortliste, so war mit einem hohen Noise-Anteil zu rechnen.

7.1.3.2 Tilde Wrapper System for CollTerm (TWSC)

Das gegenständliche Terminologieextraktionswerkzeug entstand als Teil der von der EU kofinanzierten cloudbasierten Plattform Terminology as a Service (TaaS) für die Gewinnung, Verwaltung, Pflege und den Austausch von Terminologie (vgl. Terminology as a Service o. J.). Das TaaS-Projekt entwickelte sich aufgrund des Konzepts Software as a Service (SaaS), bei dem das online zur Verfügung gestellte Terminologieverwaltungswerkzeug nur bei Bedarf in Anspruch genommen wird. Damit wird Pionierarbeit für eine neue Art der Terminologearbeit geleistet, die einen automatischen Ansatz zur Terminologiegewinnung mit linguistischem Wissen sowie Nachschlagewerken anhand verschiedener Terminologieressourcen und Webdaten umfasst (vgl. Gornostay & Vasiljevs 2014: 1943 f.).

Für die automatisierte Gewinnung von Fachtermini aus den von BenutzerInnen hochgeladenen Dokumenten kann in der Regel eine Mischform von Methoden zur Termidentifizierung eingesetzt werden. Hierbei stehen auf der TaaS-Plattform zwei Extraktionstools zur Verfügung: Das Tilde Wrapper System für CollTerm (TWSC) führt hauptsächlich eine linguistische Analyse durch, während der Kilgray Terminology Extractor (KTE) mit einem probabilistischen Modell zur simplistischen Frequenzermittlung verwendet wird (vgl. Drewer & Schmitz 2017: 50). Bei Bedarf lässt sich bestimmen, dass die Tools einzeln oder in Kombination zum Einsatz kommen. Da für den statistischen Ansatz das Werkzeug AntConc bereits untersucht wurde, sollte das linguistisch basierte Verfahren des TWSC zwecks Leistungsvergleichs im Mittelpunkt dieses Abschnitts stehen.

Das TWSC besteht grundsätzlich aus zwei Subsystemen: CollTerm-System und Wrapper-System. Ersteres wurde ursprünglich als ein Einzelwerkzeug zur Extraktion von sowohl Einwortbenennungen als auch Kollokationen der Länge von zwei bis vier Wörtern

konzipiert. Seine Methode zur Termidentifizierung basiert in erster Linie auf der linguistischen Anreicherung mithilfe von POS-Tagging in der Vorverarbeitungsphase und den morphosyntaktischen Analysen, die gemeinsam zur Erkennung einzelsprachspezifischer Musterstrukturen beitragen (vgl. Gornostay & Vasiljevs 2014: 1944). Obwohl das Werkzeug primär auf die linguistischen Ressourcen zurückgreift, durchlaufen die Termkandidaten im Anschluss noch eine statistische Überprüfung. Diese zweite Phase besteht zum einen aus einem Minimalfrequenz-Filter, bei dem alle linguistisch akzeptierten Termkandidaten, die weniger als die eingestellte Mindestfrequenz vorkommen, von der weiteren Verarbeitung ausgeschlossen werden. Zum anderen wird anhand der Assoziationsmaße (z. B. Kookkurrenz, gegenseitige Informationen) festgestellt, wie ausgeprägt und wie häufig zwei oder mehr Tokens gemeinsam auftreten (vgl. Pinnis et al. 2012b: 92). Der Vorgangsablauf des CollTerm-Systems wird in Abb. 14 zusammengefasst.



Abbildung 11: Extraktionsablauf von TWSC (vgl. Pinnis et al. 2012: 195)

Das Wrapper-System wurde hinzugefügt, um die Extraktion von Termini unterschiedlicher Länge mit CollTerm zu erweitern. Da diese sich auf die Identifizierung von Mehrwortbenennungen fester Länge (von einer bis vier Tokens) beschränkt, soll mithilfe des Wrapper-Systems diese Grenze bewältigt werden, indem dieses das CollTerm mehrfach ausführt und die Ergebnisse in einer Ausgabedatei für jedes Eingabedokument kombiniert (vgl. ACCURAT 2012: 112).

Nachdem das Projekt mit den wesentlichen Angaben, wie der Ausgangssprache, dem Fachgebiet und der Deskription, angelegt wurde, war der zu analysierende Korpus im Textformat hochzuladen. Bei der Wahl des einzusetzenden Werkzeugs wurde aus dem bereits aufgeführten Grund auf den KTE verzichtet und lediglich das TWSC angekreuzt. Auch wenn die Terminologieextraktion durch das TaaS-System grundsätzlich sämtliche EU-Amtssprachen unterstützt, ist die zusätzliche Funktion, der Term Normalizer, der die identifizierten Termkandidaten auf ihre Grundformen reduziert, derzeit jedoch nur für Englisch, Lettisch und Litauisch einsetzbar (vgl. Pinnis et al. 2012: 197). Deshalb war dieser bei den Voreinstellungen ausgelassen (Abb. 12).

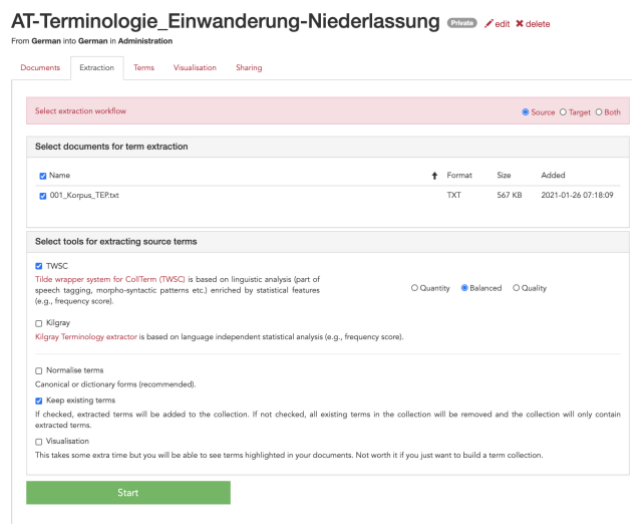


Abbildung 12: Voreinstellung der Terminologieextraktion – TWSC (TaaS)

7.1.3.3 Termsuite

Termsuite wurde im Rahmen des EU-Projekts Terminology Extraction, Translation Tools and Comparable Corpora (TTC) entwickelt, um die (semi-)automatisierte Gewinnung von Terminologien verschiedener Sprachen aus verschiedenen fachlichen Vergleichskorpora zu verwirklichen und diese in der maschinengestützten sowie maschinellen Übersetzung einzusetzen (vgl. Blancafort et al. 2010: 264). Derzeit werden sieben Sprachen unterstützt, darunter auch Deutsch, wobei einzelne Sprachen zunächst einsprachig verarbeitet werden, gefolgt von der Alignierung der resultierenden Terminologien. Zur Extraktion von Einwort- sowie Mehrwortbenennungen verwendet das Werkzeug Termsuite einen hybriden Ansatz, der Verarbeitungskomponenten umfasst: Der Spotter führt die Vorverarbeitung durch, indem der Korpus sich Segmentierung, POS-Tagging und Lemmatisierung unterzieht. Anschließend erfolgt die Terminologieextraktion mit Indexer, der auf die morphosyntaktische Musterauflage⁴ zum Abgleich mit den POS-markierten Tokens zurückgreift und nach übereinstimmenden Mustern sucht (vgl. Daille 2012: 30).

Nachdem die Einwort- und Mehrworttermkandidaten identifiziert werden, findet eine statistische Filterung statt. Dabei handelt es sich um die Berechnung der Fachspezifität mithilfe der Feststellung der relativen Frequenzen einzelner Termkandidaten. Hierzu wird im Grunde

⁴ Die integrierten Musterstrukturen des Werkzeugs stammen aus der von der EU geförderten Plattform MULTTEXT (Multilingual Text Tools and Corpora), die linguistische Datenbestände und Ressourcen verschiedener Sprachen für die sprachtechnische Forschung und Entwicklung zur Verfügung stellt (vgl. Text Encoding Initiative o. J.).

seine Auftretenshäufigkeit im untersuchten Fachkorpus mit demjenigen in einem hinterlegten gemeinsprachlichen Korpus verglichen. Nach diesem Vorgang schließt die Terminologieextraktion ab, wobei die Termkandidaten in absteigender Sortierung nach den Werten der relativen Frequenzen – von fachspezifisch bis gemeinsprachlich – geliefert werden (vgl. Gojun et al. 2012: 651 f.).

Im Zuge der praktischen Untersuchung musste vor der tatsächlichen Termextraktion durch TermSuite zunächst das zusätzliche Tool TreeTagger⁵ installiert und manuell im Termextraktionswerkzeug konfiguriert werden (Abb. 13). Dieses Tool fungiert als Spotter, der den zu untersuchenden Korpus mit den entsprechenden linguistischen Informationen annotiert, die wiederum die Erfassung der Ein- und Mehrworttermkandidaten ermöglichen. Die morphosyntaktischen Parameter verschiedener Sprachen, darunter auch die des Deutschen, sind auf der offiziellen Website des Entwicklers (<https://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/>) zum Herunterladen verfügbar.

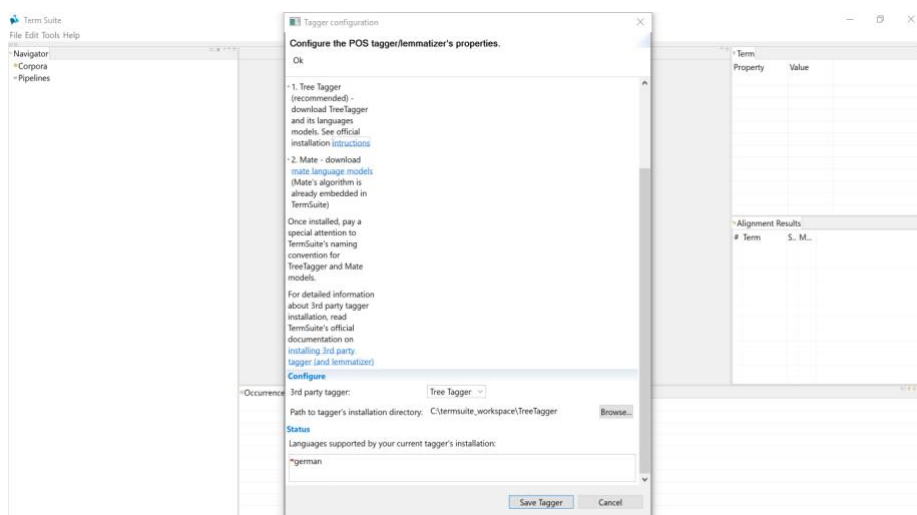


Abbildung 13: Konfiguration der linguistischen Anreicherung (TermSuite)

Daraufhin erfolgte das Importieren des Korpus, gefolgt von der Erstellung und Voreinstellung der sogenannten Pipeline, die den Verarbeitungsablauf der automatisierten Terminologieextraktion bestimmt: Wie in Abb. 14 gezeigt, sollte zunächst die

⁵ TreeTagger wurde von Helmut Schmid im Rahmen des TC-Projekts am Institut für Computerlinguistik der Universität Stuttgart entwickelt. Das Tool wurde bisher erfolgreich zur Annotation von deutschen, englischen, französischen, italienischen, dänischen, schwedischen, norwegischen, niederländischen, spanischen, bulgarischen, russischen, portugiesischen, galizischen, griechischen, chinesischen, suaheli, slowakischen, slowenischen, lateinischen, estnischen, polnischen, rumänischen, tschechischen, koptischen und altfranzösischen Texten eingesetzt (vgl. Schmid 1995).

morphosyntaktische Analyse zur Identifizierung von Termkandidaten durchgeführt werden, aus denen nur die relevanten Ergebnisse mit einem Mindestwert der relativen Frequenz von 1.0 herausgefiltert wurden.

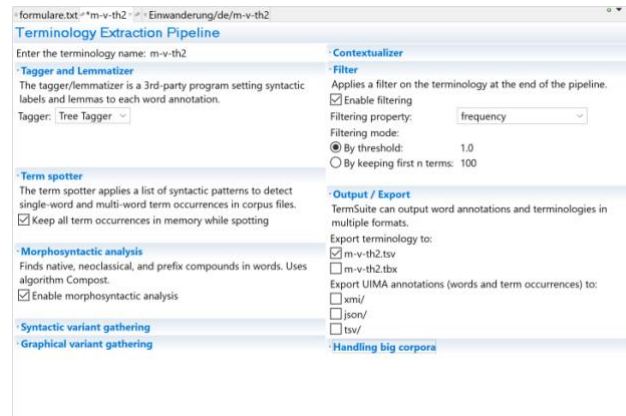


Abbildung 14: Pipeline-Einstellungen (TermSuite)

7.1.4 Quantitative Evaluierung der Extraktionsergebnisse

Nachdem die Terminologiegewinnung mit drei unterschiedlichen maschinengestützten Extraktionsverfahren abgeschlossen wurde, wurden die Extraktionsergebnisse in der nächsten Phase der praktischen Untersuchung einer quantitativen Bewertung unterzogen. Dieser Evaluierungsprozess erfolgte mithilfe des Tools Term Evaluator, dessen Eigenschaften im folgenden Abschnitt vorgestellt werden. Zudem wird die Herangehensweise anhand dieses Annotationsprogramms erläutert. Schließlich widmet sich der letzte Abschnitt der Darlegung sowie der Diskussion der Evaluierungsergebnisse.

7.1.4.1 Term Evaluator

Term Evaluator entstand im Rahmen der Studie von Inkpen et al. (2016) zur Qualitätsevaluierung der maschinengestützt extrahierten Termkandidaten und wurde nach der Entwicklung öffentlich zugänglich zur Verfügung gestellt⁶. Die Aufgabe dieses Programms besteht grundsätzlich darin, durch eine benutzerfreundliche Bedienoberfläche das Verfahren zur Validierung der Termkandidaten durch menschliche Fachpersonen zu erleichtern und zu beschleunigen, die wiederum zur vergleichenden Evaluierung der Verarbeitungsleistung von automatischen Terminologieextraktionswerkzeugen beiträgt. Dieser semiautomatische Ansatz zur Annotation wird als „Post-hoc-Evaluierung“ bezeichnet (Inkpen et al. 2016: 152).

⁶ <https://sourceforge.net/projects/termevaluator/> (Stand: 28.01.2021)

Konkret ermöglicht Term Evaluator den BenutzerInnen, einen vollständig neuen Annotationsprozess einzuleiten, durch Speicheroption eine vorherige Analyse fortzusetzen und zwei oder mehr Annotationsergebnisse zu vergleichen. Das Tool verfügt bereits über ein integriertes technisches Wörterbuch, das aus drei zusammengeführten Online-Mathematik-Wörterbüchern besteht. Zusätzlich kann gegebenenfalls eine sekundäre Liste von Benennungen, deren Relevanz und Korrektheit bereits im Zuge der früheren Annotationsverfahren bestätigt wurden, auch ins Tool importiert werden. Beschließen die BenutzerInnen jedoch, auf die Heranziehung des eingebauten Lexikons zu verzichten, so lassen sich an seiner Stelle ebenfalls andere lexikografische Ressourcen aus einem einschlägigen Fachbereich hochladen. Bei Bedarf kann so im Vorfeld eingestellt werden, dass das Tool zwei folgende Filterungsoperationen bei jedem Termkandidaten automatisch durchführt: i) Ausschließung derjenigen Termkandidaten, die als Stoppwörter gelten; und ii) Auslassung des Stoppwortanteils bei Termkandidaten, die mit einem Stoppwort beginnen oder enden. Dieses Zusatzverfahren greift auf die intrinsische Stoppwortliste zurück. Abb. 15 zeigt die Auswertungsoberfläche, in der AnnotatorInnen die extrahierten Termkandidaten bewerten können. Sie haben Einsicht in den Rang und gegebenenfalls die Auftretenshäufigkeit bzw. den durch das Extraktionswerkzeug festgestellten Termhood-Wert der jeweiligen Eingabe und können von vier Bewertungsoptionen bei der Annotation jeden Termkandidaten auswählen:

- Yes – terminologische Relevanz und Fachspezifizität liegen vor;
- Non-technical – der Termkandidat ist gemeinsprachlich bzw. weist keine Relevanz zum Fachgebiet auf;
- No – der Termkandidat wurde falsch extrahiert, z. B. durch unvollständige Wörter oder Wortsequenzen, die keine Bedeutung erzeugen;
- Not sure – die annotierende Person hat keine Gewissheit über die Fachsprachlichkeit des Termkandidaten.

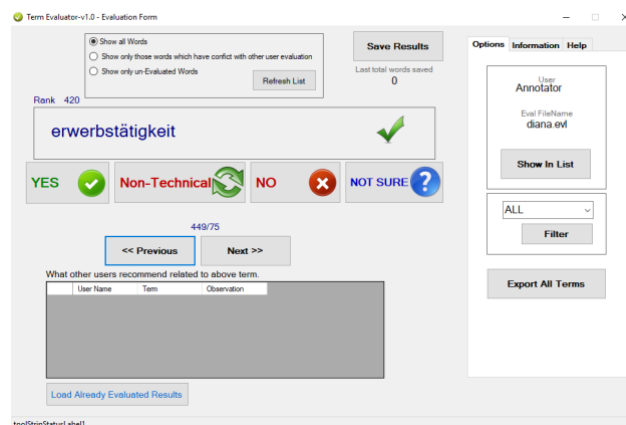
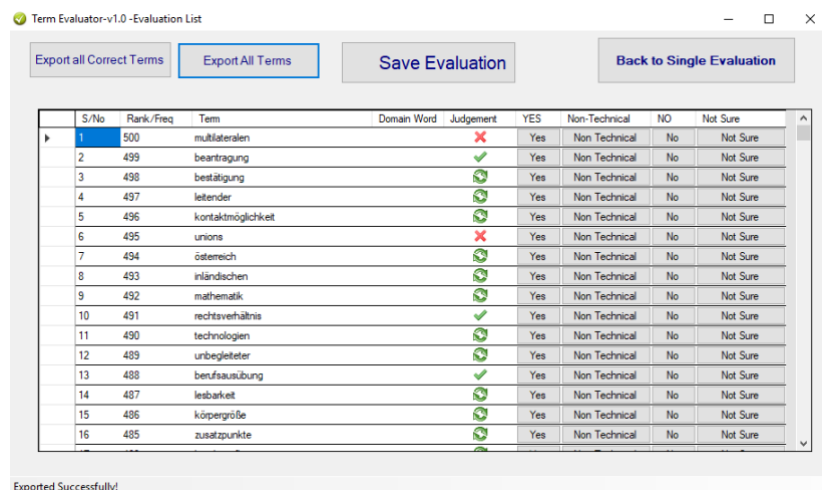


Abbildung 15: Die Auswertungsoberfläche (Term Evaluator)

Im Rahmen des Annotationsprozesses besteht für die AnnotatorInnen die Möglichkeit, sich nur die Termkandidaten anzeigen zu lassen, die zuvor noch nicht bewertet wurden, oder nur diejenigen, die im Konflikt mit den Bewertungen anderer AnnotatorInnen stehen. Sie können, wie erwähnt, die Auswertung auch speichern und zu einem späteren Zeitpunkt wieder aufrufen. Darüber hinaus steht eine Listenansicht zur Verfügung, die alle Termkandidaten einschließlich Annotationsstatus beinhaltet (Abb. 16). Als relevant bzw. korrekt annotierte Termkandidaten können während des Verfahrens jederzeit exportiert werden. Im Term Evaluator lassen sich außerdem bei mehreren AnnotatorInnen verschiedene Bewertungsergebnisse vergleichen, wobei die Anzahl der Übereinstimmungen und Unstimmigkeiten, die Schnittmenge der Annotationsentscheidungen, die Übereinstimmungsrate zwischen den AnnotatorInnen sowie einige weitere Statistiken und Vergleichseinzelheiten dargestellt werden.



S/No	Rank/Freq	Term	Domain Word	Judgement	YES	Non-Technical	NO	Not Sure
1	500	multilateral		✗	Yes	Non Technical	No	Not Sure
2	499	beartragung		✓	Yes	Non Technical	No	Not Sure
3	498	bestätigung		✓	Yes	Non Technical	No	Not Sure
4	497	leitender		✓	Yes	Non Technical	No	Not Sure
5	496	kontaktmöglichkeit		✓	Yes	Non Technical	No	Not Sure
6	495	unions		✗	Yes	Non Technical	No	Not Sure
7	494	österreich		✓	Yes	Non Technical	No	Not Sure
8	493	inländischen		✓	Yes	Non Technical	No	Not Sure
9	492	mathematik		✓	Yes	Non Technical	No	Not Sure
10	491	rechtsverhältnis		✓	Yes	Non Technical	No	Not Sure
11	490	technologien		✓	Yes	Non Technical	No	Not Sure
12	489	unbegleiteter		✓	Yes	Non Technical	No	Not Sure
13	488	berufsausübung		✓	Yes	Non Technical	No	Not Sure
14	487	lesbarkeit		✓	Yes	Non Technical	No	Not Sure
15	486	körpergröße		✓	Yes	Non Technical	No	Not Sure
16	485	zusatzpunkte		✓	Yes	Non Technical	No	Not Sure

Abbildung 16: Annotation-Listenansicht (Term Evaluator)

7.1.4.2 Evaluierungsprozess

Im Rahmen der vorliegenden Masterarbeit handelt es sich bei der quantitativen Evaluierung der Terminologieextraktion um die menschliche Annotation in Kombination mit dem Einsatz einer Referenztermliste. Damit ließen sich wenigstens die oben aufgeführten Grenzen der jeweiligen Methoden hinsichtlich der Berechnung von Precision- und Recall-Werten überwinden: Erstens sorgten die menschliche Untersuchung und manuelle Gewinnung der fachsprachlichen Benennungen zur Erstellung einer RTL aus demselben Korpus, der auch bei der maschinengestützten Terminologieextraktion verwendet wurde, zusammen mit der finalen Verifizierung möglichst dafür, dass alle fachspezifischen Benennungen erfasst wurden. So ließ sich die Gefahr minimieren, dass Termkandidaten trotz Relevanz als ungültig betrachtet

werden, weil sie sich nicht in der RTL befinden, was den Precision-Wert beeinträchtigen könnte. Zweitens konnte durch die festgelegte RTL zugleich ein bestimmter Umfang der gesamten Terminologie festgestellt werden, der die Berechnung von Recall-Werten ermöglicht. Zudem konnte die Evaluierung durch eine Person durchgeführt werden, da die RTL bereits als ein festgelegter Maßstab zur Verfügung stand, der bei der Validierung herangezogen wurde.

Nach der automatisierten Terminologieextraktion durch die drei unterschiedlichen Werkzeuge und der Übernahme im Excel-Programm wurden die Termkandidaten des jeweiligen Werkzeugs je nach Anzahl ihrer Bestandteile in fünf Kategorien eingegliedert: 1-Wort, 2-Wort, 3-Wort, 4-Wort und 5-Wort, d. h. entsprechend der manuell ermittelten Wortlängen, wobei die 1-Wort-Kategorie sowohl Simplicia als auch Komposita umfasste und die Mehrwortbenennungen sich auf die restlichen Kategorien bezogen. Durch diese Aufteilung ließ sich feststellen, ob das jeweilige Terminologieextraktionswerkzeug im Zuge der automatischen Verarbeitung unterschiedliche syntaktische Strukturen richtig erkannt hat. Allerdings warf die Evaluierung umfangreicher Termkandidaten gegen eine einsprachige RTL einige praktische Entscheidungen auf, die im Vorfeld getroffen werden sollten. Zum Beispiel wurde für jede Kategorie ein Cutoff-Wert von 500 Termkandidaten angewendet, d. h. jede Liste von Termkandidaten enthielt maximal die ersten 500 Ergebnisse. Lag die Anzahl von Termkandidaten der Kategorie unter dieser Grenze, so sollte die gesamte Liste beibehalten werden. Für jedes der drei Werkzeuge wurden der Annotatorin zwei bzw. fünf Dateien vorgelegt, die jeweils einer der vorliegenden Kategorien entsprachen. Da das Tool nur Daten mit der Kodierung Unicode UTF-8 verarbeiten kann, mussten die einzulesenden Dateien erneut in Microsoft Word bearbeitet und als einfache Texte abgespeichert werden.

Eine menschliche Annotatorin benutzte das Tool Term Evaluator zur Auswertung der maschinengestützt extrahierten Termkandidaten hinsichtlich ihrer Korrektheit und Fachspezifität. Dabei waren die Termkandidaten, die sich auch auf der RTL befanden, als terminologisch relevant zu beurteilen und mit der Option *correct* zu bewerten; bei denjenigen, die entweder Gemeinsprachlichkeit aufweisen oder deutlich anderen Fachbereichen angehören, sollte die Annotatorin die Option *non-technical* wählen. Schließlich galten die Einträge als falsch, die keine gültige semantische Bedeutung darstellten. Die Option *not sure* war nur dann zu wählen, wenn die Annotatorin auf Termkandidaten im Zuge der Annotation stieß, die zwar nicht von der RTL erfasst waren, für die jedoch eine potenzielle Verbindung zum Verwaltungsbereich der Niederlassungs- und Aufenthaltsberechtigung bestand.

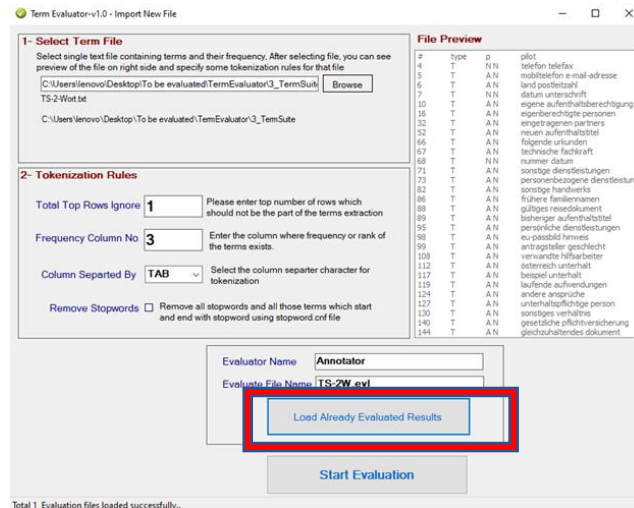


Abbildung 17: Datei importieren und Voreinstellungen (Term Evaluator)

Diese Annotationsaufgabe wurde allerdings nicht von derselben Fachperson durchgeführt, die verantwortlich für die Verifizierung der RTL war, sondern es handelte sich um eine zweite Überprüferin (siehe anonymisiertes Profil 2 im Anhang I), die gegebenenfalls die möglichen Schwachstellen der RTL feststellen und diese für die nachfolgende Ergebnisanalyse aufzeichnen sollte. Aus diesem Grund sollte die Annotatorin beim Annotationsprozess neben der Heranziehung der RTL auch auf ihre Fachkenntnisse zu entsprechenden Verwaltungsbereichen zurückgreifen. Im Falle von Unklarheiten wurden ihr die gegenständlichen Formulare sowie Links zu den behördlichen Online-Glossaren zum Fachgebiet zur Verfügung gestellt, die bei der Erstellung der RTL zurate gezogen wurden. Nachdem die Annotatorin die Beurteilung des ersten Datensatzes bzw. der Termkandidatenlisten des ersten Werkzeugs ihrer Wahl abgeschlossen und abgespeichert hat, bestand die Möglichkeit, die Funktion des Annotationsspeichers auszunutzen. Mit der Option Load already Evaluated Results konnten die Annotationsergebnisse je nach Kategorie der Wortanzahl bei der Voreinstellung für die Auswertung des nächsten Werkzeugs als Referenz importiert werden (rot eingerahmt in Abb. 17). Wenn der Termkandidat beim vorherigen Anlauf schon vorkam, wurde dieser samt seiner Bewertung in der unteren Hälfte der Auswertungsoberfläche angezeigt (Abb.18). Als Folge ließ sich der Prozess nicht nur beschleunigen, sondern fiel auch einheitlicher aus: Mithilfe dieser Funktion wurde prophylaktisch verhindert, dass die Annotatorin denselben Termkandidaten in unterschiedlichen Listen uneinheitlich beurteilen würde.

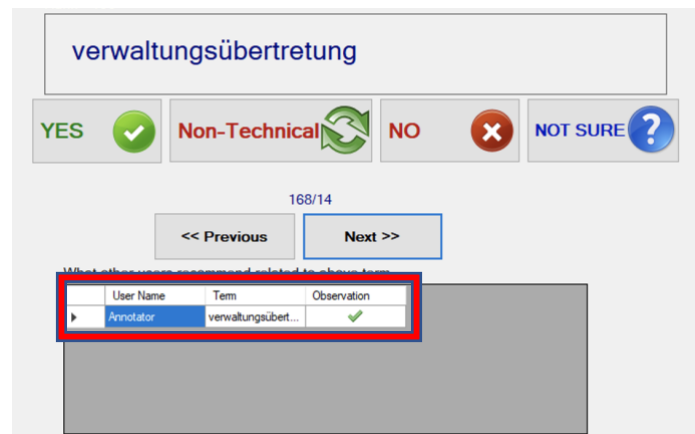


Abbildung 18: Annotationsspeicher (Term Evaluator)

7.2 Terminologieextraktionsergebnisse

Die Termkandidatenlisten von AntConc, die sich jeweils aus dem Einsatz des Wortliste- bzw. N-Gramm-Moduls ergaben, zeigten drei Kategorien von Informationen: Da im Vorfeld der Verarbeitung eingestellt wurde, dass die Ergebnisse nach ihrer Auftretenshäufigkeit im Korpus sortiert werden sollten, gibt die Spalte Rank – entsprechend dem vorbestimmten Kriterium und basierend auf der bezifferten Häufigkeitsverteilung in der mittleren Spalte Freq – die Rangfolge aller von dem Werkzeug ermittelten Termkandidaten an, die wiederum in der Spalte Word bzw. N-Gramm dargestellt werden. Insgesamt hat AntConc 1630 Einwortbenennungskandidaten durch die Funktion Wortliste identifiziert (Abb. 19) und hat 6 846 Ergebnisse für Mehrwortbenennungskandidaten erzeugt (Abb. 20).

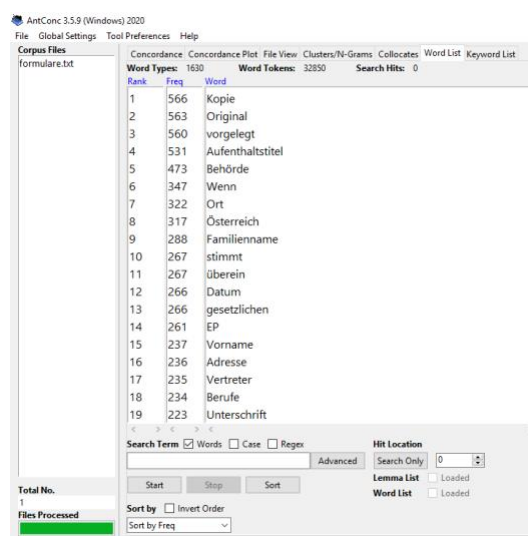


Abbildung 19: Wortliste-Ergebnis – AntConc 3.5.9

Rank	Freq	Range	N-gram
1	298	1	in Kopie
2	296	1	im Original
3	280	1	Familienname n
4	267	1	Kopie stimmt
5	267	1	Kopie stimmt mit
6	267	1	Kopie stimmt mit Original
7	267	1	Kopie vorgelegt
8	267	1	Kopie vorgelegt Kopie
9	267	1	Kopie vorgelegt Kopie stimmt
10	267	1	Kopie vorgelegt Kopie stimmt mit
11	267	1	Original vorgelegt
12	267	1	Original vorgelegt in
13	267	1	Original vorgelegt in Kopie
14	267	1	Original vorgelegt in Kopie vorgelegt
15	267	1	im Original vorgelegt
16	267	1	im Original vorgelegt in
17	267	1	im Original vorgelegt in Kopie
18	267	1	in Kopie vorgelegt
19	267	1	in Kopie vorgelegt Kopie

Abbildung 20: N-Gramm-Ergebnis – AntConc 3.5.9

Allerdings wurde im Interesse der Einheitlichkeit beschlossen, dass die festgelegte Untergrenze für Frequenzen ebenfalls den Wortliste-Ergebnissen gelten sollte. Somit waren diejenigen Termkandidaten mit weniger als fünfmaliger Auftretenshäufigkeit (wie in Inkpen et al. 2016 bei AntConc festgelegt) als irrelevant zu betrachten und wurden von der Liste ausgeschlossen, wonach schließlich noch 920 Einwortbenennungskandidaten verblieben. Um den Evaluierungsumfang einzugrenzen und da, wie angemerkt, aufgrund des dominierenden Nominalstils auf nominale Ausdrücke zu fokussieren war, wurde schließlich die Liste von Mehrworttermkandidaten möglichst von denjenigen Ergebnissen befreit, die nicht als brauchbare Nominalphrasen galten. Da der Einsatz einer Stoppwortliste beim N-Gramm-Modul nicht möglich ist, ließ sich der Präzisierungsvorgang mithilfe der Option Sortierung nach Ende des Worts realisieren, woraufhin die sich daraus ergebende Liste in der Excel-Tabelle weiterverarbeitet wurde (Abb. 21). Dabei waren die Ergebnisse, die vor allem ein Satzzeichen, eine Konjunktion oder eine Präposition am Ende enthielten, als untauglich zu betrachten und deshalb zu verwerfen. Daraus resultierte eine finale Sammlung aus 3 283 potenziellen Mehrwortbenennungen.

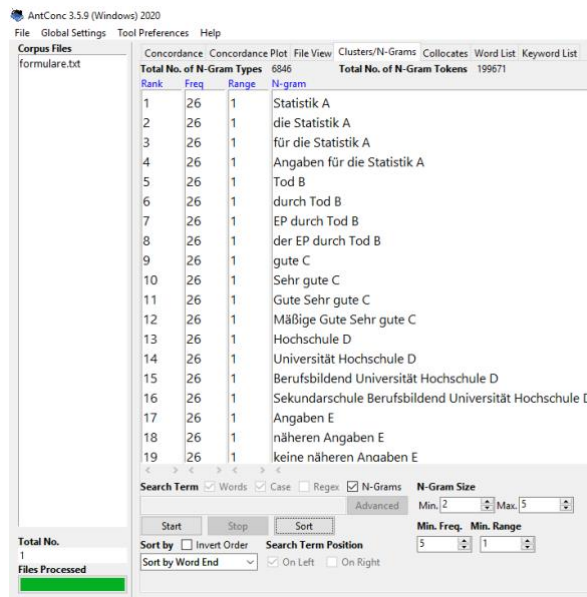


Abbildung 21: N-Gramm-Ergebnis sortiert nach Ende des Worts

Das TWSC lieferte hingegen insgesamt lediglich 181 Termkandidaten nach dem Extraktionsvorgang, die sowohl Einwort- als auch Mehrwortbenennungen umfassten und alphabetisch sortiert waren (Abb. 22). Normalerweise werden die Ergebnisse zwar vor der Weiterverarbeitung zunächst durch eine oder mehrere menschliche Fachpersonen verifiziert, jedoch wurden diese im Rahmen der vorgesehenen Leistungsevaluierung unmittelbar ohne weitere Validierung exportiert. Aufgrund der integrierten Stoppwortliste und den Musterstrukturen, die sich nicht anpassen lassen und sich deshalb eventuell insbesondere auf die Mehrwortbenennungskandidaten auswirken könnten, war mit einem höheren Silence-Anteil zu rechnen. Tatsächlich konnte das Werkzeug nur Mehrworttermkandidaten mit maximal zwei Bestandteilen erkennen.

II	Extracted terms	Approved translations	Translation candidates	Import	Export
	Aufenthaltsrecht	add translation	Other Aufenthaltsrecht		
	Aufenthaltskarte	add translation	Other Aufenthaltskarte		
	Ausbildung	add translation	Other Ausbildung		
	ausreichende Krankenversicherung	add translation	Other ausreichende Krankenversicherung		
	Aufenthalt	add translation	Other Aufenthalt		
	Angehörige	add translation	Other Angehörige		
	Aufenthaltsbeendigung	add translation	Other Aufenthaltsbeendigung		
	Adresse	add translation	Other Adresse		
	Anschrift	add translation	Other Anschrift		
	ANMELDEBESCHEINIGUNG	add translation	Other ANMELDEBESCHEINIGUNG		
	Auflösung	add translation	Other Auflösung		
	ausreichenden Existenzmittel	add translation	Other ausreichenden Existenzmittel		
	Augenfarbe	add translation	Other Augenfarbe		
	Aufenthaltstitel	add translation	Other Aufenthaltstitel		

Abbildung 22: Termkandidatenliste von TWSC (TaaS)

Mit dem hybriden Terminologieextraktionsansatz durch TermSuite wurde eine Endergebnisliste aus 1 477 Termkandidaten erzeugt, die im Termkandidaten-Ansichtsbereich dargestellt wurden (Abb. 23). Dabei handelte sich um eine Tabelle, in welcher jeder Zeile ein extrahierter Termkandidat je nach seiner Auftretenshäufigkeit im Korpus zugeordnet wurde. In den Spalten wurden die relevantesten Eigenschaften des Terms angezeigt: sein syntaktisches Muster (Pattern), seine Spezifität (Spec., d. h. relative Frequenz gegenüber dem eingebauten gemeinsprachlichen Korpus), seine Häufigkeit (Fr.) und seine Dokumenthäufigkeit (D-Freq, die in diesem Falle 1 bei jedem Eintrag aufgrund des Einzelkorpus betrug). Die vollständige Liste der Eigenschaften eines einzelnen Termkandidaten konnte zusätzlich im Navigationsbereich auf der rechten Seite eingesehen werden. Die extrahierten Termkandidaten ließen sich, wie voreingestellt, zusammen mit den Informationen zu deren Musterstrukturen und statistischen Werten im Format einer Tab-Separated-Value-Datei exportieren, die in der Excel-Tabelle leicht übernommen werden konnte.

The screenshot shows the TermSuite interface. On the left, a list of term candidates is displayed with columns for #, Pilot, Pattern, Spec., Fr., and D-Freq. The list includes terms like '1 aufenthaltstitel', '2 nur', '3 e-mail adresse', etc. On the right, a 'Term' panel shows detailed properties for a selected term, including Property, Value, Spelling rule, Pattern, Frequency, Specificity, Document Freq, Compound, and Alignment Results.

#	Pilot	Pattern	Spec.	Fr.	D-Freq
1	aufenthaltstitel	N	5.52	539	1
2	nur	R	5.25	288	1
3	e-mail adresse	N	4.90	131	1
4	telefon telefax	N N	4.89	127	1
5	mobilteléfono e-mail-adresse	A N	4.89	127	1
6	land postleitzahl	A N	4.89	126	1
7	datum unterschift	N N	4.81	106	1
8	titelnummer	N	4.80	102	1
9	erstantrag	N	4.77	96	1
0	eigene aufenthaltsberechtigung	A N	4.77	96	1
1	geburtsdatum	A	4.54	57	1
2	homepage	N	4.54	56	1
3	verwaltungübertragung	N	4.53	55	1
4	unterschrift des gesetzlichen vertreter	N A N	4.53	55	1
5	anschrift des gesetzlichen vertreter	N A N	4.52	54	1
6	eigenberechtigte personen	A N	4.52	54	1
7	wohnsitz des antragstellers	N N	4.51	53	1
8	titel	A	4.51	53	1
9	baugewerbe	A	4.50	52	1
0	maschinenbediener	N	4.50	52	1
1	montierer	N	4.50	52	1
2	nichttechnische	A	4.50	52	1

Abbildung 23: Termkandidaten-Ansichtsbereich (TermSuite)

7.3 Ergebnisse der Leistungsevaluierung

In der letzten Phase der Untersuchung wurden die relativen Precision- und Recall-Werte sowie das ausgewogene F1-Maß für das jeweilige Werkzeug hinsichtlich seiner automatisierten Extraktionsleistung berechnet. Während Precision das Verhältnis der als korrekt annotierten Termkandidaten zur Anzahl der insgesamt extrahierten Ergebnisse bestimmt, wurde das Recall-Maß im Rahmen der vorliegenden Masterarbeit anhand der Vereinigung aller Termini auf der vorbestimmten Referenzterm liste ermittelt. Es gilt vorab darauf hinzuweisen, dass je größer die Termkandidatenliste ist, desto höher fällt der Recall-Wert aus, während der Precision-Wert abnimmt, insbesondere wenn diese umfangreiche Termkandidatenliste gegenüber einer viel kürzeren RTL annotiert wurde. Obwohl bereits im Vorfeld mit einem niedrigen Precision-Wert

zu rechnen war, wurde er ebenfalls ermittelt und analysiert, da dieses Maß des Korrektheitsgrads einen bedeutsamen Indikator für die Extraktionsleistung darstellt. Dementsprechend wurden die beiden Maße im Rahmen der Berechnung des F1-Maßes gleich gewichtet, d. h. $\alpha = \frac{1}{2}$, wie dies in Abschnitt 5.7.2 dargelegt wird. Die Leistungsbewertung der drei automatischen Terminologieextraktionswerkzeuge wird nachfolgend zum Vergleich in Abbildungen 24-28 und in Tabelle 2 erläutert.

Abb. 24 zeigt die Leistung der Tools bei der Extraktion von Begriffen, die nur ein Wort enthalten. Hinsichtlich des Recall-Werts übertraf das Extraktionssystem TermSuite, das auf einem hybriden Ansatz bei der Termidentifizierung basiert, mit 79 % die anderen Tools hinsichtlich der Extraktion von Einwortbenennungen. Mit Abstand kamen TWSC (33 %) an zweiter und AntConc (29 %) an letzter Stelle. Anzumerken ist, dass TWSC trotz der linguistischen Anreicherung nicht deutlich bessere Ergebnisse als AntConc lieferte, das selbst keine ausgeklügelte Identifizierungsmethode verwendet. Dies lässt sich auf die beschränkten integrierten Ressourcen auf der TaaS-Plattform zurückführen, die keine umfassenden Fachterminologien und einzelsprachliche Auflage einschließen, was zu einem Silence-Effekt geführt hat. Gerade aufgrund des geringeren Noise-Anteils hat TWSC jedoch im Hinblick auf diesen am besten abgeschnitten, gefolgt von TermSuite. Die Tatsache, dass AntConc einen niedrigeren Recall- und Precision-Score hat als AntConc, kann ein Hinweis darauf sein, dass die bloße Häufigkeit möglicherweise nicht ausreicht, um den standardisierten Charakter von Formularen und der Verwaltungssprache auszunutzen, der sich unter anderem durch einheitliche und repetitive Verwendung des Wortschatzes auszeichnet. Der bei AntConc eingesetzte Ansatz, der ausschließlich die Frequenz ermittelt, hat einen hohen Noise-Anteil nach sich gezogen. Zum Beispiel wurden die Berufsbezeichnungen aus den Fragebögen zur statistischen Erhebung der Behörde, die jedem Formular beigelegt sind und deshalb häufig auftreten, trotz deren terminologischen Irrelevanz von AntConc erfasst und in die Termkandidatenliste inkludiert. Zahlreiche Einträge waren damit während des Annotationsprozesses zu verwerfen, was zu einem geringen Precision-Prozentsatz geführt hat.

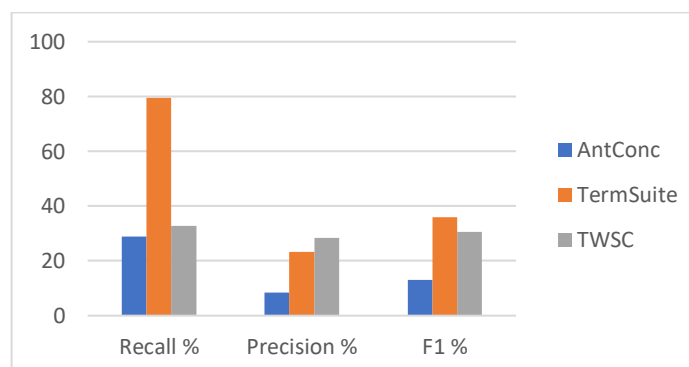


Abbildung 24: Extraktionsleistung bei Einwortbenennungen

Abb. 25 zeigt die Leistung der Werkzeuge bei der Extraktion von Zweiwortbenennungen aus dem verwaltungssprachlichen Korpus. TermSuite erhielt zwar wieder die beste Recall- sowie die gesamte F1-Punktzahl und stellt damit das leistungsstärkste Werkzeug für die Zweiwort-Termkategorie dar, die Aufmerksamkeit gilt in diesem Zusammenhang aber dem Quantensprung des Recall-Werts von AntConc, der von 29 % bei Einworttermkandidaten auf 37 Prozentpunkte bei Zweiwortergebnissen gestiegen ist. Diese hohe Leistung manifestiert sich in einer adäquaten Darstellung von Termhood und Unithood. Angesichts der Ergebnisse von Werkzeugen, welche die Daten linguistisch basiert verarbeitet haben, wurde deutlich, dass POS-Tagging umso mehr an Bedeutung gewinnt, je mehr die Anzahl der potenziellen Mehrworttermini zunimmt: TWSC konnte kaum mithalten, da die Qualität seiner Extraktionsergebnisse, wie zuvor vermutet, unter einem simplistischen POS-Tagger gelitten hat, verglichen mit TermSuite, das den weit verbreiteten und effektiven TreeTagger von Schmid (1995) einsetzte. Die konkurrenzfähige Precision-Punktzahl von TWSC, die etwas höher als die von TermSuite ausfällt, lässt sich auf die Tatsache zurückführen, dass die Termkandidatenliste lediglich zwölf Einträge umfasste.

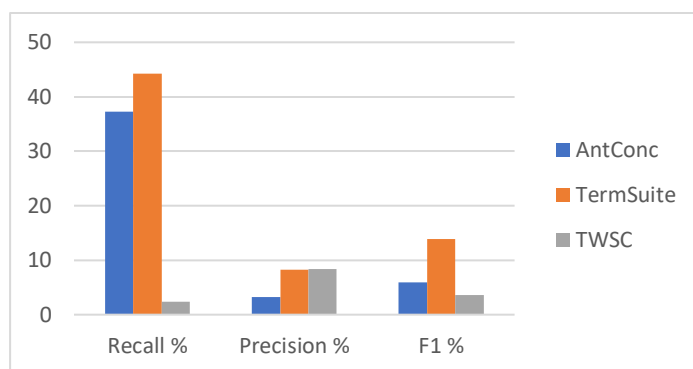


Abbildung 25: Extraktionsleistung bei Zweiwortbenennungen

Abb. 26 zeigt die Extraktionsleistung bei der automatischen Verarbeitung zur Identifizierung von Dreiworttermkandidaten. Dabei fällt der Leistungsabfall von AntConc auf – und das insbesondere im Fall von TWSC: Ersteres hat zwar Ergebnisse geliefert, mehr als die Hälfte davon waren aber als Noise zu betrachten und wurden deshalb von der Annotatorin als inkorrekt bewertet, sodass sich der Recall- und Precision-Wert von bloß 19 % bzw. 1,4 % ergeben hat. Zugleich ist Letzteres allerdings gänzlich damit gescheitert, Dreiworttermkandidaten überhaupt zu extrahieren. TermSuite weist durchweg die höchste Precision und den höchsten Recall auf, wobei die als korrekt annotierten Termkandidaten mit der morphosyntaktischen Musterstruktur (Nomen + (Genetiv-)Artikel + Nomen) die Mehrheit ausmachen. Die wesentlichen Dreiworttermini, wie Niederlassungs- und Aufenthaltsgesetz und

Sicherung des Lebensunterhalts, wurden nicht von TermSuite erkannt, sondern von AntConc, da deren Auftretenshäufigkeit dem festgelegten Schwellenwert entsprach.

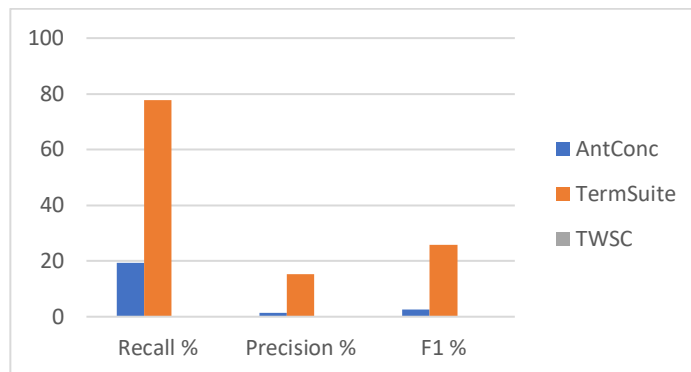


Abbildung 26: Extraktionsleistung bei Dreiwortbenennungen

Schließlich stellen Abbildungen 27 und 28 die Evaluierungsausgabe in Bezug auf die Extraktion von Vier- und Fünfworttermkandidaten dar. Anhand der Daten lässt sich deutlich feststellen, dass TermSuite am besten abgeschnitten hat, während die Termkandidatenliste von AntConc nur unbrauchbare Ergebnisse beinhaltete und die Leistung von TWSC für die beiden Kategorien weiterhin schlecht ausfiel. Während TermSuite bei der Vierwort-Kategorie einen höheren Recall-Wert als Precision hat und somit dem gleichen Trend wie in den vorigen Kategorien folgt, ergab sich im Fall der Extraktion von Fünfworttermkandidaten aber ein umgekehrtes Verhältnis. Dies wurde dadurch verursacht, dass sich nur vier Fünfworttermkandidaten aus dem maschinengestützten Extraktionsvorgang ergaben, was weniger als die gesamte Anzahl der Fünfworttermini auf der RTL entspricht.

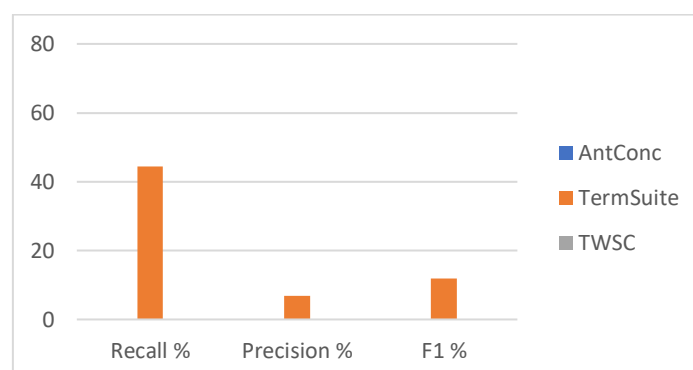


Abbildung 27: Extraktionsleistung bei Vierwortbenennungen

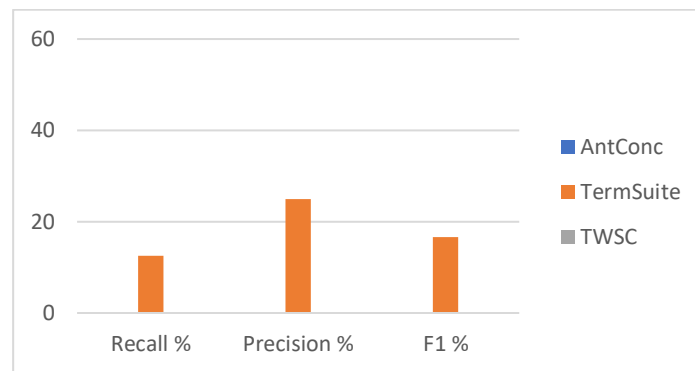


Abbildung 28: Extraktionsleistung bei Fünfwortbenennungen

Zum Schluss wurde die Gesamtleistung der einzelnen Terminologieextraktionswerkzeuge ermittelt (Tabelle 2). TermSuite erreichte mit Abstand die höchste prozentuale Punktzahl dank seiner wirksamen Analyse der Domänenspezifität von Termkandidaten gegenüber den gemeinsprachlichen Referenzkorpora, seines Einsatzes des effektiven POS-Taggers und seiner umfassenden linguistischen Ressourcen, die für den Musterabgleich von Nutzen waren. AntConc lag mit dem niedrigsten F1-Maß an letzter Stelle, obwohl sein Recall-Wert denjenigen von TWSC um mehr als das Doppelte überstieg. Dies lag an seinem deutlich geringeren Precision-Wert, denn das Tool hat einen hohen Noise-Anteil ausgegeben, während der Umfang der Termkandidatenliste von TWSC durchschnittlich um 43 % kleiner war, was wiederum zum Silence-Effekt und somit zum schlechtesten Recall-Wert beigetragen hat.

Gesamtbewertung	Recall	Precision	F1-Maß
AntConc	17,08	2,60	4,51
TWSC	7,04	7,35	7,19
TermSuite	51,67	15,70	24,10

Tabelle 2: Prozentuale Gesamtbewertung der automatischen Terminologieextraktionswerkzeuge

7.4 Diskussion

Dieses Unterkapitel lässt sich in drei Abschnitte gliedern. Zunächst wird Abschnitt 7.4.1 der Diskussion einiger wesentlicher Aspekte gewidmet, die sich im Zuge der RTL-Erstellung als problematisch herausgestellt haben. Zudem wird geschildert, wie mit solchen Schwierigkeiten umgegangen wurde. Im Abschnitt 7.4.2 werden die Ergebnisse, die im Zuge der Leistungsevaluierung der Terminologieextraktionswerkzeuge erlangt wurden, eingehend erläutert. Schließlich folgen im Abschnitt 7.4.3 die Schlussfolgerungen in Bezug auf die

Forschungsfrage sowie die Reflektionen über die möglichen Optimierungen hinsichtlich der Untersuchungsmethodologie.

7.4.1 Herausforderungen bei der RTL-Erstellung

Bei der Gestaltung der RTL musste sich mit einer Reihe von Herausforderungen auseinandergesetzt werden, die einerseits generell im Rahmen der terminologischen Praxis zu bewältigen waren und die andererseits mit der Methode zur manuellen Terminologieextraktion zusammenhingen.

Beim ersten Versuch der manuellen Terminologiegewinnung stellten Mehrwortbenennungen in Form von Wortverbindungen mit einem gemeinsamen Bestandteil, wie Aufenthalts- und Niederlassungsgesetz, Miet- oder Untermietvertrag und Pensions-Renten- oder sonstige Versicherungsleistungen, bei denen der Bindestrich eine Auslassung anzeigt und zugleich als Ergänzungszeichen in Erscheinung tritt, ein unerwartetes und einzigartiges Problem dar: Da der Bindestrich sinngemäß das Wortende markiert, steht danach sowie nach der Konjunktion ein Leerzeichen, wodurch alle Bestandteile solcher Termkandidaten dann im Zuge der Übertragung in die Excel-Tabelle in einzelne Einheiten zerlegt wurden. Da die händische Suche nach allen Einheiten mit einem anschließenden Bindestrich und ihren restlichen zusammengehörigen Bestandteilen Zeitaufwand wie auch potenzielle Silence-Effekte bedeutet hätte, wurde der Ausgangskorpus unter Berücksichtigung dieser Art von Mehrwortbenennungen erneut vorverarbeitet. Zusätzlich wurde im Texteditor nach Wörtern mit intrinsischen Teilmustern, wie x- und y und x- oder y, gesucht, um diese Mehrwortbenennungskandidaten vorerst in Komposita umzuwandeln (z. B. Aufenthalts- und Niederlassungsgesetz > Aufenthalts-und-Niederlassungsgesetz) und damit dasselbe Problem bezüglich der Segmentierung in der Excel-Tabelle zu vermeiden (vgl. Drewer & Schmitz 2017: 72). Im Zuge der Bereinigung der RTL-Liste wurde die Zusammensetzung rückgängig gemacht, sodass die aufgenommenen Mehrwortbenennungen ihrer kanonischen Form wieder entsprechen.

Der Einsatz einer Stoppwortliste mit Funktionswörtern zur Präzisierung der RTL war nicht möglich, da zahlreiche nominale Fachausdrücke im Rahmen des österreichischen Aufenthaltsrechts Artikel, Präpositionen und Konjunktionen innerhalb der Benennung enthalten. Die Entfernung solcher Elemente hätte die Identifizierung von Mehrwortbenennungen erschwert, deren syntaktische Musterstrukturen solche Funktionswörter enthalten, wie Recht auf Daueraufenthalt oder Sicherung des Lebensunterhalts. Ursprünglich wurde außerdem eine zusätzliche Auswertung der

Fachspezifität mit einer Pivot-Tabelle eingeplant, welche die Auftretenshäufigkeit einzelner Einheiten strukturiert zusammenfasst und dabei vom wiederholenden Gebrauch an Begriffen, der zu den Stilcharakteren der österreichischen Verwaltungssprache gehört, profitieren könnte. Jedoch war dieser Schritt auszulassen: Eine Menge terminologisch irrelevanter Funktionswörter, die notwendigerweise zur Vermeidung von Silence-Effekten bei Mehrwortbenennungen beibehalten werden mussten, hätten jeweils einen genauso hohen oder sogar höheren Frequenzwert wie die tatsächlichen fachspezifischen Termini erhalten, sodass die sich daraus ergebende Pivot-Tabelle keine zuverlässigen Daten geliefert hätte.

Im Rahmen der Verifizierung der Termkandidaten besteht das Hauptproblem darin, den Termhood- und Unithood-Grad eines Benennungskandidaten zu beurteilen. Wie im Unterkapitel 5.5 geschildert, lässt sich feststellen, dass die (automatisierte) Erfassung solcher abstrakter Eigenschaften keine einfache Aufgabe ist: Sie können nur indirekt durch andere konkrete und parametrisierbare Eigenschaften gemessen werden, z. B. die Auftretensfrequenz, statistische Assoziationsmaße der Kollokation oder der festgelegte linguistische Musterabgleich. Die menschliche Ermittlung, ob und inwiefern ein Terminus einem bestimmten Fachgebiet entspricht oder nicht, hat sich dementsprechend sogar als noch problematischer erwiesen, denn der Prozess ging mit spezifischen Faktoren einher. Hierzu gehört unter anderem das domänenspezifische Fachwissen der beteiligten Personen, das wiederum die individuelle Sichtweise gegenüber einzelnen Termkandidaten beeinflusst (vgl. Condamines 2010: 31 f.). In dieser Hinsicht weisen Frantziy et al. (1998: 594) sowie Rigouts Terry et al. (2020: 85) darauf hin, dass es immer noch an formalen oder präzisen Regeln mangelt, die zu einer deutlichen Unterscheidung zwischen einem Terminus und einem Nichtterminus beitragen.

Dies führte im Zuge der menschlichen Terminologiegewinnung und der RTL-Verifizierung gelegentlich zur Uneinigkeit zwischen den Beteiligten, was die terminologische Relevanz einiger Termkandidaten anbelangte. Die Hinzuziehung der lexikologischen Ressourcen auf den offiziellen Websites der zuständigen Behörden konnte, wie erwähnt, die Unklarheit bei einigen Termkandidaten beheben. Durch die Einsicht dieser Online-Glossare in Gegenprüfung mit der RTL ließ sich das Ineinandergreifen des Gebiets Einwanderung und Niederlassung in anderen Verwaltungs- und Rechtsteilbereichen bemerken: Die Aufenthaltsberechtigung einer Person aus anderen Herkunftsländern in Österreich hängt erheblich mit verschiedenen Lebensdimensionen, wie beruflicher Beschäftigung, Ausbildung, Finanziellem usw., zusammen. Dementsprechend spiegelt sich dieses Merkmal in der teils überlappenden Anwendung der Terminologie unter den verwandten (Teil-)Bereichen.

Aufgrund dieses verflochtenen Terminologiepools weisen bestimmte Termini einen Gradienten der Fachspezifität auf (vgl. Longinova et al. 2012: 8). Beispielsweise wird sich mit dem Fachausdruck Rot-Weiß-Rot-Karte unter anderem nicht nur im Rahmen des österreichischen Aufenthalts- und Niederlassungsrechts beschäftigt, sondern er findet sich auch in der Rechtsvorschrift für das österreichische Ausländerbeschäftigungsgesetz, da die Aufenthaltsberechtigung einiger AusländerInnen durch ihr Beschäftigungsverhältnis vorausgesetzt wird. So wurde beschlossen, die Aufnahme einiger relevanter Termini aus angrenzenden Teilbereichen auch zuzulassen.

Bei Fragen zu Fachspezifität konnten die von Wiesinger (2008b) formulierten Merkmale der österreichischen Verwaltungssprache darüber hinaus im Zusammenhang mit der Identifizierung und Verifizierung von Termini zurate gezogen werden. Vor allem war die Bevorzugung des Nominalstils im untersuchten Korpus besonders stark ausgeprägt und deshalb leicht zu erkennen. Dies äußert sich deutlich in der Häufigkeit der folgenden Erscheinungsformen:

- Längere nominale Satzglieder mit Aufschwemmung durch mehrfache Attribute, z. B. Verwandter in gerader aufsteigender Linie, mobiler unternehmensintern transferierter Arbeitnehmer;
- abstrakten Verbalsubstantive, die durch Substantivierung von Infinitiven gebildet werden, z. B. Bestehen der Ehe oder eingetragenen Partnerschaft, Vorliegen von ausreichenden Existenzmitteln, oder durch Ableitung auf das Suffix -ung und -keit, z. B. Innehabung, Uneinbringlichkeit, Beendigung;
- substantivierten Adjektive, z. B. Angehöriger, unbegleiteter Minderjähriger, Zusammenführender;

Neben dem nominalen Stilcharakter kamen hinsichtlich der Fachausdrücke textspezifische Adjektivkomposita vor, wobei vorwiegend auf Bildungen auf -pflichtig und -berechtigt zurückgegriffen werden, z. B. unterhaltungspflichtige Person, leistungspflichtige Krankenversicherung. Diese werden aber auch in mehreren Stellen substantiviert und damit als Nomen verwendet, was noch mehr zum Nominalstil der Verwaltungssprache beitragen.

7.4.2 Evaluierungsergebnisse

Der deutlich niedrige Precision-Wert hängt mit dem ausgelösten Noise-Effekt zusammen, der durch verschiedene Faktoren verursacht werden konnte. Einerseits waren die Ursachen für die suboptimale Extraktionsleistung offensichtlich. Zum Beispiel hatte die ausschließliche Frequenzermittlung von AntConc zur Folge, dass terminologisch irrelevante Tokens aufgrund

ihres rein statistischen Auftretens als Termkandidaten identifiziert wurden und damit einen hohen Noise-Anteil erzeugten. Diese Beschränkung des statistischen Extraktionsverfahrens ließ sich durch die funktionsbedingte Struktur von Formularen verstärken, bei der es sich im Wesentlichen um vorformulierte Erhebungsformen handelt, die wiederum insbesondere durch stichwortartige Elemente gekennzeichnet sind: Der Inhaltstext wurde bei der Konvertierung und Abspeicherung des Korpus in das für die automatische Verarbeitung erforderliche Format einer einfachen Textdatei versetzt; zudem wurden sämtliche Leerstellen zum Ausfüllen und Kästchen, welche die Textelemente in den Formularen trennen, vollständig entfernt. Infolgedessen ergaben sich daraus unter anderem mehrere zufällige Gruppierungen von Tokens, zwischen denen keine semantischen und damit keine terminologischen Zusammenhänge bestanden oder die unvollständige Satzfragmente darstellten. Dies hat sich auf die Extraktion der Mehrwortbenennung durch die Funktion N-Gramm von AntConc insbesondere nachteilig ausgewirkt, da das System die Termhood- und Unithood-Eigenschaft solcher Token-Sequenzen lediglich durch die statistisch gemeinsame Auftretenshäufigkeit bestimmt. Hinzu kommt, dass die Stoppwortliste nur bei der Funktion Wortliste verwendet werden konnte. Dementsprechend ließen sich bei der Identifizierung von Mehrworttermkandidaten potenzielle Funktionswörter bzw. gemeinsprachliche Ausdrücke nicht herausfiltern. Die vorläufige Bereinigung der Termkandidatenliste mithilfe der Sortierungsoption nach Wortende hat im Anbetracht der Gesamtbewertung von TWSC offenbar versäumt, für diese Grenze zu kompensieren.

Andererseits waren die Ursachen des Noise-Effekts in Bezug auf TWSC und TermSuite weniger eindeutig und sind vielmehr als Beobachtungen zu betrachten, die auf der Analyse der jeweiligen Termkandidatenliste basieren. In dieser Hinsicht sind zwei Gründe dafür denkbar, dass die Systeme vor allem bestimmte Mehrworttermkandidaten falsch extrahiert haben. Der erste Grund für einen syntaktischen Noise-Effekt ist die falsche Abgrenzung von lexikalischen Einheiten, die aus der mangelhaften Segmentierung resultieren könnte. Dies war möglicherweise auf dieselbe textsortenbedingte Problematik wie bei AntConc zurückzuführen und trat auf, wenn das automatische Termextraktionssystem eine lange Phrase als Termkandidat meldet, die sowohl eine gültige Benennung als auch einen Teil des unmittelbaren Kontexts der gültigen Benennung enthielt, z. B. in Österreich Risiken abdeckenden Krankenversicherungsschutz. Dabei gilt nur das Element Krankenversicherungsschutz als terminologisch relevant. Umgekehrt kommt diese Art des Noise-Faktors vor, wenn das Termextraktionssystem Elemente am Anfang oder Ende der identifizierten Nominalphrase abschneidet. Beispiele für eine fehlerhafte Token-Abgrenzung aus der von TermSuite

generierten Termkandidatenliste sind gerader aufsteigender Linie und Aufenthaltsgesetz. Hier hätte TermSuite die Einheit Verwandte bzw. Niederlassungs- mit den aufgeführten identifizierten Termkandidaten einschließen müssen, um jeweils den vollständigen Terminus zu liefern: Verwandte in gerader aufsteigender Linie bzw. Niederlassungs- und Aufenthaltsgesetz. Bei der zweiten Ursache handelt es sich um die nichtterminologische Kombination, worunter eine Art struktureller Mehrdeutigkeit zu verstehen ist. Diese ist dadurch entstanden, dass das System nicht in der Lage war, zwischen fachspezifischen Fachausdrücken und gemeinsprachlichen Kollokationen zu unterscheiden, welche die gleiche syntaktische Musterstruktur aufweisen. Zu nichtterminologischen Kombinationen wurden auch solche gezählt, die sich fachspezifisch plausibel anhörten, z. B. aktuelles Lichtbild des Antragstellers, gleichwertiger Nachweis. Obwohl diese über eine Bedeutung und eine standardisierte Nomenklatur verfügen, gelten sie nicht als Fachtermini im engeren Sinne. Entsprechend, ausreichend und aktuell gehören zu Beispielen für Adjektive, die zwar nie Teil eines gültigen Terminus waren, aber sowohl von TWSC als auch von TermSuite mehrfach extrahiert wurden.

Hingegen hat sich der Silence-Effekt, der erzeugt wird, wenn das automatische Termextraktionssystem keine gültigen terminologischen Einheiten identifizieren kann, auf den resultierenden Recall-Wert ausgewirkt. Dies hängt grundsätzlich auch mit unterschiedlichen Beschränkungen des jeweiligen Werkzeugs zusammen. Deutlich anhand der Recall-Darstellung in Abbildungen 25–28 zu bemerken war, dass alle drei Werkzeuge zunehmend Schwierigkeiten hatten, Termkandidaten zu erfassen, und die Qualität des Extraktionsergebnisses gleichzeitig stärker gelitten hat, als die Anzahl der lexikalischen Einheiten zunahm. Bei AntConc ist die Silence-Ursache eindeutig: Relevante Tokens bzw. Gruppen von Tokens wurden nicht in die Termkandidatenliste eingeschlossen, da deren Auftretenshäufigkeit dem festgelegten Grenzwert nicht entsprach bzw. diesen nicht überstieg. Damit ist es möglich, dass die hohe Leistung von AntConc bei der Zwei-Wort-Kategorie auf Zufall zurückzuführen ist, indem die relevanten 2-Gramme so häufig im Korpus aufgetreten sind, dass das Werkzeug diese als signifikant identifiziert hat. Zugleich ließ sich im Hinblick auf die Silence-Auslöser bei TWSC und TermSuite nur spekulieren, denn die umfassenden Algorithmen und alle intrinsischen Funktionen der Werkzeuge waren nicht bekannt, insbesondere die von TWSC, dessen Termkandidatenliste den größten Silence-Anteil aufweist. Da die beiden Systeme auf die linguistisch basierte Extraktionsmethode setzen, ist zu vermuten, dass das Problem in erster Linie durch begrenzte intrinsische einzelsprachliche Ressourcen bedingt wurde. Zum einen fehlten den Systemen wahrscheinlich bestimmte syntaktische Musterstrukturen, die als relevant für die österreichische Verwaltungsterminologie im Bereich

Einwanderung gelten und im Korpus auftreten. Es liegt nahe, dass TWSC deswegen nicht in der Lage war, Termkandidaten aus mehr als zwei lexikalischen Einheiten zu extrahieren – und das nur mit dem Muster (Adjektiv + Nomen). Gleichzeitig konnte TermSuite die gültigen Mehrwortbenennungen, die z. B. Konjunktionen als ein Glied enthalten, nicht erkennen, z. B. bei Niederlassungs- und Aufenthaltsgesetz, notarielle oder gerichtliche Beglaubigung und Träger von Privilegien und Immunitäten. Zum anderen zeigte sich der Silence-Effekt in der fehlerhaften linguistischen Anreicherung, z. B. erkannte TreeTagger im Zuge von POS-Tagging das Nomen Kindesstatt nicht und wies ihm eine falsche Wortart zu, sodass es als Adjektiv markiert wurde. Infolgedessen verlief die Analyse der Domänenspezifizität von TermSuite auf Basis dieses fehlerhaften Taggings, was dazu führte, dass der statistisch ermittelte Wert des Tokens unter der Grenze lag.

7.4.3 Schlussfolgerung

Anhand der Evaluierungsergebnisse lässt sich die Hypothese bestätigen und zugleich die Forschungsfrage beantworten, dass das Werkzeug *TermSuite* mit einem hybriden Terminologieextraktionsansatz bei der Gewinnung österreichischer Verwaltungsterminologie aus dem Korpus digitaler Antragsformulare im Bereich Niederlassungs- und Aufenthaltsrecht vergleichsweise das beste Ergebnis liefert. Jedoch ist anzumerken, dass die Punktzahl seiner Gesamtleistung sowie die der anderen ausgewählten Werkzeuge im Verhältnis zu den Ergebnissen existierender relevanter Studien, bei denen die Extraktionsleistung eines Systems (z. B. Vivaldi & Rodríguez 2006; Longinova et al. 2012) oder mehrerer unterschiedlicher Methoden (z. B. Love 2000; Inkpen et al. 2016; Rigouts Terryn et al. 2020) hinsichtlich der Genauigkeit und Trefferquote quantitativ ausgewertet wurde, wesentlich geringer ausfielen. Dies deutet nicht nur auf einen erheblichen Verbesserungsbedarf hin, sondern daraus lässt sich außerdem schließen, dass sich eine Terminologieextraktion mit diesen Hilfsmitteln ohne Anpassungen an mögliche Besonderheiten des Korpus, kaum in diesem Fachbereich automatisieren lässt.

Idealerweise könnten zur Behebung des niedrigen Recall-Werts bei der Verarbeitung österreichischer Verwaltungstexte im Bereich Aufenthaltsgenehmigung zusätzliche syntaktische Strukturen den linguistischen Ressourcen des Werkzeugs hinzugefügt werden, welche die typischen (längeren) Nominalphrasen der Fachsprache im Zuge der musterbasierten Analyse erkennen lassen. Die Musterstrukturen, die in der Referenztermliste enthalten sind, sollen als eine bedeutsame Ausgangsbasis in dieser Hinsicht dienen. Ferner wäre jegliche

Maßnahme von Vorteil, die darauf abzielt, einen möglichst hohen Recall-Wert zu erreichen. Bei der Durchführung der praktischen Untersuchung wurde nämlich klar, dass beispielsweise die Festlegung eines anspruchsvollen Schwellenwerts zur Verringerung möglicher Noise-Effekte in den Termkandidatenlisten, die von den automatischen Terminologieextraktionssystemen erzeugt werden, unvermeidlich zu einer entsprechenden Zunahme der Silence-Effekte führt: Je restriktiver eine Regel zur Identifizierung von Termkandidaten wirkt, desto höher ist die Wahrscheinlichkeit, dass potenzielle gültige Termkandidaten aus dem Termkandidatenbericht ausgeschlossen werden.

Um diese Einschränkung zu bewältigen, wäre der üblichen Praxis zu folgen, bei der die Erhöhung des Noise-Effekts im Rahmen der automatischen Terminologiegewinnung als ein akzeptabler Faktor in Kauf genommen wird, um den möglichen Silence-Anteil auf ein striktes Minimum zu beschränken. Die Aufgabe, eine längere Termkandidatenliste von Noise-Termkandidaten zu befreien, würde für menschliche TerminologInnen weniger Aufwand bedeuten, als den Eingabekorpus ein zweites Mal manuell nach Benennungen zu durchsuchen, die das Termextraktionssystem möglicherweise übersehen hat. Außerdem sollte der Korpus nach der Dateikonvertierung ins TXT-Format, falls diese erforderlich ist, zuerst manuell möglichst sorgfältig bearbeitet werden, um jegliche umwandlungsbedingte Sonderzeichen zu entfernen bzw. zu korrigieren. Auf diesen Schritt wurde in der durchgeführten Untersuchung verzichtet, was zur Folge hatte, dass zahlreiche Einträge in der Termkandidatenliste verworfen werden mussten, insbesondere im Fall von AntConc, da sie von fehlerhafter Zeichenkodierung betroffen waren.

Im Zuge der Untersuchung wurde darüber hinaus verdeutlicht, dass die automatische Implementierung der alleinigen syntaktischen Analyse von Eingabetexten nicht ausreicht, um absolut zu bestimmen, ob eine Phrase als fachspezifisch gilt oder nicht. Dies führt zur Erkenntnis, dass der semantische Aspekt bei der Analyse von Fachtexten zur effizienten Identifizierung von Termkandidaten nach wie vor eine entscheidende Rolle spielt, weshalb der Prozess aktuell von dem (extra-)linguistischen Wissen bzw. der Intuition der überprüfenden Fachperson abhängt. Schließlich zeigte sich anhand der Extraktionsergebnisse, dass die Verwendung der Antragsformulare als ausschließliche Quelle für die vorgesehene terminologische Auseinandersetzung nicht geeignet war. Dies lässt sich im Wesentlichen auf die Hauptfunktion der Textsorte zurückführen, die vielmehr zur Informationserhebung als zur -vermittlung dient und deshalb keinen umfangreichen und vielfältigen Textinhalt aufweist, der auf signifikante Weise zur Gewinnung domänenspezifischer Terminologie des entsprechenden Fachbereichs beitragen könnte. So könnte ein Korpus, der sowohl Formulare als auch weitere

informationsreiche Verwaltungstexte aus demselben Fachbereich beinhaltet, die Chance erhöhen, dass bei der automatisierten Terminologieextraktion aussagekräftigere Ergebnisse erzielt werden. Allerdings wäre bei der Gestaltung solch eines umfassenden Korpus maschinelle Unterstützung nötig. Beispielsweise wurde in der Studie von Longinova et al. (2012) ein *Webcrawler* eingesetzt, um domänenspezifische Texte anhand der relevanten Suchbegriffe aus dem Internet automatisch zu sammeln. Die Erstellung eines Goldstandards, d. h. einer RTL, sollte jedoch weiterhin händisch erfolgen bzw. durch Fachleute verifiziert werden, um qualitativ hochwertige terminologische Weiterverarbeitung zu gewährleisten.

8 Fazit

Die vorliegende Masterarbeit soll grundsätzlich den Horizont hinsichtlich des sprachlichen Handelns in Verwaltungsverfahren erweitern: Es besteht im Hinblick auf die bürgernahe Kommunikation bei österreichischen öffentlichen Behörden immer noch die Notwendigkeit, bei der Gestaltung verständlicher Verwaltungsschreiben nicht nur ErstsprachlerInnen, sondern auch nichtdeutschsprachige RezipientInnen zu berücksichtigen, vor allem in Verwaltungsbereichen, die im Wesentlichen ausländische ZuwanderInnen betreffen. Um die Bereitstellung einheitlicher Informationsressourcen und relevanter Verwaltungsdokumente in weitere nötige Alternativsprachen zu ermöglichen, muss sich zunächst sowohl mit sprachlichen als auch terminologischen Fragen im Zusammenhang mit der österreichischen Verwaltungssprache und ihren Spracheigenheiten auseinandergesetzt werden.

So besteht die praktische Untersuchung grundsätzlich darin, die Verwaltungssprache des Bereichs Niederlassungs- und Aufenthaltsberechtigung im Zusammenhang mit der gegenwärtigen Terminologiearbeit zu erforschen. Zu diesem Zweck wurden Formulare als Untersuchungsgegenstand ausgewählt, da diese Textsorte überwiegend als erste Schnittstelle zwischen BürgerInnen und Behörden fungiert und die typischste Erscheinungsform der Verwaltungssprache ist. Bei dem zu analysierenden Korpus handelt es sich um verschiedene Antragsformulare zum Erwerb der Aufenthaltsberechtigung, die aus den offiziellen Websites der zuständigen Behörden gesammelt wurden. Konkret soll anhand dieses Korpus die Leistung von Terminologieextraktionssystemen ausgewertet werden, die jeweils ein unterschiedliches Extraktionsverfahren eingesetzt haben, um die fachspezifische österreichische Verwaltungsterminologie im Bereich Niederlassung und Aufenthalt automatisch zu extrahieren. Nach dem Extraktionsprozess kam das Evaluierungstool Term Evaluator zum Einsatz, das entwickelt und öffentlich zugänglich gemacht wurde, um die Validierung der automatisch extrahierten Termkandidaten zu erleichtern und zu beschleunigen.

Der maschinengestützte Annotationsprozess in dieser Studie konnte insbesondere von der Funktion Annotationsspeicher profitieren, der die Bewertung von Wörtern anzeigen lässt, die bezüglich ihrer Korrektheit und ihrer terminologischen Relevanz zuvor bereits annotiert wurden. Die Referenztermliste, die im Vorfeld der automatisierten Extraktion manuell erstellt und menschlich verifiziert wurde, diente als ein Goldstandard für die maschinengestützte Annotation. Die Untersuchung soll grundsätzlich den leistungsfähigsten Ansatz zur maschinengestützten Extraktion ermitteln, der nicht nur mit den spezifischen Spracheigenheiten der österreichische Nationalvarietät effizient umgehen kann, sondern auch

mit denen der österreichischen Verwaltungssprache. Im Wesentlichen wurden die einzelnen Werkzeuge hinsichtlich ihres Ist-Zustandes analysiert, wobei ihre jeweilige innere Funktionsweise nicht manipuliert wurde. Der Schwerpunkt der Evaluierung umfasst (jedoch nicht ausschließlich) nominale terminologische Einheiten, die in der österreichischen Verwaltungssprache im entsprechenden Fachgebiet vorkommen.

Zusammenfassend ließ sich anhand der quantitativen Evaluierung feststellen, dass TermSuite, das einen hybriden Ansatz zur Terminologiegewinnung verwendet, eine bessere Gesamtleistung bei der Extraktion von Termkandidaten aus dem Korpus elektronischer Antragsformulare erbrachte. Genauer hat das Tool bei der Extraktion von Mehrwortbenennungskandidaten (mit zwei bis fünf lexikalischen Einheiten) am besten abgeschnitten, während AntConc und TWSC selbst bei der Verarbeitung von Termkandidaten mit einer lexikalischen Einheit kaum mithalten konnten. Darüber hinaus wurde ermittelt, dass die Noise-Effekte in der ausgegebenen Termkandidatenliste des jeweiligen Extraktionswerkzeuges zum einen durch die nichtterminologischen Kombinationen und zum anderen durch fehlerhafte syntaktische Segmentierung verursacht wurden. Dies hat dazu beigetragen, dass terminologisch irrelevante Elemente als Teil gültiger Termkandidaten erfasst wurden. Hingegen ließen sich keine konkrete Aussagen über die Gründe für das Auftreten der Silence-Effekte ermitteln, vielmehr wurden fundierte Vermutungen formuliert. Dies traf insbesondere auf TWSC und TermSuite zu, da es an umfassenden Informationen bezüglich der intrinsischen Funktionen und Algorithmen der beiden automatischen Terminologieextraktionssysteme mangelte.

Aufgrund der geringen Evaluierungspunktzahlen lässt sich, wie bereits in der Schlussfolgerung festgestellt, zusammenfassen, dass eine automatisierte Terminologieextraktion ohne jegliche Vorbearbeitung für diesen Fachbereich derzeit mit den ausgewählten Hilfsmitteln nicht empfehlenswert ist, da diese sowohl hinsichtlich der Genauigkeit als auch der Trefferquote nahezu unbrauchbare Extraktionsresultate liefert. Angesichts der gesamten Untersuchungsergebnisse besteht beträchtlicher Spielraum für Verbesserungen der maschinengestützten bzw. automatisierten Terminologieextraktion im Hinblick auf domänenspezifische Korpora, vor allem diejenigen, die sich durch die linguistischen Besonderheiten der österreichischen Sprachvarietät auszeichnen. Trotz der gegenwärtigen Einschränkungen, die im Zuge der ausgeführten praktischen Untersuchung dargelegt wurden, soll auf Basis des erlangten Ergebnisses – insbesondere der erstellten Referenztermliste – ein Beitrag zu den kontinuierlichen Studien zur (semi-)automatisierten Terminologieverarbeitung von Fachsprachen verschiedener Gebiete geleistet werden. Überdies

sollen die erworbenen Erkenntnisse als Anwendungsgrundlage auch für erweiterte Bemühungen hinsichtlich der Konzipierung für AdressatInnen möglichst verständlicher Verwaltungsschreiben dienen, damit die von der öffentlichen Verwaltung angestrebte BürgerInnennähe und die Kommunikationstransparenz im Verwaltungshandeln realisiert werden können.

Insbesondere durch die Erkenntnis, dass eine hybride Extraktionsmethode im Vergleich zu rein statistischen und linguistischen Ansätzen bei der Verarbeitung fachspezifischer Korpora bessere Ergebnisse liefert, lässt sich ein Weg für ihre potenziellen Weiteroptimierungen ebnen. Empfehlenswert für künftige Forschungsagenda in dieser Hinsicht wäre eine Bereitstellung eines umfassenden domänenspezifischen Referenzdatensatzes, der mit dem im Rahmen des Projekts *TermEval 2020* erstellten *ACTER* vergleichbar ist (siehe dazu Kapitel 6). Dementsprechend sollten in diesem Datensatz manuell annotierte Fachkorpora aus verschiedenen österreichischen Verwaltungs- und Rechtsbereichen enthalten sein, die als Goldstandard für die Evaluierung unterschiedlicher Methoden zur automatischen Terminologiegewinnung bzw. als linguistische Ressourcen für die Erweiterung der bestehenden Systemkonfigurationen dienen. Zusätzlich könnten diese Korpora und der Annotationsprozess um die weiteren (National-)Varietäten der deutschen Standardsprache ergänzt werden, um die Unterscheidung zwischen den terminologischen Varianten und deren Identifizierung zu ermöglichen.

9 Quellenverzeichnis

- Ahmad, Khurshid & Margaret Rogers (2001). Corpus Linguistics and Terminology Extraction. In: Wright, Sue Ellen & Budin, Gerhard (Hg.) Handbook of Terminology Management Vol. 2: Application-Oriented Terminology Management. Amsterdam: John Benjamins, 725-760.
- Albrecht, Richard (1980). Über Differenzen zwischen der „Alltagswelt“, der „Medienrealität“, der „Verrechtlichten Realität“ und der „Verwaltungswelt“. In: Grosse, Siegfried & Mentrup, Wolfgang (Hg.) Bürger-Formulare-Behörde. Wissenschaftliche Arbeitstagung zum Kommunikationsmittel ‚Formular‘. Tübingen: Narr, 76-95.
- Ammon, Ulrich (1994). Über ein fehlendes Wörterbuch "Wie sagt man in Deutschland?" und den übersehenen Wörterbuchtyp „Nationale Varianten einer Sprache“. In: Deutsche Sprache 22 (1). 51-65.
- Ammon, Ulrich (1995a). Die deutsche Sprache in Deutschland, Österreich und der Schweiz. Berlin/New York: de Gruyter.
- Ammon, Ulrich (1995b). Vorschläge zur Typologie nationaler Zentren und nationaler Varianten bei plurinationalen Sprachen - am Beispiel des Deutschen. In: Muhr, Rudolf, Schrod, Richard & Wiesinger, Peter (Hg.) Österreichisches Deutsch: Linguistische, sozialpsychologische und sprachpolitische Aspekte einer nationalen Variante des Deutschen, 2. Band. Wien: Holder-Pichler-Tempsky, 111-120.
- Ammon, Ulrich (2005). Standard und Variation: Norm, Autorität, Legitimation. In: Eichinger, Ludwig M. & Kallmeyer, Werner (Hg.) Standardvariation. Wie viel Variation verträgt die deutsche Sprache? Berlin/New York: de Gruyter, 28-40.
- Ammon, Ulrich, Bickel, Hans, Ebner, Jakob, Esterhammer, Ruth, Gasser, Markus, Hofer, Lorenz, Kellermeier-Rehbein, Birte, Löffler, Heinrich, Mangott, Doris, Moser, Hans, Schläpfer, Robert, Schloßmacher, Michael, Schmidlin & Vallaster, Günter (2011). Variantenwörterbuch des Deutschen. Die Standardsprache in Österreich, der Schweiz und Deutschland sowie in Liechtenstein, Luxemburg, Ostbelgien und Südtirol. Berlin: De Gruyter.
- Anthony, Laurence (2013). Developing AntConc for a new generation of corpus linguists. Proceedings of the Corpus Linguistics Conference (CL 2013), 14-16.

- Anthony, Laurence (2020). AntConc (Version 3.5.9) [Computer Software]. Tokyo: Waseda University. <https://www.laurenceanthony.net/software> (Stand: 15.01.2021).
- Arntz, Reiner, Picht, Heribert & Schmitz, Klaus-Dirk (2014). Einführung in die Terminologearbeit. 7. Aufl. Hildesheim: Olms Verlag.
- Becker-Mrotzek, Michael (1999). Die Sprache der Verwaltung als Institutionensprache. In: Hoffmann, Lothar, Kalverkämper, Hartwig & Wiegand, Ernst (Hg.) Fachsprachen, Handbücher zur Sprach- und Kommunikationswissenschaft 14.2. Berlin: Walter de Gruyter, 1391-1402.
- Becker-Mrotzek, Michael & Scherner, Maximilian (2000). Textsorten in der Verwaltung. In: Brinker, Klaus, Antos, Gerd, Heinemann, Wolfgang & Sager, Sven F. (Hg.) Text- und Gesprächslinguistik. Ein internationales Handbuch zeitgenössischer Forschung, 1. Halbband. Berlin: Walter de Gruyter, 628-641.
- Becker-Mrotzek, Michael (2001). Gespräche in Ämtern und Behörden. In: Brinker, Klaus. (Hg.) Text- und Gesprächslinguistik. Ein internationales Handbuch zeitgenössischer Forschung, 2. Halbband. Berlin: Walter de Gruyter, 1505–1525.
- Benrath, Daniel (2011). Die Verständlichkeit des Verwaltungsakts. Ein Beitrag zur wissenschaftlichen Diskussion der Verwaltungssprache. Verwaltungsarchiv. Zeitschrift für Verwaltungslehre, Verwaltungsrecht und Verwaltungspolitik 102, 547-569.
- Bickel, Hans & Landolt, Christoph (2012). Schweizerhochdeutsch. Wörterbuch der Standardsprache in der deutschen Schweiz. Zürich: Duden.
- Blancafort, Helena, Daille, Béatrice, Gornostay, Tatiana, Heid, Ulrich, Méchoulam, Claude & Sharoff, Serge (2010). TTC: Terminology Extraction, Translation Tools and Comparable Corpora. 14th EURALEX International Congress, Leeuwarden/Ljouwert, Netherlands. 263-268.
- Blasge, Ingrid (2014) Flüchtling, Gastarbeiter, Mensch mit „Migrationshintergrund“. Österreichische Zuwanderungsgeschichte seit 1955 aus identitätspolitischer Perspektive. In: Strobl, Philipp (Hg.) Österreich in der Zweiten Republik. Ein Land im Wandel. Hamburg: Verlag Dr. Kovac, 117-132.

- Brinckmann, Hans, Grimmer, Klaus, Höhmann, Anne, Kuhlmann, Stefan & Schäfer, Wolfgang (1986). *Formulare im Verwaltungsverfahren: Wegbereiter standardisierter Kommunikation* (Beiträge zur juristischen Informatik, 12). Darmstadt: Toeche-Mittler Verlag.
- Bundesministerium für Europäische und internationale Angelegenheiten (o. J.). Aufenthalt. <https://www.bmeia.gv.at/reise-aufenthalt/einreise-und-aufenthalt-in-oesterreich/einreise-und-aufenthalt/aufenthalt/> (Stand: 10.11.2020).
- Bundesministerium für Inneres (o. J.). Aufenthaltsverfahren für Drittstaatsangehörige. https://bmi.gv.at/312/08/start.aspx#pk_03 (Stand: 10.11.2020)
- Bundesministerium für Inneres (o. J.). Niederlassung in Österreich. <https://bmi.gv.at/312/start.aspx> (Stand: 10.11.2020)
- Burmeister, Stefan (2012). Homo migrans: Migration und die plurale Gesellschaft, eine Herausforderung für die archäologischen Museen. *Museumskunde* 77(2). 30-37.
- Bußmann, Hadumod (Hg.) (2002). *Lexikon der Sprachwissenschaft*. 3., aktualisierte und erw. Aufl. Stuttgart: Kröner.
- Cabré, María Teresa (1999). *Terminology: Theory, methods and applications*. Amsterdam/Philadelphia: John Benjamins.
- Cabré, María Teresa, Bagot, Rosa Estopá & Palatresi, Jordi Vivaldi (2001). Automatic term detection: A review of current systems. In Bourigault, Didier, Jacquemin, Christian & L'Homme, Marie-Claude (Hg.) *Recent Advances in Computational Terminology*. Amsterdam: John Benjamins, 53-88.
- Chiocchetti, Elena, Ralli, Natascia, Lušicky, Vesna & Wissik, Tanja (2013a). Spanning bridges between theory and practice: terminology workflow in the legal and administrative domain. *Comparative Legilinguistics-International Journal of Legal Communication* 16, 7-22.
- Chiocchetti, Elena, Heinisch-Obermoser, Barbara, Löckinger, Georg, Lušicky, Vesna, Ralli, Natascia, Stanizzi, Isabella & Wissik, Tanja (2013b). *Guidelines for collaborative legal/administrative terminology work*. Bolzano: EURAC.
- Clyne, Michael (1984). *Language and society in the German-speaking countries*. Cambridge: Cambridge University Press, 21-23.

- Clyne, Michael (1988). A Tendenzwende in the codification of Austrian German. *Multilingua* 7(3), 335-342.
- Clyne, Michael (Hg.) (1992). German as a pluricentric language. In: *Pluricentric languages: Different norms in different nations*. Berlin: De Gruyter Mouton, 117-147.
- Clyne, Michael (1993). Die österreichische Nationalvarietät des Deutschen im wandelnden internationalen Kontext. In: Muhr, Rudolf (Hg.) *Internationale Arbeiten zum österreichischen Deutsch und seinen nachbarsprachlichen Bezügen*. Wien: Verlag Holder-Pichler-Tempsky, 1-6.
- Clyne, Michael (1995). *The German language in a changing Europe*. Cambridge University Press.
- Cohen, Jacob (1960). A coefficient of agreement for nominal scales. *Educational and psychological measurement* 20(1), 37-46.
- Condamines, Anne (2010). Variations in Terminology: Application to the Management of Risks Related to Language Use in the Workplace. *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication* 16(1), 30-50.
- Daille, Beatrice (1994). Combined approach for terminology extraction: lexical statistics and linguistic filtering. PhD Dissertation, University Paris.
- Daille, Beatrice (1996). Study and implementation of combined techniques for automatic extraction of terminology. In Klavans, Judith L. & Resnik, Philip (Hg.) *The balancing act: Combining symbolic and statistical approaches to language*. Massachusetts: MIT Press, 49-66.
- Daille, Beatrice (2005). Variations and application-oriented terminology engineering. *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication* 11(1), 181-197.
- Daille, Beatrice (2012). Building bilingual terminologies from comparable corpora: The TTC TermSuite. *The 5th Workshop on Building and Using Comparable Corpora (LREC 2012)*, 29-32.

- Daum, Ulrich (1981). Rechtssprache–eine genormte Fachsprache. In: Radtke, Ingulf (Hg.) Der öffentliche Sprachgebrauch Bd. II. Deutsche Akademie für Sprache und Dichtung. Die Sprache des Rechts und der Verwaltung. Stuttgart: Klett-Cotta, 83-99.
- de Cilia, Rudolf (1995). Deutsche Sprache und österreichische Identität. Medien Impulse 14/1995, 4-13.
- de Groot, Gerard-René (1999). Zweisprachige juristische Wörterbücher. In: Sandrini, Peter (Hg.) Übersetzen von Rechtstexten. Fachkommunikation im Spannungsfeld zwischen Rechtsordnung und Sprache. Tübingen: Narr, 203-227.
- Diederich, Georg (1980). Das Kommunikationsmittel „Formular“. In: Grosse, S./Mentrup, W. (eds.). In: Grosse, Siegfried & Mentrup, Wolfgang (Hg.) Bürger–Formulare–Behörde. Wissenschaftliche Arbeitstagung zum Kommunikationsmittel ‚Formular‘. Tübingen: Narr, 96-111.
- DIN 2330 (2013): Begriffe und Benennungen: Allgemeine Grundsätze. Berlin: Beuth.
- DIN 2342 (2011): Begriffe der Terminologielehre. Berlin: Beuth.
- Dollinger, Stefan (2019). The pluricentricity debate: On Austrian German and other Germanic standard varieties. New York: Routledge.
- Dressler, Wolfgang & Wodak, Ruth (1983). Soziolinguistische Überlegungen zum österreichischen Wörterbuch. In: Dardano, Maurizio, Dressler, Wolfgang & Held, Gudrun (Hg.) Parallela 1. Akten des 2. österreichisch-italienischen Linguistentreffens. Tübingen: Narr, 247-260.
- Drewer, Petra & Schmitz, Klaus-Dirk (2017). Terminologiemanagement: Grundlagen - Methoden - Werkzeuge. Berlin: Springer Vieweg.
- Drewer, Petra & Siegel, Melanie (2012). Terminologieextraktion–multilingual, semantisch und mehrfach verwendbar. Tagungsband zur tekom-Frühjahrstagung 2012, 35-38.
- Duden (2020). Kanzleisprache. <https://www.duden.de/rechtschreibung/Kanzleisprache> (Stand: 27.11.2020).
- Earl, Lois L. (1970). Experiments in automatic extracting and indexing. Information Storage and Retrieval 6(4), 313-330.

- Ebert, Hermann (2011). Verwaltungssprache aus Sicht der Sprachwissenschaft. In: Blaha, Michaela Blaha & Wilhelm, Hermann (Hg.) Verständliche Sprache in Recht und Verwaltung – Herausforderungen und Chancen. Frankfurt, 13-23.
- Ebner, Jakob (1980). Duden – Wie sagt man in Österreich? Wörterbuch der österreichischen Besonderheiten. 2. vollständig überarbeitete Aufl. Mannheim/Wien/Zürich: Bibliographisches Institut.
- Ebner, Jakob (2009). Duden – Wie sagt man in Österreich? Wörterbuch des österreichischen Deutsch. 4. völlig überarbeitete Auflage. Mannheim/Wien: Dudenverlag.
- Efing, Christian (2020). Verwaltungssprache und Kommunikation mit Verwaltungsinstitutionen. In: Niehr, Thomas, Kilian, Jörg & Schiewe, Jürgen (Hg.) Handbuch Sprachkritik. Stuttgart: J.B. Metzler, 287-294.
- Ehlich, Konrad & Rehbein, Jochen (1980). Sprache in Institutionen. In: Althaus, Hans Peter, Henne, Helmut & Wiegand, Herbert Ernst (Hg.) Lexikon der Germanistischen Linguistik. Studienausgabe II. 2. Auflage. Tübingen: Max Niemeyer Verlag, 338-345.
- Europäisches Migrationsnetzwerk (2018). Glossar zu Asyl und Migration Version 5.0 – Ein Instrument zur besseren Vergleichbarkeit: https://www.emn.at/wp-content/uploads/2018/07/emn-glossar-5-0_de.pdf (Stand: 09.10.2021)
- Fassmann, Heinz & Münz, Rainer (1995). Einwanderungsland Österreich? Historische Migrationsmuster, aktuelle Trends und politische Maßnahmen. Wien: Dachs Verlag GmbH.
- Felber, Helmut & Budin, Gerhard (1989). Terminologie in Theorie und Praxis. Tübingen: Narr.
- Fisch, Rudolf & Margies, Burkhard (Hg.) (2014). Wozu eine bessere Verwaltungssprache? In: Bessere Verwaltungssprache. Grundlagen, Empirie, Handlungsmöglichkeiten. Berlin: Duncker & Humblot, 9-16.
- Fleischmann, Klaus (2006). Termextraktion: Ein Beitrag zum Aufbau einer unternehmensweiten Terminologie. Masterthese, Donau-Universität Krems.
- Fluck, Hans-Rüdiger (2008). Verwaltungssprache unter dem Einfluss der Gesetzessprache. In: Eichhoff-Cyrus, Karin M. & Antos, Gerd (Hg.) Verständlichkeit als Bürgerrecht? Die Rechts- und Verwaltungssprache in der öffentlichen Diskussion. Mannheim: Dudenverlag, 117–135.

- Fotheringham, Heinz (1980). Allgemeine Gesichtspunkte des Formulars. Gesetzliche Begriffe im Formular–Erwartungen des Benutzers an das Formular–Ansatz einer Typologie des Formulars. In: Grosse, Siegfried & Mentrup, Wolfgang (Hg.) Bürger–Formulare–Behörde. Wissenschaftliche Arbeitstagung zum Kommunikationsmittel ‚Formular‘. Tübingen: Narr, 25-43.
- Fotheringham, Heinz (1981). Die Gesetzes- und Verwaltungssprache im Spannungsfeld zwischen fachlicher Qualität und Allgemeinverständlichkeit. In: Radtke, Ingulf (Hg.) Der öffentliche Sprachgebrauch Bd. II. Deutsche Akademie für Sprache und Dichtung. Die Sprache des Rechts und der Verwaltung. Stuttgart: Klett-Cotta, 100-118.
- Frantziy, Katerina, Ananiadou, Sophia & Tsujii, Junichi (1998). The C-value/NC-value method of automatic recognition for multi-word terms. In: Nikolaou, Christos & Stephanidis, Constantine (Hg.) Research and Advanced Technology for Digital Libraries, Lecture Notes in Computer Science (1513). Berlin/Heidelberg: Springer Verlag, 585-604.
- Frantziy, Katerina, Ananiadou, Sophia & Mima, Hideki (2000). Automatic recognition of multi-word terms: the c-value/nc-value method. International journal on digital libraries 3(2), 115-130.
- Fuchs-Heinritz, Werner, Lautmann, Rüdiger, Rammstedt, Otthein & Wienold, Hans (Hg.) (1994). Lexikon zur Soziologie. 3. Auflage. Opladen: Westdeutscher Verlag.
- Fuchs-Khakhar, Christine (1987). Die Verwaltungssprache zwischen dem Anspruch auf Fachsprachlichkeit und Verständlichkeit. Tübingen: Stauffenburg Verlag.
- Georgantopoulos, Byron & Piperidis, Stelios (2000). A hybrid technique for automatic term extraction. Proceedings of the ACIDCA 2000 Conference.
- Gojun, Anita, Heid, Ulrich, Weissbach, Bernd, Loth, Carola, & Mingers, Insa (2012). Adapting and evaluating a generic term extraction tool. Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC 2012), 651-656.
- Gornostay, Tatiana & Vasiljevs, Andrejs (2014). Terminology Resources and Terminology Work Benefit from Cloud Services. Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC 2014), 1943-1948.

- Grosse, Siegfried & Mentrup, Wolfgang (Hg.) (1980). Bürger-Formulare-Behörde. Wissenschaftliche Arbeitstagung zum Kommunikationsmittel ‚Formular‘. Tübingen: Narr
- Grosse, Siegfried (1980). Allgemeine Überlegungen zur sprachlichen Fassung von Vordrucken und Formularen. In: Grosse, Siegfried & Mentrup, Wolfgang (Hg.) Bürger-Formulare-Behörde. Wissenschaftliche Arbeitstagung zum Kommunikationsmittel ‚Formular‘. Tübingen: Narr, 11-24.
- Grosse, Siegfried (1981). Vorschläge zur Verbesserung der Verständlichkeit von Verwaltungstexten. In: Radtke, Ingulf (Hg.) Der öffentliche Sprachgebrauch Bd. II. Deutsche Akademie für Sprache und Dichtung. Die Sprache des Rechts und der Verwaltung. Stuttgart: Klett-Cotta, 268-274.
- Gülich, Elisabeth (1981). Formulare als Dialoge. In: Radtke, Ingulf (ed.) Der öffentliche Sprachgebrauch Bd. II. Deutsche Akademie für Sprache und Dichtung. Die Sprache des Rechts und der Verwaltung. Stuttgart: Klett-Cotta, 322-356.
- Günther, Louis & Günther, Ludwig (1898). Recht und Sprache: Ein Beitrag zum Thema vom Juristendeutsch. Berlin: Heymann.
- Hartley, Tony (2009). Technology and translation. In: Munday, Jeremy (Hrsg.) The Routledge companion to translation studies. Routledge, 106-127.
- Heid, Ulrich (1999). Extracting terminologically relevant collocations from German technical texts. 5th International Congress on Terminology and Knowledge Engineering (TKE 99), 241-255.
- Heid, Ulrich (2004). Computergestützte Lexikographie und Terminologie. In: Carstensen, Kai-Uwe, Ebert, Christian, Endriss, Cornelia, Jekat, Susanne, Klabunde, Ralf & Langer, Hagen (Hg.) Computerlinguistik und Sprachtechnologie. Heidelberg/Berlin: Spektrum Akademischer Verlag, 471-478.
- Heid, Ulrich (2006). Extracting term candidates from recursively chunked text. In: Hacken, Pius ten (Hg.) Terminology, Computing and Translation. Tübingen: Narr, 97-115.
- Heinrich, Peter (2014). Verwaltungssprache als Element der Organisationskultur. In: Fisch, Rudolf & Margies, Burkhard (Hg.) Bessere Verwaltungssprache. Grundlagen, Empirie, Handlungsmöglichkeiten. Berlin: Duncker & Humblot, 49-62.

- Helbig, Max (1980). Der Aufbau und die Gestaltung der Vordrucke. In: Grosse, Siegfried & Mentrup, Wolfgang (Hg.) Bürger-Formulare-Behörde. Wissenschaftliche Arbeitstagung zum Kommunikationsmittel ‚Formular‘. Tübingen: Narr, 44-75.
- Hoffmann, Lothar (1985). Kommunikationsmittel Fachsprache. Eine Einführung. 2. Aufl. Tübingen: Narr.
- Hornung, Maria & Roitinger, Franz (1950). Unsere Mundarten. Eine dialektkundliche Wanderung durch Österreich. Wien: Österreichischer Bundesverlag.
- Inkpen, Diana, Paribakht, T. Sima, Faez, Farahnaz & Amjadian, Ehsan (2016). Term evaluator: A tool for terminology annotation and evaluation. *International Journal of Computational Linguistics and Applications* 7(2), 145-165.
- Justeson, John S. & Katz, Slava M. (1995). Technical terminology: some linguistic properties and an algorithm for identification in text. *Natural language engineering* (1)1, 9-27.
- Kageura, Kyo & Umino, Bin (1996). Methods of automatic term recognition: A review. *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication* 3(2), 259-289.
- Kageura, Kyo, Yoshioka, Masaharu, Takeuchi, Koichi, Koyama, Teruo, Tsuji, Keita & Fuyuki, Yoshikane (2001). Recent advances in automatic term recognition: Experiences from the NTCIR workshop on information retrieval and term recognition. *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication* 6(2), 151-173.
- Kirchhoff, Paul (1987). Die Bestimmtheit und Offenheit der Rechtssprache. Berlin: Walter de Gruyter.
- Klein, Josef (2014). Verwaltungssprache – Sprachästhetik und Funktionalität. In: Fisch, Rudolf & Margies, Burkhard (Hg.) *Bessere Verwaltungssprache. Grundlagen, Empirie, Handlungsmöglichkeiten*. Berlin: Duncker & Humblot, 17–32.
- Kloss, Heinz (1978). Die Entwicklung neuer germanischer Kultursprachen seit 1800. 2. erw. Aufl. Düsseldorf: Schwann.
- Krumm, Hans-Jürgen, Fandrych, Christian, Hufeisen, Britta & Riemer, Claudia (Hg.) (2010). *Deutsch als Fremd- und Zweitsprache: ein internationales Handbuch*. 2. Halbband. Berlin/New York: de Gruyter Mouton.

- KÜDES (Hg.) (2002). Empfehlungen für die Terminologearbeit. Konferenz der Übersetzungsdienste europäischer Staaten. Arbeitsgruppe Terminologie und Dokumentation. Bern: Schweizerische Bundeskanzlei.
- Kühn, Peter (1995). Mehrfachadressierung, Untersuchungen zur adressatenspezifischen Polyvalenz sprachlichen Handelns. Reihe Germanistische Linguistik Band 154. Berlin: Walter de Gruyter.
- Lambertz, Thomas (1999). Medium Sprache in der öffentlichen Verwaltung. In: Becker-Mrotzek, Michael & Doppler, Christine (Hg.) Medium Sprache im Beruf. Eine Aufgabe für die Linguistik. Tübingen: Guten Narr Verlag, 143–160.
- Lauriston, Andy (1994). Automatic recognition of complex terms: Problems and the TERMINO solution. Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication 1(1), 147-170.
- Lerch, Kent D. (2014). Vor dem Türhüter des Gesetzes. Verwaltungssprache und Rechtsgewährung. In: Fisch, Rudolf Fisch & Margies, Burkhard (Hg.) Bessere Verwaltungssprache. Grundlagen, Empirie, Handlungsmöglichkeiten. Berlin: Duncker & Humblot, 63-98.
- Lieske, Christian (2002). Pragmatische Evaluierung von Werkzeugen für die Term-Extraktion. In: Mayer, Felix, Schmitz, Klaus-Dirk & Zeumer, Jutta (Hg.) eTerminology – Professionelle Terminologie im Zeitalter des Internet. Akten des Symposiums. Köln: Deutscher Terminologie-Tag e.V., 109-131.
- Löffler, Heinrich (2005). Wieviel Variation verträgt die deutsche Standardsprache? Begriffsklärung: Standard und Gegenbegriffe. In: Eichinger, Ludwig M. & Kallmeyer, Werner (Hg.) Standardvariation. Wie viel Variation verträgt die deutsche Sprache? Berlin: Walter De Gruyter, 7-27.
- Love, Stacy (2000). Benchmarking the performance of two automated term extraction systems: LOGOS and ATAIO. Master's Thesis, Université de Montréal.
- Lucassen, Jan, Leo Lucassen & Manning, Patrick (Hg.) (2010). Migration history in world history: Multidisciplinary approaches. Leiden/Boston: Brill.

- Lüdenbach, Wolfgang & Herrlitz, Norbert (1981). Zur Verständlichkeit von Formularen. Ein handlungstheoretischer Versuch. In: Radtke, Ingulf (Hg.) Der öffentliche Sprachgebrauch Bd. II. Deutsche Akademie für Sprache und Dichtung. Die Sprache des Rechts und der Verwaltung. Stuttgart: Klett-Cotta, 305-321.
- Luft, Stefan (2016). Die Flüchtlingskrise. Ursachen, Konflikte, Folgen. München: Verlag C.H. Beck.
- L'Homme, Marie-Claude, Benali, Loubna, Bertrand, Claudine & Lauduique, Patricia (1996). Definition of an evaluation grid for term-extraction software. Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication 3(2), 291-312.
- MA 35 Fachbereich Einwanderung. Glossar zu den Themen Einwanderung, Aufenthalt und Staatsbürgerschaft. <https://www.wien.gv.at/verwaltung/einwanderung/glossar.html> (Stand: 09.01.2021).
- Mahlstedt, Katrin (2010). Evaluation von Terminologieextraktionsprogrammen: Ein methodologischer Überblick. Masterarbeit, Universität Wien.
- Markhardt, Heidemarie (2005). Das Österreichische Deutsch im Rahmen der EU. Frankfurt/Main: Lang.
- Maynard, Diana Gabrielle (2000). Term recognition using combined knowledge sources. PhD Dissertation. Manchester Metropolitan University.
- Maynard, Diana, Yaoyong Li & Wim Peters (2008). NLP Techniques for Term Extraction and Ontology Population. In: Buitelaar, Paul & Cimiano, Philipp (Hg.) Ontology Learning and Population: Bridging the Gap Between Text and Knowledge. Amsterdam: IOS Press, 107-127.
- Mecheril, Paul, Varela, Maria do Mar Castro, Dirim, Inci, Kalpaka, Annita & Melter, Claus (2010). Migrationspädagogik. Weinham und Basel: Beltz Verlag.
- Mecheril, Paul (2007). Migration Und Integration. In: Straub, Jürgen, Arne, Weidemann & Doris Weidemann (Hg.) Handbuch interkulturelle Kommunikation und Kompetenz. Grundbegriffe - Theorien - Anwendungsfelder. Stuttgart und Weimar: Metzler, 469-480.

- Mentrup, Wolfgang (1980a). Deutsche Sprache in Österreich und in der Schweiz. In: Althaus, Hans Peter, Henne, Helmut & Wiegand, Herbert Ernst (Hg.) Lexikon der germanistischen Linguistik. 2. völlig neu bearb. und erw. Auflage. Tübingen: Niemeyer, 527-531.
- Mentrup, Wolfgang (Hg.) (1980b). Zur Entwicklung der Groß- und Kleinschreibung im Deutschen. Gebrauchsnorm – Kodifizierung – Normierung – Einheitlichkeit – Reform. In: Materialien zur historischen Entwicklung der Groß- und Kleinschreibungsregeln. Reihe Germanistische Linguistik Band 23. Tübingen: Niemeyer, 279-333.
- Migrationsplattform der österreichischen Bundesregierung. Formen der Zuwanderung.
<https://www.migration.gv.at/de/formen-der-zuwanderung/> (Stand: 12.11.2020).
- Migrationsplattform der österreichischen Bundesregierung. Integrationsvereinbarung.
<https://www.migration.gv.at/de/leben-und-arbeiten-in-oesterreich/rahmenbedingungen-der-integration/integrationsvereinbarung/> (Stand: 17.11.2020).
- Migrationsplattform der österreichischen Bundesregierung. Mobilität innerhalb der EU.
<https://www.migration.gv.at/de/formen-der-zuwanderung/mobilitaet-innerhalb-der-eu/>
 (Stand: 12.11.2020).
- Moosmüller, Sylvia (1991). Hochsprache und Dialekt in Österreich. soziophonologische Untersuchungen zu ihrer Abgrenzung in Wien. Graz, Salzburg und Innsbruck. Wien/Köln/Weimar: Böhlau.
- Moser, Hans (1989). Österreichische Aussprachenormen–Eine Gefahr für die sprachliche Einheit des Deutschen? Jahrbuch für Internationale Germanistik 21(1), 8-25.
- Muhr, Rudolf (Hg.) (1993). Österreichisch – Bundesdeutsch – Schweizerisch: Zur Didaktik des Deutschen als plurizentrische Sprache. In: Internationale Arbeiten zum österreichischen Deutsch und seinen nachbarsprachlichen Bezügen. Wien: Verlag Holder-Pichler-Tempsky, 108-123.
- Muhr, Rudolf (1995). Grammatische und pragmatische Merkmale des österreichischen Deutsch. In: Muhr, Rudolf, Schrod, Richard & Wiesinger, Peter (Hg.) Österreichisches Deutsch. Linguistische, sozialpsychologische und sprachpolitische Aspekte einer nationalen Variante des Deutschen. Wien: Verlag Holder-Pichler-Tempsky, 208-234.

- Muhr, Rudolf (1997). Zur Terminologie und Methode der Beschreibung plurizentrischer Sprachen und deren Varietäten am Beispiel des Deutschen. In: Muhr, Rudolf & Schrod, Richard (Hg.) Österreichisches Deutsch und andere nationale Varietäten plurizentrischer Sprachen in Europa. Empirische Analysen. Wien: Verlag Holder-Pichler-Tempsky, 40-67.
- Muhr, Rudolf (2000). Österreichisches Sprachdiplom Deutsch: Lernzielkataloge zu Basisformulierungen, Lexik-Sprechhandlungen, Höflichkeitskonventionen, Diskurs und Diskursstrukturen, Deutsch als plurizentrische Sprache. Wien: öbv & hpt.
- Nakagawa, Hiroshi (2001). Experimental evaluation of ranking and selection methods in term extraction. In: Bourigault, Didier, Jacquemin, Christian & L'Homme, Marie-Claude (Hg.) Recent Advances in Computational Terminology. Amsterdam: John Benjamins, 303-325.
- Neumayer, Eric (2004). Asylum destination choice. What makes some Western European countries more attractive than others? London: LSE Research Online.
- Nuscheler, Franz (2004). Internationale Migration. Flucht und Asyl. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Oberndörfer, Peter (2001). Die Verwaltung im politisch-gesellschaftlichen Umfeld. In: Holzinger, Gerhard, Oberndörfer, Peter & Raschauer, Bernhard (Hg.) Österreichische Verwaltungslehre. Wien: Verlag Österreich, 1-80.
- Ogden, Charles K. & Richards, Ivor A. (1974). Die Bedeutung der Bedeutung – The Meaning of Meaning. Eine Untersuchung über den Einfluß der Sprache auf das Denken und die Wissenschaft des Symbolismus. Frankfurt/Main: Suhrkamp.
- Öhlinger, Theo (Hg.) (1986). Recht und Sprache. Fritz Schönherr – Gedächtnissymposium. Wien: Manzsche Verlags- und Universitätsbuchhandlung.
- Österreichische Integrationsfonds (2014). Integrationsglossar.
<https://www.integrationsfonds.at/publikationen/glossare/integrationsglossar>
 (Stand: 09.01.2021).
- Österreichisches Wörterbuch. (1997). Hg. im Auftrag des Bundesministeriums für Bildung, Wissenschaft und Kultur, 38. Aufl. Neubearbeitung. Wien.
- Österreichs digitales Amt (2020). Aufenthaltstitel für Österreich.
https://www.oesterreich.gv.at/themen/leben_in_oesterreich/aufenthalt/3/2.html

(Stand: 01.11.2020).

Österreichs digitales Amt (2020). Begriffslexikon. <https://www.oesterreich.gv.at/lexicon.html>
(Stand: 09.01.2021).

Ott, Herbert (1983). Demoskopische Untersuchung. In: Niederösterreichische Landesregierung (Hg.) Bürgernahe Gesetzestexte in Niederösterreich. Wien, 61-66.

Otto, Walter (1981). Die Paradoxie einer Fachsprache. In: Radtke, Ingulf (Hg.) Der öffentliche Sprachgebrauch Bd. II. Deutsche Akademie für Sprache und Dichtung. Die Sprache des Rechts und der Verwaltung. Stuttgart: Klett-Cotta, 44-57.

Pazienza, Maria Teresa, Pennacchiotti, Marco & Zanzotto, Fabio Massimo (2005). Terminology extraction: an analysis of linguistic and statistical approaches. In: Sirmakessis, Spiros (Hg.) Knowledge mining. Studies in Fuzziness and Soft Computing. Berlin/Heidelberg: Springer, 255-279.

Perchinig, Bernhard (2009). Von der Fremdarbeit zur Integration? (Arbeits)migrations- und Integrationspolitik in der Zweiten Republik. Österreich in Geschichte und Literatur mit Geographie 3(53), 228-246.

Pinnis, Marcis, Ljubešić, Nikola, Stefanescu, Dan, Skadina, Inguna, Tadic, Marko, & Gornostay, Tatiana (2012a, Juni). Term extraction, tagging, and mapping tools for under-resourced languages. Proceedings of the 10th Conference on Terminology and Knowledge Engineering (TKE 2012), 193-208.

Pinnis, Marcis, Ion, Radu, Stefanescu, Dan, Su, Fangzhong, Skadiņa, Inguna, Vasiljevs, Andrejs, & Babych, Bogdan (2012b, Juli). Accurat toolkit for multi-level alignment and information extraction from comparable corpora. Proceedings of the ACL 2012 System Demonstrations, 91-96).

Pohl, Heinz Dieter (1997). Gedanken zum Österreichischen Deutsch (als Teil der “pluriarealen” deutschen Sprache). In: Muhr, Rudolf & Schrod, Richard (Hg.) Österreichisches Deutsch und andere nationale Varietäten plurizentrischer Sprachen in Europa. Empirische Analysen. Wien: Verlag Holder-Pichler-Tempsky, 67-88.

Polenz, Peter (1996). Österreichisches, schweizerisches, deutschländisches und teutonisches Deutsch. Zeitschrift für germanistische Linguistik 24(2), 205-220.

- Polenz, Peter (2009). *Geschichte der deutschen Sprache*. 10. Aufl. Berlin: Walter de Gruyter.
- Portes, Alejandro & Böröcz, József (1989). Contemporary immigration: Theoretical perspectives on its determinants and modes of incorporation. *International migration review* 23(3), 606-630.
- Reiffenstein, Ingo (1977). Sprachebenen und Sprachwandel im österreichischen Deutsch der Gegenwart. In: Kolb, Herbert & Lauffer, Hartmut (Hg.) *Sprachliche Interferenz. Festschrift für Werner Betz zum 65. Geburtstag*. Tübingen: Max Niemeyer, 175-183.
- Reiffenstein, Ingo (1983). Deutsch in Österreich. In: Reiffenstein, Ingo, Brandt, Wolfgang & Freudenberg, Rudolf (Hg.) *Tendenzen, Formen und Strukturen der deutschen Standardsprache nach 1945*. Marburg: Elwert Verlag, 15-27.
- Rat für Deutschsprachige Terminologie (RaDT) (2013). *Terminologisches Basiswissen für Fachleute*. Köln: Deutscher Terminologie-Tag e. V.
- Rehbein, Jochen (1999). Die Verwendung von Institutionensprache in Ämtern und Behörden. In: Hoffman, Lothar, Kalverkämper, Hartwig & Ernst Wiegand (Hg.) *Fachsprachen, Handbücher zur Sprach- und Kommunikationswissenschaft* 14(1). Berlin: Walter de Gruyter, 660-675.
- Rigouts Terry, Ayla, Hoste, Veronique, Drouin, Patrick, & Lefever, Els (2020). Termeval 2020: Shared task on automatic term extraction using the annotated corpora for term extraction research (acter) dataset. 6th International Workshop on Computational Terminology (COMPUTERM 2020). European Language Resources Association (ELRA), 85-94.
- Ringel, Leopold & Reischauer, Georg (2019). Transparente Verwaltung. In: Veit, Sylvia, Reichard, Christoph & Wewer, Götrik (Hg.) *Handbuch zur Verwaltungsreform*. Wiesbaden: Springer VS, 225-235.
- Sager, Juan C. (2001). Terminology Compilation: Consequences and Aspects of Automation. In: Wright, Sue & Budin, Gerhard (Hg.) *Handbook of Terminology Management Vol. 2: Application-Oriented Terminology Management*. Amsterdam: John Benjamins, 761-771.
- Sandrini, Peter (1996). Terminologearbeit im Recht: Deskriptiver begriffsorientierter Ansatz vom Standpunkt des Übersetzers. Wien: International Network for Terminology (TermNet).

- Saß, Ricarda (2004). Vergleichende Untersuchung von Terminologie-Extraktions-Tools. Eine computerlinguistische Arbeit mit Englisch und Deutsch. Saarbrücken: Fachrichtung 4.6 - Angewandte Sprachwissenschaft sowie Übersetzen und Dolmetschen - Universität des Saarlandes, (Saarbrücker Studien zu Sprachdatenverarbeitung und Übersetzen, Band 21).
- Saussure, Ferdinand de (1967). Grundfragen der allgemeinen Sprachwissenschaft. Berlin: De Gruyter.
- Sauron, Véronique (2002). Tearing out the terms: evaluating terms extractors. Proceedings of the ASLIB Translating and the Computer 24.
- Schendera, Christian F. G. (2004). Die Verständlichkeit von Rechtstexten. In: Lerch, Kent D. (Hg.) Die Sprache des Rechts: Recht verstehen. Berlin/New York: De Gruyter, 321-373.
- Scheuringer, Hermann (1996). Das Deutsche als pluriareale Sprache: Ein Beitrag gegen staatlich begrenzte Horizonte in der Diskussion um die deutsche Sprache in Österreich. Unterrichtspraxis/Teaching German. Zeitschrift des amerikanischen Deutschlehrerverbandes 29(2), 147-154.
- Scheuringer, Hermann (1997). Sprachvarietäten in Österreich. In: Stickel, Gerhard (Hg.) Varietäten des Deutschen. Regional- und Umgangssprachen. Berlin/New York: de Gruyter, 332-345.
- Schmid, Helmut (1995). TreeTagger – a part-of-speech tagger for many languages.
<https://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/> (Stand: 29.01.2021).
- Schoorl, Jeannette, Heering, Liesbeth, Esveldt, Ingrid, Groenewold, George, van der Erf, Rob, Bosch, Alinda, de Valk, Helga & de Bruijn, Bart (2000). Push and Pull Factors of International Migration: A Comparative Report. Luxembourg: Office for Official Publications of the European Communities.
- Schrodt, Richard (1997). Nationale Varianten, areale Unterschiede und der „Substandard“: An den Quellen des Österreichischen Deutsch. In: Muhr, Rudolf & Schrodt, Richard (Hg.) Österreichisches Deutsch und andere nationale Varietäten plurizentrischer Sprachen in Europa. Empirische Analysen. Wien: Verlag Holder-Pichler-Tempsky, 12-40.
- Seibert, Thomas-Michael (1981). Aktenanalysen. Zur Schriftform juristischer Deutungen. Tübingen: Narr.

- Seidl, Erwin (1962). Römische Rechtsgeschichte und Römisches Zivilprozessrecht. Köln/Berlin/Bonn/München: Carl Heymanns Verlag KG.
- Seifert, Klaus (2013). Prozessmanagement für die öffentliche Verwaltung. Heidelberg: Springer Verlag.
- Sellmann, Gudrun (2014). Der Staat geht auf die Bürger zu: Verwaltungssprache, Bürokratieabbau und Dienstleistungsorientierung. In: Fisch, Rudolf Fisch & Margies, Burkhard (Hg.) Bessere Verwaltungssprache. Grundlagen, Empirie, Handlungsmöglichkeiten. Berlin: Duncker & Humblot, 177-185.
- Statistik Austria (2019). Migration & Integration. Zahlen. Daten. Indikatoren 2019. https://www.bmeia.gv.at/fileadmin/user_upload/Zentrale/Integration/Integrationsbericht_2019/Migration-Integration-2019.pdf (Stand 10.10.2020).
- Statistik Austria (2020a). Bevölkerung 2001-2020 nach Staatsangehörigkeit und Geburtsland. http://www.statistik.at/wcm/idc/idcplg?IdcService=GET_PDF_FILE&RevisionSelectionMethod=LatestReleased&dDocName=031396 (Stand 12.12.2020)
- Statistik Austria (2020b). Bevölkerung am 1.1.2020 nach detaillierter Staatsangehörigkeit und Bundesland. http://www.statistik.at/wcm/idc/idcplg?IdcService=GET_PDF_FILE&RevisionSelectionMethod=LatestReleased&dDocName=064287 (Stand 12.12.2020)
- Stickel, Gerhard (1981). Bei den kommunikativen Bedingungen und dem Sprachgebrauch der Behördenvordrucke nachgefaßt. In: Radtke, Ingulf (Hg.) Der öffentliche Sprachgebrauch Bd. II. Deutsche Akademie für Sprache und Dichtung. Die Sprache des Rechts und der Verwaltung. Stuttgart: Klett-Cotta, 284-304.
- Strehlow, Richard A. (2001). The role of terminology in retrieving information. In: Wright, Sue & Budin, Gerhard (Hg.) Handbook of Terminology Management Vol. 2: Application-Oriented Terminology Management. Amsterdam: John Benjamins, 426-441.
- Stubkjær, Flemming Talbo (1993). Zur Reihenfolge der Verbformen des Schlußfeldes im österreichischen Deutsch. In: Muhr, Rudolf (Hg.) Internationale Arbeiten zum österreichischen Deutsch und seinen nachbarsprachlichen Bezügen. Wien: Verlag Holder-Pichler-Tempsky, 39-52.

- Terminology as a Service (TaaS) (o. J.). About TaaS. <http://www.taas-project.eu/index.php-page=about-taas.htm> (Stand: 12.01.2021).
- Text Encoding Initiative (o. J.). Multilingual Text Tools and Corpora. <https://tei-c.org/activities/projects/multilingual-text-tools-and-corpora-multext/> (Stand: 28.01.2021).
- Thurmair, Gregor (2003). Making Term Extraktion Tools Usable. Proceedings of the Conference of Controlled Language Translation, EAMT-CLAW03, 170-179.
- Toebe-Albrecht, Ingrid (1989). Die Gestaltung verständlicher Formulare. Theoretische und empirische Untersuchungen zum Formularzeichensystem mit praktischen Anwendungshinweisen. Berlin: Bundesverband Druck e.V.
- UNHCR (2018). Flüchtlinge, Asylsuchende und Migranten – Was sind die Unterschiede? https://www.unhcr.org/dach/wp-content/uploads/sites/27/2018/05/CH_Karten_CH_Deutsch_WEB.pdf (Stand: 10.12.2021).
- Valaski, Joselaine, Reinehr, Sheila & Malucelli, Andreia (2015). Approaches and Strategies to Extract Relevant Terms: How are they being applied? Proceedings of the International Conference on Artificial Intelligence (ICAI 2015), 478–484.
- Van Rijsbergen, C. J. (1979). Information Retrieval: Theory and practice. Proceedings of the Joint IBM/University of Newcastle upon Tyne Seminar on Data Base Systems, 1-14.
- Vázquez, Mercè & Oliver, Antoni (2018). Improving term candidates selection using terminological tokens. Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication 24(1), 122-147.
- Vivaldi, Jorge & Rodríguez, Horacio (2006). Some notes about the evaluation of terms and term extraction systems. 4th International Conference on Language Resources and Evaluation (LREC 2006), 12-17.
- Vivaldi, Jorge & Rodríguez, Horacio (2007). Evaluation of terms and term extraction systems: A practical approach. Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication 13(2), 225-248.
- Wagner, Hildegard (1984). Die deutsche Verwaltungssprache der Gegenwart: eine Untersuchung der sprachlichen Sonderform und ihrer Leistung. In: Ballweg, Joachim, Keim,

- Inken, Steger, Hugo & Wimmer Rainer (Hg.) Sprache der Gegenwart. Düsseldorf: Schwann, 7-139.
- Wanhoff, Vanessa-Barbara (2007). Computergestützte Extraktion von Termini aus technischen Texten in italienischer Sprache. Diplomarbeit, Ruprecht-Karls-Universität.
- Wermke, Matthias (1995). Austriazismen im gemeinsprachlichen Wörterbuch des Deutschen, dargestellt an DUDEN - Deutsches Universalwörterbuch (DDUW), 2. Auflage 1989. In: Muhr, Rudolf, Schrod, Richard & Wiesinger, Peter (Hg.) Österreichisches Deutsch. Linguistische, sozialpsychologische und sprachpolitische Aspekte einer nationalen Variante des Deutschen. Wien: Hölder-Pichler-Tempsky, 198-207.
- Wiesinger, Peter (Hg.) (1988a). Die deutsche Sprache in Österreich. Eine Einführung. In: Das österreichische Deutsch. Schriften zur deutschen Sprache in Österreich. Wien: Böhlau, 9-30.
- Wiesinger, Peter (Hg.) (1988b). Zur Frage aktueller bundesdeutscher Spracheinflüsse in Österreich. In: Das österreichische Deutsch. Schriften zur deutschen Sprache in Österreich. Wien: Böhlau, 225-245.
- Wiesinger, Peter (1990). Österreich als Sprachgrenz- und Sprachkontaktraum. In: Kremer, Ludger & Niebaum, Hermann (Hg.) Grenzdialekte. Studien zur Entwicklung kontinentalwestgermanischer Dialektkontinua. Hildesheim: Olms Verlag, 501-542.
- Wiesinger, Peter (1995a). Das österreichische Deutsch in der Diskussion. In: Muhr, Rudolf, Schrod, Richard & Wiesinger, Peter (Hg.) Österreichisches Deutsch. Linguistische, sozialpsychologische und sprachpolitische Aspekte einer nationalen Variante des Deutschen. Wien: Wien: Hölder-Pichler-Tempsky, 59-75.
- Wiesinger, Peter (1995b). Die sprachlichen Verhältnisse und der Weg zur allgemeinen deutschen Schriftsprache in Österreich im 18. und frühen 19. Jahrhundert. In: Gardt, Andreas (Hg.) Sprachgeschichte des Neuhochdeutschen. Gegenstände, Methoden, Theorien. Reihe Germanistische Linguistik Band 156. Tübingen: Niemeyer, 319-367.
- Wiesinger, Peter (1996). Das österreichische Deutsch als Varietät der deutschen Sprache. Unterrichtspraxis/Teaching German. Zeitschrift des amerikanischen Deutschlehrerverbandes 2(96), 147-154.

- Wiesinger, Peter (2000). Zum „Österreichischen Wörterbuch“. Aus Anlass der 38. neubearbeiteten Auflage (1997). *Zeitschrift für Germanistische Linguistik* 28(1). 41-64.
- Wiesinger, Peter (Hg.) (2008a). Das Deutsche in Österreich. In: *Das österreichische Deutsch in Gegenwart und Geschichte*. 2. Aufl. Wien: Lit Verlag, 5-23.
- Wiesinger, Peter (Hg.) (2008b). Das österreichische Amtsdeutsch. Eine Studie zur Syntax, Stil und Lexik der österreichischen Rechts- und Verwaltungssprache der Gegenwart. In: *Das österreichische Deutsch in Gegenwart und Geschichte*. 2. Aufl. Wien: Lit Verlag, 105-132.
- Wissik, Tanja (2012). Terminologische Variation in der Rechts- und Verwaltungssprache. Eine korpusbasierte Analyse der Hochschulterminologie in den Standardvarietäten des Deutschen in Deutschland, Österreich und der Schweiz. PhD Dissertation, Universität Wien.
- Witschel, Friedrich (2004). Text, Wörter, Morpheme — Möglichkeiten einer automatischen Terminologie-Extraktion. Diplomarbeit, Universität Leipzig.
- Witschel, Friedrich (2005). Terminologie-Extraktion – Möglichkeiten der Kombination statistischer und musterbasierter Verfahren. Würzburg: Ergon Verlag.
- Wolf, Norbert Richard (1994). Österreichisches zum österreichischen Deutsch. *Zeitschrift für Dialektologie und Linguistik* 61(1), 66-76.
- Zielinski, Daniel & Ramírez-Safar, Yamile (2005). Research meets practice: t-survey 2005: An online survey on terminology extraction and terminology management. *Proceedings of the ASLIB International Conference Translating & the Computer* 27.

Anhang I

Daten	Profil 1	Profil 2
Rolle in der Untersuchung	RTL-Überprüferin	Annotatorin
Geschlecht	Weiblich	Weiblich
Alter	29	30
Muttersprache(n)	Deutsch	Russisch
Fremdsprache(n)	Französisch, Englisch	Deutsch, Ukrainisch, Englisch
Höchster Bildungsabschluss	Mag. iur.	BA
Studienrichtung	Jurastudium	Translationswissenschaft
Berufliche Tätigkeit(en) (falls vorhanden)	Richteranwärterin	Kundendienstagentin
Branche (falls beruflich tätig)	Justiz	Finanz
Erfahrungen mit Terminologiearbeit (keine / wenig / mäßig / viel / sehr viel)	Mäßig	Viel
Kenntnisse im Niederlassungs- und Aufenthaltsrechts (keine / wenig / mäßig / gut / sehr gut)	Sehr gut	Gute

Anhang II

Basisterminus	N-WB	Muster	(im Korpus auftretende) Flexionsformen	Variante	Art der Variante	N-WB (Variante)	Muster	Kollokation / Kontext im Referenzkorpus
A2 Integrationsprüfung	2-WB	N N	A2 Integrationsprüfungen	-	-	-	-	positiv abgelegte A2 Integrationsprüfungen
Ableistung des Freiwilligendienstes	3-WB	N ART N	-	-	-	-	-	Vereinbarung über die Ableistung des Freiwilligendienstes mit der aufnehmenden Organisation
Abschluss einer A2 Integrationsprüfung	4-WB	N ART N N	-	-	-	-	-	Nachweis des Österreichischen Integrationsfonds über den erfolgreichen Abschluss einer A2 Integrationsprüfung
Abschluss einer Aufnahmevereinbarung	3-WB	N ART N	-	-	-	-	-	Die Forschungseinrichtung hat vor Abschluss einer Aufnahmevereinbarung die Qualifikation der Forscherin des Forschers für das konkrete Forschungsprojekt zu prüfen.
Abschluss einer B1 Integrationsprüfung	4-WB	N ART N N	-	-	-	-	-	Nachweis des Österreichischen Integrationsfonds über den erfolgreichen Abschluss einer B1 Integrationsprüfung
Abschlussklärung	1-WB	N	-	-	-	-	-	Es kann auch sein, dass die Behörde meinen Aufenthaltstitel nicht verlängert. Das kann auch passieren, wenn ich mich nicht an die Punkte 1 bis 4 dieser Abschlussklärung halte.
Absolvierung der A2 Integrationsprüfung	4-WB	N ART N N	-	-	-	-	-	Gleichwertiger Nachweis über die erfolgreiche Absolvierung der A2 Integrationsprüfung
Absolvierung der B1 Integrationsprüfung	4-WB	N ART N N	-	-	-	-	-	Gleichwertiger Nachweis über die erfolgreiche Absolvierung der B1 Integrationsprüfung
allgemeinen Erteilungsvoraussetzungen	2-WB	ADJ N	Allgemeinen Erteilungsvoraussetzungen	-	-	-	-	Bei der Antragstellung auf Erteilung einer Niederlassungsbewilligung-Forscher sind die allgemeinen Erteilungsvoraussetzungen nachzuweisen.
Allgemeines Verwaltungsverfahrensgesetz	2-WB	ADJ N	Allgemeinen Verwaltungsverfahrensgesetzes	AVG	anaphorisch	1-WB	N	Anträge (...) gelten nur dann als Verlängerungsanträge, (...) wenn der Antrag binnen zwei Wochen nach dem Wegfall des Hindernisses gestellt wird (§ 71 Absatz 5 Allgemeines Verwaltungsverfahrensgesetz - AVG).
Angehöriger	1-WB	N	Angehörige, Angehörigen	-	-	-	-	Angehöriger eines EWR-Bürgers
Ankerperson	1-WB	N	Ankerpersonen	-	-	-	-	Dauer der Bewilligung der zusammenführenden Person (Ankerperson)
Anmeldebescheinigung	1-WB	N	Anmeldebescheinigungen	-	-	-	-	Anmeldebescheinigung für EWR-Bürger

Basisterminus	N-WB	Muster	(im Korpus auftretende) Flexionsformen	Variante	Art der Variante	N-WB (Variante)	Muster	Kollokation / Kontext im Referenzkorpus
Annahme an Kindesstatt	3-WB	N PRÄP N	-	-	-	-	-	Urkunde über die Annahme an Kindesstatt
Antrag auf Ausstellung einer Daueraufenthaltskarte	5-WB	N PRÄP N ART N	-	-	-	-	-	Der Antrag auf Ausstellung einer Daueraufenthaltskarte ist spätestens vor Ablauf der Gültigkeitsdauer der Aufenthaltskarte zu stellen
Antrag auf Erteilung einer Aufenthaltsbewilligung	5-WB	N PRÄP N ART N	-	-	-	-	-	Antrag auf Erteilung einer Aufenthaltsbewilligung für Sozialdienstleister stellen
Antrag auf Erteilung eines Aufenthaltstitels	5-WB	N PRÄP N ART N	-	-	-	-	-	Antrag auf Erteilung eines weiteren Aufenthaltstitels stellen
Antrag auf Neuausstellung	3-WB	N PRÄP N	-	-	-	-	-	Antrag auf Neustellung eines Aufenthaltstitels bzw. einer Dokumentation
antragsgegenständliche Beschäftigung	2-WB	ADJ N	antragsgegenständlicher Beschäftigung, antragsgegenständlichen Beschäftigung	-	-	-	-	Arbeitskraft für die antragsgegenständliche Beschäftigung
antragsgegenständlicher Arbeitsplatz	2-WB	ADJ N	antragsgegenständlichen Arbeitsplatzes, antragsgegenständlichen Arbeitsplatz	-	-	-	-	Qualifikation für den antragsgegenständlichen Arbeitsplatz
Antragsteller	1-WB	N	Antragstellers	-	-	-	-	Angaben über den Aufenthalt des Antragstellers in Österreich
Antragstellung	1-WB	N	-	-	-	-	-	Antragstellung auf Erteilung einer Niederlassungsbewilligung–Forscher
Anwendbarkeit	1-WB	N	-	-	-	-	-	Anwendbarkeit einer Vereinbarung zwischen zwei oder mehreren Hochschulen
Anwerbeabkommen	1-WB	N	-	-	-	-	-	Jene Arbeitskräfte, die im Zuge der Anwerbeabkommen nach Österreich kamen, wurden als Gastarbeiter bezeichnet
Anzeigebestätigung	1-WB	N	-	-	-	-	-	Anzeigebestätigung des Arbeitsmarktservice nach dem Ausländerbeschäftigungsgesetz
Anzeigebestätigung des Arbeitsmarktservice	3-WB	N ART N	-	-	-	-	-	Im Falle des Au-Pair-Dienstes: Anzeigebestätigung des Arbeitsmarktservice nach dem Ausländerbeschäftigungsgesetz
Arbeitgebererklärung	1-WB	N	-	ArbeitgeberInnenerklärung	morphologisch	1-WB	N	Arbeitgebererklärung gemäß § 20f AuslBG für die Aufenthaltstitel Aufenthaltsbewilligung ICT und Aufenthaltsbewilligung mobile ICT
Arbeitsbestätigung	1-WB	N	-	-	-	-	-	Dienstzeugnis und Arbeitsbestätigung zum Nachweise von Berufserfahrung
Arbeitsmarktservice	1-WB	N	-	AMS	anaphorisch	1-WB	N	Anzeigebestätigung des Arbeitsmarktservice

Basisterminus	N-WB	Muster	(im Korpus auftretende) Flexionsformen	Variante	Art der Variante	N-WB (Variante)	Muster	Kollokation / Kontext im Referenzkorpus
Asylgesetz	1-WB	N	Asylgesetzes	AsylG	anaphorisch	1-WB	N	Aufenthaltsberechtigung gemäß § 55 Abs. 2 Asylgesetz (AsylG)
Aufenthaltsadoption	1-WB	N	Aufenthaltsadoptionen	-	-	-	-	Ich nehme zur Kenntnis, dass (...) das Eingehen und die Vermittlung von Aufenthaltsadoptionen (...) gerichtlich strafbare Tatbestände darstellen.
Aufenthaltsbeendigung	1-WB	N	-	-	-	-	-	Erwächst eine Aufenthaltsbeendigung in Rechtskraft, ist das Verfahren über den Verlängerungsantrag auf Erteilung des Aufenthaltstitels formlos einzustellen.
Aufenthaltsberechtigung für Österreich	3-WB	N PRÄP N	-	-	-	-	-	(...) besitzt eigene Aufenthaltsberechtigung für Österreich
Aufenthaltsberechtigung Plus	2-WB	N N	-	-	-	-	-	Innehabung einer Aufenthaltsberechtigung plus für 12 Monate
Aufenthaltsbewilligung	1-WB	N	-	-	-	-	-	Antrag auf Erteilung einer Aufenthaltsbewilligung
Aufenthaltsdauer	1-WB	N	-	-	-	-	-	Daten einer in Österreich leistungspflichtigen und alle Risiken abdeckenden gesetzlichen oder privaten Krankenversicherung für die Aufenthaltsdauer
Aufenthaltssehe	1-WB	N	Aufenthaltssehen	-	-	-	-	Ich nehme zur Kenntnis, dass das Eingehen und die Vermittlung von Aufenthaltssehen bzw. -partnerschaften (...) gerichtlich strafbare Tatbestände darstellen.
Aufenthaltskarte	1-WB	N	-	-	-	-	-	Ablauf der Gültigkeitsdauer der Aufenthaltskarte
Aufenthaltsrecht	1-WB	N	Aufenthaltsrechtes	-	-	-	-	Ich nehme zur Kenntnis, dass das unionsrechtliche Aufenthaltsrecht besteht, solange die Voraussetzungen erfüllt sind.
Aufenthaltstitel	1-WB	N	Aufenthaltstitels	-	-	-	-	Aufenthaltstitel für Österreich
Aufenthaltstitel ICT	2-WB	N N	-	-	-	-	-	Gültiger Aufenthaltstitel ICT eines anderen Mitgliedstaats
Aufenthaltsverbot	1-WB	N	Aufenthaltsverbots	-	-	-	-	(...) Durchsetzung eines Rückkehrentscheidungs eines Aufenthaltsverbots einer Ausweisung einer Zurückschiebung oder der Vollziehung der Schubhaft (...)
Aufenthaltszweck	1-WB	N	Aufenthaltszwecks	-	-	-	-	eine Änderung des Aufenthaltszwecks während der Gültigkeit des Aufenthaltstitels
Aufnahmevereinbarung	1-WB	N	-	-	-	-	-	Die Forschungseinrichtung hat vor Abschluss einer Aufnahmevereinbarung die Qualifikation der Forscherin des Forschers für das konkrete Forschungsprojekt zu prüfen.
Ausbildungszeugnis	1-WB	N	-	-	-	-	-	Ausbildungszeugnis zum Nachweis spezieller Kenntnisse

Basisterminus	N-WB	Muster	(im Korpus auftretende) Flexionsformen	Variante	Art der Variante	N-WB (Variante)	Muster	Kollokation / Kontext im Referenzkorpus
Ausländerbeschäftigungsgesetz	1-WB	N	Ausländerbeschäftigungsgesetzes	-	-	-	-	Nachweis einer Berechtigung nach dem Ausländerbeschäftigungsgesetz
ausstellende Behörde	2-WB	ADJ N	-	-	-	-	-	ausstellende Behörde sämtlicher Berechtigungen
Ausstellung einer Daueraufenthaltskarte	3-WB	N ART N	-	-	-	-	-	Antrag auf Ausstellung einer Daueraufenthaltskarte
Ausweisung	1-WB	N	-	-	-	-	-	(...) Durchsetzung eines Rückkehrentscheidungs eines Aufenthaltsverbots einer Ausweisung einer Zurückschiebung oder der Vollziehung der Schubhaft (...)
B1 Integrationsprüfung	2-WB	N N	-	-	-	-	-	Abschluss einer B1 Integrationsprüfung
Beantragung	1-WB	N	-	-	-	-	-	rechtzeitige Beantragung der Daueraufenthaltskarte
Beglaubigung	1-WB	N	-	-	-	-	-	Eine notarielle oder gerichtliche Beglaubigung der Unterschriften ist nicht erforderlich.
Beglaubigung des Notars	1-WB	N ART N	-	-	-	-	-	Beglaubigung des Notars oder des Gerichts
Begründung der eingetragenen Partnerschaft	4-WB	N ART ADJ N	-	Begründung der EP	anaphorisch	3-WB	N ART N	Begründung der eingetragenen Partnerschaft mit EWR-Bürger
Behördenvermerk	1-WB	N	Behördenvermerke	-	-	-	-	-
Belehrung	1-WB	N	-	-	-	-	-	Belehrung gemäß §19 Absatz 7 NAG
Berechtigung	1-WB	N	Berechtigungen	-	-	-	-	Bisherige Berechtigungen zum Aufenthalt in Österreich
besachwaltete Person	2-WB	ADJ N	-	-	-	-	-	Der Antrag wird gestellt für eine von mir besachwaltete Person.
Beschäftigungsbewilligung	1-WB	N	-	-	-	-	-	vom AMS auf Antrag der ArbeitgeberInnen ausgestellte Beschäftigungsbewilligung
Bescheinigung des Daueraufenthaltes	3-WB	N ART N	-	-	-	-	-	Bescheinigung des Daueraufenthaltes des zusammenführenden EWR-Bürgers
Bescheinigung des Daueraufenthaltes für EWR-Bürger	5-WB	N ART N PRÄP N	-	-	-	-	-	Antrag auf Ausstellung einer Bescheinigung des Daueraufenthaltes für EWR-Bürger

Basisterminus	N-WB	Muster	(im Korpus auftretende) Flexionsformen	Variante	Art der Variante	N-WB (Variante)	Muster	Kollokation / Kontext im Referenzkorpus
besonders Hochqualifizierte	2-WB	ADV ADJ	-	-	-	-	-	Rot-Weiß-Rot-Karte für besonders Hochqualifizierte
bestandrechtliche Vorverträge	2-WB	ADJ N	-	-	-	-	-	Nachweis des Rechtsanspruches auf eine ortsübliche Unterkunft, Miet- oder Untermietvertrag, bestandrechtliche Vorverträge oder Eigentumsnachweise
Betriebsentsandter	1-WB	N	-	-	-	-	-	Eine „Aufenthaltsbewilligung – Betriebsentsandter“ kann erteilt werden, wenn eine Sicherungsbescheinigung oder eine Beschäftigungsbewilligung als Betriebsentsandter vorgelegt werden kann.
Blaue Karte EU	3-WB	ADJ N N	-	-	-	-	-	Arbeitgebererklärung gemäß § 20d AuslBG für die Aufenthaltstitel Rot-Weiß-Rot - Karte, Blaue Karte, Niederlassungsbewilligung - Künstler
Bundesgebiet	1-WB	N	-	-	-	-	-	Antrag von im Bundesgebiet aufhältigen Drittstaatsangehörigen
Daueraufenthalt	1-WB	N	Daueraufenthaltes	-	-	-	-	Bescheinigung des Daueraufenthaltes für EWR-Bürger
DAUERAUFENTHALT–EU	1-WB	N	-	-	-	-	-	-
Daueraufenthaltskarte	1-WB	N	-	-	-	-	-	Antrag auf Ausstellung einer Daueraufenthaltskarte
Daueraufenthaltsrecht	1-WB	N	Daueraufenthaltsrechtes	-	-	-	-	Recht auf Daueraufenthalt
Deutschkenntnisse	1-WB	N	-	-	-	-	-	-
Dienstgeber	1-WB	N	Dienstgebers	-	-	-	-	Bestätigung des Dienstgebers, dass der Antragsteller in einer Führungsposition tätig war
Dienstpäss	1-WB	N	-	-	-	-	-	-
Dienstvertrag	1-WB	N	Dienstvertrages	-	-	-	-	Beilage einer Kopie des Dienstvertrages
Diplomatenpass	1-WB	N	-	-	-	-	-	-
Dokumentation	1-WB	N	-	-	-	-	-	Dokumentation zum Nachweis des Aufenthaltsrechts
Dokumentation des Aufenthaltsrechtes	3-WB	N ART N	-	-	-	-	-	-

Basisterminus	N-WB	Muster	(im Korpus auftretende) Flexionsformen	Variante	Art der Variante	N-WB (Variante)	Muster	Kollokation / Kontext im Referenzkorpus
Drittstaat	1-WB	N	-	-	-	-	-	in einem Drittstaat ansässig
Drittstaatsangehörige	1-WB	N	Drittstaatsangehörigen	-	-	-	-	Drittstaatsangehörige (das sind alle Personen, die nicht EWR-BürgerInnen oder Schweizer BürgerInnen sind) sind mit erstmaliger Erteilung eines Aufenthaltstitels (...) zur Erfüllung des Moduls 1 der Integrationsvereinbarung verpflichtet.
Eigentumsnachweis	1-WB	N	Eigentumsnachweise	-	-	-	-	Nachweis des Rechtsanspruches auf eine ortsübliche Unterkunft, Miet- oder Untermietvertrag, bestandsrechtliche Vorverträge oder Eigentumsnachweise
Eingabegebühr	1-WB	N	-	-	-	-	-	Eingabegebühr entrichtet
Einwanderung	1-WB	N	-	-	-	-	-	Amt der Wiener Landesregierung Magistratsabteilung 35 für Einwanderung und Staatsbürgerschaft
Erfüllungspflicht	1-WB	N	-	-	-	-	-	Der Erfüllungspflicht ist binnen zwei Jahren ab erstmaliger Erteilung des Aufenthaltstitels nachzukommen.
Erschleichung eines Aufenthaltstitels	3-WB	N ART N	-	-	-	-	-	(...) dass das Eingehen und die Vermittlung von Aufenthaltsehen bzw.-partnerschaften (...) bzw. die Erschleichung eines Aufenthaltstitels gerichtlich strafbare Tatbestände darstellen
Erstantrag	1-WB	N	Erstanträgen	-	-	-	-	Geburtsurkunde oder ein diesem gleichzuhaltendes Dokument nur bei Erstanträgen erforderlich
Erteilung eines Aufenthaltstitels	3-WB	N ART N	-	-	-	-	-	erstmaliger Erteilung eines Aufenthaltstitels gemäß § 8 Absatz 1 Z 1,2,4,5,6,8,9 oder 10 Niederlassungs- und Aufenthaltsgesetz – NAG
Erteilung eines Visums	3-WB	N ART N	-	-	-	-	-	Erteilung eines Visums gem § 20a Abs 2 FPG
Erwerb des Rechtes auf Daueraufenthalt	5-WB	N ART N PRÄP N	-	-	-	-	-	bei vorzeitigem Erwerb des Rechtes auf Daueraufenthalt gemäß § 53a Abs. 3 - 5 NAG
Erwerbstätigkeit	1-WB	N	-	-	-	-	-	Nachweise über die Ausübung einer Erwerbstätigkeit
EU-Passbild	1-WB	N	-	-	-	-	-	EU-Passbild für Lichtbildausweis für EWR-Bürger und Aufenthaltskarte
EWR-Bürger	1-WB	N	EWR-Bürgers, EWR-Bürgern, EWR-BürgerInnen	-	-	-	-	Anmeldebescheinigung oder Bescheinigung des Daueraufenthaltes des zusammenführenden EWR-Bürgers
Existenzmittel	1-WB	N	-	-	-	-	-	Vorliegen von ausreichenden Existenzmittel

Basisterminus	N-WB	Muster	(im Korpus auftretende) Flexionsformen	Variante	Art der Variante	N-WB (Variante)	Muster	Kollokation / Kontext im Referenzkorpus
Fachkraft	1-WB	N	-	-	-	-	-	Rot-Weiß-Rot - Karte (Fachkraft in Mangelberufen)
Fachkraft im Mangelberuf	3-WB	N PRÄP N	Fachkraft in Mangelberufen	-	-	-	-	Rot-Weiß-Rot - Karte (Fachkraft in Mangelberufen)
Familienangehörige	1-WB	N	Familienangehörigen	-	-	-	-	Familienangehörige des Inhabers der Aufenthaltsbewilligung Forscher-Mobilität
Familiengemeinschaft	1-WB	N	-	-	-	-	-	Familiengemeinschaft mit unternehmensintern transferierter Arbeitnehmer
Familienstand	1-WB	N	-	-	-	-	-	-
Familienverhältnisse	1-WB	N	-	-	-	-	-	Wenn sich meine Familienverhältnisse ändern, muss ich das der Behörde melden. Das kann zum Beispiel eine Scheidung sein.
Formular	1-WB	N	Formulare	-	-	-	-	-
Forscher-Mobilität	1-WB	N	-	-	-	-	-	Familienangehörige des Inhabers der Aufenthaltsbewilligung Forscher-Mobilität
Forschungsaufenthalt	1-WB	N	-	-	-	-	-	Angaben zu einem beabsichtigten Forschungsaufenthalt in einem weiteren EU-Mitgliedsstaat
Forschungstätigkeit	1-WB	N	-	-	-	-	-	Zusage des Forschers, dass er sich bemühen wird, die Forschungstätigkeit abzuschließen
Fortbestand der Voraussetzungen	3-WB	N ART N	-	-	-	-	-	(...), solange die Voraussetzungen erfüllt sind und der Fortbestand der Voraussetzungen bei einer Meldung (...) oder aus besonderem Anlass überprüft werden kann
Freiheitsstrafe	1-WB	N	-	-	-	-	-	Ich nehme zur Kenntnis, dass ich eine Verwaltungsübertretung begehe und mit einer Geldstrafe von 50 bis 250 Euro, im Fall der Uneinbringlichkeit mit einer Freiheitsstrafe bis zu einer Woche, bestraft werden kann, wenn ich (...)
Freiwilligendienst	1-WB	N	Freiwilligendienstes	-	-	-	-	Vereinbarung über die Ableistung des Freiwilligendienstes mit der aufnehmenden Organisation
Fremde	1-WB	N	Fremden	-	-	-	-	(...) im Fall der Nichtvorlage erforderlicher Urkunden oder Nachweise, wenn deren Beschaffung für die Fremde bzw. den Fremden nachweislich nicht möglich oder nicht zumutbar war zugelassen werden kann.
Fremdpolizeigesetz	1-WB	N	-	FPG	anaphorisch	1-WB	N	Gegebenenfalls zusätzliche Nachweise bei § 41 Abs. 1 NAG (Besonders Hochqualifizierte), sofern nicht

Basisterminus	N-WB	Muster	(im Korpus auftretende) Flexionsformen	Variante	Art der Variante	N-WB (Variante)	Muster	Kollokation / Kontext im Referenzkorpus
								bereits bei Erteilung eines Visums gem. § 24a Abs. 2 Fremdenpolizeigesetz (FPG) eingereicht
Gebietskörperschaft	1-WB	N	-	-	-	-	-	(...), dass ich für den Ersatz jener Kosten hafte, die einer Gebietskörperschaft bei der Durchsetzung eines Rückkehrentscheidung (...) entstehen.
Geburtsort	1-WB	N	-	-	-	-	-	-
Geldstrafe	1-WB	N	-	-	-	-	-	Ich nehme zur Kenntnis, dass ich eine Verwaltungsübertretung begehe und mit einer Geldstrafe von 50 bis 250 Euro (...) bestraft werden kann.
gelindere Mittel	2-WB	ADJ N	gelinderer Mittel	-	-	-	-	(...), dass ich für den Ersatz jener Kosten hafte, die einer Gebietskörperschaft bei der Durchsetzung eines Rückkehrentscheidung (...) einschließlich der Aufwendungen für den Ersatz gelinderer Mittel entstehen.
gerichtlich strafbare Tatbestände	3-WB	ADV ADJ N	-	-	-	-	-	(...), dass das Eingehen und die Vermittlung von Aufenthaltsehen bzw. -partnerschaften, das Eingehen und die Vermittlung von Aufenthaltsadoptionen bzw. die Erschleichung eines Aufenthaltstitels gerichtlich strafbare Tatbestände darstellen.
gesetzgebende Körperschaften	2-WB	ADJ N	gesetzgebender Körperschaften	-	-	-	-	Angehöriger gesetzgebender Körperschaften
gesetzliche Interessensvertretung	2-WB	ADJ N	gesetzlicher Interessensvertretung	-	-	-	-	Zugehörigkeit zu gesetzlicher Interessensvertretung
gesetzliche Pflichtversicherung	2-WB	ADJ N	-	-	-	-	-	Nachweis über einen in Österreich leistungspflichtigen und alle Risiken abdeckenden Krankenversicherungsschutz, sofern keine gesetzliche Pflichtversicherung bestehen wird oder besteht.
gesetzliche Vertretung	2-WB	ADJ N	-	-	-	-	-	Ich nehme zur Kenntnis, dass die persönliche Abholung der Aufenthaltsbewilligung durch die antrag stellende Person erforderlich ist (Ausnahme gesetzliche Vertretung).
gesetzlicher Vertreter	2-WB	ADJ N	gesetzlichen Vertreters	-	-	-	-	Name und Anschrift des gesetzlichen Vertreters für nicht eigenberechtigte Person
Grenzgänger	1-WB	N	-	-	-	-	-	-
Grundversorgungsvereinbarung	1-WB	N	-	-	-	-	-	Bundes- oder Landesgesetzes, das die Grundversorgungsvereinbarung nach Art 15a B-VG BGBl. I Nr. 80/2004 umsetzt
gültiger Aufenthaltstitel	2-WB	N	-	-	-	-	-	Gültiger Aufenthaltstitel ICT eines anderen Mitgliedstaats

Basisterminus	N-WB	Muster	(im Korpus auftretende) Flexionsformen	Variante	Art der Variante	N-WB (Variante)	Muster	Kollokation / Kontext im Referenzkorpus
gültiges Reisedokument	2-WB	N	-	-	-	-	-	-
Gültigkeit	1-WB	N	-	-	-	-	-	eine Änderung des Aufenthaltszwecks während der Gültigkeit des Aufenthaltstitels
Gültigkeitsdauer	1-WB	N	-	-	-	-	-	(...), dass Verlängerungsanträge vor Ablauf der Gültigkeitsdauer eines Aufenthaltstitels, frühestens jedoch drei Monate vor Ablauf einzubringen sind
Gültigkeitsdauer der Aufenthaltskarte	3-WB	N ART N	-	-	-	-	-	Ablauf der Gültigkeitsdauer der Aufenthaltskarte
Gutachten des Arbeitsmarktservice	3-WB	N ART N	-	AMS-Gutachten	anaphorisch	1-WB	N	Gutachten des Arbeitsmarktservice (AMS-Gutachten) über den gesamtwirtschaftlichen Nutzen der beabsichtigten selbständigen Erwerbstätigkeit
Habilitation	1-WB	N	-	-	-	-	-	Nachweis einer Habilitation
Haftungserklärung	1-WB	N	-	-	-	-	-	Haftungserklärung des Zusammenführenden von einem österreichischen Notar oder einem inländischen Gericht beglaubigt mit einer Gültigkeitsdauer von mindestens 5 Jahren
Hauptzweck	1-WB	N	-	-	-	-	-	Hauptzweck des Aufenthalts
Hauptzweck des Aufenthalts	3-WB	N ART N	-	-	-	-	-	-
häusliche Gemeinschaft	2-WB	ADJ N	-	-	-	-	-	urkundlicher Nachweis einer zuständigen Behörde des Herkunftsstaates der Unterhaltsleistung des EWR-Bürgers oder des Lebens in häuslicher Gemeinschaft
Herkunftsstaat	1-WB	N	Herkunftsstaates	-	-	-	-	Nachweis des Bestehens einer dauerhaften Beziehung mit dem Zusammenführenden im Herkunftsstaat
Hochqualifizierte Schlüsselkraft	2-WB	ADJ N	-	-	-	-	-	Rot-Weiß-Rot - Karte für Hochqualifizierte Schlüsselkraft
Identitätsdaten	1-WB	N	-	-	-	-	-	(...), dass ich eine Verwaltungsübertretung begehe und mit einer Geldstrafe von 50 Euro bis zu 250 Euro zu bestrafen bin, wenn ich (...) bei Änderung der Identitätsdaten, z. B. Heirat meiner Meldepflicht nicht rechtzeitig nachkomme.
Inhaber der Aufenthaltsbewilligung	3-WB	N ART N	Inhabers der Aufenthaltsbewilligung	-	-	-	-	Familienangehörige des Inhabers der Aufenthaltsbewilligung Forscher-Mobilität
Inlandsantragstellung	1-WB	N	-	inländische Antragstellung	morphosyntaktisch	2-WB	ADJ N	Inlandsantragstellung für den Erstaufenthaltstitel
Innehabung einer Niederlassungsbewilligung	3-WB	N ART N	-	-	-	-	-	Zweckänderungsantrag (...) nach Innenhabung einer Niederlassungsbewilligung gemäß § 43 Abs. 3 NAG

Basisterminus	N-WB	Muster	(im Korpus auftretende) Flexionsformen	Variante	Art der Variante	N-WB (Variante)	Muster	Kollokation / Kontext im Referenzkorpus
Innovationstätigkeit	1-WB	N	-	-	-	-	-	Nachweis einer Forschungs- oder Innovationstätigkeit
Integrationsfond	1-WB	N	Integrationsfonds	-	-	-	-	Nachweis des Österreichischen Integrationsfonds
Integrationsgesetz	1-WB	N	-	IntG	anaphorisch	1-WB	N	(...), dass die Integrationsvereinbarung mit 1.10.2017 im Integrationsgesetz (IntG) neu geregelt wird.
Integrationsprüfung	1-WB	N	Integrationsprüfungen	-	-	-	-	(...), dass ich (...) bestraft werden kann, wenn ich bei der Integrationsprüfung zur Erfüllung des Moduls 1 oder 2 nicht erlaubte Hilfsmittel verwende.
Integrationsvereinbarung	1-WB	N	-	-	-	-	-	Drittstaatsangehörige sind mit erstmaliger Erteilung eines Aufenthaltstitels (...) zur Erfüllung des Moduls 1 der Integrationsvereinbarung verpflichtet.
Kindesstatt	1-WB	N	-	-	-	-	-	Annahme an Kindesstatt
Krankenversicherung	1-WB	N	-	-	-	-	-	Krankenversicherung für die Dauer des Aufenthalts in Österreich, die alle Risiken abdeckt
Krankenversicherungsschutz	1-WB	N	Krankenversicherungsschutzes	-	-	-	-	Vorliegen von ausreichenden Existenzmittel und umfassenden Krankenversicherungsschutzes
Ladung	1-WB	N	-	-	-	-	-	Ist die persönliche Zustellung einer Ladung oder einer Verfahrensordnung zum wiederholten Mal nicht möglich, kann das Verfahren eingestellt werden.
Lebensunterhalt	1-WB	N	Lebensunterhaltes	-	-	-	-	Sicherung des Lebensunterhaltes für die Aufenthaltsdauer
Lebensunterhaltsmittel	1-WB	N	-	-	-	-	-	Verfügbare eigene Lebensunterhaltsmittel
leitender Verwaltungsbediensteter	2-WB	ADJ N	-	-	-	-	-	Angehöriger gesetzgebender Körperschaften, leitender Verwaltungsbediensteter und Führungskraft in der Privatwirtschaft
Lichtbildausweis	1-WB	N	Lichtbildausweises	-	-	-	-	Lichtbildausweis für Personen mit EWR- oder Schweizer Staatsangehörigkeit
Lichtbildausweis für EWR- Bürger	3-WB	N PRÄP N	-	-	-	-	-	EU-Passbild für Lichtbildausweis für EWR-Bürger
Meldeverpflichtung	1-WB	N	-	Meldepflicht	morphologisch	1-WB	N	(...), dass ich (...) bestraft werden kann, wenn ich (...) meiner Meldeverpflichtung aus §§ 19 Absatz 11 und 51 Absatz 3 Niederlassungs- und Aufenthaltsgesetz (NAG) nicht rechtzeitig nachkomme.
Minderjähriger	1-WB	N	-	-	-	-	-	unbegleiteter Minderjähriger
Mitgliedsstaat	1-WB	N	-	-	-	-	-	Aufenthalt als Forscher in einem weiteren Mitgliedstaat

Basisterminus	N-WB	Muster	(im Korpus auftretende) Flexionsformen	Variante	Art der Variante	N-WB (Variante)	Muster	Kollokation / Kontext im Referenzkorpus
Mitteilung des Arbeitsmarktservice	3-WB	N ART N	-	AMS-Mitteilung	anaphorisch	1-WB	N	Ein Erstantrag "Rot-Weiß-Rot-Karte" oder "Blaue Karte EU" (...) nur dann positiv erledigt werden, wenn die Mitteilung des Arbeitsmarktservice bestätigt, dass die im AuslBG normierten Schlüsselkraftkriterien erfüllt sind.
mobiler unternehmensintern transferierter Arbeitnehmer	2-WB	ADJ N	-	mobile ICT	anaphorisch	2-WB	ADJ N	Aufenthaltsbewilligung Mobile ICT (mobiler unternehmensintern transferierter Arbeitnehmer)
Mobilitätsmaßnahmen	1-WB	N	-	-	-	-	-	Teilnahme an einem Unions- oder multilateralen Programm mit Mobilitätsmaßnahmen
Modul 1 der Integrationsvereinbarung	4-WB	N N ART N	Moduls 1 der Integrationsvereinbarung	-	-	-	-	Erfüllung des Moduls 1 der Integrationsvereinbarung
Modul 2 der Integrationsvereinbarung	4-WB	N CARD ART N	Moduls 2 der Integrationsvereinbarung	-	-	-	-	Erfüllung des Moduls 2 der Integrationsvereinbarung
natürliche Person	2-WB	ADJ N	-	-	-	-	-	eine volljährige in Österreich wohnhafte natürliche Person
Neuausstellung	1-WB	N	-	-	-	-	-	Antrag auf Neustellung eines Aufenthaltstitels bzw. einer Dokumentation
nicht eigenberechtigte Person	3-WB	PTKNEG ADJ N	nicht eigenberechtigte Personen	-	-	-	-	Name und Anschrift des gesetzlichen Vertreters für nicht eigenberechtigte Personen
Niederlassungs- und Aufenthaltsgesetz	3-WB	N KON N	Niederlassungs- und Aufenthaltsgesetzes	NAG	anaphorisch	1-WB	N	Erteilung des Aufenthaltstitels nach dem Niederlassungs- und Aufenthaltsgesetz (NAG)
Niederlassungsbehörde	1-WB	N	-	-	-	-	-	Ich nehme zur Kenntnis, dass die Tragfähigkeit der Haftungserklärung seitens der Niederlassungsbehörde überprüft wird.
Niederlassungsbewilligung	1-WB	N	-	-	-	-	-	Meine zuletzt erteilte Aufenthaltsbewilligung Forscher, Künstler, Sonderfälle unselbständiger Erwerbstätigkeit gilt seit 1.10.2017 als Niederlassungsbewilligung weiter.
notarielle oder gerichtliche Beglaubigung	4-WB	ADJ KON ADJ N	-	-	-	-	-	Eine notarielle oder gerichtliche Beglaubigung der Unterschriften ist nicht erforderlich.
örtlich zuständige Behörde	3-WB	ADV ADJ N	-	-	-	-	-	Die Forschungseinrichtung hat unverzüglich die örtlich zuständige Behörde über jede vorzeitige Beendigung einer Aufnahmevereinbarung (...) in Kenntnis zu setzen.
ortsübliche Unterkunft	2-WB	ADJ N	-	-	-	-	-	Nachweis des Rechtsanspruches auf eine ortsübliche Unterkunft
österreichisches Notar	2-WB	ADJ N	österreichischen Notar	-	-	-	-	Haftungserklärung von einem österreichischen Notar (...) beglaubigt

Basisterminus	N-WB	Muster	(im Korpus auftretende) Flexionsformen	Variante	Art der Variante	N-WB (Variante)	Muster	Kollokation / Kontext im Referenzkorpus
Pflichtversicherung	1-WB	N	-	-	-	-	-	gesetzliche Pflichtversicherung
polizeiliches Führungszeugnis	2-WB	ADJ N	-	-	-	-	-	Aktuelles polizeiliches Führungszeugnis oder ein vergleichbares Dokument
private Krankenversicherung	2-WB	ADJ N	-	-	-	-	-	Daten einer in Österreich leistungspflichtigen und alle Risiken abdeckenden gesetzlichen oder privaten Krankenversicherung für die Aufenthaltsdauer
Quote	1-WB	N	-	-	-	-	-	Durch die Quote wird die Anzahl der jährlich maximal zu erteilenden Aufenthaltstitel beschränkt.
quotenfrei	1-WB	ADJ	-	-	-	-	-	quotenfreie Rot-Weiß-Rot-Karte plus
Quotenpflicht	1-WB	N	-	-	-	-	-	Bestimmte Aufenthaltstitel unterliegen nicht der Quotenpflicht.
Rechtsanspruch	1-WB	N	Rechtsanspruches	-	-	-	-	Nachweis des Rechtsanspruches auf eine ortsübliche Unterkunft
Rechtsverhältnis	1-WB	N	-	-	-	-	-	Wird das Rechtsverhältnis zwischen Forscher und Forschungseinrichtung beendet (...), endet auch die Aufnahmevereinbarung.
Reisedokument	1-WB	N	Reisedokumentes	-	-	-	-	Art des Reisedokumentes
Rot-Weiß-Rot-Karte	1-WB	N	-	Rot-Weiß-Rot – Karte	grafisch	3-WB	N \$(N	Erteilung einer Rot-Weiß-Rot-Karte
Rot-Weiß-Rot-Karte plus	2-WB	N N	-	Rot-Weiß-Rot – Karte Plus	grafisch	4-WB	N \$(N N	quotenfreie Rot-Weiß-Rot-Karte plus
Rückkehrentscheidung	1-WB	N	-	-	-	-	-	Durchsetzung einer Rückkehrentscheidung
Schlüsselkraft	1-WB	N	Schlüsselkräfte	-	-	-	-	Selbständige Schlüsselkraft
Schriftliche Erklärung des Zusammenführenden	4-WB	ADJ N ART N	-	-	-	-	-	Schriftliche Erklärung des Zusammenführenden über Art und Umfang der Unterhaltsleistung
Schubhaft	1-WB	N	-	-	-	-	-	Vollziehung der Schubhaft
selbstständige Erwerbstätigkeit	2-WB	ADJ N	-	-	-	-	-	Nachweis über die selbstständige Erwerbstätigkeit
selbständige Mobilität	2-WB	ADJ N	-	-	-	-	-	Niederlassungsbewilligung - selbständige Mobilität

Basisterminus	N-WB	Muster	(im Korpus auftretende) Flexionsformen	Variante	Art der Variante	N-WB (Variante)	Muster	Kollokation / Kontext im Referenzkorpus
selbständige Schlüsselkraft	2-WB	ADJ N	-	-	-	-	-	Rot-Weiß-Rot - Karte für Selbständige Schlüsselkraft
Selbständigkeit	1-WB	N ART N	-	-	-	-	-	Nachweis der Selbständigkeit
Selbstständiger	1-WB	N	-	-	-	-	-	-
Sicherung des Lebensunterhaltes	3-WB	N ART N	-	-	-	-	-	Mittel zur Sicherung des Lebensunterhaltes
Sicherungsbescheinigung	1-WB	N	-	Sicherbescheinigung	morphologisch	1-WB	N	Eine vom Arbeitsmarktservice (AMS) auf Antrag der ArbeitgeberInnen ausgestellte Sicherungsbescheinigung, dass bestimmte Personen eine Beschäftigungsbewilligung erhalten werden, sofern und sobald der entsprechende Aufenthaltstitel erteilt wird
Sonderfälle unselbständiger Erwerbstätigkeit	3-WB	N ADJ N	-	-	-	-	-	Niederlassungsbewilligung für Sonderfälle unselbständiger Erwerbstätigkeit (z. B. SeelsorgerIn, WissenschaftlerIn, Medienbedienstete)
Sozialdienstleistender	1-WB	N	-	-	-	-	-	Erteilung einer Aufenthaltsbewilligung - Sozialdienstleistender
Staatsangehörigkeit	1-WB	N	Staatsangehörigkeiten	-	-	-	-	österreichische Staatsangehörigkeit nach Eheschließung
Staatsbürgerschaft	1-WB	N	-	-	-	-	-	EWR- oder Schweizer Staatsbürgerschaft
Stellungnahme des Arbeitsmarktservice	3-WB	N ART N	-	AMS-Stellungnahme	anaphorisch	1-WB	N	Ein Erstantrag "für Selbständige nach dem Europaabkommen" kann von der MA 35 nur dann positiv erledigt werden, wenn die Stellungnahme des Arbeitsmarktservice (AMS) bestätigt, dass die beabsichtigte Tätigkeit dem Europaabkommen entspricht.
strafrechtliche Verurteilungen	2-WB	ADJ N	-	-	-	-	-	Bisherige strafrechtliche Verurteilungen
tatsächliche Unterhaltsgewährung	2-WB	ADJ N	-	-	-	-	-	Nachweis über die tatsächliche Unterhaltsgewährung
Träger von Privilegien und Immunitäten	5-WB	N PRÄP N KON N	-	-	-	-	-	Nachweis über die frühere Eigenschaft als Träger von Privilegien und Immunitäten
Tragfähigkeit der Haftungserklärung	3-WB	N ART N	-	-	-	-	-	Ich nehme zur Kenntnis, dass die Tragfähigkeit der Haftungserklärung seitens der Niederlassungsbehörde überprüft wird.
Übernahme des Aufenthaltstitels	3-WB	N ART N	-	-	-	-	-	Bestätigung über die Übernahme des Aufenthaltstitels

Basisterminus	N-WB	Muster	(im Korpus auftretende) Flexionsformen	Variante	Art der Variante	N-WB (Variante)	Muster	Kollokation / Kontext im Referenzkorpus
unbegleiteter Minderjähriger	2-WB	ADJ N	-	-	-	-	-	unbegleiteter Minderjähriger, der sich nicht in Begleitung eines für ihn gesetzlich verantwortlichen Volljährigen befindet
Uneinbringlichkeit	1-WB	N	-	-	-	-	-	(...), dass ich (...) im Fall der Uneinbringlichkeit mit Freiheitsstrafe bis zu einer Woche, bestraft werden kann, wenn ich (...)
unionsrechtliches Aufenthaltsrecht	2-WB	ADJ N	-	-	-	-	-	Ich nehme zur Kenntnis, dass das unionsrechtliche Aufenthaltsrecht besteht, solange die Voraussetzungen erfüllt sind.
Unmündigkeit	1-WB	N	-	-	-	-	-	Unmündigkeit zum Zeitpunkt der Erfüllungspflicht
unselbständige Mobilität	2-WB	ADJ N	-	-	-	-	-	Rot-Weiß-Rot - Karte für Unselbständige Mobilität
Unterhaltsgewährung	1-WB	N	-	-	-	-	-	Nachweis über die tatsächliche Unterhaltsgewährung
Unterhaltsleistung	1-WB	N	-	-	-	-	-	Art und Umfang der Unterhaltsleistung
Unterhaltsmittel	1-WB	N	-	-	-	-	-	(...), dass ich für folgende Person (...) gemäß § 2 Abs. 1 Z 15 NAG für die Erfordernisse einer Unterkunft und entsprechender Unterhaltsmittel aufkomme (...)
unterhaltspflichtige Person	2-WB	ADJ N	-	-	-	-	-	Unterhaltspflichtige Person in Österreich
unternehmensintern transferierte Arbeitskraft	3-WB	ADV ADJ N	unternehmensintern transferierten Arbeitskraft	-	-	-	-	Aufnahme der unternehmensintern transferierten Arbeitskraft
unternehmensintern transferierter Arbeitnehmer	3-WB	ADV ADJ N	-	-	-	-	-	Aufenthaltsbewilligung ICT (unternehmensintern transferierter Arbeitnehmer)
Unternehmensinterner Transfer	2-WB	ADJ N	Unternehmensinternen Transfers	-	-	-	-	(...), dass der Antragsteller nach Beendigung seines unternehmensinternen Transfers in eine Niederlassung zurückkehren kann.
urkundlicher Nachweis	2-WB	ADJ N	-	-	-	-	-	urkundlicher Nachweis des Bestehens der Ehe oder eingetragenen Partnerschaft
Verfahrensanordnung	1-WB	N	-	-	-	-	-	Ist die persönliche Zustellung einer Ladung oder einer Verfahrensanordnung zum wiederholten Mal nicht möglich, kann das Verfahren eingestellt werden.
Verlängerungsantrag	1-WB	N	Verlängerungsantrags	-	-	-	-	Verzicht auf die Stellung eines weiteren Verlängerungsantrags
Vermittlung von Aufenthaltsadoptionen	3-WB	N PRÄP N	-	-	-	-	-	das Eingehen und die Vermittlung von Aufenthaltsadoptionen bzw. die Erschleichung eines Aufenthaltstitels
Versicherungsleistung	1-WB	N	Versicherungsleistungen	-	-	-	-	Bestätigungen über sonstige Versicherungsleistungen

Basisterminus	N-WB	Muster	(im Korpus auftretende) Flexionsformen	Variante	Art der Variante	N-WB (Variante)	Muster	Kollokation / Kontext im Referenzkorpus
Vertreter des Antragstellers	3-WB	N ART N	-	-	-	-	-	-
Vertretungsbehörde	1-WB	N	-	-	-	-	-	Erstanträge sind grundsätzlich im Ausland bei der örtlich zuständigen österreichischen Vertretungsbehörde zu stellen.
Verwaltungsbediensteter	1-WB	N	-	-	-	-	-	leitender Verwaltungsbediensteter und Führungskraft in der Privatwirtschaft
Verwaltungsübertretung	1-WB	N	-	-	-	-	-	Ich nehme zur Kenntnis, dass ich eine Verwaltungsübertretung begehe und mit einer Geldstrafe von 50 bis 250 Euro, (...) bestraft werden kann, wenn ich (...)
Verwandter in gerader absteigender Linie	5-WB	N PRÄP ADV ADJ N	-	-	-	-	-	Verwandter in gerader absteigender Linie (z. B. Kinder und Enkelkinder)
Verwandter in gerader aufsteigender Linie	5-WB	N PRÄP ADV ADJ N	-	-	-	-	-	Verwandter in gerader aufsteigender Linie (z. B. Eltern und Großeltern)
Verwandtschaftsverhältnis	1-WB	N	-	-	-	-	-	Urkunde über das Verwandtschaftsverhältnis
Visum	1-WB	N	Visums	-	-	-	-	Erteilung eines Visums
Vollziehung	1-WB	N	-	-	-	-	-	Vollziehung des Ausländerbeschäftigungsgesetzes
Vollziehung der Schubhaft	3-WB	N ART N	-	-	-	-	-	Durchsetzung einer Rückkehrentscheidung, eines Aufenthaltsverbots, einer Ausweisung, einer Zurückschiebung oder der Vollziehung der Schubhaft
Voraussetzung	1-WB	N	Voraussetzungen	-	-	-	-	Fortbestand der Voraussetzungen
Vorbeschäftigungszeiten	1-WB	N	-	-	-	-	-	Nachweis über ausreichende Vorbeschäftigungszeiten im Unternehmen
Wegfall	1-WB	N	-	-	-	-	-	Wegfall des Hindernisses
Wegzug	1-WB	N	-	-	-	-	-	Wegzug des zusammenführenden EWR-Bürgers
Werkvertrag	1-WB	N	-	-	-	-	-	Schriftlicher Werkvertrag über die Leistung einer bestimmten selbständigen Tätigkeit
Wohnrecht	1-WB	N	Wohnrechts	-	-	-	-	Der/Die Unterkunftgeber/in räumt dem/der Unterkunftnehmer/in ein Wohnrecht in (...) ein.
Wohnrechtsvereinbarung	1-WB	N	-	-	-	-	-	Der Wohnrechtsvereinbarung ist der dieser Unterkunft zugrunde liegende Mietvertrag der Unterkunftgeberin/ des Unterkunftgebers in Kopie anzuschließen.

Basisterminus	N-WB	Muster	(im Korpus auftretende) Flexionsformen	Variante	Art der Variante	N-WB (Variante)	Muster	Kollokation / Kontext im Referenzkorpus
Wohnrechtsverhältnis	1-WB	N	-	-	-	-	-	Das Wohnrechtsverhältnis kann nicht jederzeit widerrufen werden (...)
Zulassungsentscheidung	1-WB	N	-	-	-	-	-	die in der Zulassungsentscheidung vorgeschriebene Ergänzungsprüfung
Zurückschiebung	1-WB	N	-	-	-	-	-	Durchsetzung eines Rückkehrentscheidung, eines Aufenthaltsverbots, einer Ausweisung, einer Zurückschiebung
Zusammenführender	1-WB	N	Zusammenführenden	-	-	-	-	Haftungserklärung des Zusammenführenden
zuständige Behörde	2-WB	ADJ N	zuständigen Behörde	-	-	-	-	urkundlicher Nachweis einer zuständigen Behörde des Herkunftsstaates
Zustellungsbevollmächtigten	1-WB	N	-	Zustellbevollmächtigten	morphologisch	1-WB	N	(...), dass Sie einen Zustellbevollmächtigten in Österreich bekanntgeben, d. h. die Behörde schickt dann alle Briefe an diese Person.
Zweckänderungsantrag	1-WB	N	-	-	-	-	-	-
zwingende Erforderlichkeit	2-WB	N	-	-	-	-	-	schriftliche Erklärung des Zusammenführenden über die zwingende Erforderlichkeit der persönlichen Pflege durch den Zusammenführenden

Abstract (Deutsch)

Verwaltungssprache wird im Gegensatz zur Fachsprache in anderen Sachgebieten nicht nur in der internen Kommunikation zwischen bestimmten Berufsgruppen oder ExpertInnen verwendet, sondern erstreckt sich auch darüber hinaus auf die BürgerInnen. Im Hinblick auf Österreich stellt die Sprachbarriere, die durch die komplexen sprachlichen Merkmale der österreichischen Verwaltungsterminologie verursacht wird, eine Herausforderung in Verwaltungsverfahren dar, in denen mündliche und schriftliche Interaktionen zwischen Parteien innerhalb und außerhalb des Fachgebiets stattfinden. Vor allem im Verwaltungsbereich Einwanderung und Aufenthalt stammen die Betroffenen, die mit dieser Fachsprache in Berührung kommen, aus anderen Herkunftsländern: Dementsprechend sind diese nicht nur Laien, sondern verfügen möglicherweise auch nicht über ausreichende Sprachkenntnisse in (Österreich-)Deutsch.

Die vorliegende Masterarbeit will diese Herausforderung in Verbindung mit den Prinzipien praktischer Terminologiearbeit angehen, wobei der Schwerpunkt auf der quantitativen Leistungsevaluierung von drei unterschiedlichen automatischen Terminologieextraktionswerkzeugen bei der Verarbeitung der österreichischen Verwaltungsterminologie liegt. Das Ziel besteht darin, einen (oder mehrere) geeignete Ansätze für die Extraktion dieser fachsprachlichen Terminologie zu ermitteln, wobei insbesondere deren Sprachbesonderheiten berücksichtigt werden. Eine Sammlung von domänenspezifischen Begriffen, die manuell aus dem Korpus gewonnen wurden – Reference Term List (RTL) – dient als Goldstandard für den Annotationsprozess durch eine sachkundige Person. Zur Darstellung der vergleichenden Leistung aller drei Tools werden Recall- und Precision-Werte sowie ein durchschnittliches F1-Maß ermittelt.

Anhand der durch die Praxisanalyse untermauerten Erkenntnisse im Rahmen der vorliegenden Arbeit soll ein Beitrag zu den laufenden Bemühungen geleistet werden, Verwaltungstexte zu optimieren, um eine höhere Verständlichkeit zu erreichen und den Informationszugang für die Öffentlichkeit zu verbessern, nämlich dass sie Terminologen bei der Erarbeitung potenzieller mehrsprachiger Terminologieressourcen unterstützen könnte. Letztendlich würde damit eine Grundlage geschaffen, die sowohl menschliche als auch maschinengestützte Übersetzungen der bestehenden Antragsformulare sowie anderer relevanter Dokumente in weitere Sprachen ermöglicht.

Abstract (English)

Administrative language, contrary to specialized language in other subject fields, is not only used in internal communication among particular professional groups or experts, but also extends beyond those to citizens. In terms of Austria, a language barrier caused by the complex linguistic characteristics of the Austrian administrative terminology has been a challenge in administrative procedures where verbal and written interactions between parties in and outside of the domain occur. Especially in the area of immigration and residence, individuals encountering this specialized language are from other countries; therefore, apart from being laypeople, they might also not possess sufficient language proficiency in (Austrian-)German.

The present master's thesis aims to address this challenge in connection with the principles of practical terminology work, focusing on the performance assessment of three distinct automatic terminology extraction tools on a domain-specific corpus comprised of official forms for residence permit applications. The goal is to determine one (or more) suitable approaches for extracting Austrian administrative terminology, taking into particular account its unique linguistic properties. A set of domain-specific terms manually collected from the corpus, a Reference Term List (RTL), will serve as a gold standard for the annotation process by a human judge. The annotation outputs are then calculated in relation to the RTL to determine recall and precision measures which will reflect the comparative performance of all three tools. A subsequent calculation of F1-scores will conclude their overall performance and determine the tool achieving the highest efficiency score.

With the knowledge substantiated by the practical analysis, the present thesis hopes to contribute to the ongoing efforts to optimize administrative texts to achieve higher comprehensibility and improve informational access for the public, namely that it could facilitate terminologists in the elaboration of potential multilingual terminology resources. Eventually, this would provide a basis that enables both human and machine-assisted translations of the existing application forms as well as other relevant documents in further languages.