
Das Konzept zentraler Wortschätze

Bestandsaufnahme,
theoretisch-methodische Weiterführung
und praktische Untersuchung

Veronika Haderlein



München 2008

Das Konzept
zentraler Wortschätze

Bestandsaufnahme,
theoretisch-methodische Weiterführung
und praktische Untersuchung

Veronika Haderlein

Dissertation
an der Fakultät für Sprach- und Literaturwissenschaften
der Ludwig-Maximilians-Universität
München

vorgelegt von
Veronika Haderlein
aus München

München, Februar 2008

Erstgutachter: Prof. Dr. Elmar Seebold

Zweitgutachter: Prof. Dr. Wolfgang Schindler

Tag der mündlichen Prüfung: 30. Januar 2007

meinen Eltern

Danksagung

Mein erster Dank geht an Professor Elmar Seebold, vor allem für die kontinuierliche Unterstützung in sowohl fachlichen als auch arbeitstechnischen Fragen und die Geduld, mit der er dieses sich über viele Jahre streckende Projekt begleitet hat. Danken möchte ich ihm aber auch für den Unterricht im Grund- und Hauptstudium, den er mit viel Begeisterung betrieben hat und ohne den die Lust auf eine Promotion gar nicht entstanden wäre.

Mein Dank gilt auch Privatdozent Dr. Wolfgang Schindler für die aufmunternde Unterstützung in fachlichen und promotionstechnischen Fragen und Professor Franz Guenther für viele inspirierende Ideen und konstruktive Beiträge zu Fragen der Computerlinguistik.

Ganz herzlich danken möchte ich auch Astrid Mischke für die Geduld, mit der sie frühe Versionen des Dissertationstextes gelesen und kritisch beurteilt hat, Daniel Schnorbusch für die mühsame Arbeit des Korrekturlesens und seine umsichtige und ehrliche Kritik und Øystein Haug Olsen für all die Hilfe in datenbanktechnischen Fragen.

Zuletzt geht mein besonderer Dank an Åsne Thea Haaland, Berit Winterhalder und Carolin Daum sowie vor allem an meine Brüder Andreas und Peter und meine Eltern Anna Maria und Peter Haderlein. Ohne deren kontinuierliche emotionale und praktische Unterstützung, ihr Vertrauen und ihre Zuversicht über die vielen Jahre meiner Schul- und Studienzeit hinweg wäre diese Arbeit nicht möglich gewesen.

Inhaltsverzeichnis

1	Einführung	
1.1	Gegenstand, Ziel und Motivation dieser Arbeit	1
1.2	Vorgehensweise und Aufbau	3
1.3	Zielgruppe	3
1.4	Grenzen	4

TEIL A: BESTANDSAUFNAHME

2	Der forschungsgeschichtliche Hintergrund	
2.1	Prägende linguistische Anwendungen	6
2.1.1	Die Grundwortschatzlexikographie	6
2.1.2	Die Methode der Glottochronologie	9
2.2	Wichtige methodische Arbeiten der bisherigen Forschungsgeschichte	10
2.2.1	Kühn 1979: Die Hinwendung zur kommunikativen Funktion	11
2.2.2	Krohn 1992: Die Suche nach dem thematischen Profil verschiedener Grundwortschätze	14
2.2.3	Bonan-Garrigues 1993: Die Erarbeitung einer Wortschatzhierarchie mithilfe des Verfahrens der transparenten Introspektion	19
2.2.4	Schnörch 2002: Der Vergleich verschiedener Grundwortschätze auf der Basis lexikalischer Einheiten	21
3	Anwendungstypen des Konzepts zentraler Wortschätze	
3.1	Die Grundwortschätze der Sprachdidaktik	27
3.2	Das kontrollierte definitonische Vokabular der Fremdsprachenlexikographie	29
3.3	Zentrale Wortschätze in Anwendungen der maschinellen Sprachverarbeitung	32
3.3.1	Die Stopwortlisten des <i>Information Retrieval</i>	33
3.3.2	Der Neutralwortschatz zur Verfeinerung von Anwendungen der automatischen Textklassifikation	34
3.4	Abgrenzungen	35
3.4.1	Häufigkeitwörterbücher	35
3.4.2	Die <i>basic word list</i> der Glottochronologie	38
3.4.3	Rein quantitativ begrenzte Wortschätze der maschinellen Sprachverarbeitung	39
3.5	Eine Typologie	41
4	Probleme bei der Arbeit mit zentralen Wortschätzen – ein Überblick	
4.1	Mangelnde Einigkeit über den Begriffsapparat	43
4.2	Die Frage nach der Grundeinheit	45
4.2.1	Die Frage nach der Grundeinheit in der bisherigen Forschung.....	46
4.2.2	Die Grundeinheit zentraler Wortschätze und die Frage nach der Abgrenzung gegenüber anderen sprachlichen Einheiten.....	49
4.3	Die Frage nach der adäquaten Datengrundlage für eine Auswahl	53
4.4	Probleme der Selektion	56
4.4.1	Polyseme und Homonyme: Fragen der Bedeutungsabgrenzung.....	56
4.4.2	Der Umgang mit Fremdwörtern	60
4.4.3	Der Umgang mit Neologismen	62
4.4.4	Der Umgang mit Synonymen, regionalen Konkurrenten und Kohyponymen	64

4.4.5	Der Umgang mit Wortfeldern und ihren Grenzen	66
4.5	Das Merkmal der Zentralität und die Vielfalt der Selektionskriterien	67
4.5.1	Die Kontroverse um den Vorrang einer frequenzorientierten vs. einer pragmatischen Selektion	69
4.5.2	Einige exemplarische Selektionskriterien	71
4.5.3	Zur Relativität und Wechselwirkung der Selektionskriterien.....	87
4.6	Das Problem des Umfanges eines zentralen Wortschatzes.....	88
4.7	Fragen der Darstellung und Aufbereitung des zentralen Wortschatzes	90
4.8	Die Arbeit mit zentralen Wortschätzen im Schnittpunkt mehrerer linguistischer Disziplinen	93

TEIL B: THEORETISCH-METHODISCHE WEITERFÜHRUNG

5 Faktoren der Wortschatzschichtung

5.1	Einführung	96
5.2	Der Begriff des Lexikons	96
5.2.1	Der Terminus <i>Lexikon</i> und seine Verwendungsweisen	96
5.2.2	Die Abgrenzung von Wortschatz und Grammatik	98
5.2.3	Die Heterogenität der Einheiten des Wortschatzes.....	100
5.3	Makrostrukturelle Dimensionen der Wortschatzschichtung.....	102
5.3.1	Sprachvarietäten: die Abgrenzung von Gemeinsprache und Sondersprachen am Beispiel der Fachsprachen.....	102
5.3.2	Frequentielle Phänomene der Wortschatzschichtung	110
5.3.3	Weitere makrostrukturelle Dimensionen der Wortschatzschichtung	111
5.4	Das Kern-Peripherie-Modell des Prager Strukturalismus	111
5.5	Die Frage nach dem lexikologischen Zentrum einer Sprache gegenüber der Relativität zentraler Wortschätze	114
5.5.1	<i>coreness</i> und die Frage nach dem lexikologischen Zentrum einer Sprache	114
5.5.2	Lexikologisches Wortschatzzentrum und anwendungsspezifische zentrale Wortschätze – eine konzeptuelle Trennung	124
5.5.3	Zusammenfassender Rückblick	126

6 Entwurf eines Arbeitsmodells

6.1	Einführung	127
6.2	Das generische Wortschatzmodell	128
6.3	Das Konzept zentraler Wortschätze und seine Definition	130
6.4	Terminologische Klärungen	132

TEIL C: PRAKTISCHE UNTERSUCHUNG

7 Die praktische Untersuchung eines Lexikonausschnittes – Überblick über die Vorgehensweise

7.1	Ziel der praktischen Untersuchung	136
7.2	Besonderheiten dieser praktischen Untersuchung	137
7.3	Ausgangshypothesen und Grundsatzentscheidungen	138
7.3.1	Die allgemeine Ausrichtung der praktischen Untersuchung gegenüber der Frage nach dem Zweck eines zentralen Wortschatzes	138
7.3.2	Der Bestand des DUW als 'gute' Repräsentation des Wortschatzes des Deutschen	138
7.3.3	Die betrachteten Eigenschaften	139

7.3.4	Die Frage nach der Grundeinheit und das Problem der impliziten Selektion	140
7.3.5	Zur Trennung von Beschreibung und Gewichtung	141
7.3.6	Die Reihenfolge der Bearbeitung	142
7.4	Die praktischen Schritte im Überblick.....	143
8	Die Schritte der praktischen Untersuchung im Einzelnen	
8.1	Die Erfassung der Hauptstichwörter des DUW	145
8.2	Die Beschreibung der Einträge nach ihrer Repräsentiertheit	145
8.3	Die detaillierte Erfassung der Stichwortangaben des DUW und die Bildung von lexikalischen Einheiten	149
8.4	Die Beschreibung der Plausibilität	154
8.5	Die Beschreibung der Frequenz	155
8.5.1	Das Projekt <i>Deutscher Wortschatz</i> der Universität Leipzig	156
8.5.2	Probleme bei der Bestimmung der Frequenz.....	156
8.5.3	Konsequenzen	158
8.6	Die Beschreibung der zeitlichen Stabilität.....	159
8.6.1	Zur Eignung des Vergleichs von Wörterbüchern zur Bestimmung der Stabilität lexikalischer Einheiten	160
8.6.2	Die herangezogenen Wörterbücher	162
8.6.3	Die Frage nach der lexikalischen Identität.....	164
8.7	Die Beschreibung der Neutralität.....	171
8.8	Die Gewichtung der Eigenschaften	172
8.8.1	Vorüberlegungen	172
8.8.2	Die Bewertung der Repräsentiertheit	172
8.8.3	Die Bewertung des Grades der Worthaftigkeit	174
8.8.4	Die Bewertung der Plausibilität.....	179
8.8.5	Die Bewertung der Frequenz.....	179
8.8.6	Die Bewertung der Stabilität	181
8.8.7	Die Bewertung der Neutralität	182
8.9	Das Bewertungssystem im Überblick	185
9	Ergebnisse und Beobachtungen	
9.1	Einführung	187
9.2	Der Wirkungsgrad und die Wechselwirkung der verschiedenen Selektionseigenschaften im Vergleich	187
9.3	Die Bildung einer Referenzselektion und ihre Analyse	194
9.3.1	Die konkreten Schritte der Erarbeitung einer Referenzselektion.....	194
9.3.2	Eine Annäherung an die Frage des Umfanges eines zentralen Wortschatzes des gesamten DUW-Bestandes.....	195
9.3.3	Das Verhältnis zwischen der Extension eines Lexems und seiner kumulativen Zentralität	197
9.3.4	Die Zusammensetzung der Referenzselektion hinsichtlich der Wortarten	198
9.3.5	Ein Blick auf die semantische Zusammensetzung der Referenzselektion	199
9.4	Methodische Beobachtungen bei der praktischen Untersuchung	200
9.4.1	Die Durchführbarkeit der Trennung von Beschreibung und Bewertung	201
9.4.2	Die lexikalische Einheit als Grundeinheit.....	201
9.4.3	Die Beschreibung und Bewertung der introspektiven Eigenschaft der Plausibilität	202
9.4.4	Begrenzungen durch die Vorauswahl des Buchstaben O	202

10 Zusammenfassung und Ausblick
10.1 Erkenntnisse und Kernaussagen..... 205
10.2 Weiterführendes 207

ANHANG: TABELLEN UND VERZEICHNISSE

A: Tabellenauszüge aus der zugrundeliegenden Datenbank..... 210
B: Literaturverzeichnis..... 221
Curriculum Vitæ 235

Kapitel 1: Einführung

1.1 Gegenstand, Ziel und Motivation dieser Arbeit

Gegenstand der hier vorgelegten Arbeit ist die Neubetrachtung und Weiterführung des bisher vor allem unter dem Namen *Grundwortschatz* bekannten lexikographischen Konzeptes zentraler Wortschätze.

Dieses Konzept hat im Bereich der Didaktik des Deutschen bereits eine lange Tradition; dort wird es verwendet, um spezielle reduzierte Wortschätze für den Erst- und Zweitspracherwerb zu erstellen. Mit Hilfe dieser Wortschätze soll der Spracherwerb effektiver gestaltet werden, indem der Lernende mit den zentralsten Einheiten des Wortschatzes zuerst konfrontiert wird. Diese Versuche sind wegen des Mangels an objektiven Kriterien dafür, welche Einheiten des Wortschatzes zentral sind und wie eine Selektion für einen Grundwortschatz aussehen könnte, immer kontrovers diskutiert worden.

Auch in der Sprachverwandtschaftsforschung hat man das Konzept zentraler Wortschatzelemente diskutiert, wenn man etwa im Rahmen der Glottochronologie bzw. Lexikostatistik versuchte, die verwandtschaftliche Nähe zweier Sprachen zueinander festzustellen, indem man vergleicht, wie eine vorab definierte Liste zentraler Begriffe in den einzelnen Sprachen realisiert wird. Darüberhinaus wird in der allgemeinen linguistischen Forschung immer wieder die Frage diskutiert, ob es so etwas wie ein absolutes Zentrum des Wortschatzes einer Sprache überhaupt gibt und wie dieses definiert sein könnte.

In der Computerlinguistik schließlich findet das Konzept zentraler Wortschätze ebenfalls Anwendung, wenn auch bisher nur implizit und ohne direkten Bezug zu den entsprechenden praktischen Anwendungen in der Sprachdidaktik; so etwa in den sogenannten Stopwortlisten des *Information Retrieval* oder in den Neutralwortschätzen der maschinellen Klassifikation von Texten.

Das Konzept zentraler Wortschätze – hier vorläufig umschrieben als Wortschatzausschnitte, die die für eine spezifische linguistische Anwendung ‘zentralen’ Wörter enthalten¹ – wird also in verschiedenen Bereichen der angewandten Linguistik aktiv eingesetzt. Explizit diskutiert worden ist es bisher aber fast nur von Sprachdidaktikern und Lexikographen. Lexikologen, theoretisch orientierte Linguisten und Computerlinguisten haben sich bisher kaum oder nur am Rande dazu geäußert. Was fehlt, ist eine Betrachtung des Konzeptes von einem weiter gefassten, anwendungsübergreifenden Standpunkt aus. Die bisherigen Diskussionsbeiträge bestehen zu einem großen Teil aus der Erarbeitung konkreter Grundwortschätze für verschiedene sprachdidaktische Anwendungen und aus kritischen Beurteilungen dieser Versuche. Die wenigen Arbeiten, die sich diesem Konzept von einem theoretisch-methodischen Winkel her nähern – Kühn 1979, Krohn 1992, Bonan-Garrigues 1993 und Schnörch 2002 –, tun dies fast ausschließlich im Hinblick auf den konkreten Be-

1. Für eine eingehendere Definition vgl. Kapitel 6.

reich der Sprachdidaktik, indem sie etwa konkrete (sprachdidaktisch orientierte) Grundwortschätze analysieren und anhand dieser Ergebnisse spezifische Fragen der Sprachdidaktik diskutieren. Grundlegende methodische und lexikologische Fragen werden entweder im Lichte der Sprachdidaktik besprochen oder vernachlässigt. So fehlt etwa die explizite Einbindung von Erkenntnissen der Lexikologie zur Makrostruktur des Wortschatzes, die Zusammenstellung der verschiedenen konkreten Anwendungen in Sprachdidaktik und Computerlinguistik unter einem konzeptuellen Dach, die Etablierung einer konsistenten Terminologie oder eine konkrete Wortschatzanalyse zur Beantwortung lexikologischer Fragen unabhängig von einer praktischen Anwendung. Das Fehlen einer solchen übergreifenden Betrachtung ist zwar nachvollziehbar, weil das Konzept zentraler Wortschätze zunächst aus einer praktischen Notwendigkeit heraus in verschiedenen Bereichen entstanden ist; theoretisch-methodische Überlegungen, die über die einzelne praktische Anwendung hinausgehen, mussten dabei vorläufig in den Hintergrund rücken. Dennoch ist eine solch übergreifende Betrachtung nötig, um die Arbeit an zentralen Wortschätzen stärker lexikologisch zu verankern und Synergien zwischen den verschiedenen Anwendungsbereichen zu erzielen. Dazu soll die vorliegende Arbeit einen Beitrag leisten.

Konkret hat die Arbeit dabei folgende Ziele:

Zum einen soll sie einen Überblick über die bisherigen Anwendungsbereiche des Konzeptes zentraler Wortschätze und die wichtigsten mit ihm verbundenen Diskussionen und Probleme geben und damit zur Ausweitung und Neudefinition des Konzeptes von der reinen Sprachdidaktik auf verschiedene Anwendungsbereiche beitragen. Dabei soll die Arbeit zugleich eine Brücke schlagen zwischen traditioneller Linguistik und Computerlinguistik: einerseits, indem sie zeigt, dass das bisher unter dem Namen *Grundwortschatz* geführte Konzept nicht allein auf die Bereiche der traditionellen Linguistik beschränkt ist, sondern durchaus auch in einem größeren Rahmen betrachtet werden kann, und andererseits, indem sie ein grundlegendes Verständnis der mit diesem Konzept verbundenen Probleme, die so explizit in der Computerlinguistik nicht diskutiert werden, vermittelt.

Zum anderen soll die Arbeit einige wesentliche lexikologische Aspekte, die mit dem Konzept zentraler Wortschätze verknüpft sind, beleuchten und damit zu einer theoretisch-methodischen Weiterführung des Konzeptes beitragen.

Und schließlich soll die Arbeit anhand einer konkreten praktischen Analyse eines Wortschatzausschnittes einige spezifische lexikologische Fragen und methodische Probleme zu diesem Wortschatzkonzept diskutieren und mögliche Antworten und Lösungsvorschläge skizzieren.

Damit geht es bei dieser Arbeit also nicht darum, einen konkreten und vollständigen zentralen Wortschatz zu erstellen oder konkrete zentrale Wortschätze zu analysieren. Das Ziel ist vielmehr, den theoretisch-methodischen Rahmen dieses Konzeptes zu systematisieren,

konkrete lexikologische Fragestellungen und lexikographische Probleme zu bearbeiten und damit eine Lücke in der bisherigen Forschung zum Thema zentraler Wortschätze zu schließen.

1.2 Vorgehensweise und Aufbau

Die vorliegende Arbeit hat folgenden Aufbau:

In einem ersten Teil (Kapitel 2 bis 4) werde ich in Form einer Bestandsaufnahme einen Überblick über den bisherigen Stand der Forschung zum Konzept zentraler Wortschätze geben. Dabei soll aufgezeigt werden, in welchen Anwendungsbereichen der Linguistik das Konzept zentraler Wortschätze explizit oder implizit verwendet wird und welche Probleme mit ihm verbunden sind.

In einem zweiten Teil (Kapitel 5 und 6) werde ich in einem Versuch der theoretisch-methodischen Weiterführung wesentliche Erkenntnisse der Lexikologie zu Fragen der Makrostruktur des Wortschatzes überblickshaft zusammenstellen und darauf aufbauend ein Arbeitsmodell einer Wortschatzstruktur entwickeln, das es ermöglicht, konkrete zentrale Wortschätze an ein lexikologisches Modell anzuknüpfen. Darüberhinaus werde ich an dieser Stelle einige terminologische Klärungen vornehmen, um so zur Etablierung einer konsistenten Terminologie beizutragen.

In einem dritten Teil (Kapitel 7 bis 9) werde ich mich ausgewählten lexikologischen Problemen des Konzeptes zentraler Wortschätze zuwenden und sie anhand eines konkreten Wortschatzausschnittes untersuchen und auswerten. Im Zentrum wird dabei die Frage stehen, wie sich bestimmte Eigenschaften von Wörtern für die Erstellung zentraler Wortschätze nutzbar machen lassen, in welcher Wechselbeziehung diese Eigenschaften zueinander stehen und wie ein transparentes Verfahren zur Erstellung eines zentralen Wortschatzes aussehen könnte.

Kapitel 10 schließlich wird sich mit einer Zusammenfassung der gesamten Überlegungen und Fragen der weiteren Forschung beschäftigen.

1.3 Zielgruppe

Die Arbeit wendet sich auf der einen Seite an Linguisten der allgemeinen germanistischen Linguistik, indem sie versucht, das Konzept zentraler Wortschätze aus der Sprachdidaktik herauszulösen und in einen allgemeinen lexikologischen und lexikographischen Zusammenhang zu stellen; sie wendet sich auf der anderen Seite an Computerlinguisten, indem sie versucht, einen Überblick zu geben über die konkreten Probleme und Möglichkeiten, die mit der Arbeit an zentralen Wortschätzen verbunden sind.

1.4 Grenzen

Diese Arbeit soll einen Beitrag zur theoretisch-methodischen Forschung rund um das Konzept zentraler Wortschätze sein. Sie kann daher keinen konkreten, vollständigen und sofort einsetzbaren zentralen Wortschatz welcher Form auch immer bereitstellen, eben weil sie einen Schritt früher ansetzt als ein konkretes Wortschatzprojekt.

Darüber hinaus beschränkt sich diese Arbeit auf die deutsche Sprache als Forschungsgegenstand, auch wenn Forschungsarbeiten aus anderen Sprachen mitberücksichtigt werden. Damit sollen Ergebnisse dieser Untersuchung zunächst allein für das Deutsche Geltung haben. Inwieweit sie auf andere Sprachen übertragbar sind, muss offen bleiben und wird zu meist in der Arbeit nicht berücksichtigt.

Teil A: Bestandsaufnahme

Überblick

Im vorliegenden Teil wird es nun um den ersten der drei Schwerpunkte dieser Arbeit gehen – um eine Bestandsaufnahme der bisherigen Ergebnisse der Forschung und praktischen Anwendung zum Konzept zentraler Wortschätze. Ziel dieser Bestandsaufnahme ist es, einen Überblick über den Forschungsstand zur Arbeit mit diesem Konzept zu geben und alle wichtigen Problem- und Fragestellungen gesammelt darzustellen. Zum einen soll damit eine Grundlage für alle nachfolgenden Überlegungen der theoretisch-methodischen Weiterführung und praktischen Analyse geschaffen werden. Zum anderen soll mit dieser Bestandsaufnahme ein Beitrag dazu geleistet werden, das Thema aus dem Bereich der Sprachdidaktik herauszulösen, seine Bedeutung für andere Fachbereiche herauszuarbeiten und einen Überblick darüber zu geben, in welchen thematischen Rahmen dieses Konzept eingebettet ist.

Dabei werden wir das Thema von drei Seiten her beleuchten:

Von einem forschungsgeschichtlichen Standpunkt aus (Kapitel 2). Hier wird es um die wichtigsten Strömungen und methodischen Arbeiten der bisherigen Forschung zu zentralen Wortschätzen gehen.

Von einem anwendungstypologischen Standpunkt aus (Kapitel 3). Ziel dieses Kapitels wird es sein, darzulegen, in welchen anderen Bereichen der angewandten Linguistik als dem der Grundwortschatzlexikographie das Konzept zentraler Wortschätze mehr oder minder explizit Verwendung findet.

Von einem praktisch-lexikographischen Standpunkt aus (Kapitel 4). Schließlich gehört zu dieser Bestandsaufnahme auch ein Überblick über die Probleme, die mit dem Konzept zentraler Wortschätze bisher verbunden sind. Insbesondere dieser Überblick wird dann die Grundlage bilden für die beiden anderen Teile dieser Arbeit, die methodisch-theoretische und die praktische Weiterführung.

Kapitel 2: Der forschungsgeschichtliche Hintergrund

2.1 Prägende linguistische Anwendungen

Das Konzept zentraler Wortschätze ist aus praktischen lexikographischen Notwendigkeiten heraus entstanden. Insofern ist es auch nicht verwunderlich, dass die bisherige Forschungsliteratur dazu weniger aus rein theoretischen Arbeiten besteht, sondern vielmehr von der Suche nach Lösungen für individuelle praktische Vorhaben geprägt ist. Primus motor der Diskussionen waren dabei bisher die Grundwortschatzforschung und die Glottochronologie. Diese beiden linguistischen Anwendungen sollen im Folgenden kurz umrissen und in ihrem Bezug zum Konzept zentraler Wortschätze dargestellt werden.

2.1.1 Die Grundwortschatzlexikographie

Der Grundwortschatz als eine „didaktisch motivierte Auswahl aus dem lexikalischen System einer Sprache“ (Krohn 1992, S. 125) ist ein Konzept, das schon lange in der Sprachdidaktik Anwendung findet. Es ist die ‘Urform’ eines zentralen Wortschatzes, und die Arbeiten zu diesem Typus eines zentralen Wortschatzes bilden bis heute die Grundlage des gesamten Diskussionsmaterials.

Im Zentrum der Grundwortschatzforschung stehen dabei vor allem zwei Fragen: Wie lässt sich entscheiden, welche Wörter in einen Grundwortschatz aufgenommen werden, d.h. welche Selektionskriterien wendet man an? Und wie lässt sich die so getroffene Auswahl an Wörtern am besten darstellen? Diese beiden Aspekte – die Frage nach der Selektion und die Frage nach der Darstellung – sind grundlegende Fragestellungen der Grundwortschatzdiskussion und haben sie stark geprägt. Die Einzelheiten dieser Diskussion sind ausführlich in allen neueren Arbeiten zur Grundwortschatzlexikographie dokumentiert.² Hier soll deshalb nur ein kurzer chronologischer Abriss gegeben werden.

Ausgehend von der in Kühn 1990 ausführlich dargestellten Übersicht lassen sich die großen Züge der Forschungsgeschichte in fünf Phasen einteilen:³ Am Anfang, d.h. in den 20er- und 30er-Jahren des 20. Jahrhunderts steht ein Ansatz, den man als den *absoluten frequenzorientierten Ansatz* bezeichnen kann und bei dem Grundwortschätze als „Abfallprodukte der Frequenzforschung“ (ebda., S. 1354) entstehen und direkt aus dem Häufigkeitswörterbuch von Friedrich Wilhelm Kaeding⁴ übernommen werden.⁵ Einziges Selektionskriterium ist

2. Vgl. etwa Kühn 1979, S. 23ff., Krohn 1992, S. 55ff. und Schnörch 2002, S. 11ff.

3. Anders Krohn 1992, S. 55. Er teilt die Geschichte der Grundwortschatzlexikographie in nur drei Phasen bzw. Ansätze ein – den frequenzbasierten, den kommunikativ-pragmatischen und den frequentieil-kommunikativen Ansatz. Der Unterschied zwischen der hier vorgenommenen Einteilung und derjenigen bei Krohn beruht darauf, dass Krohn den Ansatz von Kühn 1979 und den anderer pragmatisch orientierter Ansätze zusammenfasst.

4. Kaeding 1898, eine Frequenzuntersuchung mit dem Ziel, die Häufigkeit deutscher Wortformen zu bestimmen, um so das deutsche Stenographiesystem effektiver zu gestalten.

5. Vertreter dieses Ansatzes sind etwa Morgan 1928 oder Bakonyi 1934. – Für eine Darstellung der Geschichte der Frequenzlexikographie an sich vgl. Martin 1990, S. 1319f.

die im Kaeding angegebene unmittelbare Frequenz der Wortformen unter der „verhängnisvolle[n] Gleichung“ (ebda., S. 1355), die frequentesten Wörter seien auch zugleich die didaktisch nützlichsten. Das Ergebnis sind „nackte Wortlisten“ (ebda., S. 1354), die keinen echten sprachdidaktischen Wert haben, weil die frequenzbasierte Auswahl zu inkonsistenten Ergebnissen führt und zum Beispiel kommunikativ zusammengehörige Wörter nicht unbedingt geschlossen in den Grundwortschatz aufnimmt (vgl. Kühn 1990, S. 1355f.).

Die Schwächen der Grundwortschätze, die auf der unmittelbaren Übernahme der absoluten Frequenz von Wortformen im Kaeding aufbauen, führen in einer zweiten Phase zu einem Ansatz, den ich hier als den *modifiziert-frequenzorientierten Ansatz* bezeichnen möchte. In dieser zweiten Phase, die ihren Höhepunkt in den 50er- und 60er-Jahren des 20. Jahrhunderts erreicht, geht es darum, Frequenz als Selektionskriterium zu modifizieren mit dem Ziel, die Schwächen der absoluten Frequenz auszuräumen. Man entlehnt Verfahren der allgemeinen Statistik (Distribution, Standardabweichung, Dispersionsindex) und wendet sie auf Frequenzuntersuchungen von Wortschatz an. Ergänzt wird das Selektionskriterium der Frequenz durch nicht-statistische Kriterien wie zum Beispiel das Einbeziehen sogenannter „Verfügungswörter“ (Pfeffer 1975, S. 12), die gezielte Schließung von Lücken in Wortfeldern, die durch Frequenz allein nur lückenhaft in den Grundwortschatz aufgenommen würden, oder das Entfernen von Wörtern, die den Autoren des Grundwortschatzes entbehrlich erschienen.⁶ Dieser modifizierte Ansatz vergrößert die Reihe der Probleme allerdings nur, weil trotz der Einführung nicht-statistischer Kriterien die Grundprobleme der frequenzbasierten Auswahl bestehen bleiben und diesen Problemen noch durch die Einführung mehr oder minder willkürlicher Kriterien wie zum Beispiel dem Schließen von Wortfeldlücken das Problem der mangelnden Transparenz und Nachvollziehbarkeit des Verfahrens hinzugefügt wird.

Als Gegenreaktion auf die Tatsache, dass die Arbeiten der ersten beiden Phasen der Grundwortschatzlexikographie im Wesentlichen auf der Wortfrequenz als Selektionskriterium aufbauen, geht man in einer dritten Phase von der Frequenz als Selektionskriterium völlig ab und legt pragmatische Aspekte der Auswahl und Darstellung der Einheiten des Grundwortschatzes zu Grunde. Dabei geht man „von den grundlegenden Kommunikationsintentionen, Themen und Kommunikationssituationen“ (ZERTIFIKAT DEUTSCH 1977, S. 123) aus. Zeitlich fällt dieser Ansatz, den ich in Anlehnung an Kühn 1990 als *kommunikativ-pragmatischen Ansatz* bezeichnen möchte, in die 70er-Jahre des 20. Jahrhunderts, die von der Wende der Linguistik hin zur Pragmatik geprägt sind (vgl. Krohn 1992, S. 11).⁷ Problematisch bei diesem Ansatz ist die Tatsache, dass mit der alleinigen Orientierung am

6. Vertreter dieses Ansatzes sind etwa Gougenheim / Rivenc / Michéa u. a. 1956 für das Französische und Pfeffer 1964 für das Deutsche.

7. Vertreter dieses Ansatzes sind etwa das ZERTIFIKAT DEUTSCH 1977 oder van Ek 1977. Aber auch Krohn 1992 (vgl. Abschnitt 2.2.2) kann in gewisser Weise diesem Ansatz zugerechnet werden, weil er dem Thema einer Kommunikationssituation eine steuernde Funktion für die Auswahl der Lexik zuweist und ihm in seiner Untersuchung verschiedener Grundwortschätze großes Gewicht beimisst.

vage definierten Kriterium der kommunikativen Relevanz das Selektionsverfahren wenig transparent ist, weil die Bestimmung der Themen, die die kommunikative Relevanz und damit die Wortauswahl steuern, einen großen Spielraum lässt. Darüber hinaus ist bei diesem Selektionsverfahren eine nachvollziehbare und konsistente quantitative Begrenzung der Anzahl der Wörter nicht mehr möglich.

Peter Kühn führt diesen kommunikativ-pragmatischen Ansatz in seiner Arbeit „Der Grundwortschatz. Bestimmung und Systematisierung“ (Kühn 1979) weiter und baut ihn zu einem Vorgehen aus, das den Beginn der vierten Phase der Grundwortschatzlexikographie markiert und als *funktional-pragmatischer Ansatz* bezeichnet werden kann. Kühn gibt das Bemühen um quantitative Begrenzung völlig auf und fordert stattdessen, den Grundwortschatz nicht durch quantitative, sondern durch qualitative Merkmale überschaubar zu machen (vgl. Kühn 1979, S. 62). Erstmals lenkt er die Diskussion weg von der alleinigen Konzentration auf die Selektionskriterien hin zu Fragen der Darstellung, denen er viel Platz einräumt (vgl. ebda., Kapitel 4). Er fordert, den Grundwortschatz als ein Mittel zur „lexematischen Sprachkompetenzerweiterung“ (ebda., S. 60) zu verstehen, und bezeichnet konsequenterweise den von ihm konzipierten Grundwortschatz als ein „Grundwortschatz**wörterbuch**“ (ebda., S. 120 [Hervorhebung V. H.]). Für ein solches Grundwortschatzwörterbuch fordert er auch, die rein semasiologische Anordnung zugunsten einer „integrierende[n] Konzeption des onomasiologischen und semasiologischen Wörterbuchtyps“ (ebda., S. 120) in Form eines „Begriffssystems“ aufzugeben; er geht so gar so weit, onomasiologische Aspekte der Auswahl der Lexeme zugrunde zu legen. Allerdings bleibt er es schuldig, die Operationalisierbarkeit seiner Auswahlkriterien und Anordnung genau darzulegen und sein Projekt über einen Entwurf hinaus in die Praxis umzusetzen.⁸

Die fünfte und bisher letzte Phase der Grundwortschatzlexikographie wird von Ulrich Schnörch und seiner Arbeit „Der zentrale Wortschatz des Deutschen. Strategien zu seiner Ermittlung, Analyse und lexikografischen Aufarbeitung“ (Schnörch 2002) eingeleitet. Sein Ansatz, den ich als *semantisch-integrativen Ansatz* bezeichnen möchte, geht im Unterschied zu allen bisherigen Vorschlägen vom Semem als Grundeinheit eines Grundwortschatzes aus. Er verstärkt die von Kühn bereits initiierte Verlagerung der Diskussion weg von Fragen der Selektion hin zu Fragen der Darstellung der Elemente eines Grundwortschatzes, indem er die semantischen Beziehungen der aus der Schnittmenge verschiedener Grundwortschatze gewonnenen Sememe zueinander in den Vordergrund stellt.⁹ Zugleich steht er den verschiedenen Selektionsverfahren der frequenzorientierten bzw. pragmatischen Ansätze offen gegenüber und führt sie in einer integrierenden Sichtweise zusammen, indem er sowohl frequenzbasierte als auch pragmatisch orientierte Selektionskriterien als nützlich und legitim anerkennt.¹⁰

8. Für eine detaillierte Darstellung der Arbeit von Kühn 1979 vgl. Abschnitt 2.2.1.

9. Eine Gegenströmung stellt Zillig 2000 dar, der den Grundwortschatz grundsätzlich an den Wortformen orientieren und bei der Selektion vom passiven Wortschatz des Verstehens ausgehen will (vgl. Zillig 2002, S. 93).

Fasst man die in den Diskussionen in den einzelnen Ansätzen diskutierten Frage- und Problemstellungen zusammen, dann handelt es sich dabei im Großen und Ganzen um folgende: Welche Selektionskriterien sind besser geeignet, einen Grundwortschatz für die Sprachendidaktik zu erstellen – frequenzbasierte oder pragmatisch orientierte Selektionskriterien? Ist die Frequenz eines Wortes überhaupt Ausdruck seiner Zentralität und Wichtigkeit für jemanden, der Deutsch lernt? Wie könnte ein objektives Auswahlverfahren aussehen? Wie wichtig ist Objektivität überhaupt? Wie stellt man die Einträge des Grundwortschatzes so dar, dass sie den Sprachenunterricht unterstützen und den Wortschatzerwerb erleichtern? Gibt es überhaupt einen einzigen Grundwortschatz oder muss man von verschiedenen ausgehen? Fragen danach, wie sich das Konzept, auf dem ein Grundwortschatz beruht, definieren lässt, wie der Terminus *Grundwortschatz* gegenüber anderen Begriffen abzugrenzen ist, und schließlich, wie sich dieses Konzept auch auf andere Anwendungsbereiche außerhalb der Sprachdidaktik übertragen lässt, bleiben dabei jedoch vielfach unberücksichtigt.

Heute wird das Konzept Grundwortschatz in der Sprachdidaktik durchaus noch aktiv in der praktischen Lexikographie verwendet, wie ein Blick in die Fremdsprachenabteilung der Buchhandlungen zeigt;¹¹ die linguistisch-wissenschaftliche Diskussion ist allerdings – abgesehen von einigen wenigen Arbeiten (vgl. Abschnitt 2.2) – im Vergleich zu den frühen Jahren der Grundwortschatzforschung deutlich zurückgegangen.

2.1.2 Die Methode der Glottochronologie

Die Methode der Glottochronologie scheint auf den ersten Blick auch das Konzept zentraler Wortschätze in Gebrauch zu nehmen, allerdings aus einer ganz anderen Motivation heraus, als das bei der Grundwortschatzlexikographie der Fall ist. Bei dieser Methode geht es darum, die verwandtschaftliche Nähe zweier Sprachen mit Hilfe einer gegebenen Liste von Begriffen, der sogenannten *basic word list*, zu bestimmen. Die Methode geht auf die Arbeiten von Morris Swadesh¹² zurück und lehnt sich an die in der Paläontologie übliche sogenannte C-14-Methode zur Bestimmung des Alters von biologischen Funden an, bei der man anhand der Halbwertszeit des radioaktiven Kohlenstoff-Isotops ¹⁴C und dem Restwert des im Material noch enthaltenen ¹⁴C bestimmen kann, wann der untersuchte biologische Fund abgestorben ist. In ähnlicher Weise untersucht man in der Glottochronologie die *basic word list* dahingehend, wie die in ihr enthaltenen Begriffe in den zu vergleichenden Sprachen rea-

10. Für eine detaillierte Darstellung der Arbeit von Schnörch 2002 vgl. Abschnitt 2.2.4.

11. Vgl. zum Beispiel *Thematischer Grund- und Aufbauwortschatz Englisch* (Häublein / Jenkins 2004), *Langenscheidts Grundwortschatz Latein* (Fink / Strehl 2001) oder *2000 x Minuten-Training. Russisch Grundwortschatz* (Garibiani 2000).

12. Vgl. zum Beispiel Swadesh 1950 und 1955. Andere Verfechter der Methode der Glottochronologie sind zum Beispiel Sheila Embleton (vgl. etwa Embleton 1986 und Embleton 2000) oder Isidore Dyen (vgl. etwa Dyen 1975 oder Dyen / Kruskal / Black 1992). – Für einen umfassenden Überblick zur Glottochronologie vgl. Tischler 1973. Ein Beispiel der kritischen Betrachtung dieser Methode ist der Beitrag von Bergsland / Vogt 1962. Für eine Beispieluntersuchung der Verwendbarkeit der Lexikostatistik zum synchronen Vergleich der Lexik einzelner Sprachen vgl. Haderlein 1997.

lisiert sind. Man geht dabei davon aus, dass sich die Entwicklung einer Sprache mit einer konstanten Geschwindigkeit vollzieht und dass zentrale Begriffe in der Regel durch sehr stabile Einheiten des Wortschatzes repräsentiert werden, also kaum Veränderungen der Form oder Bedeutung ausgesetzt sind. Vor diesem Hintergrund versucht man, anhand der Menge an phonetischen und morphologischen Gemeinsamkeiten Rückschlüsse zu ziehen darüber, wie nah sich zwei Sprachen lexikalisch stehen bzw. wann sich die beiden Sprachen von ihrer gemeinsamen Ursprungssprache gelöst haben.¹³

Genau diese Arbeitshypothese von der Berechenbarkeit des Trennungszeitpunktes zweier Sprachen ist es allerdings, die die Methode der Glottochronologie so umstritten gemacht hat und deretwegen sie letztlich von den meisten Linguisten aufgegeben wurde. Die Hauptkritikpunkte der Gegner dieser Methode richten sich dabei auf die Unmittelbarkeit, mit der Swadesh die Unterschiede in der Realisierung eines Begriffes in zwei Sprachen auf eine kontinuierliche historische Entwicklung zurückführt und Phänomene des Sprachkontakts ignoriert, und auf die Tatsache, dass Swadesh nur Begriffe zulässt, die im Englischen durch Simplicia realisiert sind, womit das Begriffskonzept konterkariert wird.¹⁴

Trotz ihres Status als einer Methode mit „allenfalls heurist[ischem] Wert“ (Glück (Hg.) 2000, S. 253) hat das Verfahren der Glottochronologie viel zur Diskussion rund um die Fragen nach einem stabilen Wortschatzkern beigetragen und darf deshalb bei einem forschungshistorischen Überblick zum Thema zentraler Wortschätze nicht fehlen. Ob ihre *basic word list* selbst allerdings als ein Anwendungstyp zentraler Wortschätze betrachtet werden kann, werden wir noch diskutieren (vgl. Abschnitt 3.4.2).

2.2 Wichtige methodische Arbeiten der bisherigen Forschungsgeschichte

Die beiden genannten Verfahren der angewandten Linguistik – Grundwortschatzlexikographie und Glottochronologie – haben einen reichen Bestand an Diskussionsbeiträgen rund um das Konzept zentraler Wortschätze hervorgebracht. Besonders die Diskussionen der beiden Hauptströmungen der Grundwortschatzlexikographie – dem frequenzorientierten und dem kommunikativ-pragmatischen Ansatz – nehmen in der Literatur viel Raum ein und

-
13. Die synchrone Ausrichtung dieses Ansatzes, also der Versuch der Bestimmung der synchronen lexikalischen Nähe zweier Sprachen, bezeichnet man als (sprachvergleichende) *Lexikostatistik*, während der diachrone Ansatz der Bestimmung des Trennungszeitpunktes zweier Sprachen als *Glottochronologie* bezeichnet wird (vgl. etwa Lerchner 1968, S. 259 und Wang 1994, S. 1445). Anders Crystal (1997, S. 221), der die Lexikostatistik als ein praktisches Verfahren betrachtet, das unter anderem in der Glottochronologie zum Einsatz kommt. – Im vorliegenden Text sind Lexikostatistik und Glottochronologie vereinfachend unter dem Begriff der *Glottochronologie* zusammengefasst.
 14. Für eine detaillierte Zusammenfassung der einzelnen Kritikpunkte vgl. Wang 1994, S. 1446f. – Trotz der heftigen Kritik am Verfahren von Swadesh scheinen Teilaspekte der Glottochronologie in Form einer diachronen Lexikostatistik in der sprachvergleichenden historischen Linguistik noch in Gebrauch zu sein (vgl. die Reihe von Beiträgen in Renfrew / McMahon / Trask (Hg.) 2000, die sich mit diesem Thema beschäftigen), allerdings ist man mit der Formulierung der Ergebnisse vorsichtiger und verwendet sie eher als Indizien denn als Beweise für Trennungszeitpunkte oder Verwandtschaftsgrade (vgl. Embleton 2000, S. 159).

sind zuletzt bei Schnörch (2002, S. 11ff.) ausführlich dargestellt. Doch während sich die Mehrzahl der Diskussionsbeiträge mit einzelnen Aspekten der Grundwortschatzlexikographie bzw. Glottochronologie beschäftigen, gibt es nur einige wenige Arbeiten, die das Grundwortschatzkonzept von einem umfassenden methodischen Standpunkt aus betrachten und damit eine große Bedeutung innerhalb der Forschungsgeschichte haben. Sie werden im Verlauf der weiteren Arbeit immer wieder eine Rolle spielen und sollen daher hier kurz vorgestellt werden.

2.2.1 Kühn 1979: Die Hinwendung zur kommunikativen Funktion

Peter Kühns Arbeit „Der Grundwortschatz. Bestimmung und Systematisierung“ (Kühn 1979) ist in vielerlei Hinsicht ein Klassiker der Grundwortschatzliteratur. Das hängt mit der dreifachen Rolle zusammen, die Kühn in der Grundwortschatzlexikographie übernimmt:

„Kühn liefert eine ausführliche Forschungsübersicht, er ist ein vehementer Kritiker der frequenzorientierten Grundwortschatzlexikographie und [er] lanciert schließlich ein alternatives Konzept auf kommunikativ-pragmatischer Grundlage.“ (Krohn 1992, S. 55)

Motivation der Arbeit von Kühn ist es, „die Grundwortschatzbestimmung theoretisch neu zu fundieren und den Grundwortschatz als Funktionswortschatz zu bestimmen, mit dessen Hilfe es einem Wörterbuchbenutzer möglich sein soll, mit einem Minimum an lexikalischen Zeichen ein Maximum an kommunikativer Effektivität zu erreichen.“ (Kühn 1979, S. 1). Es geht Kühn also um zweierlei – auf einer eher lexikologischen Ebene um eine theoretische Fundierung des Grundwortschatzkonzeptes und auf einer lexikographischen Ebene um die praktische Erarbeitung eines neuen, an der kommunikativen Funktion von Wörtern orientierten Grundwortschatzes.

Dabei räumt Kühn der theoretischen Fundierung das Hauptgewicht in seiner Arbeit ein. Die Arbeit besteht in weiten Teilen aus eingehenden theoretischen Diskussionen zur Lexik, ihrer Rolle in der Kommunikation (Kapitel 1), den Sinnrelationen, mit denen sich die Lexik strukturieren lässt (Kapitel 3), einer Diskussion der Dichotomie von semasiologischen und onomasiologischen Wörterbüchern (Kapitel 4) und einer Revision der Schwächen der Grundwortschatzlexikographie, die vor Kühns Arbeit entstanden war (Kapitel 2). Relativ geringen Raum nimmt dagegen die Beschreibung des von Kühn geplanten Grundwortschatzes ein (Abschnitt 2.4 und Anhang).

Mit der Überarbeitung des Grundwortschatzkonzeptes will Kühn die Grundwortschatzarbeit wegführen von einer naturwissenschaftlich orientierten Herangehensweise, wie sie die frequenzorientierten Grundwortschatze prägt, hin zu „einer verstehenden Sozialwissenschaft“ (ebda., S. 59), d.h. er will die Grundwortschatzlexikographie stärker am Konzept der kommunikativen Kompetenz orientieren, als das bis dahin der Fall war (vgl. ebda., S. 60).

Im praktisch-lexikographischen Teil will Kühn deshalb ein „Grundwortschatzwörterbuch“ (ebda., S. 1) erarbeiten, an das er im Wesentlichen vier Anforderungen stellt:¹⁵ Das Grund-

wortschatzwörterbuch soll einen auf seinen kommunikativen Zweck ausgerichteten Gebrauchswert haben, es soll eine Alternative zur subjektiven Auswahl der pragmatisch orientierten Grundwortschatzlexikographie und zur im Resultat unbefriedigenden frequenzbasierten Selektion sein, es soll theoretisch fundiert sein und sich darüberhinaus am Konzept der kommunikativen Kompetenz orientieren (vgl. ebda., S. 59-61). Dieses Vorhaben gelingt ihm im theoretischen Teil sehr gut: Die gründliche Beleuchtung des Grundwortschatzkonzepts im Lichte lexikologischer Fragestellungen wie derjenigen der Rolle der Lexik in der Kommunikation, Fragen der semasiologischen vs. onomasiologischen Darstellung oder die gründliche Analyse der herkömmlichen Selektionskriterien und der Schwächen des frequenzorientierten Ansatzes sind Aspekte, die in der Grundwortschatzdiskussion bis dahin nicht eingehend diskutiert worden waren.

Der praktische Teil der Arbeit lässt jedoch einige Fragen offen. Zum einen fehlt eine ausführliche, konkrete Beschreibung der praktischen Vorgehensweise. Zwar gibt Kühn eine Gruppe von Selektionskriterien an:¹⁶

„Welches lexikalische Material (LM) des standardsprachlichen Gebrauchs benötigt ein Sprecher/Schreiber (S), um in der Situation (Sit(x)) über das Thema (T(y)) in der Rolle (R) die kommunikative Intention (KI) mit Hilfe des Kommunikationsmodus (KM) erfolgreich durchzuführen?“ (Kühn 1979, S. 62)

Und es ist auch klar, dass am Anfang des Auswahlprozesses Kommunikationssituationen und Themen und die mit ihnen verknüpften Begriffe stehen – ein für die Diskussion zur Grundwortschatzlexikographie wichtiger Impuls, weil dadurch explizit der Begriff als Grundeinheit in die Diskussion eingeführt wird.¹⁷ Wie aber diese Begriffe ausgewählt werden und wie Kühn zu dem im Anhang der Arbeit angeführten auszugswisen Grundwortschatz kommt, was die konkreten Schritte seines Vorgehens sind, wird nicht ganz klar. Klar wird allein, dass das Thema steuerndes Element der Auswahl sein soll. Damit bricht Kühn aber mit seiner Forderung, dem nicht gangbaren Weg der subjektiven Selektion von Lexemen ein adäquateres Verfahren entgegenzusetzen. Zwar werden die einzelnen Lexeme nicht unbedingt subjektiv ausgewählt, weil sie sich an den zugrundeliegenden Themen orientieren. Die Themenauswahl selbst erscheint jedoch subjektiv. Somit verschiebt sich das Subjektivitätsproblem aber nur, gelöst wird es nicht.¹⁸

15. Dabei macht Kühn nicht explizit deutlich, was der Unterschied zwischen einem Grundwortschatz und einem Grundwortschatzwörterbuch ist. Erkennbar ist aber, dass er damit eine reichere Darstellung meint als es die bis dahin üblichen einfachen Wortlisten waren (vgl. etwa Kühn 1979, S. 131).

16. Bezeichnenderweise spricht Kühn nicht von Selektionskriterien, sondern von einem „Set der Kriterien einer funktionalen Limitation“ (Kühn 1979, S. 62). Damit spiegelt er in seiner Terminologie wider, was er in seiner Theorie annimmt: Es geht ihm nicht vornehmlich um die Selektion eines gegebenen Wortschatzes, sondern um die Begrenzung der lexikalischen Belegung gegebener Themenbereiche und Kommunikationssituationen.

17. Vgl. auch die Diskussion in Abschnitt 4.2.

18. Vgl. Krohn 1992, S. 70: Kühn ersetze das lexikalische Rekurrenzkriterium durch ein thematisches. – Ähnlich argumentiert auch Bonan-Garrigues (1992, S. 7) in Bezug auf den frequenzorientierten Ansatz.

Zum anderen scheint Kühns konzipiertes Grundwortschatzwörterbuch zu einem Bestand an Wörtern zu führen, dessen Umfang mit der Maxime der Grundwortschatzlexikographie, mit minimalen lexikalischen Mitteln eine größtmögliche lexikalische Kompetenz zu vermitteln, kollidiert. Zwar setzt Kühn sein projektiertes Grundwortschatzwörterbuch nur in Auszügen um, aber die Orientierung an der Kommunikationssituation und an Themen und Begriffen als Grundeinheit der Selektion, die nachträglich mit Lexemen belegt werden, scheint auf einen potentiell großen Umfang des Grundwortschatzwörterbuchs zu deuten. So schätzen etwa de Smet (1981, S. 364) bzw. Hendrickx (1987, S. 2) den Umfang des Grundwortschatzes, der entstehen würde, führte man Kühns Entwurf weiter, auf 8 000 bis 9 000 bzw. ca. 19 000 Einträge.¹⁹ Damit operiert Kühns Verfahren mit einem Umfang, der in anderen Wortschatzschätzungen bereits dem Gesamtwortschatz eines durchschnittlichen Sprechers des Deutschen entspricht, und setzt somit das konstitutive Merkmal zentraler Wortschätze, einen quantitativ deutlich begrenzten Ausschnitt des Gesamtwortschatzes abzubilden, außer Kraft.²⁰ Verstärkt wird das Problem des potentiell großen Umfangs des von Kühn konzipierten Grundwortschatzwörterbuchs noch durch ein Problem, das mit onomasiologischen Darstellungen von Wortschatz generell verbunden ist: Damit das Grundwortschatzwörterbuch selbständig benutzbar ist, müssen darin die zur Beschreibung der Einträge und ihrer semantischen Strukturen nötigen Wörter selbst nach Möglichkeit wiedergegeben werden. Damit aber müssen immer wieder die Grenzen der angegebenen Wortfelder erweitert werden und der Umfang des Grundwortschatzes vergrößert sich.

Dieses Problem – der wachsende, kaum zu begrenzende Umfang des von Kühn konzipierten Grundwortschatzes – mag ein Grund dafür gewesen sein, dass Kühn sein Vorhaben nicht zu Ende führte. Aber trotz des unvollendeten praktischen Teils – die Arbeit von Kühn bedeutet eine Wende in der Diskussion zur Grundwortschatzlexikographie, die sich bis dahin im Streit zwischen Anhängern des frequentiellen Ansatzes und des pragmatischen Ansatzes festgefahren hatte. Aus diesem Grund darf die Arbeit von Kühn auch in keiner Besprechung des Themas Grundwortschatz fehlen: Es ist sein Verdienst, dass dieses lexikographische Konzept erstmals in umfassender Weise in einen theoretischen Rahmen gestellt wurde. Darüber hinaus ist seine eingehende Diskussion der bis dahin üblichen Selektionskriterien, ihrer Stärken und Schwächen ein überaus wichtiger Beitrag zur Übersichtlichkeit der Forschungssituation. Und schließlich ist es Kühns Verdienst, den Begriff als Grundeinheit eines Grundwortschatzes erstmals in die Diskussion eingeführt zu haben.

19. Vgl. auch Krohn 1992, S. 67.

20. Vgl. auch Krohn 1992, S. 69. – Für eine detaillierte Diskussion der Frage nach dem Umfang zentraler Wortschätze vgl. auch Abschnitt 4.6.

2.2.2 Krohn 1992: Die Suche nach dem thematischen Profil verschiedener Grundwortschätze

Dieter Krohn nähert sich mit seiner Arbeit „Grundwortschätze und Auswahlkriterien“ (Krohn 1992) dem Thema Grundwortschatz von einer anderen Seite her: Ihm geht es nicht um konkrete Schritte zur Erarbeitung eines Grundwortschatzes, sondern um die Analyse vorhandener Grundwortschätze und Häufigkeitswörterbücher. Dabei verfolgt er mehrere Ziele (vgl. Krohn 1992, S. 8f.): Er möchte das Augenmerk auf die Konvergenzen, d.h. Gemeinsamkeiten verschiedener Grundwortschätze richten und damit die einseitige Konzentration auf die Kritik der Unterschiede durchbrechen. Des Weiteren möchte er die These, dass frequentieller und kommunikativ-pragmatischer Ansatz unvereinbar seien, auf ihre Richtigkeit überprüfen und zugleich den kommunikativ-pragmatischen Ansatz, der durch die Konzentration der Kritiker auf die Schwächen des frequentuellen Ansatzes bisher vernachlässigt wurde, einer kritischen Analyse unterziehen. Schließlich will er den Grundwortschatzbegriff präzisieren, indem er verschiedene Grundwortschätze und Häufigkeitswörterbücher auf ihre Zusammensetzung von Synsemantika, themenunspezifischen und themenspezifischen Autosemantika untersucht und das Konzept des Grundwortschatzes im Lichte von Textdeckungstests der Fremdsprachendidaktik diskutiert.²¹

Wie schon Kühn, so geht auch Krohn sein Vorhaben von zweierlei Seiten an: Er diskutiert theoretisch-methodische Rahmenbedingungen und Voraussetzungen für die Grundwortschatzarbeit auf der einen Seite und führt auf der anderen Seite eine Reihe praktischer Untersuchungen durch.

Im theoretisch-methodischen Bereich seiner Arbeit bespricht Krohn ausführlich kommunikativ-pragmatische Aspekte des Wortschatzes und – was noch wichtiger ist – die Bedeutung der „lexikalisch-pragmatischen Kompetenz“ für die Sprachdidaktik (Kapitel 2). Er gibt eine kurze Zusammenfassung der Diskussion des Für und Wider der frequenzorientierten bzw. kommunikativ-pragmatischen Ansätze der bisherigen Grundwortschatzlexikographie (Kapitel 4) und diskutiert ausführlich Begriffe wie Thema, Situation, Intention, Synsemantika, Autosemantika, Textdeckung und Themenspezifität (Kapitel 5 und 6). Und nicht zuletzt liefert er eine eingehende Darstellung der Eigenschaften verschiedener Grundwortschätze und Häufigkeitswörterbücher (Kapitel 3).

21. Für eine detaillierte Definition der Begriffe *Synsemantika*, *themenunspezifische Autosemantika* und *themenspezifische Autosemantika* vgl. Krohn 1992, S. 114f. – Unter *Textdeckung* versteht man den prozentualen Anteil, den ein Wort mit seinen verschiedenen Flexionsformen an der Gesamtmenge an laufenden Wortformen eines Textes hat. Je mehr laufende Wortformen ein Wort auf sich vereinen kann und je höher diese Anzahl im Vergleich zur Gesamtzahl an laufenden Wortformen des Textes ist, desto höher ist der Textdeckungsgrad dieses Wortes. Textdeckungstests werden in der Sprachdidaktik dazu verwendet, um etwa die Nützlichkeit der Wörter einer Wortschatzsammlung zu bestimmen: Je höher der jeweilige Textdeckungsgrad der Wörter dieser Wortschatzsammlung, so die Annahme, desto nützlicher ist sie. Umgekehrt werden Textdeckungstests auch eingesetzt, um den Schwierigkeitsgrad eines Textes in Bezug auf ein gegebenes Vokabular zu bestimmen: Enthält er viele Wortformen, die nicht durch die im Vokabular abgedeckten Wörter repräsentiert sind, ist der Text im Verhältnis zu diesem Vokabular als schwierig einzustufen.

Die praktische Untersuchung wiederum setzt sich zusammen aus einer ausführlichen Begründung, weshalb Krohn gerade die sechs Grundwortschätze und Häufigkeitswörterbücher vergleicht, die er in seiner Analyse heranzieht (Kapitel 7.2) und einer detaillierten Beschreibung der Vorgehensweise, die Krohn benutzt, um zu den verschiedenen von ihm analysierten Schnittmengen der sechs Wortschätze zu gelangen (Kapitel 7.3 bis 7.5). Die sechs Grundwortschätze, die Krohn näher untersucht, setzen sich zusammen aus zwei kommunikativ-pragmatisch orientierten Grundwortschätzen und vier frequenzorientierten Grundwortschätzen bzw. Häufigkeitswörterbüchern.²² Diese sechs Grundwortschätze bzw. Häufigkeitswörterbücher konfrontiert Krohn miteinander und konzentriert sich dabei vor allem auf den kontrastierenden Vergleich von kommunikativ-pragmatischen Grundwortschätzen auf der einen und frequenzorientierten Grundwortschätzen und Häufigkeitswörterbüchern auf der anderen Seite. Im Zentrum steht dabei die Frage nach Konvergenzen und Divergenzen zwischen den miteinander konfrontierten Wortschatzsammlungen, d.h. es geht um die Frage, welche und wieviele Einträge allen Wortschatzsammlungen gemeinsam sind (Konvergenz) und welche und wieviele Einträge nur wenigen oder einzelnen Wortschatzsammlungen eigen sind (Divergenz).

Bei den Divergenzen geht es Krohn darüber hinaus um die Frage danach, welche Einträge einer Wortschatzsammlung nur in dieser vorkommen (Krohn bezeichnet diese Fälle als „exklusive Hapaxlegomena“ (ebda., S. 189)) und welche Einträge verschiedenen Wortschatzsammlungen des gleichen Typs (kommunikativ-pragmatisch vs. frequenzorientiert) gemeinsam sind, nicht aber in Wortschatzsammlungen des anderen Typs vorkommen (Krohn bezeichnet diese Fälle als „inklusive Hapaxlegomena“). Die Untersuchung dieser Hapaxlegomena soll Aufschluss über die Besonderheiten der jeweiligen Wortschatzsammlung im einzelnen bzw. des Typs von Wortschatzsammlung geben.

Aus den Konvergenzen, d.h. der Schnittmenge aller sechs Wortschatzsammlungen bildet Krohn eine sogenannte „Kompaktliste“ (ebda., S. 128) bzw. ein „Testlexikon“. Von diesem Testlexikon nimmt er an, dass es...

„...den hauptsächlich themen- wie situationsunspezifischen Wortschatz der Synsemantika und derjenigen Autosemantika, die sowohl in standardsprachlichen wie fachsprachlichen Texten als kleinster gemeinsamer lexikalischer Nenner gelten können“ (ebda., S. 130),

umfasst. Mit Hilfe dieses Testlexikons, so Krohn, ließe sich dann anhand eines Textdekungstest der Schwierigkeitsgrad eines beliebigen Textes erfassen, indem man seinen Lemmabestand in zwei „Lernquanten“ (ebda., S. 128) gliedert, nämlich in ein „primäre[s] Lernquantum“ (ebda., S. 125), bestehend aus „den Lexemen eines Textes, die vom Grundwortschatz abgedeckt werden“, und einem „sekundäre[n] Lernquantum“, das aus den übrigen, „vom Grundwortschatz nicht erfaßten Lexeme[n] im Text“ besteht. Anders als die

22. Die beiden pragmatisch orientierten Grundwortschätze sind Baldeger (Hg.) 1980 und ZERTIFIKAT DEUTSCH 1977. Die vier frequenzorientierten Grundwortschätze bzw. Häufigkeitswörterbücher sind .Pfeffer 1984, Ortman 1979, Rosengren 1972 und 1977 und Erk 1972, 1975 und 1982. – Für eine Übersicht vgl. Krohn 1992, S. 145f.

Verwendung des Terminus *Grundwortschatz* in der hier angeführten Beschreibung des Testlexikons vermuten lässt, geht Krohn allerdings nicht von einer unmittelbaren Identität von Testlexikon und Grundwortschatz aus (vgl. ebda., S. 130).

Neben den Vielfachkonfrontationen und der Bildung des Testlexikons unternimmt Krohn auch eingehende Analysen der exklusiven Hapaxlegomena einer jeden Wortschatzsammlung, um zu sehen, ob sich mit ihrer Hilfe ein thematisches Profil der jeweiligen Wortschatzsammlung erstellen lässt.

Im Wesentlichen kommt Krohn bei den Konfrontationen und thematischen Analysen der einzelnen Wortlisten zu folgenden Ergebnissen:

Thematisch habe fast jede Wortliste ein deutliches eigenes Profil, d.h. jede der Wortlisten enthielte ausreichend viele exklusive Hapaxlegomena, die als themenspezifische Autosemantika einzustufen seien und damit Rückschlüsse über das thematische Profil der jeweiligen Wortliste zuließen. Eine Ausnahme bildeten die Wortschatzsammlungen von Erk (Erk 1972, Erk 1975 und Erk 1982); sie enthielten kaum themenspezifische Autosemantika.

Dabei sei auffällig, dass die kommunikativ-pragmatisch orientierten Wortlisten, also Baldegger (Hg.) 1980 und ZERTIFIKAT DEUTSCH 1977, einen größeren Anteil inklusiver Hapaxlegomena als die frequenzorientierten Wortlisten untereinander, dass also die Divergenzen zwischen einzelnen frequenzorientierten Wortlisten größer seien als zwischen den kommunikativ-pragmatisch orientierten.

Außerdem enthielten die kommunikativ-pragmatisch orientierten Wortschatzsammlungen mehr themenspezifische Autosemantika, d.h. ihr thematisches Profil sei ausgeprägter, als das bei den frequenzorientierten Wortlisten der Fall sei. Generell seien die themenspezifischen Autosemantika in der Mehrzahl der Fälle Substantivkomposita. Auffällig sei auch, dass die themenspezifischen Lexeme, genauer: Substantive, bei den kommunikativ-pragmatischen Wortlisten zumeist Konkreta seien, während themenspezifischen Lexeme bei den frequenzorientierten Wortlisten vor allem aus Abstrakta bestünden. Insgesamt seien die Divergenzen der frequenzorientierten gegenüber den kommunikativ-pragmatisch orientierten Wortschatzsammlungen im Bereich der Substantive am größten, im Bereich der Adverbien und Partikeln am geringsten.²³

Die Stärken der Arbeit Krohns liegen ganz sicher darin, dass er zum einen die seit der Arbeit von Kühn weitgehend ruhende Grundwortschatzdiskussion wieder aufnimmt, ihr aber zum anderen zugleich neue Impulse gibt, indem er die Aufmerksamkeit weg von den Diskrepan-

23. Zur Frage der thematischen Profile der einzelnen Wortschatzsammlungen vgl. die ausführlichen Wortlisten und Kommentare bei Krohn 1992, S. 210-300. Zur Frage von Konvergenzen und Divergenzen zwischen kommunikativ-pragmatischen Wortschatzsammlungen auf der einen und frequenzorientierten Wortschatzsammlungen auf der anderen Seite vgl. die Zusammenstellung auf S. 167-173 und die Zusammenfassung auf S. 301-303. – Sowohl Raasch 1995 als auch Wolski 1995 lasten Krohn an, seine Ergebnisse führten „über Bekanntes nicht hinaus“ (Raasch 1995, S. 435). Dennoch sind Krohns Ergebnisse als nützlich einzustufen, weil sie die Feststellung, dass Divergenzen im Wortschatz mit der Themenspezifität des zugrundeliegenden Korpus zunehmen, in der Praxis bestätigen.

zen verschiedener Grundwortschätze hin zu ihren Gemeinsamkeiten lenkt und die Konvergenzen ins Zentrum der Diskussion stellt; dass er also Analysen im Sinne von ‘Wort A ist in Grundwortschatz x aufgenommen, fehlt aber in Grundwortschatz y; dafür fehlt Wort B in Grundwortschatz x, ist aber in Grundwortschatz y vorhanden’ zugunsten einer differenzierten Suche nach Gemeinsamkeiten und Unterschieden zweier Grundwortschätze aufgibt. Außerdem liefert er erstmals eine gründliche Gegenüberstellung von frequenzorientierten und kommunikativ-pragmatisch orientierten Wortschatzsammlungen und unterzieht den kommunikativ-pragmatischen Ansatz einer ausführlichen Diskussion und praktischen Untersuchung. Dabei ist auch positiv hervorzuheben, dass Krohn viel Wert darauf legt, das eigene Verfahren, mit dessen Hilfe er die verschiedenen Ausgangsdaten kompiliert, gründlich darzulegen, wenn er etwa die Behandlung von Komposita (vgl. ebda., S. 153-163) oder Fragen der Lemmatisierung (vgl. ebda., S. 148-153) problematisiert und zugleich darum bemüht ist, die grundlegenden Grenzen des eigenen Verfahrens deutlich zu machen (vgl. ebda., S. 173).²⁴

Diesen Stärken stehen allerdings einige methodische Schwächen gegenüber:²⁵

Fragwürdig ist zum einen Krohns unbedingter Glaube, ein Text lasse sich anhand des Testlexikons und mit Hilfe von Textdeckungstests in distinkte Mengen von Synsemantika, themenunspezifischen und themenspezifischen Autosemantika einteilen, ein Verfahren, das Krohn vorschlägt, um die Kompaktliste zu einem Lernerwörterbuch auszubauen (vgl. Krohn 1992, S. 143). Diese Hoffnung ist anzweifelbar, weil es für eine eindeutige Abgrenzung von Synsemantika, themenunspezifischen Autosemantika und themenspezifischen Autosemantika, wie Schnörch (2002, S. 51 und 223) zeigt, keine klaren Kriterien gibt und bereits die Abgrenzung zwischen Synsemantika auf der einen und Autosemantika auf der anderen Seite nicht unproblematisch ist.²⁶

Eng mit diesem Kritikpunkt verflochten ist die Frage danach, wie Krohn mit Polysemien umgeht. Krohn sagt selbst, dass er polyseme Einträge auf ihre Hauptbedeutung reduziere und erst dann in den Vergleich mit aufnehme (Krohn 1992, S. 148ff.). Er versäumt es jedoch, kritisch zu diskutieren, wie man zu einer solchen Hauptbedeutung gelangt.²⁷

So kritisiert Schnörch entsprechend auch das von Krohn angewandte Verfahren zur Erstellung eines thematischen Profils für die einzelnen Wortschatzsammlungen, bei dem sich Krohn auf die kontextlos gegebenen Lexeme stützt. Schnörch (2002, S. 51) zeigt, dass an

24. Vgl. auch Raasch 1995, S. 425 und Wolski 1995, S. 342.

25. Auch in formaler Hinsicht hat die Arbeit von Krohn einige Schwächen. So fehlen zum Beispiel zu manchen von Krohn präsentierten Fakten die entsprechenden Quellenangaben. Vgl. etwa seine Angaben zum durchschnittlichen Anteil der Autosemantika von 80% eines Normaltextes (Krohn 1992, S. 120) oder die Aussage, dass in einem beliebigen Text fast jedes dritte Lexem nicht verständnistiftend sei (ebda., S. 143), zu denen jeweils Informationen zur Herkunft dieser Werte fehlen.

26. Vgl. etwa Lutzeier 1985, S. 26 oder die Diskussion zum Status von Partikeln bei Schnörch 2002, S. 141.

27. In starkem Gegensatz dazu steht Schnörchs Arbeit, der der Diskussion zur Handhabung von Polysemien viel Platz einräumt (vgl. Schnörch 2002, S. 84-94 und 97-109).

dieser Stelle eine Lesartendisambiguierung nötig gewesen wäre, und weist darauf hin, dass eine Orientierung am Semem, wie Schnörch sie vornimmt (vgl. Abschnitt 2.2.4), auch bei der von Krohn vorgenommenen Untersuchung nützlich gewesen wäre.²⁸ Allerdings ist dies weniger Krohn anzulasten als vielmehr dem Ausgangsmaterial, das er verwendet und das nur zum Teil semantische Informationen liefert.

Skeptisch steht Schnörch schließlich auch dem gesamten Bereich der Textdeckungstests gegenüber, die nur in begrenztem Maße etwas über den Schwierigkeitsgrad eines Textes oder die Wirksamkeit eines Grundwortschatzes aussagen könnten.²⁹

Eine weitere mögliche Schwäche ist die Tatsache, dass im Vergleichsmaterial Häufigkeitswörterbücher so stark vertreten sind (drei der sechs untersuchten Wortschatzsammlungen sind als Häufigkeitslisten konzipiert und nicht als Grundwortschätze). Krohn betont zwar einerseits, dass Häufigkeitswörterbücher nicht mit Grundwortschätzen gleichzusetzen seien (vgl. Krohn 1992, S. 67), geht aber andererseits davon aus, aus der Synthetisierung verschiedener Grundwortschätze und Häufigkeitslisten ließe sich ein fremdsprachendidaktischer Grundwortschatz erstellen. Dies wird vor allem von Raasch 1995 heftig kritisiert:

„Man sollte – entsprechend der heutigen Fachdiskussion – nicht in dieser Weise vorgehen, um 'die Grundlage für die Zusammenstellung eines deutsch-schwedischen Lernerwörterbuchs' [Krohn 1991, S. 308; V. H.] zu schaffen, sondern die Lernziele, die Lerner, die Lernsituation usw. definieren und aus ihnen einen zielgruppengerechten Sprachschatz ableiten. Insofern ist diese Publikation eher rückwärts gewandt.“ (Raasch 1995, S. 424f.)

Dazu gehört auch, dass Krohn nicht auf den Unterschied von einfachen Wortlisten und (Grundwortschatz-)Wörterbüchern eingeht und zum Beispiel seine aus den sechs Wortlisten synthetisierte Kompaktliste als „Lern(er)wörterbuch“ (Krohn 1992, S. 138) betrachtet wissen will. Eine Diskussion der Begriffe *Wörterbuch* und *Grundwortschatz* fehlt (vgl. Wolski 1995, S. 339-341). Dieser Mangel an theoretischer Fundierung, so Wolski, sei auch bei der Besprechung von Themen wie kommunikativer Kompetenz und der Abgrenzung von Synsemantika und Autosemantik zu finden, Begriffe, die bei Krohn unklar beschrieben und verwendet und zugleich doch „belastet“ (ebda., S. 341) seien.

Trotz dieser Schwächen und auch wenn der unmittelbare praktisch-sprachdidaktische Nutzen der Krohnschen Kompaktliste begrenzt ist, ist Krohns Arbeit ein wichtiger Beitrag zur Grundwortschatzforschung, weil er der Gegenüberstellung der praktischen Ergebnisse des kommunikativ-pragmatischen Ansatzes gegenüber derjenigen des frequentiellen Ansatzes viel Platz einräumt und damit eine Lücke in der Erforschung des Grundwortschatzkonzeptes füllt. Ein weiteres Verdienst von Krohn ist es, gezeigt zu haben, dass frequentieller und pragmatischer Ansatz sowohl im Vorgehen, zum Beispiel hinsichtlich der Subjektivität der Auswahl, als auch im Ergebnis viele Konvergenzen aufweisen.³⁰

28. Ähnlich auch Raasch 1995, S. 425.

29. Vgl. Schnörch 2002, S. 48 und S. 75. Zu einer kritischen Betrachtung von Textdeckungstests vgl. auch Carter 1998, S. 236.

2.2.3 Bonan-Garrigues 1993: Die Erarbeitung einer Wortschatzhierarchie mithilfe des Verfahrens der transparenten Introspektion

Während Krohns Arbeit ganz im Zeichen der Fremdsprachendidaktik steht, nähert sich die Arbeit von Bonan-Garrigues – „Méthode de paramétrage des dictionnaires et grammaires électroniques“ (Bonan-Garrigues 1993) – dem Konzept zentraler Wortschätze von einer ganz anderen Seite her und wendet dieses in seiner praktischen Umsetzung auf eine deutlich andere Weise an, als das bisher geschehen ist.

Ihr Ausgangspunkt sind Fragen der Aufbereitung von Sprachdaten für computerlinguistische Anwendungen wie die Unterstützung natürlichsprachlicher Abfragen von Datenbanken und die verfeinerte automatische Analyse von Texten (Bonan-Garrigues 1993, S. 6f.). Zu diesem Zweck will Bonan-Garrigues die Bestände des DELAS,³¹ eines elektronischen Wörterbuchs französischer Simplizia, hierarchisieren.³² Mit Hierarchisierung meint Bonan-Garrigues dabei die Anordnung der Einträge vom „vocabulaire le plus essentiel au moins essentiel“ (ebda., S. 6). Doch anders, als man es von computerlinguistisch geprägten Arbeiten erwarten würde, bei denen häufig die Praktikabilität der Erarbeitung von Daten und ihre eindeutige Zuordnung wichtiger ist als eine differenzierte Vorgehensweise, stützt sich Bonan-Garrigues bei dieser Hierarchisierung der Daten nicht auf vergleichsweise einfach zu erhebende Informationen wie etwa Wortfrequenzen, sondern auf ein schrittweises, transparentes Beurteilen eines jeden Lexikoneintrages hinsichtlich eines kontrollierten introspektiven Wertes der Zentralität.

Bonan-Garrigues entwickelt ihr Verfahren ausgehend von zwei Grundüberlegungen: von einer umfassenden Kritik der sogenannten objektiven Methoden (Kapitel 1.1), d.h. der frequenzbasierten Hierarchisierung des Wortschatzes; und von der problematisierenden Diskussion der Frage, was es bedeutet, ein Wort zu verstehen (Kapitel 1.2) und was die Bedeutung eines Wortes eigentlich umfasst (Kapitel 1.3). Sie kommt dabei unter anderem zu dem Schluss, dass frequenzorientierte Methoden der Wortschatzhierarchisierung nur vermeintlich objektiv seien und durchaus Elemente einer subjektiven Selektion, etwa auf der Ebene der Korpuszusammenstellung, enthielten, weshalb andere, vermeintlich rein subjektive Methoden nicht zu schlechteren Ergebnissen führten (vgl. ebda., S. 7). Auch sei der Begriff vom *Verstehen* eines Wortes nicht eindeutig zu bestimmen, weshalb die Frage da-

30. Vgl. auch die Diskussionen in Abschnitt 4.5.1.

31. „Dictionnaire Electronique du LADL [= Laboratoire d'Automatique Documentaire et Linguistique der Universität Marne-la-Vallée, V.H.] pour les mots simples du français“ (Bonan-Garrigues 1993, S. 4) – Theoretisch-methodischer Hintergrund der Arbeit von Bonan-Garrigues ist die von Maurice Gross konzipierte Lexikogrammatik („lexique-grammaire“ (Gross 1994, S. 4)) des LADL, die als Grundlage zur automatischen Analyse und Generierung von Texten dient und die sich aus elektronischen Wörterbüchern und syntaktischen Tabellen zusammensetzt. Vgl. auch Bonan-Garrigues 1993, S. 4.

32. Bonan-Garrigues versucht im zweiten Teil ihrer Arbeit auch eine Hierarchisierung der grammatischen Informationen im DELAS (vgl. Bonan-Garrigues 1993, Kapitel 2), aber sie steht in keinem unmittelbaren Zusammenhang zum Thema dieser Arbeit, deshalb sei sie hier vernachlässigt.

nach, ob ein Sprecher ein Wort verstehe oder nicht, nicht unmittelbar als Kriterium zur Beurteilung dieses Wortes dienen könne (vgl. ebda., S. 26).

Vor dem Hintergrund dieser Vorüberlegungen entwirft Bonan-Garrigues ein differenziertes Verfahren der introspektiven Beurteilung der Zentralität eines Wortes: Sie zerlegt die Vorstellung von der Zentralität eines Wortes in zwei Beschreibungskategorien – seine „représentation“ (Repräsentiertheit) und seine „plausibilité“ (Plausibilität) – und unterzieht mit Hilfe dieser Beschreibungskategorien den zu hierarchisierenden Wortbestand einem dreigeteilten Verfahren (Kapitel 1.4): Zunächst betrachtet sie alle Einträge des DELAS dahingehend, ob sie in ihrem mentalen Lexikon irgendeine Reminiszenz hervorrufen, d.h. ob sie dort in irgendeiner Weise repräsentiert sind (vgl. ebda., S. 40f.) und ordnet ihnen daraufhin einen Wert der Repräsentiertheit von 0 oder 1 zu. In einem zweiten Schritt betrachtet sie jeden Eintrag des DELAS dahingehend, wie plausibel es ihr erscheint, dass dieser Eintrag, d.h. die Wortform, in irgendeinem sprachlichen Zusammenhang auftreten wird, und ordnet dem Eintrag entsprechend einen von drei Plausibilitätswerten zu (vgl. ebda., S. 41-44). In einem dritten Schritt dann werden die Werte der Repräsentiertheit und der Plausibilität für jeden Eintrag zusammengelegt und darauf aufbauend einer von drei Schichten zugeordnet (vgl. ebda., S. 44f. und 57).³³

Um ihr eigenes Verfahren zu überprüfen, unterzieht sie die so gewonnene Hierarchisierung des Wortbestandes des DELAS verschiedenen Tests, darunter einer Sprecherbefragung, bei der sie muttersprachliche Sprecher bittet, ebenfalls eine Gruppe von Wörtern hinsichtlich ihrer Repräsentiertheit und Plausibilität zu beschreiben (vgl. ebda., S. 50-52). Dieses Testverfahren ist, wie sich zeigt, nicht unproblematisch, zum einen, weil es sich als schwierig erweist, den Testpersonen klar zu machen, dass es bei diesen Tests nicht um ihre sprachliche Leistung geht, sondern darum, den Wortschatz der Sprechergruppe zu erfassen, der sie angehören (vgl. ebda., S. 50), und zum anderen, weil es nicht immer einfach ist, den Testpersonen den Unterschied und die Bedeutung der beiden Kriterien von Repräsentiertheit und Plausibilität klar zu machen. Vor allem aber die Tatsache, dass die Ergebnisse der Beurteilung durch die Sprecher sehr stark voneinander abweichen, führt Bonan-Garrigues zu der bemerkenswerten Schlussfolgerung, dass nur linguistische geübte Sprecher, also solche, die Sprache als Objekt betrachten können, zur Beurteilung von Wörtern hinsichtlich ihrer Repräsentiertheit und Plausibilität herangezogen werden könnten (vgl. Bonan-Garrigues 1993, S. 52).³⁴

Nun ist es leicht, das von Bonan-Garrigues entwickelte und angewandte Verfahren auf den ersten Blick als nicht relevant abzulehnen, weil es auf der Introspektion eines einzigen Be-

33. Für eine detaillierte Diskussion dieses Selektionsverfahrens vgl. Abschnitt 4.5.2.2.

34. Bemerkenswert ist diese Schlussfolgerung insofern, als andere Versuche, die Zentralität von Wörtern introspektiv zu bestimmen, wie etwa Gougenheims Erfassung von sog. *mots disponibles* (vgl. Gougenheim / Rivenc / Michéa u. a. 1956), gerade die Befragung von Sprechern als Legitimation ihres Verfahrens herausstellen.

urteilenden beruht. Andererseits stellt Bonan-Garrigues' Arbeit einen besonderen Beitrag zur Arbeit mit zentralen Wortschätzen dar: Zum einen ist Bonan-Garrigues die erste, die die introspektive Beurteilung der Zentralität eines Wortes ausführlich und systematisch analysiert und anwendet und sie in ein operatives und transparentes Verfahren umsetzt.³⁵ Andererseits ist ihre Arbeit eine der wenigen, die das Konzept von Wortschatzschichtungen in computerlinguistischen Anwendungen zur Sprache bringt und zugleich mit grundlegenden Fragen der Linguistik und Kognitionswissenschaft verknüpft.³⁶ Damit erweitert sie die Diskussion, die in den anderen zentralen methodischen Arbeiten auf rein linguistische Aspekte beschränkt bleibt, und trägt damit viel zu einer Komplettierung des Themas zentraler Wortschätze bei.

2.2.4 Schnörch 2002: Der Vergleich verschiedener Grundwortschätze auf der Basis lexikalischer Einheiten

Ähnlich wie Krohn geht es Schnörch in seiner Arbeit „Der zentrale Wortschatz des Deutschen. Strategien zu seiner Ermittlung, Analyse und lexikografischen Aufarbeitung“ (Schnörch 2002) nicht unmittelbar um das Erstellen eines konkreten zentralen Wortschatzes, sondern um die Analyse einer Reihe gegebener Grundwortschätze und Häufigkeitswörterbücher zum Zwecke des Erkenntnisgewinns rund um das Konzept des Grundwortschatzes. Motivation der Arbeit von Schnörch ist es dabei, die Grundwortschatzlexikographie wiederzubeleben (vgl. ebda., S. 8) und den Grundwortschatz als Forschungsgegenstand und nicht allein als praktisches Objekt der Lexikographie in den Mittelpunkt zu rücken.³⁷ Im Gegensatz zu Krohn ist Schnörch allerdings eher an qualitativen Detailanalysen interessiert als an quantitativen Erkenntnissen (vgl. ebda., S. 187). Zwar führt auch Schnörch eine quantitative Analyse der von ihm untersuchten Grundwortschätze und Häufigkeitswörterbücher durch, tut dies aber vor allem, um Krohns Grundthesen nachzuzeichnen und um den Umfang der von ihm erarbeiteten Schnittmenge zahlenmäßig abzustechen (vgl. ebda., S. 219).

Kern der Arbeit von Schnörch sind dabei zwei Bereiche:³⁸

Zum einen steht die gründliche Beschreibung und Diskussion von sieben verschiedenen Korpora im Vordergrund, d.h. derjenigen Grundwortschätze und Häufigkeitswörterbücher, die in einem weiteren Schritt die Grundlage für lexikologische Analysen darstellen (vgl. Schnörch 2002, Kapitel 2).³⁹ Im Rahmen dieser Darstellung beleuchtet Schnörch verschiedene grundlegende lexikologische und lexikographische Fragen, die nicht unbedingt spezi-

35. Für einen Überblick über die Vorläufer dieses Verfahrens vgl. Bonan-Garrigues 1992, S. 20

36. Andere, allerdings bei weitem nicht so umfangreiche Arbeiten sind Lehmann 1991 und Schulz 1991.

37. Eine Motivation, die Schnörch Krohn zuschreibt (Schnörch 2002, S. 183), die aber durchaus auch aus Schnörchs Arbeit herauszulesen ist.

38. Hinzu kommt noch eine ausführliche, vor allem an der Arbeit von Kühn orientierte Darlegung der bisherigen Grundwortschatzdiskussion, die den beiden Hauptteilen vorangestellt ist. Vgl. Schnörch 2002, Kapitel 1.1.

fisch für den Grundwortschatz sind, aber wichtige Aspekte dieses Konzeptes darstellen. Zu den von Schnörch diskutierten Themen gehören die Auflösung bzw. Unterscheidung von Polysemen und Homonymen (S. 84-94 und 97-109), Fragen der Lexikalisierung (S. 280-286), die Unterscheidung von gesprochener vs. geschriebener Sprache (S. 76-79), die Binnenstruktur von Lexikoneinträgen (S. 95-97), die lexikographische Behandlung von gebundenen Ausdrücken (S. 109-114), Fragen der Lemmatisierung von Korpusdaten (S. 135-142), die Vergleichbarkeit und Interpretierbarkeit von Frequenzwörterbüchern (S. 142f.), Fragen der semasiologischen gegenüber der onomasiologischen Wörterbuchstruktur (S. 152-156 und 161-164) und Fragen zum kindlichen Wortschatz (S. 156-160).

Der zweite Kernteil der Arbeit von Schnörch stellt zum anderen die Erarbeitung und Analyse einer Materialbasis dar. Diese Materialbasis – Schnörch nennt sie „Gesamtkorpus“ (ebda., S. 57) – gewinnt Schnörch aus der Zusammenführung der sieben zuvor beschriebenen Korpora. Sie hat einen Umfang von 26847 Einheiten, die sich aus je ca. 4000 Einheiten des jeweiligen Korpus zusammensetzen und um Dubletten bereinigt sind (vgl. ebda., S. 218). Sie dient ihm als Untersuchungsgegenstand für eine an Krohn angelehnte quantitative Oberflächenuntersuchung.⁴⁰

Allerdings geht Schnörch einen Schritt weiter als Krohn: Er greift aus seinem „Gesamtkorpus“ eine Teilmenge heraus (er nennt sie „Untersuchungsschnittmenge“ (ebda., S. 226)) und unterzieht sie einer qualitativen Tiefenuntersuchung hinsichtlich semantischer, grammatischer und morphologischer Eigenschaften der in der Schnittmenge enthaltenen Wortschatzelemente.⁴¹ Anders als Krohn geht es Schnörch also nicht um die vielfache Konfrontation der Schnittmengen verschiedener Kombinationen von Korpora, sondern um die detaillierte qualitative Analyse einer spezifischen Kombination und Schnittmengenbildung von Einzelkorpora (vgl. Schnörch 2002, S. 58). Dabei ist es Schnörch, ähnlich wie Krohn, wichtig, „eine Änderung der Blickrichtung vorzunehmen“ (ebda., S. 56): Nicht die Auswahlkriterien stehen im Vordergrund, sondern der Vergleich bereits reduzierter Wortschätze. Die Zusammenstellung verschiedener Korpora dient der Ermittlung einer Teilmenge des Wortschatzes, die dann gründlich analysiert wird. Es geht Schnörch darum, den Wortschatz als solchen besser zu verstehen, seine Architektur und Zusammensetzung (vgl.

39. Die sieben Korpora sind: Kosaras 1970, Pfeffer 1970, Ruoff 1981, Rosengren 1972 und 1977, Augst 1985, *Zertifikat Deutschals Fremdsprache* 1992, Krohn 1992. Für einen tabellarischen Überblick vgl. Schnörch 2002, S. 192-194.

40. D.h. die Erstellung von thematischen Profilen anhand der Analyse von Hapaxlegomena und die Bildung von Schnittmengen zum Nachweis von Gemeinsamkeiten (Schnörch 2002, S. 219-226). Schnörch zeigt, dass die Bildung von thematischen Profilen allein anhand der Wortform, also ohne Bedeutungsangaben, zwar in eingeschränkter Weise für einen ersten Überblick über die „thematische[n] Auffächerung“ (ebda., S. 222) eines Wortschatzes herangezogen werden kann, dass sie aber für eine detailliertere Bearbeitung eines Wortschatzes kaum von Nutzen ist. Die Bildung von Schnittmengen könne zwar Gemeinsamkeiten zwischen zwei Wortschätzen aufdecken, diese Gemeinsamkeiten seien aber vor allem Gemeinsamkeiten der Form (ebda., S. 227f.)

41. Zur Untersuchungsschnittmenge gehören alle diejenigen Einheiten, die den Korpora von Kosaras 1970, Pfeffer 1970, Ruoff 1981 und Rosengren 1972 und 1977 gemeinsam sind.

ebda., S. 382), und die Grundwortschatzlexikographie nicht als Arbeit an Listen von Wörtern zu begreifen, sondern als „integrative“ (ebda., S. 8) Wortschatzarbeit.⁴²

Doch nicht nur in der Tiefenuntersuchung unterscheidet Schnörch sich von Krohn, auch in der Wahl der Grundeinheit sind die beiden Ansätze verschieden. Während Krohn mit Lexemen arbeitet, bildet das Semem die Grundlage aller quantitativen und qualitativen Untersuchungen bei Schnörch:

„Zähleinheit sind nicht Lexeme, sondern Sememe. [...] Als Ergebnis wird ein Kernwortschatz angestrebt, dessen Präsentation sich nicht in Form einer bloßen Liste von Wörtern erschöpft; diese wird vielmehr ergänzt um semantische, morphologische und grammatische Angaben.“ (Schnörch 2002, S. 187f.)

Ähnlich wie bei der eingehenden Vorstellung der sieben Ausgangskorpora verwendet Schnörch die Schnittmengenbildung auch, um allgemeine Fragen der Grundwortschatzlexikographie zu diskutieren. Zu den dabei angesprochenen Themen gehören Fragen zum Vergleich als Methode (S. 189-191 und 215f.), das Problem der Lemmatisierung (S. 216f.), die Frage nach dem Beitrag von Hapaxlegomena zum Profil eines Wortschatzes (S. 219-226), die Bewahrung semantischer Informationen bei der Schnittmengenbildung (S. 226-230), Fragen der Bedeutungsanalyse und der Wortartenzuordnung (S. 231-235), die Valenz von Verben (S. 236-253), das Verhalten von Substantiven und Adjektiven (S. 253-257), das Vorkommen von Sinnrelationen in der Schnittmenge (S. 257-262), die Diskussion des Sinns quantitativer Auswertungen wie Textdeckungstests (S. 277f.), die Unterscheidung von Syn- und Autosemantika (S. 278-280), die Frage nach der kleinsten Zähleinheit im Zusammenhang mit Komposita (S. 281-295) und die Zusammensetzung der Untersuchungsschnittmenge in Bezug auf Komposita (S. 293f.), Nomina agentis (S. 295-297), Movierungen und Diminutivbildungen (S. 297-300), Verben (S. 301-305) und Adjektive (S. 305f.). Schließlich rundet er seine Untersuchungen ab mit einer allgemeinen Diskussion der Grenze zwischen Grammatik und Lexikon, zwischen Wortbildung und Lexem in der lexikographischen Praxis (S. 306-311), die Frage nach dem Umgang mit Fremdwörtern (S. 311-335), die Frage nach dem Vorkommen und der Rolle von Sinnrelationen in der (Grundwortschatz-)Lexikographie (S. 336-371) und schließlich die Frage nach der richtigen Organisation eines (Grundwortschatz-)Wörterbuchs (S. 371-377).

Damit stellt Schnörchs Arbeit wie die Beiträge von Krohn und Kühn eine an den methodischen Grundlagen orientierte Untersuchung dar. Schnörch erweitert den Blick der bis dahin geführten Diskussion und führt ihn weg von der alleinigen Konzentration auf die Diskussion der Selektionskriterien hin zu Fragen der Darstellung, zu Fragen der semantischen Unterscheidung und Fragen der Verwendung und fordert statt der bisherigen auf Selektionskriterien fixierten Grundwortschatzarbeit eine „echte[n] Grundwortschatzlexi-

42. Ähnlich auch Schnörch 2002, S. 380.

kografie“ (ebda., S. 380).⁴³ Auch betont er immer wieder die Relativität, d.h. die Abhängigkeit eines Grundwortschatzes von dem Zweck, zu dem er erstellt wird.⁴⁴

Hand in Hand mit dieser Gründlichkeit der Betrachtung geht die Tatsache, dass Schnörch viele verschiedene Aspekte der Grundwortschatzarbeit, die über die Selektionskriterien hinausgehen, mit Hilfe der Erkenntnisse anderer linguistischer Teildisziplinen beleuchtet und damit zu einer ‘Verwissenschaftlichung’ der Diskussion beiträgt.

Eine weitere Stärke der Arbeit von Schnörch ist ihre angemessene Mischung aus kritisch-wissenschaftlicher Diskussion und einer guten Portion Sinn für die lexikographische Praxis und ihre Begrenzungen. Das zeigen etwa die Diskussion zur Lösung der Frage, wann ein Wort lexikalisiert ist und wann es durch Wortbildungsregeln erklärt werden kann, oder die Diskussion zum Streit um die Objektivität von Selektionskriterien und zu Fragen der Darstellung: Anstatt den Mangel an Objektivität bei der Auswahl von Lexemen in monotoner Weise zu kritisieren, betont Schnörch immer wieder, dass Objektivität gar nicht erreichbar sei und stattdessen durch größtmögliche Transparenz ersetzt werden müsse.⁴⁵

Weitere Stärken der Arbeit von Schnörch sind darüber hinaus die Ausgewogenheit, mit der er bisherige Arbeiten der Grundwortschatz- und Häufigkeitslexikographie beurteilt, sowie die gründliche und umfassende Darstellung der sieben untersuchten Korpora.⁴⁶

Auf der anderen Seite bestehen einige Unklarheiten: So wird zum Beispiel nicht klar, ob Schnörch davon ausgeht, dass es ein bestimmtes Zentrum des Deutschen, also einen zentralen Wortschatz des Deutschen, überhaupt gibt. Der Titel seiner Arbeit – „Der zentrale Wortschatz des Deutschen. Strategien zu seiner Ermittlung, Analyse und lexikografischen Aufarbeitung“ – deutet darauf hin, dass er von der Existenz eines solchen Zentrums ausgeht. Zugleich betont er aber immer wieder die Relativität des Grundwortschatzes und dass es den einen Grundwortschatz nicht geben könne (s. oben). Nun ist es durchaus möglich, dass Schnörch die beiden Begriffe *Grundwortschatz* und *zentraler Wortschatz* nicht als synonym versteht und verwendet wissen will; eine explizite Darlegung dieser Auffassung fehlt jedoch ebenso wie eine Definition des Begriffs *Grundwortschatz*. Zwar wird implizit klar, dass Schnörch mit dem Terminus *Grundwortschatz* nicht nur den Grundwortschatz der Fremdsprachendidaktik meint, sondern auch den Rechtschreibgrundwortschatz (Schnörch

43. Vgl. etwa auch Schnörch 2002, S. 23, S. 74, S. 336 oder S. 380.

44. Vgl. etwa Schnörch 2002, S. 38, S. 52, S. 74, S. 80 und S. 81.

45. Im Falle der Lexikalisierungsfrage schlägt Schnörch eine Fall-zu-Fall-Entscheidung anhand der Daumenregeln von regelhafter Paraphrasierung vs. blockhafter Bedeutung vor (vgl. Schnörch 2002, S. 286). Für die Rolle der Transparenz des Verfahrens gegenüber der Objektivität vgl. etwa Schnörch 2002, S. 56, 96f., 103, 124f., 132, 150, 202, 243, 258 und den gesamten Abschnitt 3.2, in dem Schnörch bei der gründlichen Darstellung seiner Vorgehensweise seiner eigenen Forderung nach Transparenz folgt.

46. Vgl. etwa Schnörchs klare Darstellung und ausgewogene Beurteilung des methodisch schwierigen Verfahrens, das Pfeffer (1970) anwendet (Schnörch 2002, S. 63-68), Schnörchs Diskussion der Argumente der Gegner des Frequenzansatzes (Schnörch 2002, S. 74) und ihre vergebliche Forderung nach absoluter Objektivität und Konsistenz der Selektion (Schnörch 2002, S. 361f.) oder die abschließende Beurteilung des Lemmatisierungsverfahrens bei Rosengren (1972 und 1977) (Schnörch 2002, S. 140). – Ähnlich beurteilt auch Schulz (2004, S. 332) die Arbeit von Schnörch.

2002, S. 23) und bis zu einem gewissen Grad auch das von Herbst (1986) entwickelte definitorische Vokabular (vgl. Schnörch 2002, S. 39-44). Zudem referiert Schnörch auch die von Hoffmann (1984) entwickelte Unterscheidung der Begriffe *häufigster Wortschatz*, *Grundwortschatz* und *lexikalisches Minimum* (vgl. Schnörch 2002, S. 26f.); es fehlt aber eine eindeutige Stellungnahme dazu, was genau Schnörch mit dem Begriff *Grundwortschatz* bezeichnen wissen möchte.

In engem Zusammenhang mit der Frage, was Schnörch mit den Begriffen *Grundwortschatz* und *zentraler Wortschatz* abdecken möchte, steht auch der von Matthias Schulz vorgebrachte Zweifel, ob ein Vergleich mehrerer Grundwortschätze und Häufigkeitswörterbücher einen Rückschluss auf den Gesamtwortschatz des Deutschen überhaupt zulässt:

„Kann das Konglomerat aus Lexemen [...] aus sieben disparaten Korpora tatsächlich als Untersuchungsbasis für die Frage nach dem zentralen Wortschatz sinnvolle Verwendung finden? U. Schnörch hat diesen Einwand vorausgesehen und begegnet ihm zunächst mit Zitaten wissenschaftstheoretischer Argumentationen: ein *tertium comparationis* legitimierte auch scheinbar sinnlose Vergleiche, man könne also sehr wohl, so U. Schnörch, Äpfel mit Birnen vergleichen (S. 189). Das überzeugt für den vorliegenden Fall aber nicht unbedingt: natürlich können Äpfel und Birnen verglichen werden (etwa hinsichtlich der Saftausbeute), es ist aber fraglich, ob als Ergebnis zwangsläufig brauchbare Aussagen über 'typische' Früchte von Obstbäumen erreicht werden.“ (Schulz, M. 2004, S. 333)

Tatsächlich erscheint der Transfer von den Einzelkorpora auf den Gesamtwortschatz etwas gewagt, nicht weil ein solcher Transfer überhaupt nicht möglich wäre, sondern weil die Auswahl der Korpora, aus denen die Schlussfolgerungen über den Gesamtwortschatz gezogen werden sollen, nicht unter Aspekten der Repräsentativität – wie etwa Zeitraum und Umfang der Erhebungsgrundlage – geschehen ist, sondern anhand von Kriterien der gründlichen Darstellung und Nachvollziehbarkeit (vgl. Schnörch 2002, S. 57) und vermutlich auch einfach der Zugänglichkeit. Das alles sind legitime Kriterien, anhand derer man Wortschätze für einen Vergleich zusammenstellen kann, sie führen aber nicht zu einer Korpusgrundlage, die unmittelbare Schlussfolgerungen für den Gesamtwortschatz und einen eventuellen allgemeingültigen Kernwortschatz zulässt.

Ein weiterer Einwand von Schulz gegen das Verfahren von Schnörch bezieht sich auf die automatische Übernahme des Eintrages eines Wörterbuches als ein Semem; diese sei irreführend (vgl. Schulz 2004, S. 333), weil die Gliederung eines Wörterbuches, d.h. die Regeln, die bestimmen, wann ein Lexem als ein Stichwort angesetzt wird und wann es in mehreren Stichwörtern erfasst wird, in hohem Maße von dem konkreten lexikographischen Verfahren des jeweiligen Wörterbuches abhängen. Das stimmt sicherlich, andererseits ist es verständlich, dass Schnörch, um die schiere Menge von rund 27 000 Datensätzen bewältigen zu können, ein solches vereinfachendes Verfahren anwenden musste.

Es lassen sich also auch an Schnörchs Verfahren Schwächen feststellen. Sie fallen jedoch kaum ins Gewicht, wenn man bedenkt, welchen Beitrag zur Grundwortschatzforschung Schnörch mit dieser Arbeit leistet: Er diskutiert ausführlich wesentliche, bis dahin vernachlässigte Fragen der Grundwortschatzlexikographie und verknüpft sie mit Erkenntnissen aus

anderen Teildisziplinen, wenn er etwa Fragen der adäquaten Zähleinheit mit der Diskussion zur Auflösung von Mehrdeutigkeiten und der Bildung von Sememen verbindet. Dadurch gelingt es ihm, die bis dahin an der lexikographischen Praxis orientierte Diskussion an lexikologische Themen anzuschließen, wenn er auch auf die sprachdidaktische Anwendung zentraler Wortschätze beschränkt bleibt. Hinzu kommt noch die gründliche Analyse der Einzelkorpora, die Schnörch liefert und die einen überaus wertvollen Beitrag zur bisherigen Geschichte der Grundwortschatzlexikographie darstellt. Wenn man also den von Schulz kritisierten Anspruch auf die Übertragbarkeit der Ergebnisse der auf den sieben Korpora beruhenden Schnittmenge auf den Gesamtwortschatz ein wenig in den Hintergrund stellt, dann wird deutlich, dass Schnörchs Arbeit neben der Arbeit von Kühn (1979) den bisher wichtigsten und umfassendsten Beitrag im Rahmen der Erforschung des Konzepts zentraler Wortschätze darstellt.

Kapitel 3: Anwendungstypen des Konzepts zentraler Wortschätze

In der bisherigen Literatur wurde das Konzept zentraler Wortschätze zumeist auf den auf die Didaktik von Fremdsprachen ausgerichteten Grundwortschatz bezogen und darauf begrenzt. Wenn man sich aber in anderen Bereichen der angewandten Linguistik umsieht, dann wird deutlich, dass es dort Wortschatzabgrenzungen gibt, die dem des Grundwortschatzes ähneln, wenn sie auch unterschiedliche Zielsetzungen haben und sich nicht explizit auf ein Konzept zentraler Wortschätze beziehen. Wir nehmen also zunächst einmal an, dass diese Wortschatzabgrenzungen auch Kandidaten für Instanzen des Konzeptes zentraler Wortschätze sind, und wollen sie hinsichtlich der ihnen gemeinsamen Eigenschaften näher betrachten.

Ziel dieser Zusammenschau ist es, das Konzept zentraler Wortschätze auf eine breitere Basis zu stellen und von der spezifisch didaktischen Anwendung gelöst zu betrachten. Zwar haben auch andere eine solche Begriffserweiterung bereits angedeutet.⁴⁷ Was jedoch noch fehlt, ist eine systematische Zusammenstellung aller bisher explizit oder implizit vorgeschlagenen Anwendungstypen des Konzeptes zentraler Wortschätze.⁴⁸ Aus den durch den Vergleich der verschiedenen Anwendungen herausdestillierten gemeinsamen Eigenschaften wollen wir in Kapitel 6 eine genauere Definition des Konzeptes zentraler Wortschätze herausarbeiten.

3.1 Die Grundwortschätze der Sprachdidaktik

Die Grundwortschätze der Sprachdidaktik stellen den ursprünglichen Typ eines zentralen Wortschatzes dar, also das, was in dieser Arbeit spezifisch als *Grundwortschatz* bezeichnet wird. Aus den Diskussionen rund um diesen Anwendungstyp stammen die bisherigen Erkenntnisse und Vorgehensweisen der Arbeit an zentralen Wortschätzen.

Zweck der Anwendung des Konzeptes zentraler Wortschätze in der Sprachdidaktik ist das Bemühen um eine effektivere Gestaltung des Sprachunterrichts: Indem man die wichtigsten Wörter zuerst lernt, ist man schneller im Stande, richtig zu kommunizieren. Eine Rolle spielt bei diesen Überlegungen auch die Tatsache, dass der Wortschatz als ein offenes System mit potentiell unendlich vielen Elementen in der Praxis der Sprachvermittlung eine quantitative Reduzierung erfordert.⁴⁹

47. Vgl. etwa Bonan-Garrigues (1993, zum Beispiel S. 6f.), die explizit die Verwendung eines hierarchisierten elektronischen Wörterbuchs auf viele verschiedene Anwendungen bezieht, oder Schnörch (2002, zum Beispiel S. 39), der in seine Betrachtungen zum Grundwortschatz auch das definitorische Vokabular miteinbezieht.

48. Eine Bemerkung zum Terminus *zentraler Wortschatz*: Bereits Schnörch 2002 verwendet den Terminus *zentraler Wortschatz*, allerdings in einer offensichtlich engeren Bedeutung. Schnörch gibt keine klare Definition der von ihm verwendeten Begriffe, doch wird implizit klar, dass er die Bezeichnung *zentraler Wortschatz* vor allem mit dem spezifischen Typus eines sprachdidaktischen Wortschatzes (Grundwortschatz) gleichsetzt.

49. Vgl. etwa Erk 1972, S. 9.

Das Konzept des Grundwortschatzes wird dabei in allen Bereichen der Sprachdidaktik angewandt: Die ersten Grundwortschätze entstanden im Rahmen der Fremdsprachendidaktik. Beim Grundwortschatz der Fremdsprachendidaktik handelt es sich um den Grundwortschatz im engeren Sinn, also um denjenigen Typus eines zentralen Wortschatzes, für den sich der Begriff *Grundwortschatz* etabliert hat und der sich an Studierende des Deutschen als Fremdsprache wendet. Später folgten Grundwortschätze und Grundwortschatzuntersuchungen für den Rechtschreibunterricht von muttersprachlichen Schülern, bei denen es darum ging, den Rechtschreibunterricht effizienter zu gestalten, indem man sich auf Wörter konzentriert, die in vielen Zusammenhängen eine wichtige Rolle spielen.⁵⁰ Schließlich gibt es auch Bemühungen, Grundwortschätze bzw. Vokabularien für die Didaktik von Fachsprachen zu entwickeln.⁵¹ Beim Grundwortschatz der Fachsprachendidaktik handelt es sich um einen dem Grundwortschatz im engeren Sinn verwandten zentralen Wortschatz, bei dem es um die Vermittlung der zentralen Begriffe einer Fachsprache geht, also um die Einführung in eine 'fremde' Sprache im Sinne einer Terminologie, die dem Lernenden zunächst unbekannt ist.⁵²

Treibende Kraft der Wortschatzauswahl in der Grundwortschatzlexikographie ist der kommunikative Nutzen, den die ausgewählten Wörter dem Sprachlernenden bieten. Dabei ist umstritten, wie dieser kommunikative Nutzen praktisch erfasst werden kann. Die meisten bisherigen Ansätze lassen sich dabei zwei Ausrichtungen zuordnen – der frequenzorientierten Ausrichtung und der kommunikativ-pragmatisch orientierten (vgl. Abschnitt 2.1.1).

In Verbindung mit der Erstellung solcher didaktisch orientierter zentraler Wortschätze werden folgende Themen besonders intensiv diskutiert: die grundsätzliche Frage nach der Rolle der Wortschatzauswahl und Wortschatzarbeit in der Fremdsprachendidaktik, die Frage danach, wie aussagekräftig und nützlich sogenannte Textdeckungstests sind, um die Adäquatheit eines Wortschatzes bzw. den Schwierigkeitsgrad eines Textes beurteilen zu können, und die Frage danach, welche Rolle die Unterscheidung von rezeptivem gegenüber produktivem und potentielltem Wortschatz bei der Grundwortschatzlexikographie spielt und welchen Einfluss die Eigenheiten von geschriebener gegenüber gesprochener Sprache auf die Grundwortschatzlexikographie haben.⁵³

50. Vgl. etwa die Grundlagenforschung bei Augst 1983 und Augst 1985 und Plickat 1983 als ein Beispiel für einen onomasiologisch aufbereiteten Rechtschreibgrundwortschatz. – Anzumerken ist auch, dass Wortschätze in der muttersprachlichen Didaktik nicht nur für die Rechtschreibung eingesetzt werden, sondern auch zur Wortschatzerweiterung. Ob man dabei jedoch noch von einem Grundwortschatz sprechen kann, scheint fraglich, weil es bei der Wortschatzerweiterung ja offensichtlich nicht um einen quantitativ begrenzten Wortschatz geht, sondern um den Ausbau eines existierenden Wortschatzes.

51. Vgl. die Grundlagenforschung von Heinrich Erk (Erk 1972, Erk 1975, Erk 1982).

52. Am Beispiel der fachsprachlich ausgerichteten Grundwortschätze lässt sich besonders gut zeigen, wie schnell sich die Rolle eines Grundwortschatzes von einem Hilfsmittel zum effektiven Wortschatzerwerb zu einem Mittel der Normgebung ändern kann. Viele fachsprachliche Grundwortschätze haben die Rolle einer Einführung in das Fachgebiet (vgl. z. B. Wieland 2000), sie bilden nicht nur den Grundwortschatz eines Fachgebietes ab, sondern auch das dazugehörige Grundwissen, und definieren damit, was zu diesem Grundwissen gehört.

Zusätzlich geht es im Bereich der Grundwortschatzarbeit für Fachsprachen um Fragen der richtigen Abgrenzung bzw. Einteilung nach Fachgebieten, um die Abgrenzung von lexikosemantischen und enzyklopädischen (sachlichen) Informationen, um die Abgrenzung von Fachsprache und Gemeinsprache und um die Frage, welches Wortschatzmodell das Verhältnis der einzelnen Sprachvarietäten zueinander am besten abbildet.⁵⁴ Darüberhinaus scheint ein Konsens darüber zu bestehen, dass das Selektionskriterium der Frequenz bei der Ermittlung von wichtigen Fachtermini weniger problematisch ist als bei der Erarbeitung eines allgemeinsprachlichen Grundwortschatzes.⁵⁵

Diese Übersicht über den Anwendungstyp Grundwortschatz zeigt bereits, dass es dabei nicht allein um Grundwortschatz-spezifische Fragen geht, sondern auch um solche Fragen, die mit dem Gebiet der Fachsprachen und der Lexikologie im allgemeinen zu tun haben.

3.2 Das kontrollierte definitonische Vokabular der Fremdsprachenlexikographie

Beim kontrollierten definitonischen Vokabular der Fremdsprachenlexikographie handelt es sich um einen begrenzten Wortschatz, der für die Erklärung von Begriffen in einem für fremdsprachige Nutzer gedachten einsprachigen Wörterbuch verwendet wird. Der Zweck definitonischer Vokabularien ist es, eine verständliche und konsistente Definition der Stichwörter des Wörterbuches zu gewährleisten, indem dem Lexikographen eine begrenzte, d.h. überschaubare Anzahl von Lexemen mit einer wohldefinierten Bedeutung an die Hand gegeben werden, mit deren Hilfe er die Stichwörter des Wörterbuchs konsistent und in verständlicher Weise beschreiben kann. Der Vorteil dieses Verfahrens liegt einerseits in einer erhöhten Konsistenz der Definition von Stichwörtern und somit einer besseren Verständlichkeit für den Benutzer, weil die Definitionen in weiten Strecken auf die gleichen Definitionsbegriffe zurückgreifen, andererseits in einem erhöhten Bewusstsein des Lexikographen dafür, in welchen Fällen er 'ungewöhnliche', d.h. nicht im definitonischen Vokabular verzeichnete Begriffe zur Definition eines Stichwortes heranzieht.

Ein bekanntes und ausführlich besprochenes Beispiel der praktischen Umsetzung eines definitonischen Vokabulars ist das des *Longman Dictionary of Contemporary English*

53. Zur Rolle der Wortschatzauswahl in der Sprachdidaktik vgl. etwa Schnörch 2002, S. 28-38. Zur Diskussion von Textdeckungstests vgl. etwa Kaufmann 1968, S. 13, Krohn 1992, S. 57 und S. 116-124 oder Schnörch 2002, S. 48 und 80. Zur Rolle der Unterscheidung von rezeptivem und produktivem Wortschatz vgl. etwa Bonan-Garrigues 1993, S. 22, Kühn 1979, S. 23 oder Krohn 1992, S. 22f. Zur Frage des Einflusses von geschriebener vs. gesprochener Sprache auf die Grundwortschatzlexikographie vgl. schließlich etwa Schnörch 2002, S. 78-80.

54. Zur Frage der Abgrenzung von Fachgebieten vgl. etwa Erk 1972, S. 30. Zur Frage der Abgrenzung von lexikalisch-semantischem gegenüber enzyklopädischem Wissen vgl. etwa Jahr 1996, S. 25 oder Kucera 1986, S. 145. Zur Frage der Abgrenzung von Fachsprache vgl. etwa Krohn 1992, S. 46 oder 139f. und insbesondere Abschnitt 5.3.1 in dieser Arbeit. Zur Frage des adäquaten Wortschatzmodells zur Abbildung von Fachsprachen vgl. schließlich etwa Rosengren 1986, S. 158f. und insbesondere Kapitel 6 in dieser Arbeit.

55. Vgl. etwa Kühn 1979, S. 53.

(LDOCE 1984). Die darin praktizierte Anwendung des Konzepts eines definatorischen Vokabulars ist bei Herbst 1986 ausführlich analysiert und besprochen. Herbst beschreibt die Funktion des definatorischen Vokabulars folgendermaßen:

„[The] defining vocabulary provides an instrument to enforce lexicographical discipline and [...] it contributes to giving 'the meanings of the words in plain, simple, unambiguous language' [= Sweet 1899, S. 147; V.H.].“ (Herbst 1986, S. 113).

Da das definatorische Vokabular von Lexikographen zur Definition von Stichwörtern herangezogen werden soll, sind Fragen der Darstellung besonders wichtig. Es geht darum, die Bedeutungen und vor allem die Anwendungsmöglichkeiten der einzelnen Elemente des Vokabulars möglichst präzise anzugeben. Fragen der Selektion treten hier eher in den Hintergrund, weil der Umfang des definatorischen Vokabulars gegenüber der Darstellung und vor allem der Verwendbarkeit des Vokabulars eine untergeordnete Rolle spielt. Denn...

„...it is far more important for the defining vocabulary to be sufficiently large to provide clear and semantically accurate definitions than to keep the number of items artificially low.“ (ebda., S. 114).

Das definatorische Vokabular des LDOCE 1984 beruht auf einer Mischung aus Frequenzlisten und didaktischen Grundwortschätzen des Englischen. Der Grund, warum man sich hier nicht nur auf Frequenzlisten gestützt hat, liegt zunächst darin, dass es sich bei dem LDOCE 1984 um ein Wörterbuch handelt, das sich spezifisch an Fremdsprachlernende des Englischen wendet, und da lag es nahe, andere didaktische Hilfsmittel der Wortschatzvermittlung als Grundlage heranzuziehen. Aber auch die spezifischen Schwächen der Frequenz als alleinigem Selektionskriterium werden so kompensiert, wie Herbst (1986, 102f.) zeigt: Für die Definition von Stichwörtern eines Wörterbuchs seien oft abstrakte Begriffe nötig, die in Frequenzlisten zumeist einen niedrigen Rang haben. Diese Abstraktionen – Herbst nennt als Beispiele *atom*, *item*, *unit* – seien notwendig, um Definitionen bestimmter Wörter nicht unnötig kompliziert und vielschichtig zu machen. Darüberhinaus liefern Frequenzlisten häufig Wörter mit ähnlicher Bedeutung in einer ähnlichen Frequenzstufe, während andere Wörter gänzlich in den Frequenzlisten fehlen.

Ein für das definatorische Vokabular spezifisches Problem ist das des Konfliktes zwischen der Präzision, die das definatorische Vokabular bei der Definition der Stichwörter zulässt, und der Konsistenz und Verständlichkeit bzw. Einfachheit des in den Definitionen verwendeten Vokabulars. Die Präzision der Definition erfordert in der Regel, dass dem Lexikographen eine reiche Auswahl an Lexemen zur Verfügung steht, mit denen er die spezifischen unterscheidenden Merkmale des zu definierenden Stichwortes beschreiben kann. Die Konsistenz und Verständlichkeit der in den Definitionen verwendeten Begriffe wiederum verlangt in der Regel ein möglichst geringes Maß an Variation. Diesem Problem wird im LDOCE 1984 Rechnung getragen, indem man sogenannte „Mischrepräsentationen“ (Petöfi 1977, S. 303) zulässt, d.h. man erlaubt Definitionen, die aus Lexemen des definatorischen Vokabulars und aus externen Lexemen bestehen, solange diese externen Lexeme an anderer

Stelle im selben Wörterbuch definiert sind und als Verweis in der Definition gekennzeichnet sind (vgl. Herbst 1986, S. 102).⁵⁶ Herbst kommt zu dem Schluss:

„The number of definitions which would gain in precision from the use of words not in the defining vocabulary would certainly not be large enough to question the principle of using a controlled defining vocabulary altogether.“ (ebda., S. 111)

Zu unterscheiden ist das kontrollierte definitorische Vokabular von kontrollierten Vokabularien bestimmter Fächerdisziplinen, wie etwa der Technik. Kontrollierte Vokabularien in Fachsprachen sind Sets von Fachtermini, die mit einer wohldefinierten Bedeutung versehen sind. Sie stellen eine Art Glossar oder Terminologiedatenbank dar und können beliebig groß sein.⁵⁷ Damit aber gehören solche kontrollierten Vokabularien nicht in den Bereich zentraler Wortschätze, weil es bei ihnen nicht um die mengenmäßige Begrenzung geht. Beim kontrollierten definitorischen Vokabular hingegen geht es durchaus um die mengenmäßige Begrenzung der Einheiten, weil durch sie die konsistente und vor allem verständliche Definition der Stichwörter gewährleistet werden soll.

Auch bleibt anzumerken, dass die Verwendung eines definitorischen Vokabulars auch auf kritische Stimmen gestoßen ist, wie etwa bei Carter (1998, S. 153): Er führt gegen das definitorische Vokabular an, dass man zum einen nicht davon ausgehen könne, dass die Wörter, die zur Definition herangezogen werden, dem Fremdsprachenlerner bekannt sind, zumal einige der verwendeten Wörter nicht unbedingt gebräuchlich seien.⁵⁸ Darüberhinaus gehe die Einfachheit der Definition oft auf Kosten der Genauigkeit, und die Beschränkung auf wenige Wörter könne in manchen Fällen dazu führen, dass das Ergebnis unnötig umständlich werde. Darüberhinaus funktionierten definitorische Vokabularien bei der Beschreibung von Konkreta besser als für die Beschreibung von Abstrakta. Insgesamt kommt Carter allerdings zu dem Schluss, dass die Vorteile der Verwendung eines definitorischen Vokabulars seine Schwächen überwiegen, zumal fremdsprachige Wörterbuchbenutzer auf definitorischen Vokabularien aufbauende Definitionen bevorzugten.

Für die hier vorzunehmende grundlegende Betrachtung verschiedener Anwendungsmöglichkeiten des Konzeptes zentraler Wortschätze sind die bei Herbst angeführten allgemeinen Prinzipien der Erstellung eines definitorischen Vokabulars von besonderem Interesse. Nach Herbst (1986, S. 113ff.) muss ein definitorisches Vokabular folgende Kriterien erfüllen:

-
56. Allerdings zeigt Herbst in seiner Analyse der Verwendung externer Lexeme in Definitionen des LDOCE 1984, dass die Regel der Kreuzreferenz, d.h. des Anführens und Verweisens auf externe Lexeme der Definition nicht immer konsequent durchgehalten ist (vgl. Herbst 1986, S. 105ff.). Besonders Eigennamen und Zahlen durchbrechen diese Regel.
 57. Vgl. zum Beispiel EURODICAUTOM, die Terminologiedatenbank der europäischen Kommission (online abrufbar unter <http://europa.eu.int/eurodicautom/Controller>).
 58. Vgl. die bei Carter 1998, S. 153 angeführten Beispiele.

- Seine Einträge müssen dem Wörterbuchbenutzer vertraut sein.
- Die Einträge müssen konkrete Lexeme sein, nicht abstrakte semantische Komponenten; zugleich sollen ihnen spezifische Bedeutungen zugeordnet werden.
- Die Verwendung von Morphemen soll nur bei durchsichtigen Bildungen zugelassen sein.
- Der Umfang des definitorischen Vokabulars soll gegenüber seiner Eigenschaft, präzise Definitionen zuzulassen, eine untergeordnete Rolle spielen.
- Die Selektion seiner Einträge soll nach den Kriterien der Vertrautheit und der lexikographischen Nützlichkeit erfolgen, sodass zum Beispiel Internationalismen (wie etwa das Lexem *Prinzip*) der 'einheimischen' Variante (wie etwa dem Lexem *Grundsatz*) und Abstrakta, die der vereinfachten Definition eines Stichwortes dienen, vorgezogen werden.
- Und schließlich soll es erlaubt sein, in einer Wörterbuchdefinition die Einheiten des definitorischen Vokabulars mit externen Lexemen, die an anderer Stelle im Wörterbuch definiert sind, zu kombinieren.

Interessant an dieser Liste von Prinzipien ist zum einen, dass Herbst ganz deutlich die lexikalische Einheit von einer Wortform und einer Wortbedeutung zur Grundeinheit des definitorischen Vokabulars macht. Damit stellt er die Frage nach der Semantik in den Mittelpunkt. Auch die Betonung der lexikographischen Nützlichkeit als Selektionskriterium ist interessant, macht sie doch deutlich, wie sehr Selektionskriterien von der Anwendung des zu erstellenden zentralen Wortschatzes abhängig sind und wie arbiträr ihre Anwendung sein kann.⁵⁹

3.3 Zentrale Wortschätze in Anwendungen der maschinellen Sprachverarbeitung

Sieht man sich die verschiedenen Anwendungen an, die aus der maschinellen Sprachverarbeitung erwachsen sind, so entdeckt man, dass das Konzept zentraler Wortschätze dort mehrfach angewandt wird, auch wenn die explizite Anknüpfung an seine linguistischen Wurzeln und eine Rückbindung zwischen Computerlinguistik und traditioneller Linguistik oft fehlen.⁶⁰

59. Vgl. auch Herbst 1986, S. 102.

60. Eine Ausnahme bildet hier die Arbeit von Bonan-Garrigues. Sie schlägt explizit eine Brücke zwischen Linguistik und Computerlinguistik und umreißt die Anwendungsmöglichkeiten eines hierarchisierten Wörterbuchs in der computergestützten Sprachverarbeitung folgendermaßen:
„Les dictionnaires électroniques constituent des outils formels indispensables à tout traitement automatique de la langue. Les différentes applications ne nécessitent cependant pas d'utiliser des outils aussi volumineux. Il est donc souvent souhaitable, voire nécessaire, de pouvoir effectuer un refaçonnage 'sur mesure'. Ainsi, des lexiques hiérarchisés, allant du vocabulaire le plus essentiel au moins essentiel, seraient utiles à de nombreuses applications informatisées: élaboration d'interfaces en langage naturel pour l'interrogation de bases de données grand public, mise en oeuvre de services télématiques, compression de dictionnaires pour l'analyse automatique de textes, etc.“ (Bonan-Garrigues 1993, S. 6f.)

3.3.1 Die Stopwortlisten des *Information Retrieval*

Ein Bereich, in dem das Konzept zentraler Wortschätze angewandt wird, ist der Bereich des *Information Retrieval*, einem Verfahren, das es einem Benutzer erlauben soll, Informationen in elektronisch gespeicherten Daten, z. B. Texten, mit Hilfe der Eingabe von Suchwörtern zu durchsuchen und abzurufen.⁶¹ Ein Beispiel einer solchen Anwendung sind Internet-Suchmaschinen.

Beim Verfahren des *Information Retrieval* werden, vereinfacht betrachtet, Texte zur Suche aufbereitet, indem ein Index über die in ihnen enthaltenen Informationseinheiten (in der Regel Wörter) angelegt wird. Bei diesem Indexierungsprozess wird unter anderem gezählt, wie häufig ein Wort in der Textmenge insgesamt vorkommt und in wievielen verschiedenen Texten es verwendet wird, wie verbreitet es also ist. Wörter, von denen man annimmt, dass sie in einer beliebigen Textmenge besonders häufig vorkommen und zugleich sehr verbreitet sind, werden in sogenannten Stopwortlisten registriert (vgl. Fox 1992, S. 113) und beim Indexieren eines Textes bzw. bei einer Suchanfrage, die eines dieser Wörter enthält, ignoriert oder als wenig relevant im Verhältnis zur Suchanfrage bewertet.⁶² Im Hintergrund steht dabei die Annahme, dass diese Wörter aufgrund ihrer Häufigkeit und ihres Verbreitungsgrades nichts über die spezifischen Charakteristika individueller Texte aussagen und daher nichts zu ihrer gegenseitigen Unterscheidung beitragen. Während es in den Anfängen des *Information Retrieval*, als leistungsstarke Prozessoren und Speicherplatz noch teure Ressourcen darstellten, beim Einsatz von Stopwortlisten noch darum ging, durch das Ignorieren der Stopwörter beim Indexieren Speicherplatz und Prozesszeit zu sparen (vgl. Fox 1992, S. 113), geht es heute mit der geringen Bewertung von Stopwörtern hinsichtlich ihrer Relevanz für die Suchanfrage um ein möglichst präzises Suchergebnis.

Traditionell setzt sich dabei die Liste der Stopwörter aus Funktionswörtern zusammen. Während in typischen Lehrbüchern zum *Information Retrieval* dieses Verfahren nur insofern problematisiert wird, als auf seine Schwächen hingewiesen wird,⁶³ diskutiert Fox (1992, S. 113) die Relativität dieser Stopwortliste ausführlicher und weist unter anderem darauf hin, dass innerhalb von fachsprachlichen Textsammlungen, wie etwa der Informatik, auch autosemantische Wörter wie *computer*, *program* oder *language* Stopwörter sein können. Damit rückt er den Begriff *Stopwortliste* deutlich in die Nähe des Konzepts zentraler Wortschätze. Neuere Verfahren des *Information Retrieval* scheinen sich allerdings vom Konzept der Stopwortliste zu lösen und stattdessen auf die Anreicherung des Indexes mit semantischen Informationen zu setzen (vgl. Tzoukermann / Klavans / Strzalkowski 2003, S. 538f.).

61. Für eine detailliertere Darstellung des Verfahrens des *Information Retrieval* vgl. Tzoukermann / Klavans / Strzalkowski 2003, S. 530.

62. Deshalb werden Stopwortlisten oft auch als „negative dictionary“ (Fox 1992, S. 113) bezeichnet.

63. Vgl. etwa das Nicht-Auffinden von Dokumenten, die das Shakespeare-Zitat *To be or not to be* enthalten (Jackson / Moulinier 2002, S. 67).

3.3.2 Der Neutralwortschatz zur Verfeinerung von Anwendungen der automatischen Textklassifikation

Ein weiterer Bereich, in dem das Konzept zentraler Wortschätze zumindest theoretisch in Betracht gezogen wird, ist der Bereich der automatischen Klassifikation von Texten nach Sachgebieten. Stark vereinfacht formuliert beruht dieses Verfahren darauf, dass man für das zu klassifizierende Dokument eine mathematische Repräsentation erstellt, einen sogenannten Dokumentenvektor, der aus allen Wörtern des Dokuments berechnet wird und so viele Dimensionen hat, wie das Dokument Wörter enthält. Dieser Dokumentenvektor wird dann mit den Repräsentationen der verschiedenen Sachgebiete verglichen, so dass schließlich das Dokument demjenigen Sachgebiet zugeordnet wird, dem der Dokumentenvektor am meisten gleicht.

Dabei gilt, dass das Verfahren desto effektiver und präziser wird, je mehr Wörter man aus der Berechnung dieser Repräsentation herausnehmen kann. An dieser Stelle kommt eine Variante eines zentralen Wortschatzes ins Spiel: Um das Verfahren der automatischen Klassifikation zu beschleunigen, lassen sich Wörter abgrenzen, die allen Texten eines Korpus gemeinsam sind und damit nichts zur thematischen Unterscheidung der Texte beitragen. Indem man sie bei der Berechnung der Textrepräsentation herausnimmt, lässt sich das Verfahren der automatischen Textklassifikation beschleunigen (vgl. Sebastiani 2006, S. 460).⁶⁴

Verankert ist diese Annahme in der Erkenntnis der Lexikologie und Sprachdidaktik, dass erst autosemantische Lexeme („content words“ (Hendrickx 1987, S. 11)) zur Unterscheidung zwischen der Ausdrucksweise verschiedener Gruppen bzw. Texte beitragen. Das von Rosengren (1986, S. 158f.) skizzierte Wortschatzmodell beschreibt dabei treffend den theoretischen Hintergrund:

„Die Standardsprache [...] besitzt ein Lexikon, das aus einem Kern und einer Reihe von Speziallexika besteht [...]. Der Kern besteht aus Wörtern, die in mehreren Situationen wiederkehren und die auch die Bausteine der verschiedenen Speziallexika sind.“

Um die Abgrenzung eben dieser „Bausteine“ geht es beim Neutralwortschatz der automatischen Textklassifikation.

Wie die Stopwortlisten des Information Retrieval wird auch der Neutralwortschatz als Hilfsmittel der automatischen Klassifikation von Texten zumeist auf die Funktionswörter beschränkt, was insofern verständlich ist, als man in der maschinellen Sprachverarbeitung auf eindeutige und möglichst sichere und kontextunabhängige Aussagen angewiesen ist

64. Anzumerken ist dabei allerdings, dass die alleinige (quantitative) Auswertung des in einem Text enthaltenen Vokabulars nur ungefähre Angaben über die Fachzugehörigkeit eines Textes machen kann; eine wirklich gesicherte Einordnung eines Textes danach, ob er ein Fachtext ist und wenn ja, welchem Fach er angehört, kann erst eine detaillierte Analyse aller sprachlichen Aspekte geben, also auch derjenigen, die unter- und oberhalb der Wortebene liegen (gehäuftes Vorkommen bestimmter Wortbildungen, die Verwendung bestimmter syntaktischer Muster etc.). Vgl. die Diskussion zur Fachsprachlichkeit in Abschnitt 5.3.1.

und man mit der Beschränkung auf die relativ klar abgrenzbare Gruppe der Funktionswörter eine solche sichere Aussagengrundlage hat. Von einem streng linguistischen Standpunkt aus gesehen wäre es allerdings sinnvoller, sowohl die Stopwortliste als auch den Neutralwortschatz als relative Größen zu verstehen, die abhängig sind davon, auf welche Textgrundlage sie bezogen werden sollen.⁶⁵

Wie diese kurze Übersicht zeigt, spielt bei beiden Typen von zentralen Wortschätzen in der maschinellen Sprachverarbeitung, d.h. den Stopwortlisten des *Information Retrieval* und dem Neutralwortschatz der automatischen Textklassifikation, die Abgrenzung ‘neutraler’, d.h. themenneutraler bzw. dokumentunabhängiger Wörter eine Rolle. Für diese Abgrenzung hat man sich bisher zumeist auf die Funktionswörter bzw. Synsemantika konzentriert. Was bisher jedoch fehlt, ist eine Einbindung semantischer Aspekte, die die Relativität der Zentralität von Wörtern je nach Thema und Textsorte berücksichtigt.

3.4 Abgrenzungen

Bei der Frage danach, in welchen Bereichen der angewandten Linguistik zentrale Wortschätze vorkommen, stößt man auch auf Wortschätze, die auf den ersten Blick die Charakteristika eines zentralen Wortschatzes zu tragen scheinen, die jedoch bei genauerem Hinsehen nur einige wenigen Eigenschaften mit zentralen Wortschätzen teilen. Wir wollen im Folgenden einen kurzen Blick auf diese Wortschätze werfen.

3.4.1 Häufigkeitswörterbücher

In der bisherigen Grundwortschatzdiskussion werden Grundwortschätze und Häufigkeitswörterbücher oft gemeinsam besprochen und bearbeitet. Das hängt sicherlich mit der Rolle der Häufigkeit als Selektionskriterium zusammen, aber auch mit den Problemen ihrer jeweiligen Erstellung. Denn genau wie bei der Erstellung eines zentralen Wortschatzes geht es auch bei der Erarbeitung von Frequenzwörterbüchern um grundlegende lexikographische Fragen etwa nach der Grundeinheit oder nach der Behandlung von Lemmatisierung und Polysemie. Diese beiden Konzepten gemeinsamen Fragestellungen sind es wohl auch, die zu einer terminologischen Vermengung der beiden Konzepte geführt haben (vgl. Abschnitt 4.1). Wenn man genau hinsieht, sind allerdings die Unterschiede zwischen Häufigkeitswörterbüchern und zentralen Wortschätzen zahlreicher als die Gemeinsamkeiten. Wir wollen daher hier einen kurzen Vergleich vornehmen.

Zunächst einmal lassen sich Frequenzwörterbücher definieren als „books that give a systematic representation of the frequential aspect(s) of a vocabulary, making use of a two-dimensional structure, viz. a macro- and a microstructure“ (Martin 1990, S. 1314).⁶⁶ Dabei

65. Vgl. auch Sebastiani 2006, S. 460. – Eine Bestätigung der Vermutung, dass der Neutralwortschatz einer Sammlung von Texten abhängig ist vom Thema und Genre der Textsammlungen ist etwa das bei Carter (1998, S. 108) zitierte Beispiel eines Vokabularvergleichs verschiedener Texte, die das gleiche Thema haben, aber verschiedenen Genres zuzuordnen sind, und bei dem deutlich wird, dass das für das jeweilige Genre richtige Vokabular von Genre zu Genre stark variiert.

gibt es grundsätzlich zwei Typen von Häufigkeitswörterbüchern: Auf der einen Seite gibt es Häufigkeitswörterbücher zur Erforschung der Häufigkeit von Wörtern in einem gegebenen, relativ breit gefassten Korpus. Hierbei handelt es sich um Häufigkeitswörterbücher im herkömmlichen Sinn. Daneben gibt es noch Häufigkeitswörterbücher, die das lexikalische Profil einer bestimmten Sprachvarietät, etwa einer Fachsprache, abbilden sollen, wie etwa die Arbeiten von Erk (1972, 1975, 1982). Dabei wird der Häufigkeitsuntersuchung bereits ein eng gefasstes, themenspezifisches Textkorpus zugrundegelegt. Ziel ist es, herauszufinden, welche Wörter für ein bestimmtes Themen- oder Fachgebiet typisch sind. Insofern lässt sich eine Verbindung zwischen den Grundwortschätzen zu Fachsprachen und Häufigkeitswörterbüchern zur Abbildung des lexikalischen Profils einer Sprachvarietät herstellen. Während jedoch ein fachsprachlicher Grundwortschatz wie alle zentralen Wortschätze quantitativ begrenzt ist, sind Häufigkeitswörterbücher in der Regel so groß wie der Wortbestand des Korpus, den sie abbilden sollen, also theoretisch so groß wie der Bestand aller erfassbaren Wörter einer Sprache oder Sprachvarietät.

Ein weiterer Unterschied zwischen zentralen Wortschätzen und Häufigkeitswörterbüchern besteht darin, dass Häufigkeitswörterbücher einen Zustand, d.h. eine sprachliche Eigenschaft beschreiben, während zentrale Wortschätze zu einem bestimmten Zweck eingesetzt werden, also Werkzeuge sind (vgl. Rosengren 1986, S. 156). Genau genommen handelt es sich beim Wörterbuchtyp eines Häufigkeitswörterbuchs also eher um eine besondere Form der Darstellung und Anordnung von Lexemen als um einen besonderen Wortschatz, ähnlich wie auch onomasiologische, semasiologische oder rückläufige Wörterbücher eher einen Darstellungstyp repräsentieren als einen spezifischen Wortschatz. Das Häufigkeitswörterbuch ordnet Lexeme nach ihrer Frequenz bezüglich eines bestimmten Korpus und stellt Informationen über die Frequenz dieser Lexeme bereit, genauso wie ein rückläufiges Wörterbuch Lexeme bezüglich ihrer Endungen anordnet und Informationen darüber gibt, welche Bestimmungsglieder ein Endglied eingehen kann.

Ein weiteres Unterscheidungsmerkmal zwischen zentralen Wortschätzen und Häufigkeitswörterbüchern liegt in der Art ihrer Entstehung: Während zentrale Wortschätze aus Textkorpora und Wörterbüchern gewonnen werden (können), bauen Häufigkeitswörterbücher in der Regel allein auf Textkorpora auf.⁶⁷

Schließlich unterscheiden sich zentrale Wortschätze und Häufigkeitswörterbücher auch hinsichtlich der Rolle der Frequenz: Während sie bei der Erstellung und Darstellung eines Häufigkeitswörterbuchs eine konstituierende Funktion hat, ist Frequenz bei der Erarbeitung von zentralen Wortschätzen ein Selektionskriterium unter vielen, jedenfalls, wenn man den jüngsten Entwicklungen und dem integrativen Ansatz von Schnörch 2002 folgt.⁶⁸

66. Martin weist darauf hin, dass auch andere Frequenzuntersuchungen, etwa von Morphemen, manchmal als Frequenzwörterbücher bezeichnet werde (ebda., S. 1314).

67. Vgl. auch Krohn 1992, S. 30f.

Auch vom Standpunkt der Diskussionen über die adäquate Selektion für einen zentralen Wortschatz ist eine Gleichsetzung von Häufigkeitwörterbüchern und zentralen Wortschätzen nicht angebracht, wie Carter (1998, S. 233-235) deutlich macht: Zum einen, so Carter, bestehe kein direktes Verhältnis zwischen der Frequenz eines Wortes und seiner Nützlichkeit. Des Weiteren habe die Mehrzahl aller Wörter in einem Korpus neben einer geringen Frequenz oft auch eine sehr geringe Distribution, also Verteilung über verschiedene Texte des Korpus hinweg. Gerade diese Wörter aber seien innerhalb eines bestimmten Themenbereiches oft sehr wichtig und hätten innerhalb eines Textes, der dieses Thema behandelt, eine höhere Frequenz. Ausserdem hätten Wörter mit hoher Frequenz oft homonyme Varianten, weshalb man bei Frequenzangaben und Lemmatisierung der Wörter des Korpus gründlich vorgehen müsse. Ebenso sei es wichtig, bei der Arbeit mit einem gegebenen Korpus nachzuprüfen, welche Definition von Wort bzw. Zählleinheit ihm zugrunde liege, ob beispielsweise Phraseologismen und orthographisch getrennt dargestellte Wörter als Einheiten erfassbar seien; dies, so Carter, sei besonders deshalb wichtig, als gerade zentrale Wörter tendenziell viele verschiedene Kollokationen eingehen könnten, und zwar gerade mit Wörtern, die selbst keine allzu hohe Frequenz aufwiesen. Auch ist es wichtig, dass das herangezogene Korpus eine breite Auswahl an Textarten erfasse.⁶⁹

Über der Diskussion zur Abgrenzung von zentralen Wortschätzen gegenüber Häufigkeitwörterbüchern soll nun nicht der Eindruck entstehen, Häufigkeitwörterbücher seien nutzlos. Zum einen stellen Häufigkeitsuntersuchungen von einem systemlinguistischen Standpunkt aus eine legitime und naheliegende Form der Beschreibung von Sprache dar:

„Die Frequenz von Erscheinungen ist in allen Bereichen der Sprachwissenschaft von oft bedeutendem Belang. Das gilt nicht nur für Relationen innerhalb des Wortschatzes, es gilt ebenso im phonetisch-phonologischen wie im morphologischen oder syntaktischen Bereich: die meisten Vorkommnisse des Sprachwandels hängen in irgendeiner Weise mit der Vorkommenshäufigkeit von Ausgangs- oder Endform zusammen; kein sprachliches System ist hinreichend beschrieben, wenn die Vorkommenshäufigkeit seiner einzelnen Teile nicht berücksichtigt ist.“ (Ruoff 1981, S. 8)

Zum anderen, aus einer praktisch-lexikographischen Perspektive betrachtet, stellen Frequenzen eine von mehreren Eigenschaftskategorien dar, anhand derer sich Wörter beschreiben lassen. Frequenz ist eine Eigenschaft von Wörtern, die sich für die lexikographische Praxis verwerten lässt, wenn man weiß, wie Frequenzbeschreibungen zu deuten sind. Gerade an dieser Stelle wurde jedoch in der Grundwortschatzlexikographie lange Zeit der Fehler gemacht, Ergebnisse von Frequenzerhebungen unkritisch mit dem Inhalt eines

68. Dagegen ist beim frequenzorientierten Ansatz (sowohl dem absoluten als auch dem modifizierten, vgl. Abschnitt 2.1.1) Frequenz tatsächlich ein konstituierendes Element der Erstellung eines zentralen Wortschatzes. Vgl. etwa die Fachwortschätze von Hoffmann (Hg.) 1970, 1970a, 1973, 1976, 1978, die auf reinen Frequenzuntersuchungen aufbauen. Bei den beiden pragmatisch orientierten Ansätzen (vgl. Abschnitt 2.1.1) wiederum ist Frequenz zumindest formal vom Verfahren völlig ausgeschlossen.

69. Die Forderung nach einem möglichst breit angelegten Korpus gilt wohl vor allem vor dem Hintergrund der Fremdsprachendidaktik, denn andere Besprechungen der Korpusarbeit im allgemeinen fordern eben gerade eine schmale Materialgrundlage, weil eine repräsentative Erfassung der ganzen Breite ohnehin nicht möglich sei (vgl. etwa Schnörch 2002, S. 132 oder Bergenholtz / Pedersen 1994, S. 163)

Grundwortschatzes gleichzusetzen, indem man etwa die frequentesten Wörter eines Häufigkeitswörterbuches unmittelbar in einen Grundwortschatz übernahm:

„Häufigkeitszählungen sollten als Grundvoraussetzung aufgefasst werden, auf deren Basis man ganz allgemeine Einblicke in ein Sprachsystem gewinnen kann. Werden nun von außen falsche oder überhöhte Implikationen von einem Wissenschaftszweig (Sprachdidaktik) auf einen anderen (Lexikometrie) übertragen, und kommt es dabei fast zwangsläufig zu unbefriedigenden Ergebnissen, so ist dieser Umstand weniger der Lexikometrie anzulasten, sondern eher der Sprachdidaktik; (zu) lange Zeit hatte man versäumt, sich ein eigenes, den speziellen Belangen angepasstes Methodengerüst zu erstellen. Stattdessen griff man weitgehend unreflektiert auf Auswahlverfahren zurück, die ursprünglich zu anderen Zwecken entwickelt wurden. Die Ergebnisse ließen zu wünschen übrig. Und aus dieser Enttäuschung heraus ist es zwar verständlich, jedoch nur wenig konstruktiv, die Methode der Frequenzzählung in der Folge gänzlich abzulehnen.“ (Schnörch 2002, S. 25)

Frequenzerhebungen leisten also durchaus einen nützlichen und legitimen Beitrag zur Erarbeitung eines zentralen Wortschatzes, nur gilt es dabei eben, einige Aspekte der Eigenarten von Frequenzlisten allgemein und der spezifischen Liste, die man für seine Arbeit heranziehen will, zu beachten (vgl. Martin 1990, S. 1315f.):⁷⁰ Zu diesen Aspekten gehören Fragen nach dem Ausgangsmaterial der Frequenzerhebung, also wie sich das Ausgangskorpus zusammensetzt und wieviele der im Ausgangskorpus enthaltenen Einheiten im Frequenzwörterbuch tatsächlich auch wiedergegeben werden. Damit verknüpft ist die Frage nach der Größe des Korpus, die Art der Darstellung der Einträge, also etwa die Frage danach, welche Frequenzinformationen (absolute Frequenz, Distribution etc.) und welche semantischen und morphologischen Informationen das Frequenzwörterbuch zu jedem Eintrag enthält, und wie alle diese Informationen erhoben werden.

Vor dem Hintergrund der hier betrachteten grundlegenden Unterschiede zwischen zentralen Wortschätzen und Häufigkeitswörterbüchern und der damit verbundenen bisherigen terminologischen Vermischung der Begriffe *Häufigkeitswörterbuch* und *Grundwortschatz* wollen wir hier festhalten: Den Terminus *Grundwortschatz* verwenden wir nur für die Fälle, in denen es um die Erarbeitung eines didaktisch orientierten zentralen Wortschatzes für den Rechtschreib-, Fremdsprachen- oder Fachsprachenunterricht geht. Häufigkeitswörterbücher sollen nicht zu den zentralen Wortschätzen gerechnet werden, genausowenig wie wir die häufigsten Wörter eines Häufigkeitswörterbuches automatisch mit einem Grundwortschatz gleichsetzen, sondern einfach als die n häufigsten Wörter des in diesem Häufigkeitswörterbuch erfassten Korpus x bezeichnen.

3.4.2 Die *basic word list* der Glottochronologie

In Abschnitt 2.1 haben wir gesehen, dass sowohl die Grundwortschatzlexikographie als auch die Glottochronologie die Diskussion zum Konzept zentraler Wortschätze in prägender Weise beeinflusst haben. Vergleicht man jedoch die *basic word list* der Glottochronologie

70. Vgl. auch Rosengren 1968, S. 21 oder Schnörch 2002, S. 142. Ähnlich auch Ruoff 1981, S. 15; er liefert anschauliche Beispiele dafür, wie wichtig die Kenntnis der Gesamtzusammensetzung des Korpus für die Interpretation von Frequenzen ist.

logie mit den unter der Bezeichnung *Grundwortschatz* erarbeiteten didaktischen Wortschätzen, so sieht man: Beide arbeiten zwar mit reduzierten Wortschätzen und zentralen Einheiten des Wortschatzes, die Rolle der Reduzierung bzw. der Zentralität ist jedoch eine andere.

Während bei der Grundwortschatzlexikographie die Reduzierung des Wortschatzes und die Konzentration auf seine zentralen Einheiten den Zweck des Grundwortschatzes an sich darstellt, weil es gerade die zentralen Einheiten des Wortschatzes sind, die man zuerst vermitteln will, ist die Begrenztheit der *basic word list* der Glottochronologie nur ein Seiteneffekt: Zweck der *basic word list* ist es, eine Grundlage für den Vergleich zweier Sprachen zur Verfügung zu stellen. Dass man dabei mit einer begrenzten Anzahl an Begriffen arbeitet und versucht, möglichst zentrale Begriffe zu erfassen, hängt nicht damit zusammen, dass hier Zentralität etwa so wichtig wäre. Vielmehr spielt die Annahme eine Rolle, dass die Wahrscheinlichkeit, dass zwei Sprachen ein Wort gemeinsam haben, im Wortschatzzentrum höher ist als im Rest des Wortschatzes, weil man annimmt, dass zentrale Begriffe weniger sprachlichen Veränderungen ausgesetzt sind als nicht zentrale Begriffe. Ob es sich bei diesen zentralen Begriffen um wenige oder viele handelt, ist dabei grundsätzlich von untergeordneter Bedeutung. Dass die *basic word list* nur 200 Einträge enthält, ist auf die Annahme zurückzuführen, dass es nur wenige wirklich zentrale und damit stabile Begriffe in einer Sprache gibt.

Daraus folgt auch, dass die Selektion der Einheiten, die in den jeweiligen zentralen Wortschatz aufgenommen werden sollen, auf unterschiedlichen Kriterien beruht: Während man in der Grundwortschatzlexikographie versucht, den Begriff der Zentralität anhand unterschiedlichster linguistischer Kriterien, von denen Stabilität nur eines von vielen ist, operationalisierbar zu machen (vgl. Abschnitt 4.4), zählt in der Glottochronologie allein das Merkmal der Stabilität – die Einheiten, die in die *basic word list* aufgenommen werden, sollen so stabil wie möglich sein, so dass die zu vergleichenden Einheiten des Wortschatzes so nahe wie möglich am lexikalischen Ursprung der Sprache liegen. Weil Zentralität und quantitative Begrenzung in der *basic word list* der Glottochronologie nur eine Nebenrolle spielen, wollen wir sie nicht als einen Vertreter des Konzepts zentraler Wortschätze betrachten.

3.4.3 Rein quantitativ begrenzte Wortschätze der maschinellen Sprachverarbeitung

Neben den in 3.3 besprochenen Anwendungen zentraler Wortschätze gibt es in der maschinellen Sprachverarbeitung noch andere Typen von Teilwortschätzen. Im Unterschied zur Stopwortliste des *Information Retrieval* und dem Neutralwortschatz der automatischen Textklassifikation handelt es sich dabei aber um Wortschätze, die allein aus praktischen Gründen quantitativ begrenzt sind, in der Regel anhand des Kriteriums der Frequenz. Die Neutralität bzw. Zentralität der darin enthaltenen Wortschatzeinheiten spielt für ihre Auf-

nahme in den jeweiligen Wortschatz keine Rolle. Um diesen Unterschied deutlich zu machen, sollen diese Typen von Wortschätzen hier kurz vorgestellt werden.

3.4.3.1 Elektronische Wortlisten zur Unterstützung der automatischen Rechtschreibkorrektur in Textverarbeitungsprogrammen

Hierbei handelt es sich um eine Liste der gebräuchlichsten Wörter einer Sprache, die in Kombination mit einer Liste morphologischer Regeln zur automatischen Überprüfung der Rechtschreibung in elektronischen Texten herangezogen werden (vgl. Trost 2003, S. 37). In der einfachsten Form wird dabei der zu überprüfende Text mit den Kombinationen abgeglichen, die die Wortliste und die dazugehörigen morphologischen Regeln erlauben. Alle im Text vorkommenden Wortformen, zu denen sich keine Entsprechungen finden, werden als möglicherweise falsch markiert. Fortgeschrittenere Korrektursysteme berücksichtigen dabei auch die Umgebung des zu untersuchenden Wortes und entdecken komplexere Schreibfehler wie wiederholte Wörter. Semantische Aspekte der einzelnen Wörter scheinen dabei nicht weiter für das Verfahren genutzt zu werden. Über die Aufnahme eines Wortes in den hierfür verwendeten Wortschatz entscheidet allein seine Frequenz, die aufgrund oft nicht offengelegter Korpusdaten ermittelt wird. Der Grund für die quantitative Beschränkung dieses Wortschatzes ist dabei ein rein praktischer – weil es nicht möglich ist, alle möglichen Wortformen einer Sprache zu erfassen, wird ein willkürlich festgesetzter Ausschnitt herangezogen.

3.4.3.2 Eine elektronische Wortliste zur Unterstützung der Worteingabe auf kleinen elektronischen Geräten (*predictive texting*)

Beim *predictive texting* handelt es sich um ein Verfahren, mit dessen Hilfe man versucht, die Eingabe von Wörtern auf kleinen Bildschirmen und mit Hilfe kleiner Tastaturen, wie das etwa bei Mobiltelefonen der Fall ist, zu erleichtern:⁷¹

„Language prediction techniques strive to reduce the input burden by predicting (or rather disambiguating) what the user is entering. To do this, the system maps key presses against an embedded database of language statistics. The system analyses a [...] corpus [...] to establish the relative frequency of characters, bigrams (pairs of characters), trigrams, words, or phrases in the language of interest. These statistical properties are used to suggest or predict letters or words as text is entered. [...] The software compares the sequence of key presses with the most frequently occurring word with that particular sequence of key presses in the dictionary stored in the memory of the [respective device], and presents them as the most likely choices to the user for verification [...].“ (Hård af Segerstad 2003, S. 130)

Basis dieses Verfahrens ist also ein Frequenzwörterbuch, das aufgrund zumeist nicht offengelegter Korpusdaten erhoben wird.⁷² Der Grundgedanke des Verfahrens des *predictive*

71. Eine generelle Darstellung dieser Anwendung findet sich bei Hård af Segerstad 2003 oder auch beispielsweise bei Hasselgren / Montnemery / Svensson u. a. 2003. – Das Verfahren des *predictive texting* ist auch bekannt unter der (eigentlich kommerziell genutzten) Produktbezeichnung T9[®], die sich auf die 9 Zifferntasten, die die Tastatur eines Mobiltelefons typischerweise enthält, bezieht.

72. Für eine kritische Besprechung der Eignung dieser Korpusdaten für das Verfahren des *predictive texting* vor dem Hintergrund der Benutzungssituation vgl. Hård af Segerstad 2003, S. 133-135.

texting beruht dabei auf der Erkenntnis von der degressiven Struktur des Wortschatzes: Die in der dem *predictive texting* zugrundeliegende Wortliste ist so zusammengesetzt, dass sie in 95% der Fälle dem Benutzer das von ihm beabsichtigte Wort anzeigt (vgl. ebda., S. 130). Auch hier sind es nicht die individuellen Eigenschaften einzelner Wörter, die über ihre Aufnahme in das dem *predictive texting* zugrundeliegende Wörterbuch entscheiden, sondern allein ihre Frequenz, die wiederum durch eine relativ arbiträre Verwendung von Korpusdaten erhoben wird. Auch die quantitative Begrenzung des Wörterbuchs ist arbiträr bzw. praktisch bestimmt. Wäre es möglich, alle möglichen Wortformen einer Sprache zu erfassen und in effizienter Weise zu speichern, wäre eine quantitative Begrenzung gar nicht anzustreben.

Wenn man sich diese beiden Wortschatztypen nun rückblickend ansieht, dann kann man überspitzt formulieren: Bei diesen Wortschätzen handelt es sich um Wortsammlungen, die, hätte man die praktischen Möglichkeiten dazu, am besten den gesamten Wortschatz einer Sprache abdecken sollten; weil das aber aus linguistischen und speichertechnischen Gründen nicht möglich ist, begrenzt man sie quantitativ, indem man anhand des Frequenzkriteriums eine willkürliche Grenze setzt. In gewisser Weise stellen also diese beiden Wortschatztypen Varianten der in Abschnitt 3.4.1 besprochenen Häufigkeitswörterbücher dar, indem sie den Wortschatz einer Sprache allein hinsichtlich seiner Frequenz darstellen. Wir wollen sie deshalb nicht zu den zentralen Wortschätzen zählen.

3.5 Eine Typologie

Abschließend wollen wir, um den Bezug auf die einzelnen Typen zentraler Wortschätze zu vereinfachen und um eine Grundlage für eine klare Terminologie zu schaffen, die bisher betrachteten Beispiele für Anwendungen eines zentralen Wortschatzes gruppieren und in einer Typologie zusammenfassen:

Zum einen gibt es den *didaktischen Typus* eines zentralen Wortschatzes. Seine Adressaten sind Sprecher oder Lernende des Deutschen, ihr Verwendungszusammenhang ist die Sprachdidaktik bzw. die Sprachvermittlung. Zu ihnen gehören zunächst einmal die in Abschnitt 3.1 vorgestellten Grundwortschätze. Aber auch das kontrollierte definitonische Vokabular (vgl. Abschnitt 3.2) soll dazugerechnet werden. Zwar wird es nicht in einer direkten Unterrichtssituation verwendet, dient aber doch der Sprachvermittlung mittels eines Wörterbuchs.

Zum anderen gibt es zentrale Wortschätze, deren Adressaten nicht Sprecher, sondern Rechenprogramme sind. Sie dienen zudem nicht der Vermittlung von Sprachinformationen, sondern der Analyse von Sprachdaten. Diesen Typus zentraler Wortschätze, wie er in Abschnitt 3.3 beschrieben wird, wollen wir im weiteren als den *analytischen Typus* eines zentralen Wortschatzes bezeichnen.

Nicht zu den zentralen Wortschätzen gerechnet werden dagegen Häufigkeitswörterbücher (vgl. Abschnitt 3.4.1), die *basic word list* der Glottochronologie (vgl. Abschnitt 3.4.2) und Wortschätze für bestimmte Anwendungen der maschinellen Sprachverarbeitung, die allein auf einer quantitativen Begrenzung beruhen und bei denen die Zentralität der Wörter nur eine untergeordnete Rolle spielt (vgl. Abschnitt 3.4.3).

Kapitel 4: Probleme bei der Arbeit mit zentralen Wortschätzen – ein Überblick

Nachdem wir uns angesehen haben, wo überall das Konzept zentraler Wortschätze mehr oder minder explizit angewendet wird, soll es nun im Folgenden um einen Überblick über die Probleme gehen, die in der bisherigen Forschung zu den verschiedenen praktischen Anwendungen dieses Konzeptes aufgetaucht und diskutiert worden sind.

Mit diesem Überblick wollen wir zum einen der Frage nachgehen, warum es eigentlich so schwierig ist, einen zentralen Wortschatz zu erstellen, und weshalb dieses Konzept so kontrovers diskutiert wird. Zum anderen soll es auch darum gehen, zu zeigen, mit welchen konkreten Problemkreisen man sich bei der Arbeit mit einem zentralen Wortschatz auseinandersetzen muss. Und schließlich soll es bei diesem Überblick vor allem auch darum gehen, die Grundlage für die in der praktischen Untersuchung in Teil C zu bearbeitenden Fragestellungen zu erarbeiten, damit deutlich wird, mit welchen spezifischen Fragestellungen sich die praktische Untersuchung vorrangig beschäftigen soll, und welche Erkenntnisse es zu den jeweiligen Fragestellungen bereits gibt.

Basis ist dabei die Analyse der einschlägigen Literatur, wie wir sie in Abschnitt 2.2 mit der Besprechung von Kühn 1979, Krohn 1990, Bonan-Garrigues 1993 und Schnörch 2002 vorgenommen haben. In allen diesen Arbeiten kehren bestimmte Fragestellungen immer wieder; und eben um diese wiederkehrenden Themen soll es in diesem Überblick gehen.⁷³ Dabei orientiert sich dieser Überblick an den drei großen Teilbereichen der Arbeit mit zentralen Wortschätzen – Fragen der Selektion, Fragen der Darstellung und Fragen des Umfangs eines zentralen Wortschatzes – berücksichtigt aber auch Aspekte, die neben dieser Dreiteilung bestehen.

4.1 Mangelnde Einigkeit über den Begriffsapparat

Ein grundlegendes Problem der bisherigen Arbeit mit zentralen Wortschätzen ist die Tatsache, dass man bisher mit keinem übergreifenden oder auch nur innerhalb einer Arbeit klar und konsistent definierten Begriffsapparat gearbeitet hat.⁷⁴

Das ist nicht weiter verwunderlich, da es in der Forschung zunächst um die unmittelbare Lösung von Fragen der lexikographischen Praxis ging und grundsätzliche methodische Aspekte wie die Etablierung eines Begriffsapparates zurückgestellt werden mussten. Die implizite und variierende Verwendung von gleichlautenden, aber semantisch unterschiedlich belegten Termini bereitet jedoch Probleme, weil sie zu Missverständnissen führt. Dies

73. Anzumerken ist an dieser Stelle, dass alle hier vorgestellten Bereiche Teil größerer linguistischer Themen sind – eben weil Fragen der Erstellung eines zentralen Wortschatzes im Schnittpunkt verschiedener Teilbereiche liegen (vgl. auch Abschnitt 4.8). Die hier vorgenommene Darstellung der mit der Arbeit an zentralen Wortschätzen verknüpften Problembereiche soll daher nur ein Überblick sein. Für Details zu den jeweiligen Fragen wird jeweils auf die entsprechende, detailliertere Fachliteratur verwiesen.

74. Vgl. auch Krohn 1992, S. 8.

gilt vor allem für den Terminus *Grundwortschatz*. Er wird in der Literatur für die Bezeichnung verschiedener Dinge herangezogen.

Zum einen ist da seine Verwendung als “didaktisch motivierter Wortschatzausschnitt” (Krohn 1992, S. 8), die zum Teil noch weiter danach verfeinert wird, ob der jeweilige Grundwortschatz mit einem Erklärungsteil arbeitet und ob er den rezeptiven oder den produktiven Wortschatz darstellt.⁷⁵

Zum anderen wird der Begriff *Grundwortschatz* statt für spezifische Anwendungen auch als Oberbegriff für verschiedene Typen zentraler Wortschätze verwendet, wie das zum Beispiel Schnörch tut, wenn er den Begriff *Grundwortschatz* nicht nur auf den didaktisch ausgerichteten Anwendungstyp eines zentralen Wortschatzes bezieht, sondern ihn auch auf das definatorische Vokabular ausweitet.⁷⁶

Ähnlich vielseitig wird auch die englische Entsprechung des *core vocabulary* und des konkurrierenden Terminus’ *basic vocabulary* gehandhabt. Lehmann beispielsweise gibt dem Begriff *core vocabulary* drei Bedeutungen:

„There are several ways to define **core vocabulary**, I can think of the following three:

1. The core vocabulary consists of the *n* most frequent words of a language.
2. The core vocabulary is that vocabulary which is common to all native speakers of a language.
3. The **semantic** core vocabulary consists of those words which suffice to define all of the remaining vocabulary of a language.“ (Lehmann 1991, S. 303)

Bußmann (1996, S. 49) dagegen setzt den Terminus *core vocabulary* mit *basic vocabulary* gleich und beschränkt ihn auf den didaktischen Anwendungstypus eines zentralen Wortschatzes, parallel zu ihrer Definition des Terminus’ *Grundwortschatz* (vgl. Bußmann 1990, S. 297f.).

Auch die Verwendung des Terminus *basic vocabulary* für eine willkürlich angesetzte Menge hochfrequenter Wörter trägt zur terminologischen Verwirrung bei. Ein Beispiel hierfür ist die Arbeit von Carlson u. a. 1985, bei der es um die phonetische Untersuchung der 10000 frequentesten Wörter verschiedener Frequenzuntersuchungen von fünf europäischen Sprachen geht und bei der das *basic vocabulary* einfach aus diesen 10 000 Wörtern besteht.⁷⁷

Ähnlich missverständlich wirkt es, wenn unter dem Begriff *Grundwortschatz* Grundwortschätze und Häufigkeitwörterbücher gemeinsam betrachtet werden. Viele der im Streit um frequenzorientierte vs. kommunikativ-pragmatische Selektionskriterien angeführten Argumente, etwa die Lückenhaftigkeit der Ergebnislisten (vgl. Abschnitt 4.5.2.1) können nämlich durchaus darauf zurückgeführt werden, dass man sich bei der Kritik an frequenzorientierten Grundwortschätzen eigentlich auf die allgemein begrenzte Aussagekraft von Häufigkeitwörterbüchern bezog:

75. Kühn unterscheidet entsprechend „Grundwortschatzbücher“ und „Wortlisten“ (Kühn 1990, S. 1359) bzw. „passives Minimum“ und „aktives Minimum“ (Kühn 1979, S. 23).

76. Vgl. etwa Schnörch 2002, S. 25 und 39.

77. Weshalb die Grenze gerade bei 10 000 Wörtern gezogen wurde, ist dort nicht erklärt.

„Wir müssen der ebenso massiven wie pauschalen Kritik am Frequenzkriterium und einer damit einhergehenden Abwertung qualitativ-introspektiver Kriterien in einigen wichtigen Punkten begegnen, wobei die wörterbuchtypologische Unterscheidung von Frequenzwörterbüchern und didaktischen Wortlisten eine wichtige Rolle spielt. Denn Kühn [= Kühn 1979, V. H.] schert in seiner kritischen Auswertung KÄDING [= Kaeding 1898, V. H.] [...] und die didaktischen Wortlisten [...] alle über einen Kamm.

Zunächst erscheint es wenig sinnvoll, in einem Frequenzwörterbuch wie KÄDING das frequentielle Auseinanderklaffen von Lemmata, die gleichen Wortfeldern angehören, dem Frequenzwörterbuch als Mangel anzulasten, ohne zuvor die Prinzipien zu diskutieren, nach denen frequentielle Wortkorpora für didaktische Wortlisten nutzbar gemacht werden können. Das 'Fehlen' bestimmter Lexeme in einer bestimmten Häufigkeitszone ist für sich genommen noch kein Argument gegen Frequenzuntersuchungen und ihre Auswertung für die Erstellung eines Grundwortschatzes.“ (Krohn 1992, S. 61f.)⁷⁸

Die terminologischen Unklarheiten werden noch dadurch verstärkt, dass im Zusammenhang mit dem Begriff *Grundwortschatz* auch diskutiert wird, ob es so etwas wie einen absoluten Kern und damit **den** Grundwortschatz des Deutschen gibt. Obwohl diese Frage noch nicht geklärt ist, wird dieses angenommene lexikalische Zentrum gern automatisch mit dem anwendungsorientierten Typus eines Grundwortschatzes gleichgesetzt.⁷⁹

Wir haben es also mit einer Reihe von terminologischen Ungeklärtheiten, Überlappungen und divergierenden Verwendungsweisen zu tun. Solche terminologischen Ungeklärtheiten sind so lange unproblematisch, als man mehrere Termini für ein und denselben Sachverhalt verwendet, solange es sich also nur um formale Varianten desselben Begriffs handelt, wie das etwa Bußmann mit den beiden Termini *basic vocabulary* und *core vocabulary* tut. Was die Sache hier jedoch problematisch macht, ist, dass ein und derselbe Terminus mit verschiedenen Bedeutungen belegt wird. Das gilt, wie wir gesehen haben, vor allem für den Terminus *Grundwortschatz*. Hinzu kommt, dass in der Mehrzahl der einschlägigen Arbeiten eine explizite Definition der in der Arbeit verwendeten Begriffe fehlt, dass also nicht nur eine allgemein etablierte Terminologie fehlt, sondern dass auch innerhalb einer einzelnen Arbeit terminologische Unklarheiten bestehen.

Die terminologischen Ungeklärtheiten tragen also viel zu einer erschwerten Diskussion bei. In Abschnitt 6.4. wollen wir uns deshalb einigen terminologischen Klärungen widmen.

4.2 Die Frage nach der Grundeinheit

Die Frage, um welche Einheiten es sich überhaupt handelt, über deren Aufnahme in einen zentralen Wortschatz man entscheiden muss, berührt ein wesentliches lexikologisches Problem: Was ist die Grundeinheit des Wortschatzes, also derjenigen Menge, die wir nach zentralen und nicht zentralen Einheiten gruppieren wollen? Im vorliegenden Fall geht es dabei insbesondere um die Frage, was man zählt, wenn man die Selektion für einen zentralen Wortschatz beispielsweise auf Frequenzerhebungen aus einem Korpus aufbaut – wie zählt man die laufenden Wortformen des Korpus und wie führt man sie auf eine Grundform zu-

78. Ähnlich auch Schnörch 2002, S. 25.

79. Vgl. zum Beispiel Andreev 1980, S. 2f., Rosengren 1986, S. 158f. oder Schnörch 2002, S. 7. – Vgl. auch die entsprechende Diskussion in Abschnitt 5.5.

rück? Wie behandelt man Phraseologismen? Betrachtet man sie überhaupt als Teil des Wortschatzes, und wenn ja, zählt man ihre Einzelteile als Ganzes oder zählt man ihre einzelnen Glieder separat? Und wie geht man mit reihenbildenden Komposita und Ableitungen um? Die Antwort auf diese Fragen wird einen großen Einfluss auf das Selektionsverfahren und auf die Art der Darstellung haben, vor allem aber beeinflusst sie den Umfang des zentralen Wortschatzes. Denn je nach dem, was man als Grundeinheit ansetzt – ob man also etwa die Einzelbedeutungen polysemer Lexeme als separate Einträge ansetzt oder aber in einem einzigen Eintrag gruppiert –, kann der Umfang eines solchen zentralen Wortschatzes stark schwanken.

Es handelt sich dabei also vor allem um Probleme der Abgrenzung dessen, was zu einem Wortschatz gehört, gegenüber Einheiten, die zu anderen Teilsystemen der Sprache gehören. Es geht zum einen um die Abgrenzung des Wortschatzes gegenüber der Syntax (Behandlung von Phraseologismen) und gegenüber der Morphologie (Fragen der Lemmatisierung) und gegenüber der Wortbildungsgrammatik (Fragen der Lexikalisierung). Alle diese Fragen spielen eine zentrale Rolle in der Lexikologie im allgemeinen und sind nicht spezifische Probleme des hier besprochenen Konzeptes.⁸⁰ Wir wollen uns daher im Folgenden auf zwei Aspekte konzentrieren: Wie wurden diese Probleme bisher bei der Arbeit mit zentralen Wortschätzen behandelt und gelöst? Und wie wirken sich diese Probleme generell auf die verschiedenen Aspekte der Arbeit mit zentralen Wortschätzen aus?

4.2.1 Die Frage nach der Grundeinheit in der bisherigen Forschung

Die Frage danach, was eigentlich ein Wort ist, ist eine Diskussion, die immer kontrovers geführt worden ist (vgl. Glück (Hg.) 2000, S. 792f.). Dass die Definition des Begriffs *Wort* bzw. dessen, was die Grundeinheit des Lexikons sein soll, „je nach Bestimmung des Gegenstandsbereichs [...] und theoret[ischen] und/oder prakt[ischen] Zielsetzungen notwendig differier[t]“ (ebda., S. 793), zeigt ein kurzer Vergleich dessen, wie diese Frage in verschiedenen Arbeiten zu zentralen Wortschätzen diskutiert und konkret gelöst wird. In den hier besprochenen Referenzarbeiten wird eine der vier folgenden Alternativen gewählt: Grundeinheit ist entweder

- eine Zeichenkette zwischen zwei Begrenzungszeichen,
- ein mehr oder minder explizit definiertes Lexem,
- ein Semem, d.h. eine einzelne Lesart eines Lexems,
- oder ein Begriff bzw. ein Feld von Begriffen.⁸¹

80. Vgl. etwa die Diskussion der Frage, was ein Wort sei, bei Carter 1998, S. 3-33.

81. In der computerlinguistischen Literatur ist man noch genauer und unterscheidet zwischen der lexikalischen Einheit als Einheit des Lexikons im allgemeinen und dem Verfahren, mit dem man diese Einheit im Lexikon auffindet. Vgl. zum Beispiel Handke 1995. Er diskutiert ausführlich die Frage nach dem Auffinden einer Lexikoneinheit im Zusammenhang mit dem Begriff „lexical access“ (ebda., S. 23) und verwendet den Begriff *access unit*.

Dem ersten Prinzip folgt die Arbeit von Bonan-Garrigues. Diese Arbeit bewegt sich im Rahmen der Bedingungen und Notwendigkeiten der maschinellen Sprachverarbeitung und geht aufgrund des verwendeten Materials von einer strikt formalen Abgrenzung der zu betrachtenden und zu zählenden Einheiten aus: Das von Bonan-Garrigues als Ausgangsmaterial verwendete elektronische Lexikon DELAS definiert ein Wort und damit einen Lexikoneintrag als eine Kette von Zeichen zwischen zwei Begrenzungszeichen (vgl. Bonan-Garrigues 1993, S. 22). Dieses Vorgehen ist zwar in der Praxis vergleichsweise einfach durchzuführen, bringt aber vielfältige Probleme der Bedeutungsunterscheidung, wie Bonan-Garrigues deutlich macht (vgl. ebda., S. 58).

Dem zweiten Prinzip – einer mehr oder minder expliziten Hinwendung zum Lexem als Grundeinheit – folgt Krohn 1992, auch wenn er das nur implizit deutlich macht.⁸² Diese Entscheidung, bei der jeder Eintrag der von ihm untersuchten Korpora durchaus verschiedene Bedeutungen haben kann, erlaubt es Krohn auch, zu den jeweiligen Korpora ein thematisches Profil zu erstellen. Eine ausführliche Diskussion des Problems der Grundeinheit fehlt jedoch.

Schnörch hingegen widmet der Diskussion der kleinsten abzugrenzenden Einheit des Grundwortschatzes viel Raum. Er macht explizit deutlich, dass Überlegungen dazu, was man als Grundeinheit betrachtet, ein notwendiger Teil der Arbeit an einem zu erstellenden zentralen Wortschatz ist:

„Speziell im Hinblick auf die Reduzierung des Gesamtwortschatzes ist es ein vorrangiger Aspekt, die Zähleinheit zu definieren, ob man sie in diesem Kontext nun ‘Lexem’ oder auch allgemeiner ‘Wort’ nennt. Die Zähleinheit ist als operationalisierbare Größe zu begreifen, zumal die Obergrenzen für Minimalwortschätze stets mit einer mehr oder minder konkreten Zahl von ‘Wörtern’ (meist 2000) angegeben wird.“ (Schnörch 2002, S. 51)

Er diskutiert ausführlich Aspekte der Unterscheidung von Polysemen gegenüber Homonymen (S. 84ff. und 97ff.), die Behandlung von Phraseologismen (S. 109ff. und 126f.) und Fragen der Lemmatisierung (S. 136ff. und 216f.). Für seine eigenen Untersuchungen setzt Schnörch schließlich das Semem als Zähleinheit an (vgl. ebda., S. 187), wobei er Sememe implizit gleichsetzt mit Lesarten (vgl. ebda., S. 54) und damit an den Begriff der lexikalischen Einheit (*lexical unit*), wie er bei Cruse (1986, S. 23-83) definiert und ausführlich diskutiert wird.⁸³

Kühn 1979 schließlich geht von einem gemischten Ansatz aus: Für die Selektion für das von ihm projektierte Grundwortschatzbuch wählt er den Begriff bzw. ein Thema im Sinne eines Feldes von Begriffen als Grundeinheit aus. Damit unterscheidet sich Kühns Ansatz deutlich von den bis dahin entstandenen Grundwortschatzprojekten. Während man in der Mehrzahl aller Versuche, einen Grundwortschatz zu erarbeiten, von einer Liste von Wör-

82. Z.B. Krohn 1992, S. 172.

83. Die Parallele zwischen dem Semem bei Schnörch und der lexikalischen Einheit bei Cruse ist insbesondere in der Abgrenzung gegenüber dem Lexem zu sehen, das Cruse als „family of lexical units“ (Cruse 1986, S. 81) gegenüber der lexikalischen Einheit, die nur eine Bedeutung an sich bindet, abgrenzt.

tern oder Lexemen in Form eines Korpus oder eines Wörterbuches ausgeht, geht Kühn den umgekehrten Weg: Er wählt unter einer Menge von Themenbereichen, Kommunikationssituationen, Handlungsmustern und Kommunikationsmodi aus (vgl. ebda., S. 61ff.) und ordnet erst nach diesem Selektionsprozess die dazugehörigen Wortschatzeinheiten zu (vgl. ebda. S. 120 und 78f.). Grundeinheit der Selektion ist bei Kühn also der Begriff, dem im Verlauf der Arbeit an der Darstellung des Grundwortschatzes Lexeme zugeordnet werden.

Von allen diesen Vorgehensweisen ist in der Theorie sicher die Orientierung am Begriff als einer „außersprachliche[n]“ (Henne 1972, S. 128) Einheit die richtige, weil erst der Begriff einen ‘neutralen Dritten’ zwischen der Form und der Bedeutung eines Wortes darstellt:

„[Der Begriff ist ein] Aggregat kategorialer oder relationaler Merkmale, das die Gegenstände, Zustände, Prozesse (etc.), denen die Merkmale zukommen, zu einer [...] Klasse zusammenfaßt und das mit einem kommunizierbaren, i.d.R. verbalen Ausdruck verknüpft ist; wer eine Entität mit diesem Ausdruck benennt, >begreift< sie als Instanz der durch das Merkmalsaggregat (die [...] Intension) gekennzeichneten Klasse.“ (Glück (Hg.) 2000, S. 99)

Nun ließe sich argumentieren, dass ja auch ein Semem eine ‘formneutrale’ Repräsentation eines Begriffs und damit eine geeignete Grundeinheit darstellen könne. Allerdings sind Sememe im Sinne von Lesarten terminologisch häufig eng mit einem Lexem in Form eines einfachen Wortes assoziiert, während der Terminus *Begriff* deutlicher macht, dass auch komplexere Konstruktionen wie etwa Phraseologismen als Repräsentationen einer Bedeutung möglich sind. Eine Selektion für einen zentralen Wortschatz, die sich am Begriff als Grundeinheit ausrichtet, würde daher eine bewusstere Auswahl der in den Wortschatz aufzunehmenden Wortschatzeinheiten erlauben, als es eine unmittelbare Orientierung am Lexem bzw. Semem ermöglicht. Denn durch die Orientierung am Begriff als Grundeinheit wird es möglich, die Selektion in eine semantische Auswahl, also eine Selektion der Begriffe, die man repräsentiert sehen möchte, und eine lexikalische Auswahl, also die Auswahl der Lexeme bzw. Sememe, die die gewählten Begriffe lexikalisch repräsentieren sollen, aufzuteilen. Eine solche Trennung von Begriff und lexikalischer Repräsentation würde schließlich auch ein transparenteres Selektionsverfahren fördern.

Eben ein solches Verfahren stellt Kühns Grundwortschatzprojekt dar. Allerdings sind mit dieser Vorgehensweise zwei Schwierigkeiten verbunden: Zum einen ist die praktische Verwendung von Begriffen als Grundeinheit dadurch erschwert, dass eine operationalisierbare Vorstellung des Konzepts *Begriff* nach wie vor fehlt und eine systematische Zusammenstellung bzw. Hierarchisierung von Begriffen nicht möglich ist. Das macht es für den Lexikographen schwierig, sich einen Überblick über die Menge von Begriffen zu verschaffen, aus der er auswählen soll. So etwa ist nicht klar, welche Gruppen von Begriffen überhaupt anzusetzen sind und in welchem Verhältnis sie zueinander stehen. Außerdem lassen sich nicht alle Begriffe einer bestimmten Begriffsgruppe zuordnen. Auch eine Unterscheidung zwischen einem Begriff und den ihn repräsentierenden Wörtern ist oft nur schwer möglich, weil wir Wörter verwenden, um auf Begriffe zu referieren. Und manche Begriffe, vor allem Oberbegriffe, die eine Begriffsgruppe benennen sollen, sind nicht durch etablierte Lexeme

repräsentiert, sondern können nur durch Ad-hoc-Bildungen bezeichnet werden.⁸⁴ Zum anderen scheint eine Orientierung am Begriff als Grundeinheit zu einem großen Lexemebestand zu führen, wie das bei Kühn 1979 skizzierte Grundwortschatzprojekt und die Hochrechnungen, die De Smet (1981, S. 364) und Hendrickx (1987, S. 2) dazu angestellt haben, zeigen.

Betrachtet man nun alle vier Ansätze hinsichtlich ihrer Eignung als Grundeinheit für einen zentralen Wortschatz, dann wird schnell deutlich, dass die Orientierung am Semem, wie sie Schnörch vornimmt, unter den gegebenen Umständen die beste Lösung ist: Die Orientierung an der Zeichenkette, wie sie Bonan-Garrigues 1993 vornimmt bzw. wegen der Datenerhebung vornehmen muss, erweist sich als nicht optimal, weil sie alle Aspekte der Bedeutungsseite von Wörtern außer Acht läßt. Der von Krohn 1992 vertretene Ansatz vom Lexem als Grundeinheit bringt wiederum insofern Schwierigkeiten mit sich, weil damit Phänomene der Polysemie unberücksichtigt bleiben, da auf Lexemebene nicht zwischen Lesarten unterschieden wird. Zieht man schließlich noch die praktischen Schwierigkeiten in Betracht, die eine Orientierung am Begriff als Grundeinheit mit sich bringt, bleibt die Orientierung am Semem bzw. der lexikalischen Einheit, wie sie Schnörch 2002 vornimmt, als gute Kompromisslösung stehen.

4.2.2 Die Grundeinheit zentraler Wortschätze und die Frage nach der Abgrenzung gegenüber anderen sprachlichen Einheiten

Im Folgenden wollen wir uns nun den eingangs erwähnten Problemen der Abgrenzung zuwenden und untersuchen, wie diese Abgrenzungsentscheidungen in der bisherigen Literatur besprochen wurden. Wir wollen mit der Abgrenzung von Einheiten des Wortschatzes gegenüber Einheiten auf der Ebene der Syntax, d.h. der Behandlung von Phraseologismen beginnen.⁸⁵

Traditionell wurden Phraseologismen in der Grundwortschatzlexikographie und auch bei der Arbeit mit Häufigkeitserhebungen bisher von einer Aufnahme in den jeweiligen Bestand ausgeschlossen.⁸⁶ Dies ist so lange unproblematisch, so lange man eher die Wortform als Merkmal zur Abgrenzung der Grundeinheit des Wortschatzes heranzieht. Eine konsequente und umfassende Berücksichtigung der Bedeutungsseite der Wortschatzeinträge ist

84. Für eine kritische Diskussion der Verwendung von Begriffen als Grundeinheiten bei der Grundwortschatzarbeit vgl. auch Buchbinder 1973, S. 304 und Kühn 1979, S. 120-122. – Vgl. auch die Diskussion zur onomasiologischen Darstellung von Wortschätzen in Abschnitt 4.7.

85. Dabei soll mit Schindler 2002, S. 38f. der Terminus *Phraseologismus* hier als Sammelbezeichnung für alle Erscheinungen oberhalb der Wortebene und unterhalb der Satz(glied)ebene stehen, ohne dass dabei die Tatsache, dass es unterschiedliche Typen von Phraseologismen und unterschiedliche Grade der Festigkeit der Wortverbindungen gibt, geleugnet werden soll. Abgrenzungsfragen dieser Art sind bei Schindler 2002 oder auch bei Carter (1998, S. 68-72 und 158-162) ausführlich dargestellt.

86. Eine Ausnahme bildet Ruoff 1981. Er berücksichtigt Phraseologismen bei seinen Häufigkeitsuntersuchungen, zerlegt sie aber in ihre jeweiligen Bestandteile und zählt diese. Für eine kritische Besprechung der Behandlung von Phraseologismen bei Ruoff 1981 vgl. Schnörch 2002, S. 126f.

damit allerdings nicht ohne weiteres möglich, weil Bedeutungseinheiten nicht an Ein-Wort-Lexeme gebunden sind.

Betrachtet man nun die hier vorgestellten vier Referenzarbeiten, so sind sie in ihrem konkreten Umgang mit Phraseologismen zwar zum Teil von ihrem Ausgangsmaterial abhängig, aber es zeichnen sich doch Tendenzen ab:

Da Kühns gesamter Ansatz am Begriff orientiert ist, ist die Form und damit auch die Frage danach, ob der Begriff mehrgliedrig oder einfach realisiert ist, irrelevant. Allerdings enthält der Beispiel-Grundwortschatz, den er skizziert, nur einfache Substantive, so dass schwer zu erkennen ist, ob er auch Phraseologismen aufgenommen hätte.

Ähnlich verhält es sich bei Krohn: Er diskutiert zwar Phraseologismen bzw. Idiome im Zusammenhang mit der phraseologisch-idiomatischen Kompetenz als einer Teilfähigkeit der lexikalisch-pragmatischen Kompetenz, die nötig ist, um Phraseologismen zu verstehen und zu produzieren (vgl. Krohn 1992, S. 20). Die konkrete Handhabung von Phraseologismen im Grundwortschatz, also die Frage, ob und wie sie in einem Grundwortschatz berücksichtigt werden sollten, diskutiert Krohn dabei nicht. In seiner Kompaktliste sind keine Phraseologismen verzeichnet, was aber wohl am Ausgangsmaterial, den Grundwortschätzen und Häufigkeitslisten, die Krohn als Grundlage seiner Kompaktliste heranzieht, liegt; sie enthalten keine Phraseologismen.

Bei Bonan-Garrigues wiederum werden Fragen zur Behandlung von Phraseologismen nicht behandelt, weil ihr Ausgangsmaterial ohnehin nur aus Simplicia besteht. Allerdings diskutiert sie Fragen der Zusammengehörigkeit von mehrgliedrigen Einheiten im Rahmen der Behandlung von Komposita in der französischen Grundwortschatzlexikographie (vgl. Bonan-Garrigues 1993, S. 32.); durch die Linksköpfigkeit und Auseinanderschreibung französischer Komposita gibt es dort ähnliche Probleme der Abgrenzung wie im Deutschen bei den Phraseologismen.

Erst Schnörch widmet sich ausführlich der Diskussion und Behandlung von Phraseologismen. Er diskutiert das Problem im Zusammenhang mit der Besprechung von Pfeffer 1970 und Ruoff 1981 und kommt zu dem Schluss, dass Phraseologismen nicht aufgelöst werden dürfen, sondern als Einheit verzeichnet werden müssen.⁸⁷ Damit verhält er sich konsequent zu der von ihm selbst eingeführten Forderung, die Arbeit an einem zentralen Wortschatz auf Sememen aufzubauen.

Wenden wir uns nun dem zweiten Bereich der Abgrenzung zu, nämlich der Abgrenzung des Wortschatzes gegenüber der Morphologie, also der Abgrenzung von Wörtern (Lexemen) gegenüber Wortformen und Bestandteilen von Wörtern (gebundenen Lexemen). Hier geht es um Fragen der Lemmatisierung, also etwa um die Einordnung von Suppletivformen, die Zuordnung der richtigen Wortart, die Abgrenzung von Partikeln und Adverbien. Dabei

87. Vgl. Schnörch 2002, S. 109-114 und 126f.

kommen Fragen der Lemmatisierung dann zum Tragen, wenn der zu erstellende Wortschatz aus einem Korpus oder mehreren Korpora bzw. Wörterbüchern gewonnen werden soll.

Dementsprechend spielen Fragen der Lemmatisierung und Zuordnung von Wortformen verschiedener Quellen bei Kühn keine Rolle, da sein Grundwortschatzkonzept onomasio-logisch orientiert ist, d.h. von einer Reihe von Inhaltsbereichen ausgeht und nicht von einem gegebenen Korpus. Ähnliches gilt für Bonan-Garrigues, die in ihrer Arbeit von der gegebenen Liste von Simplizia des DELAS ausgeht und daher Zuordnungen von Wortformen zu Lemmata nicht mehr vornehmen muss.⁸⁸ Aber sowohl Krohn 1992 als auch Schnörch 2002 sehen sich mit diesem Problem konfrontiert, weil sie ihre Untersuchungsmenge aus einer Reihe von Grundwortschätzen bzw. Häufigkeitslisten erstellen und deshalb die jeweiligen Einträge in konsistenter Weise zu einer gemeinsamen Grundform vereinheitlichen müssen. Auch wenn Krohn der Diskussion der Frage der Lemmatisierung allgemein viel Raum widmet (vgl. Krohn 1992, Kapitel 7.4), greift er in der Praxis zu einer Vereinfachung: Krohn nimmt keine Bedeutungszuordnung der in seinem Ausgangsmaterial enthaltenen Einträge vor und orientiert die Zuordnung der Einträge unter demselben Lemma allein an deren Form. Dieses Vorgehen, so Krohn, sei nur bei Ortmann, einem der von ihm analysierten Grundwortschätze, problematisch, weil dessen Häufigkeitsliste auf dem Ausgangsmaterial von Kaeding beruhe, bei dem wiederum nicht zwischen Groß- und Kleinschreibung unterschieden wurde. Die Zuordnung musste daher intuitiv-willkürlich vorgenommen werden (vgl. Krohn 1992, S. 148f.).

Auch Schnörch diskutiert Fragen der Lemmatisierung ausführlich in Zusammenhang mit der Besprechung von Rosengren (1972, 1977) (vgl. Schnörch 2002, S. 136-142) und seiner eigenen Materialbearbeitung (vgl. ebda., S. 214-217). Er konkludiert damit, dass das Vorgehen von Rosengren, die die Zuordnung zu Flexionsreihen als grundlegendes Werkzeug der Bildung von Lemmata verwendet, zwar theoretisch wenig verankert, aber sehr praktikabel sei, auch wenn damit unflektierte Wortformen nicht unterschieden werden könnten (vgl. ebda., S. 140).⁸⁹ In seiner eigenen Analyse geht Schnörch vereinfachend vor, indem er sich auf Substantive, Verben und Adjektive konzentriert und alles andere in einer Kategorie *Sonstige* zusammenfasst.

Neben der Abgrenzung der Einheiten des Wortschatzes gegenüber Einheiten der syntaktischen und der morphologischen Ebene müssen die Einheiten des Wortschatzes auch gegenüber Einheiten der Grammatik im Sinne von Wortbildungsregeln abgegrenzt werden. Dabei geht es vor allem um die Frage, ob und in wie weit Wortbildungsprodukte, also Komposita

88. Damit ist der Schritt der Lemmatisierung allerdings nur auf die Ebene der Wortlistenstellung verschoben. Wie Bonan-Garrigues andeutet (1993, S. 58-61), wurden die Zuordnungen bei der Erstellung des DELAS offensichtlich nur auf der Ebene rein syntaktischer Unterscheidungskriterien, etwa Unterschieden im Valenzverhalten, vorgenommen.

89. Zu den Details des Vorgehens der Lemmatisierung bei Rosengren vgl. Rosengren 1968, S. 16 und Rosengren 1969, S. 103f.

und Ableitungen in den Wortschatz gehören, weil sie lexikalisiert sind, oder ob sie durchsichtige Bildungen darstellen, die durch das Aufführen ihrer Bestandteile und die Angabe von Wortbildungsregeln ersetzt werden sollen.⁹⁰ Die Beantwortung dieser Frage hat sehr praktische Folgen – je mehr Komposita und Ableitungen durch die Angabe ihrer Bestandteile und der zugehörigen Wortbildungsregeln erfasst werden, desto geringer ist der Umfang des so erarbeiteten Wortschatzes. Traditionell ist man daher bei der Grundwortschatzlexikographie auch davon ausgegangen, dass durchsichtige Wortbildungen von der Aufnahme in den Grundwortschatz ausgeschlossen werden.

Dabei stellt sich allerdings das Problem, dass die Unterscheidung, was eine durchsichtige Bildung und was lexikalisiert ist, nicht trivial ist. So zeigt sich etwa bei Kühn 1979, dass er zwar bei der Frage nach der Aufnahme von durchsichtigen Bildungen in den Grundwortschatz im theoretischen Teil seiner Arbeit einen klaren Standpunkt bezieht: Durchsichtige Formen seien nicht in den Grundwortschatz aufzunehmen, sondern in der Grammatik zu erklären (vgl. Kühn 1979, S. 56.) Im praktischen Teil seiner Arbeit wird jedoch klar, dass diese Abgrenzung nicht einfach ist, denn der Entwurf seines Grundwortschatzes enthält Komposita, die man durchaus als durchsichtig betrachten könnte.⁹¹

Das traditionelle Verfahren der Unterscheidung von durchsichtigen und lexikalisierten Wortbildungen, bei dem man versucht, die Wortbildung in freie Lexeme zu zerlegen, lehnt Schnörch ab. Er begründet seine Meinung damit, dass es auch Komposita gebe, die zwar in freie Lexeme zerlegt werden könnten, die in ihrer Gesamtbedeutung aber lexikalisiert seien, wie etwa *Fahrrad* oder *Bahnhof* (vgl. Schnörch 2002, S. 282).⁹² An Stelle des Kriteriums, ob sich ein Wortbildungsprodukt in freie Lexeme zerlegen ließe, schlägt Schnörch einen von Rettig entwickelten Ansatz der Unterscheidung vor, der davon ausgeht, dass ein Kompositum dann lexikalisiert ist, wenn seine Bedeutung eine „blockhafte Einheit“ (Rettig 1989, S. 644) darstellt, also nicht in Einzelkomponenten zerlegt werden kann. Damit operiert Rettig mit einer Eigenschaft, die komplementär zum Merkmal der Durchsichtigkeit steht:

90. Mit Lexikalisierung beziehen wir uns auf die Definition bei Glück ((Hg.) 2000, S. 409), der Lexikalisierung definiert als „Prozess und Ergebnis der Aufnahme eines sprachl[ichen] Ausdrucks ins Lexikon einer Spr[ache]. Durch L[exikalisierung] kann ein komplexer Ausdruck in formaler und/oder semant[ischer] Hinsicht seine Motivation verlieren.“

91. Vgl. etwa *Mähmaschine* und *Dreschmaschine* (Kühn 1979, S. 144, S. 163 und S. 165).

92. Ähnlich auch das Beispiel *Wohnzimmer* bei Krohn:
„Der lexikalische Ausschluß bestimmter Komposita [...] erweist sich [...] in kontrastiver Hinsicht als problematisch, nämlich immer dann, wenn aus der Sicht der Muttersprache des Deutschlerner ein Kompositum nicht vorhersehbar/vorhersagbar ist. Beispielsweise entspricht dem Kompositum *Wohnzimmer* im Schwedischen *vardagsrum*, das – ebenfalls ein Kompositum – aufgrund der Semantik des Bestimmungsglieds als ‘Alltagszimmer’ im Deutschen zu paraphrasieren wäre. Hier erscheint uns die muttersprachliche Perspektive insofern zu eng, als aus der Bekanntheit der Grundglieder nicht automatisch geschlossen werden darf, daß der Schwede beispielsweise ohne vorgängige Lernprozeß spontan in *Wohnzimmer* das semantische Äquivalent zu *vardagsrum* erblickt.“ (Krohn 1992, S. 156).

„Nicht nur, was als semantisch fest angesehen werden kann, sondern alle Abschnitte der Sprechkette, die einem Sprecher oder Hörer vertraut sind, die ihm für den Sprachäußerungs- oder Verstehensvorgang als ganzes, als sprachliches Stereotyp zur Verfügung stehen, ihm 'blockverfügbar' sind, sind Grundeinheiten des Lexikons und werden hier als Lexeme bezeichnet.“ (Rettig 1989, S. 642)

Einen anderen Weg geht Krohn (vgl. Krohn 1992, S. 155f.): Auch er lehnt die traditionelle Dichotomie von Motiviertheit vs. Nichtmotiviertheit einer Wortbildungskonstruktion zur Entscheidung über die Zugehörigkeit eines Kompositums oder einer Ableitung zum Grundwortschatz als zu grob ab. Denn gerade ad hoc gebildete Formen könnten auch eine Bezeichnungslücke schließen können. Stattdessen schlägt er vor, die Lexikalisiertheit eines Wortes anhand von Frequenzuntersuchungen festzustellen, ein Verfahren, das Schnörch wiederum ablehnt (vgl. Schnörch 2002, S. 284). Eine Einigkeit über den Umgang mit Komposita scheint also nicht zu bestehen.

Betrachten wir nun rückblickend die hier angesprochenen Fragenkomplexe, dann wird deutlich, dass die Art und Weise, wie man diese Fragen jeweils praktisch entscheidet, vor allem den Umfang der Menge an Wortschatzeinheiten beeinflusst, aus der man für einen zentralen Wortschatz auswählen muss. Entscheidet man sich etwa für das Semem als Grundeinheit, ist man mit einer deutlich größeren Menge an Wortschatzeinheiten konfrontiert, aus denen man auswählen muss, als wenn man das Lexem als Grundeinheit des Wortschatzes betrachtet. Ähnliches gilt umgekehrt für die Abgrenzung von Wortschatzeinheiten gegenüber der syntaktischen bzw. wortbildungsgrammatischen Ebene, also den Umgang mit Phraseologismen und Wortbildungen – je großzügiger man den Begriff dessen, was noch zum Wortschatz gehört, auslegt, um so größer wird die Anzahl der potentiellen Einträge für den zentralen Wortschatz.

Die Wahl der Grundeinheit legt also zunächst einmal fest, wie groß die Menge an Wortschatzeinheiten werden kann, aus der man auszuwählen hat, um einen zentralen Wortschatz zu erstellen. Die Abschnitten 4.4 und 4.5 werden die Details dieses Auswahlprozesses und der damit verbundenen Probleme beschreiben; zunächst wollen wir uns jedoch der Frage der Datengrundlage widmen, aus der im Rahmen der Auswahl einzelne Einheiten ausgewählt werden sollen.

4.3 Die Frage nach der adäquaten Datengrundlage für eine Auswahl

Bei der Fragen nach der adäquaten Datengrundlage für eine Auswahl geht es hauptsächlich darum, ob man als Grundlage der Auswahl den Bestand von Wörterbüchern oder den eines Textkorpus heranziehen soll bzw. welche Aspekte bei der jeweiligen Datengrundlage zu berücksichtigen sind.⁹³

Vom Standpunkt der Grundwortschatzlexikographie aus gesehen ist dabei die Frage nach der adäquaten Datengrundlage nicht immer Bestandteil der Diskussionen gewesen. Die anfänglichen Grundwortschatze waren fast ausnahmslos Übernahmen der bei Kaeding verzeichneten Häufigkeitslisten oder nahmen keinen expliziten Bezug auf ihnen

zugrundeliegende Wörterbücher oder Textkorpora. Erst mit Arbeiten wie der von Pfeffer 1970 fing man an, Fragen der Datengrundlage explizit zu diskutieren bzw. das Verfahren der Korpuserstellung in die Diskussion mit aufzunehmen. In der maschinellen Sprachverarbeitung und ihren Anwendungen hingegen haben Fragen der Datengrundlage im Rahmen der Verfahren der Korpuslinguistik früh eine Rolle gespielt und wurden von dort aus auch in andere Bereiche der Linguistik, etwa die Frequenzforschung in der Lexikologie und die korpusgestützte Kollokationsanalysen übernommen.

Bei der Arbeit mit zentralen Wortschätzen hat man grundsätzlich zwei Alternativen für die Wahl der Datengrundlage – den Bestand eines oder mehrerer Wörterbücher oder den Bestand eines Textkorpus.⁹⁴ Dabei bestehen zwischen beiden Alternativen deutliche Unterschiede im Grad der methodischen Adäquatheit auf der einen und im Grad der praktischen Zugänglichkeit und dem Reichtum an Informationen auf der anderen Seite. Der Bestand eines Wörterbuches bietet den Vorteil, dass er in der Regel leicht zugänglich ist und zudem bereits mit vielen Informationen zu den einzelnen Einträgen versehen ist. Ein Textkorpus dagegen kann zunächst nur Rohdaten über die in einer Gruppe von Texten vorhandenen Wortformen bieten; die Annotation mit weiteren Informationen ist mit einigem Aufwand verbunden. Während man also bei der Verwendung eines Wörterbuchs auf die Ergebnisse früherer lexikographischer Bearbeitungen zurückgreifen kann, muss man sich diese Informationen bei Textkorpora erst erarbeiten. Andererseits kommt die Verwendung von Daten, die direkt aus authentischen Texten erhoben worden sind, gerade bei der Arbeit mit zentralen Wortschätzen, die ja gerade auf die praktische Anwendungssituation zugeschnitten sind, der Qualität des Ergebnisses zugute. Denn während die Daten eines Textkorpus direkt die Eigenschaften der ihm zugrundeliegenden Texte widerspiegeln (im Kontext der theoretischen Linguistik also die Ebene der *parole* wiedergeben), stellen die Daten in einem Wörterbuch bereits eine Abstraktion von Sprachdaten (also die Ebene der *langue*) dar und führen durch die lexikographische Bearbeitung bereits ein Element der Subjektivität in das Verfahren ein.⁹⁵

Nun ist es allerdings nicht so, dass die Erstellung von Textkorpora völlig problemlos zu bewerkstelligen ist und unmittelbar zu eindeutigen und objektiven Ergebnissen führt. Im Laufe der Entwicklung der Verfahren der Korpuslinguistik hat man zwar einige grundlegende Richtlinien dafür erarbeitet, wie ein gutes Korpus erarbeitet werden und aufgebaut sein soll,

93. Mit *Textkorpus* ist hier ganz allgemein und zugleich vereinfachend eine größere oder kleinere Sammlung von Texten im Sinne geschriebener oder gesprochener Sprachäußerungen, die hinsichtlich ihrer sprachlichen Eigenheiten ausgewertet werden und in dieser ausgewerteten Form vorliegen. In der Korpuslinguistik hingegen unterscheidet man grundsätzlich vielfach zwischen verschiedenen Typen von Korpora, etwa danach, ob sie die einzelnen Texte in ihrer Gesamtlänge erfassen oder auf eine standardisierte Länge kürzen. Für eine Beschreibung verschiedener Typen von Korpora vgl. zum Beispiel Bergenholtz / Pedersen 1994, S. 162-165, Sinclair 1998, S. 121-137 oder Teubert 1998, S. 147-160.

94. Die hier angestellten Überlegungen gelten natürlich nicht für die Arbeit mit zentralen Wortschätzen allein, sondern auch für die Erstellung von Wörterbüchern im allgemeinen. Vgl. etwa auch Bergenholtz 1990.

95. Vgl. etwa Kennedy 1998, S. 7 – Vgl. auch Abschnitt 9.2 für ein praktisches Beispiel.

aber insbesondere die Frage nach der Repräsentativität eines Textkorpus und der Objektivität seiner Ergebnisse ist ein immer wieder stark diskutiertes Thema.⁹⁶ Dabei scheint sich ein Konsens darüber abzuzeichnen, dass sich für ein Textkorpus eine (statistische) Repräsentativität hinsichtlich des Wortschatzes einer Sprache allenfalls für sehr spezifische Typen von Texten, etwa fachsprachliche, erreichen lässt.⁹⁷ Dies hat seinen Grund darin, dass statistisch gesehen die Repräsentativität eines Ausschnitts aus einer Gesamtmenge, wie ihn ein Textkorpus darstellt, voraussetzt, dass man die Größe und Zusammensetzung der Gesamtmenge kennt, was jedoch beim Wortschatz mit seinem offenen Charakter nicht möglich ist.⁹⁸ Erschwerend kommt hinzu, dass Repräsentativität ein relativer Begriff ist, der abhängig ist davon, worauf man ihn bezieht, etwa auf die Sprache, einen durchschnittlichen Sprecher oder die Art der Kommunikationssituation (vgl. Lehmann 1991, S. 304).

Neben dem Problem der Repräsentativität ist auch die Frage nach der Objektivität der Ergebnisse stark umstritten. So macht etwa Bonan-Garrigues deutlich, dass auch die Erstellung und statistische Auswertung eines Korpus als einem vermeintlich objektiven Verfahren der Erhebung von Daten über Sprache ein Moment der Subjektivität enthält:

„Si, dans les méthodes reposant sur les statistiques, les relevés de mots des listes fréquentielles sont bien objectifs, on oublie souvent que la subjectivité n'est pas éliminée pour autant. Elle se trouve déplacée en amont de ce calcul, c'est-à-dire au moment du choix du corpus de textes.“ (Bonan-Garrigues 1992, S. 7)⁹⁹

Darüberhinaus ist aber vor allem nicht zu vergessen, dass die Ergebnisse einer Korpuserhebung nicht unmittelbar in einen zentralen Wortschatz übernommen werden können, wie etwa die Diskussion der Schwächen rein frequenzbasierter Wortschätze zeigt.¹⁰⁰ Und schließlich ist man bei der Arbeit mit Korpora oft mit dem ganz praktischen Dilemma konfrontiert, dass es auf der einen Seite Korpora gibt, deren Ergebnisse zwar methodisch nicht ganz einwandfrei sind, die aber öffentlich zur Verfügung stehen, während methodisch 'saubere' Korpora oft nicht öffentlich zugänglich sind.

Welche Schlussfolgerungen kann man aus diesem Überblick für die Arbeit mit zentralen Wortschätzen ziehen? Zum einen wird deutlich, dass die Tatsache, dass es sich bei zentralen Wortschätzen um lexikographische Produkte handelt, bei deren Erstellung vielfältige praktische Aspekte berücksichtigt werden müssen, einigen Einfluss auf die Entscheidung für oder gegen ein Wörterbuch oder ein Textkorpus als Datengrundlage hat. Während ein Wörterbuch als Datengrundlage viele praktische Vorteile bietet, sind die durch ein Textkorpus

96. Für eine Diskussion der Anforderungen für die Erarbeitung eines Korpus vgl. etwa Sinclair 1998, S. 112.

97. Vgl. Bergenholtz / Pedersen 1994, S. 163 und 166. Ähnlich auch Erk 1972, S. 26f, Rosengren 1972, S. XII und Schnörch 2002, S. 132. – Zur Frage danach, ob fachsprachliche Korpora von Sprechern, die mit dem entsprechenden Fach nicht vertraut sind, erstellt werden können, vgl. Bergenholtz / Pedersen 1994, S. 167f.

98. Vgl. auch die Diskussion zur theoretischen Verankerung von Frequenzen als Selektionskriterium in Abschnitt 4.5.2.1.

99. Ähnlich auch Kühn 1979, S. 44.

100. Vgl. Abschnitt 2.1.1 und 4.5.2.1 und zum Beispiel Erk 1972, S. 27

erhobenen Daten potentiell näher am unmittelbaren, lexikographisch noch nicht bearbeiteten oder gefilterten Sprachgebrauch, allerdings auch schwerer zugänglich. Für beide Arten der Datengrundlage gilt aber, dass eine genaue Kenntnis ihrer Entstehung und Aufbereitung bzw. Darstellung wichtig ist, weil sie für ein transparentes und bewusstes Vorgehen und vor allem für die Deutung von Ergebnissen unerlässlich ist.¹⁰¹

4.4 Probleme der Selektion

Was in einen zentralen Wortschatz aufgenommen werden soll, wie man also unter den Elementen des Wortschatzes auswählt, ist eine für die Arbeit mit zentralen Wortschätzen entscheidende Frage. Entsprechend viel Raum nimmt sie auch in der bisherigen Literatur ein. Dabei dreht sich die Diskussion vornehmlich um die Wahl der adäquaten Selektionskriterien, aber auch andere grundsätzliche Fragen sind Gegenstand der Diskussion, etwa die nach dem Umgang mit besonderen Teilbeständen des Wortschatzes wie etwa Fremdwörtern und Neologismen und danach, welche Rolle bestimmte Strukturen innerhalb des Wortschatzes, wie etwa Wortfelder, spielen. Nicht zuletzt sind auch insbesondere mit der Arbeit von Schnörch 2002 Fragen der Unterscheidung und Abgrenzung von Bedeutungen ins Zentrum der Diskussion gerückt.

Wir wollen uns zunächst allgemeinen, für das Verfahren der Selektion relevanten Eigenschaften des Wortschatzes und seiner Einheiten widmen, bevor wir dann in einem weiteren Schritt einige konkrete Selektionskriterien betrachten.

4.4.1 Polyseme und Homonyme: Fragen der Bedeutungsabgrenzung

Ein wesentlicher Fragenkomplex, der bei Überlegungen zur adäquaten Selektion von Einheiten des Wortschatzes eine Rolle spielt, ist der nach dem Umgang mit mehrdeutigen Lexemen. Es geht dabei um die Frage danach, wo man die Grenze zwischen verschiedenen Teilbedeutungen eines Lexems zieht, welche der Teilbedeutungen eventuell in einen zentralen Wortschatz übernommen werden und ob sich etwa eine Grundbedeutung ausmachen lässt. Dazu gehört auch die Frage danach, ob es sich bei dem mehrdeutigen Lexem um ein Polysem im Sinne eines Lexems mit einem „Gefüge ‘zusammenhängender’ Teilbedeutungen [...]“ (Glück (Hg.) 2000, S. 537) handelt oder um zwei homonyme Formen, die „bei Nichtübereinstimmung ihres Inhalts ausdrucksseitig [...] identi[isch] sind“ (ebda., S. 280).

Solche und ähnliche Aspekte der Semantik spielten lange Zeit in der Grundwortschatzlexikographie bei der Auswahl von Einträgen gegenüber operationalen Fragen der Selektion eine nur untergeordnete Rolle. Der Grund dafür ist vermutlich im Ursprung der Grundwortschatzlexikographie zu suchen: die ersten Grundwortschätze entstanden aus den nicht-se-

101. Ein treffendes Beispiel dafür, wieviel die genaue Kenntnis der Datengrundlage und ihrer Entstehung zur Interpretation von Ergebnissen beitragen kann, gibt Schnörch (2002, S. 299). Er macht deutlich, dass das gehäufte Vorkommen der Diminutivendung *-lein* bei der Häufigkeitserhebung von Ruoff 1981 nur erklärbar ist, wenn man weiß, wie und wo die Daten bei Ruoff erhoben wurden (nämlich durch Tonaufnahmen mit Sprechern aus dem südwestdeutschen Sprachraum).

mantisierten Wortlisten von Kaeding. Erst in neueren Arbeiten, beginnend mit Kühn 1979, und vor allem konsequent umgesetzt von Schnörch 2002, treten Fragen der Semantik deutlich in den Vordergrund. Die Gründe für diese Hinwendung zur Semantik liegen in den Schwächen der frequenzorientierten Grundwortschatzlexikographie, die kontextlose Wortlisten hervorbrachte, in dem mangelnden Nutzen dieser „nackten Wortlisten“ (Kühn 1990, S. 1354) im Sprachunterricht, und in einer generellen Neuorientierung, bei der man Lexikon und Grammatik enger miteinander integriert und die sich auch im Konzept der Lexikogrammatik (vgl. Abschnitt 2.2.3) bemerkbar macht.¹⁰² An dieser Stelle wollen wir uns in einem kurzen Exkurs ansehen, weshalb die Berücksichtigung semantischer Informationen bei der Arbeit mit zentralen Wortschätzen so wichtig ist.

Zunächst einmal ist von einem theoretischen Standpunkt aus klar, dass man die Bedeutung von Wortschatzelementen bei der Auswahl und Darstellung eines zentralen Wortschatzes berücksichtigen muss, wenn man die zweigeteilte Struktur sprachlicher Zeichen, die sich aus einer Form und einer Bedeutung zusammensetzt, ernst nehmen will. Auch und insbesondere von einem sprachdidaktischen Standpunkt aus, also vom Standpunkt des Verfassers eines Grundwortschatzes im engeren Sinne, ist es nötig, Semantik zu berücksichtigen, denn die Einheit, die sich aus einer Form und einer Bedeutung zusammensetzt, ist es gerade, die man im Sprachunterricht vermitteln will. Und schließlich sind Aspekte der Semantik unentbehrlich, wenn es darum geht, Befunde aus einem Korpus laufender Wortformen bzw. Befunde aus vielfachen Quellen auf eine Einheit zusammenzuführen.

Im Folgenden wollen wir uns ansehen, wie Aspekte der Bedeutungsunterscheidung und des Umgangs mit Polysemen bzw. Homonymen in den vier Referenzarbeiten diskutiert werden. Im Vordergrund steht dabei, wie wir sehen werden, die Abgrenzung von Polysemen gegenüber Homonymen

Kühn 1979 geht mit dem von ihm projizierten Grundwortschatzbuch vom Begriff als Grundeinheit aus (vgl. Abschnitt 4.2) und kann damit grundsätzlich die praktischen Schwierigkeiten der Auflösung von Polysemen und ihrer Abgrenzung gegenüber Homonymen umgehen. Dennoch widmet Kühn der theoretischen Unterscheidung von Polysemen und Homonymen viel Raum (vgl. Kühn 1979, S. 89-93). Er trennt Polyseme von Homonymen strikt danach, ob sie ein gemeinsames Bedeutungsmerkmal haben – dann sind sie Polyseme – oder nicht – dann sind sie Homonyme. In dem von ihm skizzierten Grundwortschatz löst er Polysemien bzw. Homonymien durch spezifische Bedeutungsangaben und Merkmalsanalysen auf.

Krohn hingegen diskutiert die Abgrenzung von Homonymie und Polysemie nicht. Er gibt nur an, er müsse die polyseme Auffächerung bei dreien seiner Wortschatzsammlungen¹⁰³ rückgängig machen, weil die entsprechenden semantischen Informationen und Auffäche-

102. Für eine detaillierte Darstellung der Rolle der Semantik in der Geschichte der Grundwortschatzlexikographie vgl. Haderlein 2005.

103. Erk (1972, 1975, 1982), Baldegger (Hg.) 1980 und ZERTIFIKAT DEUTSCH 1977

rungen in den anderen von ihm untersuchten Korpora fehlten (vgl. Krohn 1992, S. 149f.). Damit geht er praktisch den umgekehrten Weg zu dem von Schnörch: Während Schnörch den nicht-semantisierten Einträgen seiner Ausgangskorpora Semantik zuordnet, um Konsistenz über seine Korpora hinweg herzustellen, entfernt Krohn die semantischen Angaben.

Bonan-Garrigues wiederum diskutiert zwar nicht explizit die Problematik der Unterscheidung von Homonymie vs. Polysemie, diskutiert aber sehr wohl die Frage nach der Unterscheidung der Bedeutungen von Homographen bzw. Homonymen (vgl. Bonan-Garrigues 1993, S. 58-61). In ihrem praktischen Verfahren allerdings unterscheidet Bonan-Garrigues bei der Beurteilung von Repräsentation und Plausibilität bewusst nicht zwischen Bedeutungen, sondern kalkuliert vielmehr die Vagheit von Bedeutungen als einen tragenden Faktor bei der Betrachtung der Repräsentiertheit bewusst mit ein.¹⁰⁴

Schnörch schließlich diskutiert das Problem ausführlich im Zusammenhang mit der kritischen Besprechung von Pfeffer (vgl. Schnörch 2002, S. 84-94 und 97-109). Den traditionellen Kriterien der Abgrenzung – eine gemeinsame Etymologie der zu vergleichenden Wortbedeutungen, semantische Nähe und ein ähnliches morphosyntaktisches Verhalten – steht Schnörch dabei skeptisch gegenüber:

„Keines der drei Kriterien führt für sich genommen [...] zu wirklich befriedigenden Ergebnissen. Dies hat zur Folge, dass sie in der Wörterbuchpraxis zumeist in nicht genauer spezifizierter Weise vermischt werden. Dabei wird allerdings stillschweigend darüber hinweggesehen, dass der Versuch, ein semantisches Problem zu lösen, bei (a) auf eine sprachhistorische Ebene verlagert wird und bei (c) auf eine grammatische.“ (Schnörch 2002, S. 89)

So kommt Schnörch zu dem Schluss, dass die Unterscheidung von Polysemie und Homonymie und die Auflösung eines polysemen Lexems in seine Einzelteile ein relativ zu lösendes Problem sei, d.h. in Abhängigkeit vom Zweck des zu erstellenden zentralen Wortschatzes (vgl. Schnörch 2002, S. 102f.). In seiner praktischen Analyse von sieben Grundwortschatz- bzw. Häufigkeitslisten arbeitet Schnörch direkt auf Semem-Ebene, indem er die Einträge von Kosaras u. Pfeffer als Grundlage verwendet. Der durch sie vorgegebene Bedeutungsfächer je Wort wird auf die Frequenzlisten von Ruoff 1981 und Rosengren 1972 und 1977 übertragen, so dass Ruoff und Rosengren im Nachhinein monosemiert werden können. Dabei ist Schnörch sich der methodischen Schwierigkeiten, die die Annahme, die semantische Auffächerung in einer Wortschatzdarstellung ließe sich ohne weiteres auf andere Wortschatzdarstellungen übertragen, liegen, durchaus bewusst.¹⁰⁵

104. Allerdings werden zu den angegebenen Wörtern durchaus auch grammatische Angaben mitgeliefert (vgl. Bonan-Garrigues 1993, S. 61). Bonan-Garrigues weist dabei auch auf die "retroaktive" Wirkung von grammatischem Zusatzwissen hin (*jaune* (Adj.) – *jaune* (Subst.)) und wertet dies als Indiz für die Existenz einer Lexikongrammatik. Und sie diskutiert auch Ambiguität (vgl. Bonan-Garrigues 1993, S. 57) und verweist auf andere Arbeiten, die mit Hilfen von lokalen Grammatiken (= Grammatiken von Kollokationspartnern und Phraseologismen) Ambiguitäten auflösen helfen sollen.

105. Vgl. etwa auch die entsprechende Kritik bei Schulz 2004, S. 333.

Aber:

„Entscheidend ist, dass die bislang stillschweigend vorausgesetzte Bedeutungsidentität von Wortkörpern durch die Angabe von in zwei Korpora vorgegebenen Sememen transparent gemacht wird: 'Das Gespenst der Mehrdeutigkeit wird gewissermaßen materialisiert'.“ (Schnörch 2002, S. 229)

An dieser Stelle wollen wir einen kurzen Blick auf eine alternative Betrachtung des Polysemie-Homonymie-Problems werfen, die auf einen Ansatz von Katz 1967 und Lehrer 1974 zurückgeht.¹⁰⁶ Sie nehmen an, dass die Unterscheidung von Polysemie und Homonymie weniger eine Frage der Dichotomie als vielmehr der einer graduellen semantischen Ähnlichkeit sei. Lehrer 1974 unterzieht diese Hypothese einem Experiment durch eine Sprecherbefragung, bei der Testpersonen in wiederholten Testrunden gebeten werden, Bedeutungen, die eine gemeinsame Wortform haben, auf einer Skala der semantischen Ähnlichkeit anzuordnen. Dabei zeigt sich, dass die Bestimmung der semantischen Ähnlichkeit bei solchen Bedeutungspaaren am konsistentesten ist, bei denen die Bedeutungen entweder sehr ähnlich oder sehr verschieden sind, dass also Sprecher intuitiv sehr wohl eine Unterscheidung zwischen polysemen und homonymen Bedeutungsbeziehungen (vgl. Lehrer 1974, S. 36). Daraus zieht Lehrer den Schluss: „A dictionary should distinguish between homonymy and polysemy. Speakers' judgements might be used in drawing this distinction in lexicology [...].“ (Lehrer 1974, S. 38). Auch wenn also die Abgrenzung von Polysemen gegenüber Homonymen bisher nur unbefriedigend gelöst werden kann, kann es nützlich sein, zu wissen, dass Sprecher sehr wohl zwischen semantischer Unabhängigkeit und semantischer Verwandtschaft bei Gleichheit der Form unterscheiden, dass aber Homonymie und Polysemie nur zwei Pole einer Skala der Bedeutungsähnlichkeit darstellen.

Die Abgrenzung polysemer Lexeme gegenüber homonymen Lexemen löst aber zunächst nur ein Problem, nämlich die Frage, ob zwei Bedeutungen gleicher Form als getrennte Lexeme betrachtet werden sollen oder als Teilbedeutungen ein und desselben Lexems. Im Falle eines Lexems mit mehreren Teilbedeutungen wird es dann in einem weiteren Schritt nötig, zu entscheiden, wie man zwischen den einzelnen Teilbedeutungen unterscheidet und – wenn man sich für das Semem, also die einzelne Lesart bzw. Teilbedeutung eines Lexems als Grundeinheit entschieden hat – welche dieser Teilbedeutungen in den zentralen Wortschatz aufgenommen wird (vgl. Schnörch 2002, S. 86). Dabei liegt es aufgrund der Forderung nach Zentralität der Wortschatzeinheiten nahe, nach einer Bedeutung zu suchen, die als eine Art Grundbedeutung die Ableitung aller anderen Teilbedeutungen zulässt. Ob es jedoch so etwas wie eine zentrale Grundbedeutung gibt, ist in der Literatur umstritten bzw. wird als nicht möglich erachtet.¹⁰⁷

In den Bereich der Abgrenzung von Bedeutungen fällt auch die ganze Thematik der Erfassung metaphorischer Bedeutungen und ihrer Berücksichtigung in einem zentralen Wort-

106. Vgl. auch den Hinweis bei Carter 1998, S. 35.

107. Vgl. etwa Werner 1989, S. 922 und Schnörch 2002, S. 97.

schatz. Intuitiv möchte man denken, dass der metaphorische, also übertragene Gebrauch von Wörtern aufgrund seines Charakters als einer von der Grundbedeutung abgeleiteten Bedeutung nicht zu deren zentraler Verwendung gehört, dass also etwa Sememe, die einen metaphorischen Gebrauch eines Lexems darstellen, nicht in den zentralen Wortschatz aufzunehmen sind. Dem gegenüber steht die Idealforderung, dass ein solcher Wortschatz zentrale Begriffe und nicht zentrale Wörter enthalten soll (vgl. Abschnitt 4.2), womit die sprachliche Form, in der ein Begriff zum Ausdruck kommt, in den Hintergrund tritt und es keinen Grund mehr gibt, metaphorische Formen von der Selektion auszuschließen. Hinzu kommt die relativ junge Auffassung, viele Bereiche des Sprachgebrauchs seien als metaphorisch aufzufassen, so dass eine strenge Regel, die jeden metaphorischen Gebrauch von der Aufnahme in einen zentralen Wortschatz ausschließt, wenig zweckmäßig wäre.¹⁰⁸

Die Frage nach der Abgrenzung von Bedeutungen, d. h. letztlich das Problem der Bildung von Sememen, ist also ein Problem nicht nur der Arbeit an zentralen Wortschätzen, sondern auch und vor allem der allgemeinen Lexikologie und Lexikographie, das auch dort noch nicht befriedigend gelöst ist. Für die Arbeit mit zentralen Wortschätzen kann man mit Schnörch (2002, S. 102f.) nur auf ein möglichst transparentes, für den jeweiligen Anwendungsfall spezifisches Vorgehen setzen.

4.4.2 Der Umgang mit Fremdwörtern

Ausgangspunkt der Diskussionen zum Umgang mit Fremdwörtern ist die Tatsache, dass man Fremdwörter intuitiv nicht zum Normalbestand des deutschen Wortschatzes rechnet.¹⁰⁹ Das zeigt sich zum Beispiel an der reichen Auswahl an Fremdwörterbüchern und der langen Tradition einer Fremdwörterbuchlexikographie, die es für das Deutsche gibt. Bei der Arbeit mit zentralen Wortschätzen stellt sich aufgrund dieser intuitiven Einschätzung daher schnell die Frage, ob man etwa Wörter wie *Outsider* oder *Outback* in den zentralen Wortschatz aufnehmen soll oder nicht.

Gegen einen pauschalen Ausschluss sprechen jedoch mehrere Argumente. Zum einen ist die Abgrenzung von Fremdwort vs. Erbwort nicht trivial (vgl. etwa Glück (Hg.) 2000, S. 220), so dass manchmal sogar die Existenz bzw. praktische Bedeutung von Fremdwörtern angezweifelt wird (vgl. Augst 1988, S. 4). Zum anderen spielt die Frage nach der Abgrenzung von Fremdwort vs. Erbwort vor allem auf der formalen Seite eine Rolle, denn das Merkmal 'Fremdwort' ist vor allem ein Attribut der Formseite und sagt nichts über die Zentralität der dadurch repräsentierten Bedeutung aus (vgl. Augst 1988, S. 4). Allenfalls ließe

108. Vgl. etwa Glück (Hg.) 2000, S. 437 oder Lakoff / Johnson 2003.

109. Fremdwörter und fremde Wörter sind dabei nicht miteinander zu verwechseln: „Fremde Wörter [...] und Fremdwörter sind zwei unterschiedliche Teilmengen aus dem Gesamtwortschatz aufgrund unterschiedlicher Kriterien. Sie haben unterschiedliche komplementäre Begriffe: bekannte Wörter – fremde Wörter und deutsche Wörter – Fremdwörter. Der Gegensatz fremdes Wort vs. bekanntes Wort bezieht sich auf das signifié, die Inhaltsseite, der Gegensatz Fremdwort vs. deutsches Wort auf das signifiant, die Ausdrucksseite.“ (Augst 1988, S. 4).

sich argumentieren, aus der fremdwortartigen Form eines Wortes könne man schließen, dass die gesamte lexikalische Einheit noch ziemlich jung ist, dass also ein Fremdwort eine Art Neologismus darstellt. Das stimmt aber nicht unbedingt, wie man in Bußmann (1990, S. 253) nachlesen kann.

Allerdings spielt der Status eines Lexems als Fremdwort dann eine Rolle, wenn es zu ein und derselben Bedeutung ein Fremdwort und ein Erbwort gibt, die miteinander in Konkurrenz stehen. In einem solchen Fall muss man von Fall zu Fall, eventuell anhand von Frequenzen, entscheiden, durch welche der beiden konkurrierenden Formen die Bedeutung im zentralen Wortschatz vertreten sein soll.¹¹⁰ Darüber hinaus kann eine als fremdwortartig aufgefasste Form ein Hinweis auf die stilistische Markiertheit des jeweiligen Lexems sein, eine Information, die sich für das Selektionsverfahren hinsichtlich des Kriteriums der stilistischen Neutralität (vgl. Abschnitt 4.5.2.5) ausnutzen ließe.

In der bisherigen Forschungsliteratur ist das Problem unterschiedlich ausführlich behandelt worden. Während Kühn 1979 und Krohn 1992 diesem Fragenkomplex kaum Aufmerksamkeit widmen, diskutiert Schnörch das Problem ausführlich. Er untersucht Fremdwörter und Internationalismen in der von ihm erstellten Schnittmenge und diskutiert die Problematik der Auswahl und Abgrenzung von Fremdwörtern ganz allgemein aus der Notwendigkeit heraus, ihren gegenwartssprachlichen Gebrauchswert und Status zu beurteilen (Schnörch 2002, S. 311-331). Dabei kommt er zu dem Schluss, dass Fremdwörter aus dem Wortschatz allgemein und auch aus dem Grundwortschatz nicht mehr wegzudenken sind (ebda., S. 321f.). Die Frage von Fremdwort oder Nicht-Fremdwort sei ein diachroner Aspekt, der bei der Grundwortschatz-Arbeit nicht relevant sei (ebda., S. 323), zumindest für die Selektion nicht; außerdem könnten sogenannte Fremdwörter allein schon wegen ihrer Anzahl nicht von vornherein für die Erarbeitung eines zentralen Wortschatzes ignoriert werden (ebda., S.335). Darüber hinaus könne gerade bei der Darstellung im didaktischen Grundwortschatz die Verwendung von Internationalismen von Vorteil sein (ebda., S. 330f.). Ähnlich argumentieren Pfeffer (1973, S. 425) und Augst (1988, S. 4), die beide verlangen, sowohl im Gesamtwortschatz als auch im Grundwortschatz Fremdwörter den Erbwörtern gleichzustellen.¹¹¹

Insgesamt wird hier deutlich, dass die Diskussion zur Frage nach der Aufnahme eines Fremdwortes in einen zentralen Wortschatz ein Beispiel dafür ist, dass viele Aspekte der

110. Das gilt im übrigen nicht nur für Fremdwort-Erbwort-Paare, sondern für alle Synonyme. Vgl. dazu auch Abschnitt 4.4.4.

111. Vgl. folgendes Zitat:

„Fremdwörter lassen sich also nicht aus der synchronen Sprache verbannen [...]. Für das gesunde Verhältnis zwischen Erb- und Lehnwort oder Fremdwort sorgt der Sprachgebrauch selbst [...]. Das Wörterbuch muß feststellen und angeben [...], von wem diese Wörter benutzt werden, gegenüber welchen anderen Sprachteilhabern, in welchen Sprech- und Schreibsituationen, mit welchem Sachbezug, in welchem Kontext, mit welcher Stilfärbung und vor allem mit welcher Bedeutung im Verhältnis zu den Bedeutungen der anderen Wörter des Wortfeldes, in dem die Sprachgemeinschaft dem Wort seinen Platz zugewiesen hat.“ (Pfeffer 1973, S. 425)

Arbeit mit zentralen Wortschätzen allgemeine lexikographische Probleme widerspiegeln und zuspitzen. Denn die Frage nach der Aufnahme von Fremdwörtern in einen zentralen Wortschatz ist im Grunde genommen auch eine Frage nach der Aufnahme in den etablierten Gesamtwortschatz.

4.4.3 Der Umgang mit Neologismen

Eine der am längsten diskutierten Anforderungen an Kandidaten für einen zentralen Wortschatz ist ihre Stabilität, d.h. die Anforderung, dass ein Lexem über einen gewissen Mindestzeitraum hinweg wohletabliert im Wortschatz vorhanden gewesen sein muss, damit es überhaupt in einen zentralen Wortschatz aufgenommen werden kann.¹¹² Entsprechend wird Stabilität häufig auch als ein elementares Selektionskriterium bezeichnet (vgl. Abschnitt 4.5.2.3.), und das nicht nur bei der Arbeit mit zentralen Wortschätzen, sondern auch in der allgemeinen Lexikographie (vgl. Bergenholtz 1989, S. 774).

Aber Stabilität gilt nicht nur als eine grundlegende Anforderung an Wortschatzeinheiten, sie wird auch, wenn auch nicht unumstrittenerweise, als konstituierendes Merkmal des Zentrums eines Wortschatzes betrachtet.¹¹³ Als Indiz für die Richtigkeit dieser Annahme von der Stabilität des (inneren) Kerns des Wortschatzes des Deutschen führt Krohn beispielsweise eine Schnittmengenbildung zwischen zwei konventionellen Grundwortschätzen und eine Schnittmengenbildung zwischen einem konventionellen Grundwortschatz und dem Häufigkeitswörterbuch des Kaeding durch. Dabei stellt sich heraus, dass die Divergenzen zwischen den beiden Grundwortschätzen kaum geringer sind als zwischen dem Grundwortschatz und Kaedings Häufigkeitswörterbuch:

„Es läßt sich [...] nachweisen, daß die im direkten Vergleich verschiedener Wortlisten auftretenden Wortschatzdivergenzen nicht geringer sind als bei solchen Wortschatzvergleichen, bei denen ein entsprechend limitiertes KÄDING-Material der eine Vergleichspartner war [...]. Damit soll aber nicht geleugnet werden, daß vor allem auf niedrigeren Frequenzstufen [bei KÄDING, V.H.] tatsächlich veralteter Wortschatz repräsentiert ist, eben weil ein Großteil des exzerpierten Textmaterials aus den letzten beiden Jahrzehnten des vergangenen Jahrhunderts stammt.“ (Krohn 1992, S. 41)

Wenn aber die Stabilität eines Lexems eine so tragende Rolle spielt, dann sind Neologismen, d.h. Neubildungen und Importe aus anderen Sprachen, grundsätzlich von der Aufnah-

112. Vgl. etwa Kühn 1979, S. 24. – Der Begriff *Stabilität* wird bei Kühn auch im Zusammenhang mit der Konsistenz der Frequenz eines Wortes, d.h. seiner gleichmäßigen Distribution über die Texte eines Korpus hinweg verwendet (vgl. Kühn 1979, S. 34), eine Bedeutung, auf die hier nicht Bezug genommen werden soll.

113. Vgl. etwa Kühn 1979, S. 43:

„Die Annahme einer Stabilität und Unveränderlichkeit der Sprache, auf die sich die Lexikographen vieler Wortlisten berufen, kann jedoch nur für die Strukturwörter gelten, die sich auf den ersten 1000 Plätzen etwa gleichrangig verteilen und nachweisen lassen. Der Rangwert der Inhaltswörter unterliegt jedoch in besonderem Maße den Änderungen und Anforderungen des Sprachgebrauchs an die Zeit.“
Dabei gilt, hält man sich an das Zentrum-Peripherie-Modell des Prager Strukturalismus (vgl. Abschnitt 5.4) Stabilität nicht nur als Merkmal des Zentrums des Wortschatzes, sondern auch als Merkmal des Zentrums anderer Teilsysteme der Sprache, etwa der Phonologie.

me in einen zentralen Wortschatz ausgenommen. Verbunden mit dieser Regel sind dabei drei Schwierigkeiten:

Zum einen enthält der Wortschatz als offenes System, das eng an die Veränderungen der außersprachlichen Welt geknüpft ist und sich ihnen viel stärker anpassen kann und muss als die anderen Teilsysteme der Sprache, viele Neubildungen. Schließt man also alle Neubildungen grundsätzlich von der Aufnahme in einen zentralen Wortschatz aus, schränkt man die Anzahl der Lexeme, unter denen man auswählen kann, drastisch ein.

Zum anderen ist unter gewissen Umständen, d. h. im speziellen Fall der zentralen Wortschätze des didaktischen Typs, die Stabilität bzw. Etabliertheit eines Lexems nicht unmittelbar identisch mit seiner Wichtigkeit für den Benutzer des Wortschatzes. Wie weit ist etwa das Lexem *Apfel* wichtiger als das Lexem *Kiwi*? Geht man streng nach der Stabilitätsregel vor, dann müsste das Lexem *Apfel* vor dem Lexem *Kiwi* den Vorrang haben. Dennoch kann es sein, dass *Kiwi* einen wichtigeren Bestandteil des Wortschatzes für jemanden, der Deutsch lernen soll, darstellt und etwa auch eine höhere Frequenz aufweist als *Apfel*.¹¹⁴ An dieser Stelle wird deutlich, dass bei der Selektion für einen zentralen Wortschatz nicht ein Selektionskriterium den Ausschlag geben kann, sondern mehrere Selektionseigenschaften zusammen betrachtet werden müssen.

Vor allem aber ist die Frage danach, wann ein Wort, das von außen neu in eine Sprache kommt oder neu gebildet wird, wohletabliert ist und sich stabilisiert hat, nicht einfach zu beantworten. Insbesondere muss man bei der Frage danach, ob ein Lexem 'neu' ist oder nicht, zwischen zwei Eigenschaften unterscheiden – der Eigenschaft der faktischen Neuheit eines Lexems in einem Wortschatz und der Eigenschaft des „Neuheitseffektes“ (Barz 1998, S.11), den ein Lexem bei den Sprechern einer Sprache haben kann:

„Obwohl jede lexikalische Einheit irgendwann neu aufkommt, wird ihre Neuheit nicht von allen Sprachbenutzern gleichermaßen erkannt, d.h. das Wort wird nicht von allen eine Zeitlang 'als neu empfunden' (Herberg 1988, S. 9). Vielmehr kann es nahezu unauffällig bleiben, auch wenn es sich nicht über die einmalige Verwendung hinaus verbreitet. Spontane Urteile über die Neuheit lexikalischer Einheiten, aber auch die Stichwortlisten einsprachiger Wörterbücher offenbaren erhebliche Normunsicherheiten bei der Entscheidung neu / nicht neu. Dieser Sachverhalt legt es nahe, in der Neologie zwischen der objektiven Neuheit der Wörter einerseits und deren Neuheitseffekt, ihrer diesbezüglichen Wirkung auf die Rezipienten andererseits, zu unterscheiden.“ (Barz 1998, S. 12f.)

Diese beiden Eigenschaften, die faktische Neuheit und der Neuheitseffekt eines Wortes, müssen getrennt voneinander beurteilt werden und sind im Allgemeinen nicht ohne eine Korpusanalyse oder eine Sprecherbefragung messbar.¹¹⁵

Für unsere Zwecke wichtig ist hier die Frage, ob die Neuheit bzw. Nicht-Neuheit oder der Neuheitseffekt bei der Auswahl einer Wortschatzeinheit für einen zentralen Wortschatz eine Rolle spielen soll. Grundsätzlich soll ein Kandidat für einen zentralen Wortschatz in

114. Ich danke Elmar Seebold für den Hinweis auf diesen Aspekt und das Beispiel.

115. Für eine eingehende Diskussion der adäquaten Verfahren zur Messung der Neuheit bzw. des Neuheitseffektes eines Lexems vgl. Barz 1998 und Schulz 2000.

dem jeweiligen Kommunikationskontext – dem Fachgebiet, der Kommunikationssituation – nicht neu sein. Bisher wurde aber der Unterschied zwischen faktischer Neuheit und dem Neuheitseffekt eines Wortes in der Literatur zu zentralen Wortschätzen nicht thematisiert. Wir wollen im Weiteren vereinfachend von der Neuheit bzw. Nicht-Neuheit eines Wortes im Wortschatz ausgehen, wenn von Neologismen und der Stabilität eines Wortes die Rede ist.

4.4.4 Der Umgang mit Synonymen, regionalen Konkurrenten und Kohyponymen

Während sich die Problemkreise von Fremdwörtern und Neologismen um einzelne Wörter drehen, geht es beim Problembereich des Umgangs mit Synonymen, regionalen Konkurrenten und Kohyponymen um Gruppen von Wörtern, die in einem gewissen synonymischen bzw. konkurrierenden Verhältnis zueinander stehen, weil sie den gleichen oder einen ähnlichen semantischen Raum besetzen, sich aber ihrer Form nach oder im regionalen Gebrauch voneinander unterscheiden. Dabei geht es konkret um die Frage, wie man unter solchen miteinander konkurrierenden Lexemen auswählen soll, denn:

„**Synonymische Relationen** stellen in gewisser Weise einen Problem- bzw. Sonderfall innerhalb eines zentralen Wortschatzes dar. Stark vereinfacht gesagt müssten Identität und Similarität im Kernwortschatz eigentlich vermieden werden, da ansonsten die Gefahr eines hohen Redundanzgrades bestünde, und folglich der Raum für zusätzliche 'elementare' Wörter fehlen würde.“ (Schnörch 2002, S. 356)¹¹⁶

Wenn wir hier von konkurrierenden Lexemen sprechen, dann sind damit drei Typen von Konkurrenten gemeint:

Der eine Typus ist der der unmittelbaren Synonyme.¹¹⁷ Nach welchen Regeln entscheidet man etwa, ob man das Wort *Sofa* oder eher *Couch* in den zentralen Wortschatz aufnehmen soll?

Der zweite Typus ist der der lokalen Konkurrenten: Bestimmte Begriffe sind regional unterschiedlich realisiert, wie etwa der Begriff 'Metzger', der regional durch verschiedene Bezeichnungen realisiert ist (vgl. etwa *Fleischer* vs. *Metzger* vs. *Fleischhauer* oder *Samstag*

116. Vor dem Hintergrund der Fremdsprachendidaktischen Verwendung zentraler Wortschätze wenden sich sowohl Krohn (1992, S. 152) als auch Zillig (2000, S.102) gegen ein grundsätzliches Ausschließen synonymen Lexeme. Sie argumentieren damit, dass Synonyme gerade beim Aufbau und der Erweiterung des rezeptiven Wortschatzes nützlich sein können. Dem steht allerdings die Frage gegenüber, ob man mit einer solchen Erweiterung des Zwecks des Grundwortschatzes nicht bereits die Grenze zwischen Grundwortschatz und genereller Wortschatzarbeit überschreitet.

117. Ich schließe mich dabei der Sicht von Schnörch an, der unter Synonymie vor allem eine Ähnlichkeitsrelation versteht:
„Synonymie steht in diesem Kontext einer Ähnlichkeitsrelation weit näher als einer Identitätsrelation. Völlig identische Wörter kommen als Folge des sprachlichen Ökonomieprinzips außer im Grenzbereich zwischen Fach- und Gemeinsprache (z.B. *Meningitis* = *Hirnhautentzündung*) ohnehin relativ selten vor. Will man den Synonymie-Begriff auf Grund dessen nicht als solchen in Frage stellen, so muss er auf Ähnlichkeitsbeziehungen ausgedehnt werden (können).“ (Schnörch 2002, S. 358). – Eine ausführliche Diskussion der theoretischen Grundlagen von Synonymie und Hypo- bzw. Hyperonymie findet sich im übrigen bei Schnörch 2002, S. 257ff. (Synonymie) und 345ff. (Hypo- bzw. Hyperonymie).

vs. *Sonnabend*). Hier muss man Regeln aufstellen, die es erlauben, eine nachvollziehbare Entscheidung darüber zu treffen, wann man sich für die eine oder andere regionale Variante entscheidet.¹¹⁸ Dabei macht das mit diesem Konkurrenztyp verknüpfte Problem der nötigen Auswahl einen besonderen Aspekt der Arbeit mit zentralen Wortschätzen sehr deutlich: Sie kann schnell eine normierende Kraft bekommen, denn indem man eine regionale Variante einer anderen vorzieht, setzt man eine Priorität und weist der nicht gewählten Variante einen weniger prominenten Platz zu. Das gilt allerdings vor allem für die didaktischen Wortschätze, nicht unbedingt für die analytischen, denn dort kann man eventuell regionale Varianten nebeneinander verzeichnen.

Der dritte Typus der Wortkonkurrenz ist auf einer etwas abstrakteren Ebene anzusiedeln. In einer horizontalen Dimension geht es hier um die Konkurrenz von Kohyponymen. Nachdem ein zentraler Wortschatz nur eine begrenzte Anzahl von Lexemen aufnehmen soll, ist es beispielsweise eine berechnete Frage, welche der Lexeme *Pudel*, *Collie*, *Schäferhund* und *Hund*, die alle zur Klasse 'Hund' gehören, in den zentralen Wortschatz aufgenommen werden sollen. Ähnliches gilt in einer vertikalen Dimension auch für Hyperonym-Hyponym-Beziehungen wie etwa *Pudel* – *Hund* – *Säugetier*, bei denen die Frage berechnete ist, welcher der Begriffe der zentrale, d.h. begrifflich neutrale ist. An dieser Stelle leuchtet wieder ein Problem auf, das sich durch die gesamte Diskussion der adäquaten Selektion zieht: Was bedeutet eigentlich die Eigenschaft der Zentralität und wie lässt sie sich messen? In Abschnitt 5.5 werden wir uns näher mit dieser Frage beschäftigen.¹¹⁹

Ein weiteres Beispiel für die Schwierigkeiten im Umgang mit Hyponymie stellt die Tatsache dar, dass es nicht immer ein Hyperonym gibt, das eine Reihe von Hyponymen zusammenfassen könnte und zugleich gebräuchlich ist. Kühn (1979, S. 25) führt hier das Beispiel der regionalen Konkurrenten *Fleischer*, *Metzger*, *Schlachter* etc. an, zu denen es keine allgemein verbindliche hochdeutsche Variante gebe. Dies wird insbesondere bei dem bei Kühn 1979 konzipierten begriffsorientierten Grundwortschatzwörterbuch deutlich; dort dient der Begriff im Sinne einer semantischen Basis als Grundeinheit der Selektion und wird erst nach der Selektion mit Lexemen besetzt. Schwierig wird dieses Verfahren, wenn es zu dem Begriff, also der allen Lexemen gemeinsamen semantischen Basis, kein eigenes etabliertes Lexem gibt und eine 'künstliche' Wortbildung nötig ist, um den Begriff zu repräsentieren.

Zugleich legt der Ansatz vom Begriff als Grundeinheit der Selektion auch eine Lösung des Synonymieproblems des ersten Typs nahe: Der Begriff lässt sich auch als eine Variable bzw. ein Platzhalter betrachten, der unter verschiedenen (regionalen, fachsprachlichen vs. gemeinsprachlichen) Bedingungen verschiedenartig realisiert werden kann. Solange man

118. Eine damit verwandte Problemstellung ist die der Auswahl unter verschiedenen, dem Inhalt nach synonymen, von der Stilfärbung her aber unterschiedlich gebrauchten Lexemen, also etwa das Problem der Auswahl von fachsprachlichem gegenüber umgangssprachlichem Lexem

119. Vgl. auch das Beispiel der Lexeme *Apfelbaum*, *Birnbaum*, *Kirschbaum*, *Nadelbaum* und *Laubbaum* bei Schnörch 2002, S. 362.

den Begriff nur methodisch als Grundeinheit betrachtet, bei der Selektion und insbesondere bei der Darstellung aber auf das Semem zurückgreift, ließe sich damit das Problem des Umgangs mit Synonymen zumindest theoretisch-methodisch lösen. Praktisch müssen sich dann allerdings Verfahren finden, die helfen, zunächst einmal alle diejenigen Sememe zu finden, die den zugrundeliegenden Begriff realisieren, und dann in einem weiteren Schritt Entscheidungshilfen dafür geben, wie zwischen den konkurrierenden Sememen zu wählen ist. Hier können Korpusanalysen nützlich sein.

4.4.5 Der Umgang mit Wortfeldern und ihren Grenzen

Eine mit der Frage des Umgangs mit konkurrierenden Lexemen verwandte Problemstellung ist die des Umgangs mit Wortfeldern. Einer der Hauptkritikpunkte an den frequenzorientierten Grundwortschätzen war die Tatsache, dass, legt man der Selektion allein die Frequenz eines Lexems als Kriterium zugrunde, Wortfelder aufgebrochen und möglicherweise nicht geschlossen in den Grundwortschatz aufgenommen werden. Daher wurde die Forderung laut, geschlossene Wortfelder als ganze in den Wortschatz zu übernehmen.¹²⁰ Entsprechend füllt beispielsweise Pfeffer 1970 paradigmatische Lücken, die durch die Auswahl anhand der Frequenz entstanden sind, bewusst auf (vgl. Schnörch 2002, S. 65). Ähnlich verfährt man beim ZERTIFIKAT DEUTSCH 1977 (vgl. Schnörch 2002, S. 179).

Das Problem, das sich hier verbirgt, ist zunächst ein terminologisches. Denn hier ist zu unterscheiden zwischen geschlossenen Wortfeldern im Sinne von Nomenklaturen, wie sie etwa die Wochentage darstellen, und eher offenen Wortfeldern, wie sie etwa die Gruppe der Farbbezeichnungen oder Gruppen von Antonymen darstellen. In der Grundwortschatzliteratur wird oft ungenau von ‘geschlossenen Wortfeldern’ gesprochen, wenn in Wahrheit eher offene Wortfelder gemeint sind.¹²¹ Dabei wird vernachlässigt, dass Nomenklaturen im Gegensatz zu den eher offenen Wortfeldern eine klar abgegrenzte Gruppe von Wörtern darstellen, und zwar sowohl intern (jedes Wort einer Nomenklatur hat einen gegenüber einem anderen Wort der Nomenklatur klar abgegrenzten semantischen Bezugsbereich), als auch extern (ein Wort *x* gehört eindeutig zu einer Nomenklatur, nicht aber das Wort *y*), während Wortfelder eher weniger deutliche Abgrenzungen aufweisen.

Das Problem, das sich dabei stellt, ist jedoch nicht unbedingt das der Nomenklatur, auf die sich die Kritik in der Regel bezieht, sondern die Tatsache, dass die Mehrzahl aller Wortfel-

120. Vgl. etwa Bergenholtz 1989, S. 774 und Erk 1963, S. 87f. Für ein Beispiel vgl. auch Krohn 1992, S. 51. – Diese Forderung wird in der englischsprachigen Literatur mit *linguistic closure* bezeichnet (vgl. etwa Lehmann 1991, S. 304) und teils in der deutschen Grundwortschatzliteratur unter dem Terminus *Geschlossenheit* übernommen (vgl. Erk 1963, S. 88).

121. Vgl. etwa: „If one takes the *n* most frequent words of some frequency count one will not doubt discover that these words will not exhibit a **linguistic closure** in the sense that natural sentences can be formed with all and only the words in the set. Further one will see that semantic relations will be incomplete. Thus finds one in Oehler (1980) [= Oehler 1980, V. H.] which is based on the old Kaeding count that *weiblich* (*female*) occurs but not its antonym *männlich* (*male*). For a core vocabulary to be set up for a natural language system, I think, one must strive for linguistic closure, since otherwise, one ends up with words one cannot use.“ (Lehmann 1991, S. 304).

der zwar in gewisser Weise als solche zu erkennen sind, aber nicht unbedingt aus einer festen Anzahl von Lexemen bestehen, wie das zum Beispiel beim Wortfeld der Farbbezeichnungen der Fall ist. Hier stellt sich also die Frage, wie man um ein solches nicht eindeutig abgrenzbares Wortfeld eine Grenze zieht.¹²² Nachdem eine solche Grenzziehung immer in gewissem Maße arbiträr sein wird, ist die Lückenhaftigkeit eines zentralen Wortschatzes unvermeidlich.¹²³ Darüberhinaus führt der Mangel an allgemein anerkannten Regeln zum Umgang mit Wortfeldern zu unterschiedlichen Ergebnissen bei der Lemmatisierung von Korpusmaterial (vgl. Schnörch 2002, S. 216). Die Ergänzung von mutmaßlich lückenhaften Wortfeldern ist daher auch ein hoch subjektives Verfahren:

„So nehmen sowohl Pfeffer (1964:69f.) als auch Kühn und das ZERTIFIKAT unter Hinweis auf Wortfeldzugehörigkeit bestimmte Lexeme in den Grundwortschatz auf, um subjektiv empfundene lexikalische Lücken zu schließen. Damit wird aber ein semantisch-logisches Vorverständnis von Vollständigkeit zu einem Kriterium der Wortauswahl. Man geht nämlich von der stillschweigenden Voraussetzung aus, daß strukturelle Vollständigkeit und kommunikative Nützlichkeit sich in dem Sinne gegenseitig bedingen, daß die Wortauswahl aus Gründen der lexikalischen Vollständigkeit in jedem Falle auch kommunikativ zu rechtfertigen ist.“ (Krohn 1992, S. 85)

Das Hauptproblem beim Umgang mit Wortfeldern ist, zumindest im Falle des Grundwortschatzes, allerdings, wie Krohn (1992, S. 85-87) deutlich macht, dass ein Ergänzen vermeintlich lückenhafter Wortfelder zur Überrepräsentation eines Wortfeldes auf Kosten der lexikalischen Unterrepräsentation eines anderen Wortfeldes führen kann, insbesondere wenn man dem zu erstellenden zentralen Wortschatz eine strenge quantitative Grenze auferlegt. Eine frequentiell gestützte Auswahl an wichtigen Vertretern des jeweiligen Wortfeldes kombiniert mit einer onomasiologischen Darstellung der Einträge kann dabei helfen, solche Über- und Unterrepräsentationen aufzudecken.

4.5 Das Merkmal der Zentralität und die Vielfalt der Selektionskriterien

Neben den bisher dargelegten allgemeinen Problemen der Selektion stellt die Frage nach den adäquaten Selektionskriterien eines der grundlegenden Themen der Arbeit mit zentralen Wortschatzen dar. Ihm liegt die wesentliche Frage zugrunde, wie sich das allgemeine Merkmal der Zentralität in konkret fassbaren Eigenschaften von Wörtern niederschlägt und wie sich diese Eigenschaften als Selektionskriterien nutzen lassen. Weil der Begriff der Zentralität ein sehr vieldeutiger und relativer Begriff ist, gibt es auch entsprechend viele Selektionskriterien. Dabei steht die Diskussion der Selektionskriterien in fast allen Arbeiten im Zentrum, eben weil es bei der Erstellung eines zentralen Wortschatzes um eine Selektion von Einheiten aus der potentiell unendlichen Menge von Einheiten des Wortschatzes einer

122. Ähnliche Probleme der Grenzziehung ergäben sich auch, wenn man eine Selektion anhand von Wortfamilien (*Fahrer, fahren, Fahrt, Fahrzeug*) vornimmt. Wortfamilien sind im Prinzip ein gut geeignetes Mittel der Wortschatzvermittlung, weshalb man Wortfamilien theoretisch zumindest für die Erstellung didaktisch orientierter zentraler Wortschatze heranziehen könnte. Doch gerade bei Wortfamilien stellt sich die Frage, welche der zur selben Familie gehörigen Lexeme man in den zentralen Wortschatz aufnimmt.

123. Vgl. Krohn 1992, S. 85 und Schnörch 2002, S. 74.

Sprache geht und daher den Fragen der Selektion aus offensichtlichen Gründen viel Aufmerksamkeit gewidmet werden muss.

Zu den bisher vertretenen Selektionskriterien gehören dabei vor allem folgende Kriterien:¹²⁴

- die Verwendbarkeit,¹²⁵
- die Gebräuchlichkeit,¹²⁶
- die Relevanz eines Wortes für den potentiellen Sprachbenutzer und die Benutzungssituation,¹²⁷
- der semantische Wert eines Wortes,¹²⁸
- seine Frequenz und Streuung,¹²⁹
- seine Vertrautheit und Disponibilität,¹³⁰
- die Eigenschaft der Repräsentiertheit und Plausibilität,¹³¹
- das Kriterium der zeitlichen Stabilität,¹³²
- das Kriterium der Zugehörigkeit zu den zentralen Wortarten,¹³³
- die morphologische Produktivität,¹³⁴
- die Fähigkeit zur Strukturbildung,¹³⁵
- die stilistische Neutralität,¹³⁶
- die Grundbegrifflichkeit,¹³⁷
- die Zugehörigkeit zu einem geschlossenen Wortbereich¹³⁸ und
- das semantische Expansionspotential.¹³⁹

124. Für einen Überblick über die Selektionskriterien und ihre Beurteilung in der bisherigen Literatur vgl. Kühn 1979, S. 23-40 (für die grundlegende Darlegung der Selektionskriterien) und S. 40-59 (für die Kritik an den Selektionskriterien); Krohn 1992, S. 59-61 (für eine kritische Besprechung des Selektionskriteriums der Frequenz), Bonan-Garrigues 1993, S. 8-20 (für eine kritische Besprechung des Selektionskriteriums der Frequenz) und Schnörch 2003, S. 16-23.

125. Vgl. Kühn 1979, S. 58.

126. Vgl. Buchbinder 1973, S. 302.

127. Vgl. Kühn 1979, S. 68.

128. Vgl. Buchbinder 1973, S. 303.

129. Vgl. zum Beispiel Muller 1977, Andreev 1980, Geens 1978 u. a. – Vgl. auch Abschnitt 4.5.2.1.

130. Vgl. Gougenheim / Rivenc / Michéa u. a. 1964 und Michéa 1964.

131. Vgl. Bonan-Garrigues 1993 und Abschnitt 4.5.2.2.

132. Vgl. Abschnitt 4.5.2.3.

133. Vgl. Abschnitt 4.5.2.4.

134. Vgl. etwa Kühn 1979, S. 26 oder Krohn 1992, S. 97-100.

135. Vgl. Buchbinder 1973, S. 302.

136. Vgl. Abschnitt 4.5.2.5.

137. Vgl. etwa Riesel / Schendels 1975, S. 61.

138. Vgl. etwa Bergenholtz 1989, S. 774 oder Schnörch 2002, S. 179.

4.5.1 Die Kontroverse um den Vorrang einer frequenzorientierten vs. einer pragmatischen Selektion

Die oben angeführte kurze Auflistung zeigt, dass die bisher verwendeten Kriterien sehr unterschiedlich sind, sowohl was die Eigenschaften betrifft, auf die sie sich beziehen, als auch darauf, wie konkret fassbar sie sind. Es besteht daher auch kein Konsens darüber, wie ein allgemein anerkanntes und bewährtes Kriterienset für die Selektion zu einem zentralen Wortschatz beispielsweise aussehen könnte. Insbesondere die grundsätzliche Opposition von frequenzorientierten gegenüber pragmatisch orientierten Kriterien dominierte lange Zeit die Diskussionen.¹⁴⁰ Erst mit den Arbeiten von Krohn 1992 beginnt eine Auflösung der bis dahin deutlich gezogenen Grenze zwischen frequenzorientiertem und pragmatischem Ansatz, so dass heute eine integrative Sicht, die Frequenz als eine von vielen Eigenschaften eines Wortes betrachtet, die ursprüngliche Dualität der Ansätze ersetzt.¹⁴¹ Die Gründe für diese Änderung der Sichtweise sind in Folgendem zu finden:

Zum einen sind, wie Krohn 1992 zeigen kann, die beiden früher als gegensätzlich betrachteten Ansätze der frequenzorientierten vs. pragmatisch orientierten Selektion einander im praktischen lexikographischen Ergebnis ähnlicher als es der bis dahin heftig geführte Streit vermuten ließ:

„Die Beschreibung von Selektionskriterien zur Ermittlung von Grundwortschätzen gründet sich entweder auf der Frequenz/Vorkommenshäufigkeit und der Streuung von Lexemen oder auf der textkorpusgestützten oder intuitiv vorgenommenen Themen-Situationsauswahl und hat mehr in der Theorie und weniger in der Praxis der Grundwortschatzlexikographie zu stark divergierenden, um nicht zu sagen antithetischen Wissenschaftsparadigmen in diesem Forschungsbereich geführt. Dabei hat die Bilanzierung der reinen Frequenzlexikographie und der reinen kommunikativen Pragmalexikographie verdeutlicht, daß der gegenseitige Vorwurf, keine wohldefinierten, intersubjektiv nachprüfbaren Kriterien aufweisen zu können, insofern nicht berechtigt ist, als zumindest für den Zweck eines didaktisch verwertbaren Grundwortschatzes die Frequenzlexikographie nicht ohne Themen und Situationen und die Pragmalexikographie, wenn auch weniger offenkundig, nicht ohne Häufigkeitsangaben [...] auskommt.“ (Krohn 1992, S. 74f.)¹⁴²

Beispielsweise kann Krohn anhand der Auswahl und Berücksichtigung hoch polysemer Lexeme in dem von ihm zur Analyse herangezogenen Wortbestand des ZERTIFIKAT DEUTSCH, einem Grundwortschatz, der sich ausschließlich an pragmatischen Kriterien der Selektion orientiert, zeigen, dass die Frequenz des Gebrauchs und die introspektiv bestimmte Nützlichkeit eines Lexems durchaus als zwei Seiten ein und derselben Medaille betrachtet werden können, d.h. dass eine hohe Frequenz nur Ausdruck der ausgeprägten Vielseitigkeit eines Wortes ist.¹⁴³ In den Ausführungen zum Selektionsverfahren für das ZERTIFIKAT DEUTSCH 1977...

139. Vgl. Hasan 1974, S. 182 oder Buchbinder 1973, S. 305.

140. Vgl. auch Abschnitt 2.1.1.

141. Vgl. etwa Krohn 1992, S. 75 und Schnörch 2002, S. 26, aber auch in Ansätzen bereits bei Rosengren 1976, S. 332.

142. Ähnlich auch Krohn 1992, S. 55 und 71.

„...wird polysemen Lexemen nicht nur die (erwartbare) Eigenschaft der relativ hohen Frequenz zuerkannt, sondern auch eine mehr oder minder deutlich erkennbare Beziehung zwischen kommunikativer Nützlichkeit und textbedingter Häufigkeit von Lexemen in dem Sinne angenommen, daß zwischen Frequenz und Nützlichkeit ein relativ hoher Deckungsgrad besteht. Diese implizit erschließbare Konvergenz von frequentiellen und kommunikativ-pragmatischen Kriterien ist ein deutliches Indiz dafür, daß die theoretischen Prämissen für die Verwerfung des Frequenzkriteriums nicht stichhaltig sind.“ (Krohn 1992, S. 49f.)

Ähnliches zeigt Krohn auch bei der Diskussion der Frage, wo überall Selektion eine Rolle spielt: Während die Vertreter des kommunikativ-pragmatischen Ansatzes in ihrem Verfahren von der Selektion bestimmter Themenbereiche ausgingen und damit die Häufigkeit des Vorkommens eines Themas zum grundlegenden, eigentlichen Auswahlkriterium machten, ginge es beim frequentiellen Verfahren um die Häufigkeit des Vorkommens einzelner Wörter. Damit seien aber die beiden Verfahrensweisen gar nicht so weit voneinander entfernt; vor allem das von Kühn vorgeschlagene Verfahren unterscheide sich in dieser Hinsicht kaum vom frequenzorientierten Ansatz:

„Mit dieser Definition [kommunikationsrelevante Themen = Themen, die den meisten schriftlichen und mündlichen Kommunikationsformen zugrundeliegen, V.H.] wird das so heftig kritisierte Frequenzkriterium in modifizierter Form aber wieder eingeführt, weil die Wichtigkeit oder die kommunikative Relevanz von Themen als Funktion ihres Vorkommens in Texten betrachtet wird. Anders formuliert operiert Kühn nicht mit einem lexikalischen Rekurrenzkriterium, d.h. mit der Vorkommenshäufigkeit von Wörtern/Lexemen, sondern mit einem thematischen Rekurrenzkriterium, d.h. mit der Vorkommenshäufigkeit von Themen in Texten als Mittel der Wortauswahl.“ (Krohn 1992, S. 70)

Neben Krohns Beobachtungen zu Konvergenzen im Resultat und im Auswahlverfahren bei frequenz- bzw. pragmatisch orientierten Verfahren spielt auch die Frage nach der Objektivität eine ausschlaggebende Rolle für die Annahme eines integrativen Ansatzes: Die genaue Betrachtung frequentieller Verfahren, die gerade wegen ihrer angeblichen objektiven Vorgehensweise verteidigt werden, zeigt nämlich, dass Objektivität auch durch sie nicht zu erreichen ist, sondern nur auf eine andere Ebene – die der Korpuszusammenstellung und Datenauswertung – verschoben wird.¹⁴⁴

Umgekehrt schließlich ist auch eine absolute semantische Vollständigkeit und Geschlossenheit des zentralen Wortschatzes, um die man sich im pragmatischen Ansatz bemüht, nicht erreichbar, wie Schnörch zeigen kann:

„Aus der Vielzahl möglicher Ko-Hyponyme führt Pfeffer [...] Birke, Buche, Eiche, Fichte, Linde und Tanne an. Weshalb nur und gerade diese, lautet folgerichtig die Frage; weshalb fehlen Ahorn, Kiefer, Lärche, Pappel und auch das sind noch längst nicht alle zumindest heimischen Bäume. Es dürfte klar sein, dass Vollständigkeit weder angestrebt werden kann noch muss und es immer Lücken geben wird. Aus diesem Grund ist die Auswahl gegenüber der Strukturierung [nach Hyponymien bzw. Hyponymien, V. H.] auch zweitrangig.“ (Schnörch 2002, S. 361f.)

143. Vgl. Krohn 1992, S. 49f. – Wie schon weiter oben erklärt, gilt dies jedoch nur, wenn man das Lexem und nicht die lexikalische Einheit von einer Form und einer Bedeutung als Grundeinheit heranzieht. Denn die Vielseitigkeit eines Wortes ist an seine semantische Vielseitigkeit gebunden, und die ist von der Mehrdeutigkeit abhängig, die es nur auf der Ebene des Lexems gibt, nicht aber auf der Ebene der lexikalischen Einheit. Vgl. dazu auch Kaufmann 1968, S. 17.

144. Vgl. etwa Bonan-Garrigues 1993, S. 6 und Schnörch 2002, S. 150.

Dabei soll nicht geleugnet werden, dass in bestimmten Fällen ein hauptsächlich frequentiell orientiertes Verfahren oder ein vor allem pragmatisch ausgerichtetes Vorgehen einem integrativen Ansatz vorzuziehen sind, etwa wenn es um die Ermittlung von Fachwortschatz geht (vgl. Buchbinder 1973, S. 301f.) oder um die Erstellung von Wortlisten, die in hauptsächlich statistisch orientierten Anwendungen wie etwa der Stopwortliste des *Information Retrieval* (vgl. Abschnitt 3.3) eingesetzt werden sollen. Gerade bei analytischen Wortschätzen wird ein frequenzorientiertes Selektionsverfahren als Abbildung des Gebrauchs von Wörtern – idealerweise unter Berücksichtigung der Bedeutungsseite der betrachteten Wörter – eine gute Grundlage sein.¹⁴⁵

Dass der Stellenwert des frequentiellen Verfahrens stark vom gewünschten Ergebnis bzw. Zweck des zu erstellenden zentralen Wortschatzes abhängt, illustriert auch Krohn, wenn er über sein „Testlexikon“, die von ihm erstellte Schnittmenge mehrerer Grundwortschätze und Häufigkeitswörterbücher sagt:

„Inwiefern Struktur und Funktion des Testlexikons als eine Art von Grundwortschatz interpretiert werden kann, hängt weitgehend davon ab, ob der Terminus 'Grundwortschatz' in Übereinstimmung mit dem frequentiellen Ansatz als der häufigste oder in Übereinstimmung mit dem kommunikativen Ansatz als der Wortschatz aufgefaßt werden kann, der im Bereich der Lexik den Bedingungen des kommunikativen Minimums genügt. Aus frequentieller Sicht stellt das Testlexikon faktisch einen auf 1.300 Lexeme beschränkten Grundwortschatz dar, aus kommunikativer Sicht dagegen nicht, eben weil der kommunikativ relevante themenspezifische Wortschatz weitgehend ausgeblendet ist.“ (Krohn 1992, S. 130)

Aus Krohns Aussage wird aber umgekehrt auch deutlich, dass für die didaktisch orientierten Wortschätze, also den Grundwortschatz und das definitonische Vokabular, mehr als nur frequentielle Aspekte für ein sinnvolles Ergebnis nötig sind. Vor dem Hintergrund der hier zusammengefassten Diskussion wollen wir daher vereinfachend annehmen, dass zwar analytische Wortschätze tendenziell mehr von frequenzorientierten Verfahren profitieren, dass aber umgekehrt für didaktische zentrale Wortschätze ein integrativer Ansatz, der viele Aspekte der Wortschatzstrukturierung, also sowohl quantitative als auch qualitative Selektionskriterien in Betracht zieht und die Transparenz des Vorgehens in den Mittelpunkt stellt, einer einseitigen Konzentration auf einen der beiden traditionellen Ansätze vorzuziehen ist.

4.5.2 Einige exemplarische Selektionskriterien

Im Folgenden wollen wir uns nun einigen exemplarischen Selektionskriterien zuwenden und sie vor dem Hintergrund der bisherigen Diskussionen in der Literatur zu zentralen Wortschätzen ausführlicher besprechen. Der Grund für diese detaillierte Darlegung gerade dieser Selektionskriterien liegt darin, dass sie im praktischen Teil als exemplarische Selektionskriterien

145. Insbesondere der Neutralwortschatz für die automatische Textklassifikation wird vor allem von einer hauptsächlich frequentiell orientierten Auswahl abhängig sein, wie etwa auch die Untersuchung themenspezifischer und themenunspezifischer Autosemantika bei Krohn (1992, S. 51f.) zeigt.

tionskriterien eingesetzt werden sollen, wozu wiederum eine genaue Kenntnis der Stärken und Schwächen der einzelnen Kriterien nötig ist.

4.5.2.1 Frequenz und Streuung

Die Merkmale der Frequenz eines Wortes und seiner Streuung im Korpus stellen die ältesten Selektionskriterien dar. Frequenz definiert sich dabei im Ausgangspunkt als die absolute Häufigkeit, mit der ein Wort, d.h. zunächst eine laufende Wortform, und, in aggregierter Form, die zugehörige Grundform in einem Textkorpus vorkommt. Da sich bei großen Korpora schnell viele verschiedene Häufigkeiten ergeben können, werden Wörter oft in Häufigkeitsgruppen zusammengefasst.¹⁴⁶

Neben der einfachsten Form der absoluten Häufigkeitszählung gibt es auch verschiedene Versuche, die Häufigkeitsmessung zu verfeinern, indem man etwa die Häufigkeit eines Wortes in zwei verschiedenen Korpora untersucht und die Gesamthäufigkeit des Wortes in beiden Korpora zueinander in Beziehung setzt.¹⁴⁷

Ein anderes Mittel der Verfeinerung und Ergänzung des reinen Frequenzkriteriums ist das der Streuung. Mit Streuung, auch als Distribution bezeichnet, versucht man zu erfassen, wie weit verbreitet ein Wort ist, d.h. in wie vielen verschiedenen Texten bzw. bei wievielen verschiedenen Autoren es verwendet wird. Denn die absolute Frequenz eines Wortes ist ein an sich ambivalenter Wert: Die Frequenz eines Wortes kann entweder dadurch entstehen, dass ein Wort im Gesamtkorpus gleichmäßig verteilt so und so oft vorkommt oder dass es in einigen wenigen Einzeltexten des Korpus besonders häufig ist. Damit lässt eine hohe absolute Frequenz aber keinen unmittelbaren Schluss auf die hohe Gebräuchlichkeit des Wortes generell zu, sondern kann auch Ausdruck für seine Gebräuchlichkeit in bestimmten (themenspezifischen) Texten sein. Ähnlich sagt Andreev,...

„...daß die Auftretenshäufigkeit eines Wortes in Texten eine Funktion zweier Argumente ist. Erstens hängt sie ab von der Anzahl der Autoren, die dieses oder jenes Lexem gebrauchen. [...] Zweitens hängt die Auftretenshäufigkeit auch davon ab, mit welcher Intensität einzelne Autoren ein gegebenes Lexem verwenden.“ (Andreev 1980, S. 84)

Als Konsequenz daraus entwickelt Andreev beispielsweise ein Verfahren zur Erstellung eines Distributionswörterbuchs, bei dem die breite Verteilung eines Wortes über verschiedene Texte eines Korpus hinweg stärker zählt als die absolute Frequenz.¹⁴⁸

Distribution als ein Mittel zur Beschreibung der Zentralität eines Wortes wird auch im *Information Retrieval* verwendet, allerdings unter umgekehrten Vorzeichen: Dort gilt die sogenannte *Inverted Document Frequency*, also die möglichst geringe Anzahl von

146. Vgl. etwa das Verfahren des Leipziger Projekts *Deutscher Wortschatz*, bei dem die gezählten Wörter mit einer Häufigkeitsklasse angegeben werden, die relativ zum häufigsten Wort des Korpus – in diesem Falle *der* – angegeben wird (vgl. Abschnitt 8.5). Ähnliches schlägt Buchbinder vor: Je nachdem, in welcher Frequenzbreite ein Wort sich befindet, werden ihm Bewertungspunkte zugeordnet (vgl. Buchbinder 1973, S. 306f.).

147. Vgl. den sog. „Balanced Difference Coefficient“ bei Geens 1978, S. 49.

Dokumenten, in denen das Suchwort in dem zu durchsuchenden Dokumentenbestand vorkommt, zusammen mit einer hohen Einzelfrequenz des gesuchten Wortes als Indikator für die Relevanz der gefundenen Dokumente.¹⁴⁹ Damit wird deutlich, dass das Kriterium der Streuung bei genauem Hinsehen nur ein symptomatisches Kriterium für die Eigenschaft der Themenspezifität ist: Je breiter die Streuung eines Lexems, desto geringer, so könnte man annehmen, ist seine Themenspezifität.

Wie wir bereits bei der Darstellung der Geschichte der Grundwortschatzlexikographie (vgl. Abschnitt 2.1.1) gesehen haben, stellt Frequenz allerdings ein heftig umstrittenes Selektionskriterium dar. Die Hauptkritikpunkte seien hier kurz zusammengefasst:¹⁵⁰

Zunächst einmal muss die automatische Gleichsetzung von hoher Frequenz und kommunikativer Nützlichkeit, zu der die Verwendung von Frequenzlisten zur Bildung eines zentralen Wortschatzes verführen kann, mit Skepsis betrachtet werden.¹⁵¹ Das zeigt die Tatsache, dass die häufigsten Wörter in der Regel Funktionswörter sind, die innerhalb eines Textes nur wenig zur Semantik der Aussage beitragen, während Autosemantika als die eigentlichen Träger der Bedeutung einer Aussage in der Regel eine viel geringere Frequenz aufweisen.¹⁵² Bei dieser Argumentation muss man allerdings berücksichtigen, dass sie vor dem Hintergrund des Fremdsprachenunterrichts geführt wird, wo die kommunikative Nützlichkeit der zu selektierenden Wortschatzeinheiten im Vordergrund steht. Die dort getroffene Schlussfolgerung, dass Frequenzuntersuchungen grundsätzlich als zweckfrei zu betrachten seien, während die Erarbeitung eines Grundwortschatzes in höchstem Maße zweckorientiert sei (vgl. Erk 1963, S. 88f.), muss jedoch als für alle Typen von zentralen Wortschatzen gültig betrachtet werden.

Ein weiterer Kritikpunkt an der Frequenz als Selektionskriterium ist die mangelnde theoretische Verankerung. So ist umstritten, ob das dem Frequenzmodell und den daraus mancherorts gezogenen Schlussfolgerungen zur Wahrscheinlichkeit des Auftretens einer Einheit zugrundeliegende mathematische Modell der Urne zum Funktionieren von Sprache im allgemeinen und zur Wortwahl im Besonderen passt. Beim Urnenmodell geht man da-

148. Andreev entwickelt darüber hinaus ein komplexes Beschreibungssystem für die Distribution von Wörtern, auf die einzugehen hier kein Raum ist. – Zur Distribution vgl. auch Muller 1977 und 231-243 und Endres-Niggemeyer (1975, S. 3); sie bezeichnet die Distribution allerdings als „den Gebrauchswert eines Wortes“. Auch Kühn referiert die Distribution als ein Kennzeichen der „Stabilität der Häufigkeit“ (Kühn 1979, S. 34). – Im übrigen entsteht hier eine Verbindung zwischen dem quantitativen Kriterium der Frequenz bzw. Distribution und dem qualitativen Kriterium der Neutralität (vgl. Abschnitt 4.5.2.5): Ein Kennzeichen der stilistischen und thematischen Neutralität ist es, dass das jeweilige Wort in thematisch und stilistisch verschiedenen Kommunikationsszusammenhängen vorkommt.

149. Vgl. die Diskussionen bei Harman 1992.

150. Vgl. etwa auch Kühn 1979, S. 41-48 oder Carter 1998, S. 46.

151. Vgl. Erk 1963, S. 87, Kühn 1984, S. 244 und Bonan-Garrigues 1993, S. 18. – Eine eher der Frequenz zugeneigte Diskussion der Frage, ob häufige Wörter auch zugleich nützliche Wörter sind, findet sich bei Krohn 1992, S. 56-58.

152. Vgl. Erk 1963, S. 86f. – Augst nennt das die „degressive Struktur des Wortschatzes“ (Augst 1987, S. 117). – Vgl. auch Abschnitt 5.3.2.

von aus, dass die Urne eine gegebene, d.h. endliche Anzahl von Einheiten, etwa Kugeln, enthält, die voneinander unabhängig sind und unabhängig voneinander aus der Urne gezogen werden können. Aus der Anzahl der Einheiten in einer Urne und aus dem Wissen darüber, wieviele gleichartige Einheiten in der Urne vorhanden sind, kann man dann vorhersagen, mit welcher Wahrscheinlichkeit man eine Einheit eines bestimmten Typs aus der Urne ziehen wird. Genau diese Eigenschaften treffen aber, so etwa Kühn (1979, S. 42f.), auf das Lexikon einer Sprache nicht zu. Zum einen sei ihr Inhalt nicht endlich, überschaubar und stabil, sondern unendlich groß und veränderlich, da Wortbildungsprozesse eine Inventarisierung des Bestandes erschweren; zum anderen seien die im Lexikon vorhandenen Einheiten nicht voneinander unabhängig, weil die Wahl der ersten Einheit aufgrund ihrer syntaktischen, semantischen und pragmatischen Eigenschaften bestimmt, welche anderen Einheiten als nächste gewählt werden können, und damit die Auswahlmöglichkeiten einschränkt.

Gegen diese Einwände führt Hendrickx (1987, S. 9ff.) wiederum folgende Argumente an: Die von Kühn in Frage gestellte Endlichkeit und Stabilität des Wortschatzes sei in einem abgegrenzten Zeitraum betrachtet durchaus gegeben; zwar sei der Wortschatz theoretisch unendlich, praktisch sei er aber begrenzt, weil sich nicht jede Einheit mit jeder kombinieren ließe. Das einzige wirkliche Problem seien semantische Veränderungen, aber auch diese seien durch ihre Verbindung zur ursprünglichen Bedeutung erfassbar. Dem Einwand von Kühn, beim Produzieren von Text seien die Einheiten eben nicht von einander unabhängig hält Hendrickx die de Saussuresche Unterscheidung von *langue* und *parole* entgegen: „The frequency of words, which are parts of the language system, can only be measured in the discourse, the actual use.“ (Hendrickx 1987, S. 9).¹⁵³

In gewisser Weise muss beiden Standpunkten Recht gegeben werden. Die Wahrscheinlichkeit, dass ein bestimmtes Wort in einer beliebigen Kommunikationssituation verwendet wird, lässt sich sicherlich nicht direkt aus seiner Frequenz ableiten. Andererseits ist auch Hendrickx recht zu geben, wenn er Frequenz als messbaren Ausdruck des tatsächlichen Sprachgebrauchs bezeichnet. Wenn man also Frequenz als eine Messung des Ist-Zustandes betrachtet und alle Versuche der Vorhersage der Wahrscheinlichkeit des Gebrauchs eines Wortes außer Acht lässt, können Frequenzbeobachtungen sehr nützliche Aussagen über den Gebrauch von Wörtern liefern.

Kritisiert werden Frequenzerhebungen auch dafür, dass sie, weil sie auf unterschiedlichen Grundeinheiten beruhen, nicht miteinander vergleichbar seien und keine echten Anhaltspunkte für den Umfang eines zentralen Wortschatzes liefern könnten (vgl. Kühn 1979, S. 46). Schwerwiegender ist jedoch, dass sich bisherige Frequenzzählungen oft auf die Zählung von Wortformen beschränken, semantische und syntaktische Aspekte jedoch vernachlässigen. In der einfachsten Variante der Frequenzermittlung werden dabei formal

153. Für eine eingehendere Diskussion der Frage, ob das Urnenmodell auf den Wortschatz anzuwenden sei vgl. Martin 1990, S. 1316f.

identische Einheiten als Instanzen ein und derselben zu zählenden Einheit gerechnet, ohne dass dabei zum Beispiel auf syntaktische Unterschiede, wie etwa das unterschiedliche Verhalten verschiedener Wortarten, Rücksicht genommen wird. Das führt zu verzerrten Häufigkeitsangaben, weil gleichlautende, funktional aber unterschiedliche Einheiten zusammengelegt und ihre Frequenz addiert wird. Aber selbst unter Berücksichtigung syntaktischer und morphologischer Merkmale sind Häufigkeitszählungen bisher nicht immer zuverlässig, weil bisher eine wirkliche Einbeziehung semantischer Informationen zur Unterscheidung von Homonymen identischer Wortart nach Bedeutungen (wie etwa das bekannte Beispiel von *Bank1* ('Möbelstück') und *Bank2* ('Geldinstitut')) nicht möglich ist. So erhalten homonyme Formen tendenziell höhere Frequenzwerte, als ihnen eigentlich zustünde. Diese Orientierung an der laufenden Wortform anstatt an der Grundform und die Vernachlässigung der Einheit von Semantik, Syntax und Lautform führen in der Regel dazu, dass man die Ergebnisse einer Frequenzerhebung sehr gründlich interpretieren muss und nicht immer unmittelbar übernehmen kann.¹⁵⁴

Ein weiterer Kritikpunkt an der Frequenz als Selektionskriterium richtet sich gegen das bereits in Abschnitt 4.4.5 besprochene Problem der lückenhaften Aufnahme von Wortfeldern. Schnörch tritt dieser Kritik vehement entgegen, denn diese kritisiere...

„...Grundwortschatzsammlungen auf Grund von vielfach beliebigen, d.h. häufig wenig systematischen Stichproben [...], nach dem Motto: Die aufgenommenen Wörter *a,b,c* sind überflüssig, dafür fehlen die Wörter *x,y,z*. Hier wird eine Objektivität vorgespiegelt bzw. erwartet, die es trotz aller Bestrebungen nicht gibt, nicht geben kann. [...] Eine solche Kritik ist ebenso subjektiv gefärbt wie die Vorgehensweise und die Ergebnisse des Kritisierten.“ (Schnörch 2002, S. 74)

Wenn man die hier angeführten Kritikpunkte insgesamt betrachtet, dann wird, wie wir bereits in Abschnitt 4.1 gesehen haben, deutlich, dass viele der gegen den frequenzorientierten Ansatz ins Feld geführten Argumente darauf zurückzuführen sind, dass Grundwortschätze mit Frequenzlisten gleichgesetzt werden, wie beispielsweise Krohn (1992, S. 61f.) deutlich macht.¹⁵⁵

Geht man aber, wie das im integrativen Ansatz von Schnörch der Fall ist, davon aus, dass bei der Selektion für einen zentralen Wortschatz vielfältige Eigenschaften, von denen Frequenz nur eine ist, herangezogen werden, ändert sich das Bild: Dann werden die Aspekte der Frequenz, die von den Gegnern des frequenzorientierten Ansatzes als kritisch beurteilt werden, einfach zu Einschränkungen des Selektionskriteriums der Frequenz, das wie andere Selektionskriterien auch Stärken und Schwächen aufweist. Die Schwächen müssen bei der Verwendung der Frequenz zwar berücksichtigt werden, schließen aber ihre Verwendung nicht von vorneherein aus. Zudem haben die Stärken, die den Einschränkungen des Kriteriums der Frequenz gegenüberstehen, einiges Gewicht: So ist etwa die Frequenz aus

154. Zum Problem verzerrter Häufigkeitsangaben aufgrund von nicht unterschiedenen Homonymen vgl. z.B. auch Erk 1963, S. 93 und Erk 1972, S. 11.

155. Vgl. auch die Diskussionen in den Abschnitten 3.4.3 und 4.1.

einer praktisch-lexikographischen Perspektive ein nützlicher Untersuchungsaspekt, weil sie trotz der mit ihrer Erhebung verbundenen praktischen Schwierigkeiten wie etwa der Lemmatisierung oder der richtigen Korpuszusammensetzung im Vergleich zu anderen Eigenschaften noch relativ problemlos festgestellt werden kann. Aus einer systemlinguistischen Sicht wiederum ist Frequenz eines von mehreren etablierten Merkmalen, hinsichtlich derer sich Sprache beobachten lässt; und es spielt in anderen linguistischen Disziplinen, etwa der kognitiven Linguistik, eine wesentliche Rolle als Untersuchungs- und Erklärungsinstrument, etwa wenn es darum geht, zu erklären, weshalb bestimmte Wörter besser erinnert werden als andere.¹⁵⁶

Schließlich spricht für die Anwendung der Frequenz im Rahmen eines integrativen Ansatzes auch, wie bereits in Abschnitt 4.5.1 besprochen, die Tatsache, dass die Resultate rein frequenzorientierter Selektionsverfahren und die rein pragmatisch orientierter Selektionsverfahren einander vergleichsweise ähnlich sind. Vor allem darf nicht der pauschale Schluss gezogen werden, rein kommunikativ-pragmatische Verfahren seien für die Grundwortschatzlexikographie vorzuziehen, weil Fragen der lexikalischen Belegung eines gegebenen Themenkatalogs auf der Basis eines Textkorpus noch gar nicht ausreichend geklärt sind (vgl. Krohn 1992, S. 80f.).

Wenn wir nun das bisher Gesagte zusammenfassen, dann wird deutlich, dass Frequenz eine Eigenschaft ist, die bei der Erstellung eines zentralen Wortschatzes in jedem Fall mitberücksichtigt werden muss. Hauptargument dafür ist die Tatsache, dass Frequenz ein messbarer Ausdruck des Gebrauchs von Wörtern in einem gegebenen Kontext ist. Die Schwierigkeiten, die mit der Erhebung von Frequenzen verbunden sind, müssen bei der Interpretation der Ergebnisse berücksichtigt werden, sind aber kein Grund dafür, die Frequenzanalyse als solche aufzugeben. Wichtig ist ein differenzierter Umgang mit der Eigenschaft der Frequenz. Es gilt, sie weder völlig zu verwerfen noch sie zu monopolisieren, indem man sie zum alleinigen Selektionskriterium macht, anhand dessen Ja-Nein-Entscheidungen über die Aufnahme von Wörtern in einen zentralen Wortschatz gefällt werden. Dabei ist auch eine klare Trennung zwischen der Anwendungsorientiertheit zentraler Wortschätze und dem lexikologisch-beschreibenden Charakter von Häufigkeitswörterbüchern im Auge zu behalten. Auch muss man bedenken, dass Frequenzuntersuchungen eher geeignet sind, Aussagen über allgemeine Regularitäten und Tendenzen in der Sprache zu machen als über das Verhalten eines konkreten einzelnen Wortes spezifische und globale Aussagen zu machen.¹⁵⁷

Zu einem differenzierten Umgang gehört auch die genaue Kenntnis dessen, wie die konkreten Häufigkeitsangaben zustande gekommen sind, woraus sich das zugrundeliegende Korpus zusammensetzt, weshalb es so zusammengestellt wurde, ob sich die

156. Vgl. etwa Glück (Hg.) 2000, S. 197, Schindler 2002, S. 42 oder beispielsweise die Erklärungsmodelle der Verarbeitung von visuellen bzw. phonetischen Informationen bei Seidenberg 1989 bzw. Dell 1989.

Häufigkeitsangaben auf das gesamte Korpus beziehen oder nur auf Teile etc. (vgl. Erk 1963, S. 89).

Schließlich gehört zu einem differenzierten Umgang auch die Frage, wieviel Gewicht der Frequenz im Kontext eines spezifischen zentralen Wortschatzes beigemessen werden soll. Beispielsweise können Frequenzuntersuchungen bei der Erarbeitung von Grundwortschätzen für bestimmte Fachbereiche zur ersten Grobsortierung derjenigen Einheiten, die zum Verständnis eines Textes dieses Fachbereiches nötig ist, von Nutzen sein. Dies räumt sogar Kühn, der sonst ein erklärter Gegner der frequenzorientierten Wortschatzarbeit ist, ein (vgl. etwa Kühn 1979, S. 53). Demgegenüber scheint eine einseitige Bevorzugung der Frequenz als Selektionskriterium bei allgemeinsprachlichen Grundwortschätzen, wie wir gesehen haben, weniger geeignet zu sein. Für die Erstellung eines definitiven Vokabulars ist Frequenz sicherlich nicht das erste und wichtigste Selektionsmerkmal, sondern kann allenfalls hilfreich sein, wenn es darum geht zu entscheiden, welches von zwei alternativen Wörtern man verwenden soll (vgl. Herbst 1986, S. 102).

Wir wollen also festhalten: Die Frequenz eines Wortes ist eine von mehreren Eigenschaften, die seine Zentralität für eine bestimmte Anwendung bestimmen, und muss daher in einer Gesamtbeurteilung berücksichtigt werden. Dabei ist es möglich, innerhalb der Gesamtbeurteilung bestimmte Gewichtungen vorzunehmen, indem man etwa für analytische Wortschätze der Frequenz eines Wortes mehr Gewicht beimisst als einer anderen Eigenschaft. Prinzipiell sollte aber immer die Berücksichtigung mehrerer Eigenschaften eines Wortes Ausgangspunkt der Bestimmung seiner Zentralität sein:

„Frequenz und Distribution sind [...] wichtige Faktoren, um einem Lexem das Attribut 'zentral' zu bescheinigen, obgleich ggf. natürlich noch weitere (sprachimmanente) Kriterien, etwa der Gebrauchswert bzw. die 'Reichweite' [...], zu berücksichtigen wären.“ (Schnörch 2002, S. 26)¹⁵⁸

4.5.2.2 Transparente Introspektion

Wie bereits in Abschnitt 2.2.3 kurz beschrieben, entwickelt Bonan-Garrigues 1993 in ihrer Arbeit ein Verfahren der transparenten Introspektion, das zwar von seiner Grundidee her teilweise nicht neu, in der Art der praktischen Durchführung aber bisher einzigartig ist.¹⁵⁹

157. So läßt sich zum Beispiel aus der Feststellung von Ruoff (1983, S. 21), der Frequenzen der gesprochenen Sprache untersucht und feststellt, dass in dem von ihm untersuchten Material 60% der ermittelten Wörter Substantive sind, zugleich aber nur 10% der laufenden Formen von Substantiven gestellt werden, einiges ableiten; etwa dass Substantive eine offenere Klasse darstellen als Verben oder dass Verben und Adverbien stärker 'wiederverwendet' werden als etwa Adjektive oder Substantive. Man kann aber nicht sagen, dass ein bestimmtes Substantiv in der Regel immer eine geringere Frequenz hat als ein bestimmtes Verb. – Ähnlich argumentiert auch Schnörch (2002, S. 25 und 121f.).

158. Ähnlich auch bei Schnörch 2002, S. 372.

159. Für eine Besprechung der Introspektion als Mittel der Selektion bei früheren Arbeiten der Grundwortschatzforschung vgl. Krohn 1992, S. 63f.

Das Verfahren der transparenten Introspektion beruht dabei auf der Beschreibung und Bewertung zweier Eigenschaften – der Repräsentiertheit („représentation“) und der Plausibilität („plausibilité“).¹⁶⁰

Bonan-Garrigues nähert sich dem Konzept der *Repräsentiertheit*, indem sie zunächst den Begriff des *Verstehens* („comprendre“ (Bonan-Garrigues 1993, S. 24) ausführlich diskutiert (ebda., S. 24-26) und zeigt, dass dieser Begriff kein geeignetes Werkzeug ist, um die Zentralität eines Wortes zu erfassen. Stattdessen wendet sie sich dem Phänomen des Wiedererkennens von Wörtern zu und analysiert, welche Typen von Wiedererkennung es gibt (vgl. ebda., S. 25-29). In ihrer Analyse kommt sie zu den folgenden Typen des Wiedererkennens: „mots [...] connus“ (ebda., S. 26), „mots [...] reconnus“, „mots [...] reconstitués“ (ebda., S. 28) und „mots inconnus“ (ebda., S. 26). Zwar ließen sich, so Bonan-Garrigues, zwischen diesen verschiedenen Stufen des Wiedererkennens keine genauen Grenzen ziehen (Bonan-Garrigues 1993, S. 26f.), dennoch seien diese verschiedenen Grade der Wiedererkennung eine gute Grundlage für die Beschreibung der Repräsentiertheit: Wörter, die man wiedererkennt, sei es ob sie bekannt („connus“), wiedererkannt („reconnus“) oder erschließbar („reconstitués“) sind, sind im internen Lexikon repräsentiert. Unbekannte Wörter („mots inconnus“) gelten als nicht repräsentiert (vgl. Bonan-Garrigues 1993, S. 41). Diese Phänomene des Wiedererkennens sind Grundlage der Eigenschaft der Repräsentiertheit. Im Gegensatz zum Verstehen eines Wortes, das von seiner Verwendung in einem konkreten Kontext abhängig ist, beruht die Eigenschaft der Repräsentiertheit allein auf der strikten Anwesenheit des Wortes im internen Lexikon eines Sprechers, denn was zählt ist allein die Frage, ob ein Wort irgendeine Reaktion hervorruft, eine Reminiszenz. Dieses Konzept erlaubt es, bei der Selektion von Wörtern eines Lexikons über das Verständnis einzelner, isolierter Wörter hinauszugehen und die damit verbundenen Probleme zu umgehen.

Entsprechend definiert Bonan-Garrigues Repräsentiertheit folgendermaßen:

„Nous dirons donc que, pour nous, la représentation est une **‘image mentale – précise ou vague, forte ou faible, émotionnellement positive ou négative – signalant la simple présence du mot dans le lexique mental d’un individu’**.“ (Bonan-Garrigues 1993, S. 40)

Mit dieser Definition sind die Probleme, die mit der Frage danach, was es bedeutet, ein Wort zu verstehen bzw. was es bedeutet, ein Wort zu kennen, für die Praxis erst einmal gelöst:¹⁶¹ Die von Bonan-Garrigues beschriebenen „mots connus“, „mots reconnus“, „mots partiellement connus“ und „mots virtuellement connus“ rufen eine Reaktion hervor, werden

160. Der Terminus *Repräsentiertheit* stellt dabei den Versuch einer möglichst bedeutungsnahen Übersetzung des von Bonan-Garrigues verwendeten Terminus der *représentation* dar. Seine zunächst von der Form her näher liegende Übersetzung durch das Wort *Repräsentation* würde zwar die Präsenz eines Wortes im Lexikon bezeichnen, nicht aber eine graduelle Form der Präsenz zulassen, wie wir sie bei der praktischen Untersuchung anwenden werden (vgl. Abschnitt 8.2).

161. Vgl. die Abschnitte „Comprendre un mot a-t-il un sens?“ und „Inconnus, connus, re-connus: étude de cas“ bei Bonan-Garrigues 1993, S. 24-26 und S. 26-29.

als repräsentiert beschrieben und erhalten den Wert R1; Wörter, die keinen Wiedererkennungswert haben („mots inconnus“), gelten als nicht repräsentiert und erhalten den Wert R0.

Nun ist es für jemanden, der Französisch nur als Fremdsprache beherrscht, durchaus schwierig, die Einteilungen Bonan-Garrigues' nach Wiedererkennung und Nicht-Wiedererkennung nachzuvollziehen.¹⁶² Und es stellt sich auch die Frage, ob es ein absolutes Nicht-Wiedererkennen überhaupt gibt, denn denkbar ist auch, dass jede Lautkombination irgendwelche, wenn auch vielleicht rein lautmalerischen Assoziationen hervorruft. Dennoch scheint Bonan-Garrigues Vorgehensweise einleuchtend und in jedem Fall transparenter als eine grobe Einteilung des Wortschatzes nach gebräuchlichen und weniger gebräuchlichen Lexemen.

Das zweite Merkmal, auf das hin Bonan-Garrigues jedes Wort des DELAS untersucht, ist das der Plausibilität. Mit Plausibilität meint Bonan-Garrigues dabei die intuitive Beurteilung der Möglichkeit, dass dem Beurteilenden ein bestimmtes Wort im Ganzen der schriftlichen und mündlichen Äusserungen, denen wir gewöhnlich ausgesetzt sind, begegnet.

Aus offensichtlichen Gründen werden dabei nur Wörter, die bei der Beschreibung der Repräsentiertheit als repräsentiert eingestuft wurden, auf ihre Plausibilität hin untersucht. Über Wörter, die im mentalen Lexikon des Beurteilenden nicht vorhanden sind, kann der Beurteilende keine Aussage hinsichtlich ihrer Plausibilität machen. Die Beurteilung der Plausibilität dient daher der feineren Unterscheidung der Lexeme, die im mentalen Lexikon repräsentiert sind.

Bei der Beschreibung der Plausibilität wird jedem Lexem ein Grad an Plausibilität zugeordnet: Lexeme, deren Auftreten in einem kommunikativen Zusammenhang dem Beurteilenden sehr plausibel erscheint, erhalten den Grad P1, Lexeme mittlerer Plausibilität den Grad P2 und Lexeme niedriger Plausibilität den Grad P3.¹⁶³ Dabei wird das Kriterium der Plausibilität auf jedes Wort einzeln angewandt, ohne dabei benachbarte Wörter eines Paradigmas oder eines Wortfeldes miteinzubeziehen.

Bonan-Garrigues betont bei diesem Verfahren, dass Plausibilität nicht mit statistischer Wahrscheinlichkeit zu verwechseln sei. Während die statistische Wahrscheinlichkeit eine Extrapolation ist, die aus der Häufigkeitszusammensetzung eines Textkorpus gewonnen wird, ist die Plausibilität eine Vorhersage über den Gebrauch eines Wortes, die sich aus der kommunikativen Erfahrung des Beurteilenden heraus speist. Eine wichtige Rolle spielt dabei auch der Wissenshintergrund des Beurteilenden. Wie wir bereits in Abschnitt 2.2.3 kurz besprochen haben, vertritt Bonan-Garrigues die Ansicht, dass nur Linguisten die Plausibi-

162. Vgl. etwa die Unterscheidung zwischen *salsepareille*, die Bonan-Garrigues als „partiellement connu“ bezeichnet, und *journade*, die Bonan-Garrigues als „inconnu“ einstuft (Bonan-Garrigues 1993, S. 41).

163. Das ursprüngliche Verfahren sah nur zwei Grade von Plausibilität vor – plausibel und nicht plausibel. Beispiele für plausible vs. nicht plausible Lexeme sind: *dollar – dirham*, *arriver – arrimer*, *mignon – mignard*, *danger – dangerosité*, *mensualisation – menstruation* (vgl. Bonan-Garrigues 1993, S. 43)

lität eines Wortes beschreiben können. Denn im Gegensatz zu dem bei Gougenheim / Rivenc / Michéa u. a. 1956 entwickelten Kriterium der Vertrautheit, das sich auf die umfangreiche Befragung von nicht linguistisch ausgebildeten Sprechern bezieht, ist nach Ansicht Bonan-Garrigues' gerade die Fähigkeit eines Sprechers mit linguistischer Ausbildung, Sprache als Objekt von einer Metaebene her zu betrachten, eine Grundvoraussetzung für das Funktionieren des Beschreibungsverfahrens der Plausibilität.

Jedes der Wörter des DELAS durchläuft damit mindestens zwei Beschreibungsrunden:

In einem ersten Schritt wird jedes Wort nach dem Kriterium der Repräsentiertheit beschrieben. Je nach dem, ob das Wort im mentalen Lexikon des Beurteilenden repräsentiert ist oder nicht, bekommt es den Wert R1 bzw. R0.¹⁶⁴

In einem zweiten Schritt werden alle Wörter, die den Wert R1 haben, nach ihrer Plausibilität beschrieben; Wörter mit geringer Plausibilität bekommen das Merkmal P3, Wörter mit mittlerer Plausibilität bekommen das Merkmal P2, und Wörter mit hoher Plausibilität bekommen das Merkmal P1.¹⁶⁵ Dabei spielt die Reihenfolge eine Rolle: Nachdem die Repräsentiertheit eines Wortes im mentalen Lexikon des Sprechers möglichst spontan und ohne allzuviel aktiviertes Kontextwissen geschehen soll, muss die Eigenschaft der Repräsentiertheit für alle Wörter der Untersuchungsmenge zuerst beschrieben werden. Erst dann können die Wörter hinsichtlich ihrer Plausibilität beschrieben werden.

Die so beschriebenen Wörter werden dann in die Gruppen R1 bzw. R0 einerseits und die Gruppen P1 bzw. P2 bzw. P3 andererseits eingeteilt. Aus der Kombination dieser Gruppen ergibt sich dann in einem dritten Schritt eine Schichtenzuordnung: Wörter, die sowohl in Gruppe R1 als auch in Gruppe P1 vorkommen, gehören zur grundlegenden Schicht von Wörtern, der sogenannten Schicht 1 („couche essentielle haute“ (Bonan-Garrigues 1993, S. 45)). Wörter, die sowohl in Gruppe R1 als auch in Gruppe P2 vorkommen, gehören zur nächst grundlegenden Schicht, der sogenannten Schicht 2 („couche essentielle basse“). Alle anderen Wörter, d.h. diejenigen, die sowohl in Gruppe R1 als auch in Gruppe P3 vorkommen und diejenigen, die der Gruppe R0 angehören, werden der Peripherie, der sogenannten Schicht 3 („couche résiduelle“) zugeordnet. Das Ergebnis dieser Beurteilung ist die Einteilung der Wörter des DELAS nach dem folgenden Verhältnis: 30 % der Wörter des DELAS gehören der Schicht 1 an, 18 % der Schicht 2 und 52 % der Schicht 3 (vgl. Bonan-Garrigues 1993, S. 57).

Bonan-Garrigues beschreibt zwei hauptsächliche Problembereiche dieses Verfahrens:

164. Das von Bonan-Garrigues angeführte Beispiel einer solchen Einteilung umfasst die Einträge *délétion*, *vaigrage*, die nach Bonan-Garrigues Beurteilung das Repräsentiertheitsmerkmal R0 haben, und die Einträge *eau*, *vin*, *pléonasme*, *hypallage*, *balsamine*, *salsepareille*, *carénage*, die nach Bonan-Garrigues Beurteilung das Repräsentiertheitsmerkmal R1 haben (vgl. Bonan-Garrigues 1993, S. 45).

165. Das von Bonan-Garrigues angeführte Beispiel der unterschiedlichen Zuordnung von Plausibilitäten umfasst die Einträge *eau*, *vin* und *pléonasme*, denen sie die Plausibilität P1 zuordnet, die Einträge *salsepareille* und *carénage*, denen sie die Plausibilität P2 zuordnet und die Einträge *balsamine* und *hypallage*, denen sie die Plausibilität P3 zuordnet (vgl. Bonan-Garrigues 1993, S. 45).

Zum einen das Problem graphischer Varianten ein und derselben Bedeutung wie etwa *curry* vs. *carry* oder *tzar* vs. *czar* vs. *tsar* etc. (vgl. Bonan-Garrigues 1993, S. 46), die in der alphabetischen Anordnung des Lexikons teilweise weit von einander entfernt liegen. Als grundlegende Regel ordnet Bonan-Garrigues dabei alle Varianten ein und derselben Schicht zu, macht jedoch bei der Beurteilung der Plausibilität einen Unterschied, wenn ihr eine graphische Variante weitaus weniger plausibel erschien als eine andere.¹⁶⁶

Der andere Problembereich sind Einträge des DELAS, die von der Form her freistehend sind, aber keine freien Lexeme darstellen, wie etwa *ping pong*, *parce que*, *cahin caha*, *nec plus ultra* (vgl. Bonan-Garrigues 1993, S. 46). Dieses Problem rührt wohl vom Entstehungsverfahren des DELAS her, das durch Extraktion von Wörtern aus Texten entstanden ist und daher Wörter nach ihrer Form abgrenzt (Lexeme, die aus mehreren von einander abgesetzten Formen bestehen, werden nicht als eine Einheit berücksichtigt) und an verschiedenen Stellen einordnet. Bonan-Garrigues begegnet diesem Problem, indem sie versucht, zuerst die mit dem unfreien Morphem, das sie zur Beurteilung vor sich liegen hat, verbundene Konstruktion zu rekonstruieren und dann die Einzelteile der Gesamtkonstruktion mit den jeweils gleichen Repräsentiertheits- und Plausibilitätsmerkmalen zu belegen.¹⁶⁷

In einem abschließenden Schritt evaluiert Bonan-Garrigues die Ergebnisse ihres Verfahrens, und das auf dreierlei Weise:

Zum einen stellt sie einen Vergleich mit einer rein frequenzbasierten Methode an: Sie stellt ihrer Klassifikation der Einträge des DELAS der Frequenzliste des *Dictionnaire de fréquence des mots du français* (Beauchemin / Martel / Theoret 1992) gegenüber und vergleicht, wie sehr die von ihr erarbeiteten Schichten C1 und C3 mit den hohen bzw. niedrigen Frequenzrängen des *Dictionnaire de fréquence des mots du français* übereinstimmen. Das Ergebnis sei, so Bonan-Garrigues, dass die von ihr vorgenommene Zusammenstellung homogener sei als diejenige, die durch die alleinige Konzentration auf die Frequenz entstanden wäre (vgl. Bonan-Garrigues 1993, S. 63f.). Eine genauere Begründung bzw. nachvollziehbare Angabe für die Gründe dieser Schlussfolgerung fehlt allerdings.

Zum anderen testet sie die „couverture lexicale par nos trois couches sur un large corpus de textes de sources diverses“ (Bonan-Garrigues 1993, S. 64), d.h. sie führt einen Textdekungstest durch, indem sie die von ihr erstellte Wortschatzschichtung mit einem Textkorpus bestehend aus Zeitungs- und Zeitschriftenartikeln, Maschinenanleitungen, technischen Texten, Texten der französischen Presseagentur AFP etc. abgleicht.¹⁶⁸ Dieser Textkorpus enthält ca. 20 000 verschiedene Simplizia, von denen sich 15 000 in der von Bo-

166. Vgl. *enquiquineur* und *enquiquiner*, die der Schicht 1 zugeordnet werden, gegenüber *enkikiner* und *enkikineur*, die der Schicht 2 zugeordnet werden (vgl. Bonan-Garrigues 1993, S. 46)

167. Vgl. Bonan-Garrigues 1993, S. 46. – Es ist dabei offensichtlich, dass die Tatsache, dass der Beurteilende imstande ist, anhand eines einzelnen Bestandteils die Gesamtkonstruktion zu rekonstruieren, eine Repräsentiertheit von 0 ausschließt.

168. Der Grund für diese Auswahl ist vermutlich ein praktischer: Dieses Korpus war zum Zeitpunkt der Untersuchungen von Bonan-Garrigues bereits am LADL vorhanden und aufbereitet.

nan-Garrigues erstellten Schicht 1 wiederfinden, 2 500 in der Schicht 2 und 2 000 in der Schicht 3.¹⁶⁹ D.h. die von Bonan-Garrigues ermittelte Schicht 1 deckt die große Mehrzahl der im Korpus vorkommenden Simplizia ab, worin Bonan-Garrigues eine Bestätigung ihres Verfahrens sieht (vgl. Bonan-Garrigues 1993, S. 64f.).

Schließlich führt sie auch eine Sprecherbefragung durch, bei der sie Sprechern eine Auswahl an Wörtern des DELAS vorlegt und sie bittet, sie nach den Eigenschaften von Repräsentiertheit und Plausibilität zu beurteilen. Diese Sprecherbefragung führt zu sehr großen Unterschieden in der Beurteilung, weshalb Bonan-Garrigues zu dem oben genannten Schluss kommt, dass die Schichtung von Wortschatz anhand der Eigenschaften der Repräsentiertheit und Plausibilität durch Linguisten erarbeitet werden muss (vgl. Bonan-Garrigues 1993, S. 52).

Betrachtet man nun das Verfahren von Bonan-Garrigues in seiner Gesamtheit, dann wird deutlich, dass es im Grunde in zwei Hauptschritte eingeteilt ist: In einem ersten Schritt werden die unbekanntesten Wörter des zu untersuchenden Bestandes durch die Beschreibung der Repräsentiertheit der Peripherie der Wortschatzschichtung zugeordnet; in einem zweiten Schritt geht es um die Frage, welcher Schicht die bekannten Wörter eines Wortbestandes zuzuordnen sind. Weil alle bekannten Wörter den gleichen Repräsentiertheitswert (R1) haben, geschieht diese Zuordnung allein anhand des Merkmals der Plausibilität. Damit erhält die Eigenschaft der Plausibilität viel Gewicht.

Was die Praktikabilität dieses Kriteriums anbelangt, so ist darüber hinaus klar, dass ein solches Verfahren aufwändig ist, zumal die Beurteilung von Lexemen auf ihre Repräsentiertheit und Plausibilität hin möglichst mehrere Beurteiler und Beurteilungsrunden erfordert und mit Tests kombiniert werden sollte, wie sie Bonan-Garrigues beispielhaft durchführt.

Zugleich stellt das von Bonan-Garrigues entwickelte Verfahren einen Kompromiss dar zwischen der Notwendigkeit, andere als nur mechanische Kriterien wie Frequenz und Streuung als Selektionskriterien zuzulassen, und der Kritik an der willkürlichen introspektiven Selektion, wie sie Vertreter der pragmatischen Richtung praktiziert haben und die auf so diffusen Kriterien wie Nützlichkeit, Gebräuchlichkeit etc. beruht.¹⁷⁰ Gerade dadurch, dass Bonan-Garrigues den Begriff der *Introspektion* in konkrete Beschreibungsschritte zerlegt, verhilft sie ihm zum Status eines transparenten Beschreibungsmerkmals und trägt so zur Erweiterung des Repertoires an Merkmalen bei, die man betrachten kann, um zu einer Schichtung des Wortschatzes zu kommen. Zwar steht der Wert der Ergebnisse für einen spezifischen Anwendungstypus eines zentralen Wortschatzes noch aus, d.h. man müsste dieses Verfahren für einen konkreten zentralen Wortschatz einmal durchführen und dann diesen Wortschatz in seiner konkreten Anwendung evaluieren. Aber es ist bereits ersichtlich, dass die Brückenfunktion, der Kompromiss, den dieser Versuch darstellt, viel wert ist. Ein Zweck

169. Die restlichen waren im DELAS nicht verzeichnet und daher in der von Bonan-Garrigues erstellten Wortschatzschichtung nicht vertreten.

170. Vgl. etwa auch (Krohn 1992, S. 49).

der in dieser Arbeit vorgenommenen praktischen Untersuchung wird es daher auch sein, das Kriterium der Introspektion anderen Eigenschaften gegenüberzustellen und zu vergleichen, welche Auswirkung es auf die Selektion, d.h. zum Beispiel auf den Umfang des resultierenden zentralen Wortschatzes hat.

4.5.2.3 Zeitliche Stabilität

Das Kriterium der Stabilität eines Lexems, d.h. der Unveränderlichkeit seiner Form und Bedeutung über einen gegebenen Zeitraum hinweg ist eines der zentralen Auswahlkriterien der bisherigen Arbeit mit zentralen Wortschatzen.¹⁷¹ Ursprünglich aus dem Zentrum-Peripherie-Modell der Prager Strukturalisten stammend (vgl. Abschnitt 5.4) hat dieses Kriterium schnell eine wesentliche Rolle in der Grundwortschatzlexikographie übernommen:

„Obwohl sich das lexikalische Potential einer Sprache in einer permanenten Entwicklung befindet und sich in qualitativer und quantitativer Hinsicht den wechselnden Kommunikationsbedürfnissen anpaßt, gibt es auch einen Bestandteil der Lexik, der sich 'trotz der vielfältigen Beeinflussung durch die gesellschaftliche Entwicklung nur verhältnismäßig langsam verändert.' [= Schmidt 1955, S. 533, V. H.]. Aus diesem Grunde kann zwischen einem labilen, sich in ständiger Entwicklung befindlichem [sic!] Allgemeinwortschatz [...] und einem relativ stabilen Grundwortschatz unterschieden werden. Der stabile Grundwortschatz hat dabei die Funktion, 'Mittel der Verständigung zu sein' [= Schmidt 1955, S. 533, V. H.] und zwar nicht innerhalb eines kurzen Zeitabschnittes, sondern über einen sehr langen Zeitraum hinweg.“ (Kühn 1979, S. 24)

Unterstützt wird die Forderung nach Stabilität auch in der frequenzorientierten Grundwortschatzarbeit, denn die Stabilität und Kontinuität des Ausgangsmaterials, d.h. der Textkorpora, aus denen Frequenzlisten erstellt werden, ist eine Grundvoraussetzung für konsistente Ergebnisse (vgl. Kühn 1979, S. 43).¹⁷²

Bei dem Kriterium der Stabilität ist man also darum bemüht, Neologismen, die nur vorübergehender Bestandteil des Wortschatzes sind, von der Aufnahme in einen zentralen Wortschatz auszuschließen. Dabei hilft die punktuell erhobene Frequenz eines solchen Wortes nur wenig: Neologismen können eine Zeit lang die Frequenzlisten zu Zeitungen zum Beispiel stark dominieren,¹⁷³ müssen damit aber noch nicht zum dauerhaft zentralen Bestandteil des Wortschatzes gerechnet werden. Erst eine über einen gewissen Zeitraum hinweg gleichbleibende Präsenz eines Wortes im Wortgebrauch kann ein Hinweis auf seine Zentralität sein.

Während das Kriterium der Stabilität also an sich ein relativ etabliertes Selektionskriterium darstellt, ist seine praktische Erfassung umstritten. Eine Möglichkeit ist der Blick in alte Wörterbücher. Deren Verwendung ist, wie Matthias Schulz (2000) darlegen kann, nicht un-

171. Es spielt auch bei der Frage nach dem lexikologischen Zentrum des Wortschatzes einer Sprache eine Rolle. Vgl. etwa Kühn 1979, S. 24 oder Andreev 1980, S. 82f.

172. Vgl. auch Krohn 1992, S. 41, der bei Kaeding veralteten Wortschatz vor allem bei den niedrigfrequenten Einträgen feststellt.

173. Vgl. etwa das Wort *Elchtest*, das vor einigen Jahren aufgrund einer besonders spektakulären Nachricht aus der Autoindustrie in vielen Zeitungsartikeln häufig auftauchte.

problematisch, weil das Vorhandensein oder Fehlen eines Lexems in einem Wörterbuch zunächst einmal nur etwas über die lexikographische Praxis des jeweiligen Wörterbuches aussagt und nicht über das tatsächliche Vorhandensein oder Fehlen des Lexems in der Sprache.¹⁷⁴ Ideal wäre daher eine Korpusuntersuchung über einen längeren Zeitraum hinweg. Nachdem es dazu, speziell auch für Zeiträume vor der maschinellen Erschließung von Sprachdaten, an ausreichendem Korpusmaterial fehlt, das zum einen vergleichbar und zum anderen mit detaillierten semantischen Informationen angereichert wäre, bleibt der Wörterbuchvergleich die bisher einzige praktikable Lösung.

Das Kriterium der zeitlichen Stabilität ist also ein in gewisser Weise paradoxes Kriterium: Auf den ersten Blick erscheint es als ein gut greifbares, linguistisch gesehen sehr handfestes Kriterium, dessen Relevanz für die Auswahl durchaus einleuchtet. Schwierig ist aber die praktische Umsetzung, weil die Grenzziehung zwischen stabilen und nicht stabilen Einheiten schwierig ist.

Trotz dieser praktischen Schwierigkeiten ist Stabilität ein sehr wichtiges Kriterium der Selektion, weil es von einem theoretischen Standpunkt ein tragender Aspekt der Trennung von zentralen und peripheren Wortschatzeinheiten darstellt und weil es von einem linguistischen Standpunkt aus verglichen mit vielen anderen Selektionskriterien noch ein relativ hohes Maß an Intersubjektivität mitbringt.

4.5.2.4 Zugehörigkeit zu den zentralen Wortarten

In manchen Arbeiten zum Grundwortschatz wird auch diskutiert, ob man lexikalische Einheiten, die einer bestimmten, als zentral angesehenen Wortart wie etwa der der Substantive, Adjektive oder Verben angehören, anderen Wortschatzeinheiten, die nicht diesen Wortarten angehören, vorziehen sollte (vgl. etwa Köpke 1964, S. 70). Argumentiert wird dabei ähnlich wie bei Buchbinder (1973, S. 303), der den semantischen Wert eines Wortes als wichtigen Teil seiner Zentralität betrachtet: Substantive, Verben und Adjektive haben häufig eine ausgeprägtere lexikalische Semantik als das etwa Präpositionen oder Konjunktionen haben, ausserdem tragen sie die eigentliche Funktion des Prädizierens und Referierens in einem Satz, und diese Funktion soll eine prominente Rolle bei der Auswahl von Lexemen für einen Grundwortschatz darstellen.¹⁷⁵

Fraglich ist allerdings, ob die pauschale Selektion aufgrund von Wortartenzugehörigkeit ein kluger Schritt ist. Denkbar ist zum Beispiel, dass die gehäufte Verwendung von Präpositionen in einem Textkorpus einen wichtigen Hinweis dafür liefern kann, dass es sich bei der Textsammlung um Fachtexte handelt, in denen erfahrungsgemäß viel mit präpositionalen Konstruktionen gearbeitet wird. Darüber hinaus ist nicht unmittelbar erkennbar, wie die Bevorzugung einer bestimmten Wortart die Erstellung eines zentralen Wortschatzes quan-

174. Vgl. Schulz 2000, S. 64. Eine detaillierte Besprechung des Verfahrens des Wörterbuchvergleichs findet sich in Abschnitt 8.6.

175. Vgl. etwa Krohn 1992, S. 84 oder 167f.

titativ beeinflusst. Wichtiger als die generelle Bevorzugung oder Unterdrückung von Lexemen einer bestimmten Wortart scheint es mir daher zu sein, bei der Erstellung eines zentralen Wortschatzes die Wortarten in die Beschreibung der Eigenschaften der Lexeme, aus denen ausgewählt werden soll, aufzunehmen und erst in einem zweiten Schritt zur Selektion heranzuziehen, so dass man mit verschiedenen Graden der Bevorzugung experimentieren kann.

Nützlich kann es allerdings sein, existierende Wortschätze auf ihre Zusammensetzung hinsichtlich der Wortarten zu untersuchen, wie es etwa Schnörch und Krohn getan haben, und daraus etwas über die untersuchten Wortschätze zu lernen. Das Kriterium der Zugehörigkeit zu einer bestimmten Wortart ist also weniger ein Mittel der Selektion als ein Mittel der Analyse.

4.5.2.5 Stilistische Neutralität

Stilistische Neutralität als die Eigenschaft eines Wortes, sich „gegenüber territorialen, generationalen, sozialen oder professionalen Merkmalen, die den Sprecher / Schreiber charakterisieren, neutral“ (Kühn 1979, S. 27) zu verhalten, gilt neben der Stabilität als grundlegende Anforderung an die Wortschatzelemente, die in einen Grundwortschatz aufgenommen werden sollen.¹⁷⁶

Die Definition und Operationalisierung dieses Merkmals fällt dabei in der (Grundwortschatz-)Literatur unterschiedlich aus. So arbeitet beispielsweise Andreev mit einer thematisch orientierten Neutralität, indem er die Distribution eines Wortes über verschiedene Texttypen und Themen hinweg betrachtet – je größer das Spektrum an Texttypen und Themen ist, bei denen ein Wort verwendet wird, desto größer ist seine (thematische) Neutralität (vgl. Andreev 1980, S. 85). Ähnlich argumentiert Buchbinder, wenn er darauf hinweist, dass Neutralität häufig mit anderen Merkmalen der Zentralität (hohe Frequenz, ausgeprägter semantischer Wert) einhergeht.¹⁷⁷ Damit sei Neutralität möglicherweise überflüssig; aus praktischen Gründen sei sie aber wichtig, weil sie leichter zu fassen sei als der semantische Wert (vgl. Buchbinder 1973, S. 307).

Weniger an der unmittelbaren Operationalisierbarkeit orientiert, sondern an der grundlegenden theoretischen Unterscheidung interessiert ist Kühn. Er geht von stilistischer Neutralität aus und definiert sie darüber, ob ein Wort „symbolfunktionalen“ (Kühn 1979, S. 4), „signalfunktionalen“ oder „symptomfunktionalen“ Charakter hat, d.h. ob das Lexem vorwiegend ein Mittel der Referenz, ein Mittel der Illokutionsanzeige oder ein Element einer Sondersprache darstellt. Demnach sind symbolfunktionale Elemente des Wortschatzes,

176. Vgl. zum Beispiel Hendrickx 1987, S. 11. – Für eine übergeordnete Behandlung der stilistischen Markiertheit im Rahmen der Stilfeorschung vgl. Sowinski 2002.

177. Ein Beispiel für die Wechselwirkung einzelner Selektionskriterien, wie sie in Abschnitt 4.5.3 diskutiert ist.

also Lexeme, die vornehmlich eine referentielle, d.h. präzisierende Funktion haben, stilistisch neutral.¹⁷⁸

Dies unterstützt auch Krohn, präzisiert dabei aber auch, dass nicht-referentielle Bedeutungen nicht etwa nebensächlich wären, dass sie aber erst durch eine referentielle Bedeutung möglich werden:

„Damit wollen wir nicht einer Wiedereinführung von traditionellen Wortlisten das Wort reden, sondern nur nochmals darauf hinweisen, daß der pragmatische Gehalt einer Äußerung den regelhaft lexikalisch gebildeten propositionalen Gehalt der Äußerung voraussetzt oder präsupponiert. Mit anderen Worten sind situations- und themenunabhängige Kenntnisse von Wortbedeutungen als lexikalische Kompetenz im engeren Sinne eine unabdingbare Voraussetzung für den systematischen Aufbau und Ausbau der pragmatischen Aspekte einer lexikalischen Kompetenz im weiteren Sinne.“ (Krohn 1992, S. 26f.)

Ähnlich wie Buchbinder und Andreev argumentiert auch Schnörch mit der Distribution bzw. thematischen Breite des einer Auswahl zugrundezulegenden Textkorpus und einem strengen Frequenzfilter als Mitteln der Ermittlung stilistischer Neutralität (vgl. Schnörch 2002, S. 78).

Hinter dem Begriff der Neutralität verbergen sich also zwei Schattierungen von Neutralität, eine thematische Neutralität, die sich über die Distribution eines Wortes bestimmt, und eine stilistische Neutralität, die sich über das Fehlen einer stilistischen bzw. pragmatischen Markiertheit definiert. Beide sind eng miteinander verknüpft, etwa wenn es um die stilistische Markierung eines Wortes als ‘fachsprachlich’ geht.

Problematisch bei der gesamten Diskussion zur Frage der Neutralität ist die Tatsache, dass Neutralität keine absolute Größe ist: Ein Wort, das in einem standardsprachlichen Zusammenhang als fachsprachlich markiert ist, kann in einem fachsprachlichen Zusammenhang unmarkiert sein¹⁷⁹.

Ähnlich argumentiert auch Schnörch in Verbindung mit der Frage, ob ein neutraler Grundwortschatz gleichermaßen für die geschriebene Sprache wie für die gesprochene Sprache eingesetzt werden kann: Neutralität im Sinne der gleichmäßigen Verteilung eines Lexems über einen breit gefächerten Korpus hinweg garantiert keine gleichmäßige Anwendbarkeit des Lexems in Situationen schriftlicher gegenüber mündlicher Kommunikation (vgl. Schnörch 2002, S. 79).

Aber es geht nicht allein um die Frage, ob ein Lexem stilistisch oder neutral oder markiert ist. Um die Grundlage für eine differenzierte Beurteilung und Auswahl zu schaffen, kann es sich lohnen, verschiedene Arten von Markiertheit zu unterscheiden. Denn es ist denkbar, dass man etwa umgangssprachlich und als der Fachsprache der Medizin angehörig markier-

178. Insgesamt betrachtet Kühn referentielle Wortschatzelemente als die typischen Kandidaten für einen Grundwortschatz, weil sie eine zentrale Rolle im wichtigsten Sprechakt – der Prädikation – spielen (vgl. auch Kühn 1979, S. 71). – Für eine detaillierte Beschreibung der von Kühn vorgenommenen Dreiteilung vgl. Kühn 1979, S. 5-7.

179. So mag etwa die Bezeichnung *Karies* in nicht-fachlichen Zusammenhängen als Fachausdruck gelten, in einem zahnärztlichen Fachartikel aber also völlig unmarkiert erscheinen. – Vgl. auch Kühn 1979, S. 27.

te Lexeme sehr wohl in den zentralen Wortschatz aufnehmen möchte, nicht aber veraltete Lexeme oder Fachtermini des Maschinenbaus, weil etwa medizinische Fachbegriffe in der alltäglichen Kommunikation eher Verwendung finden als Fachbegriffe des Maschinenbaus. Daher kann es sich lohnen, bei der Beschreibung von Lexemen eine feinere Untergliederung vorzunehmen als nur die Zweiteilung von ‘markiert vs. unmarkiert’.

Wenn wir nun die Diskussion zur Forderung nach stilistischer Neutralität zusammenfassen, dann sind folgende Punkte wichtig:

Neutralität ist ein relativer Begriff. Je nach dem, welchen Sprachbereich man mit einem zentralen Wortschatz repräsentieren möchte, können verschiedene Lexeme neutral bzw. zentral für den aktuellen Sprachbereich sein, wie das zum Beispiel bei Kühn 1979, S. 27 diskutiert ist.

Darüber hinaus kann es nützlich sein, bei der Beschreibung der Markiertheit von Lexemen von einer feineren Unterscheidung als nur derjenigen von markiert vs. unmarkiert auszugehen. Wie fein diese Unterscheidung sein soll, hängt von dem zu erstellenden zentralen Wortschatz ab.

4.5.3 Zur Relativität und Wechselwirkung der Selektionskriterien

Unabhängig von der konkreten Frage, wie die verschiedenen in der Literatur angeführten Selektionskriterien in die Praxis umzusetzen und welche Selektionskriterien für die Erstellung eines konkreten zentralen Wortschatzes optimal sind, ist grundsätzlich noch nicht ganz klar, wie die verschiedenen Selektionskriterien zusammenwirken, ob und wie sie einander beeinflussen und wie sehr sie mit individuellen Informationen über Lexeme zum Selektionsprozess beitragen. Buchbinder (1973, S. 307) führt einige Beispiele für eine solche vermutete Wechselbeziehung an: Starke Polysemie sei oft mit einer ausgeprägten Fähigkeit zum Eingehen verschiedenster Kollokationen verbunden, während die Fähigkeit zur Bildung syntaktischer Strukturen, wie sie Funktionswörter besitzen, oft an eine hohe Frequenz bzw. Distribution gekoppelt seien. Umgekehrt gehe eine hohe Frequenz oft auch mit starker Polysemie und ausgeprägter stilistischer Neutralität einher. Dies sind bisher aber nur sehr allgemeine Vermutungen, für die ein differenzierter Nachweis noch erbracht werden muss.

Natürlich kann man in Anbetracht dieser vermuteten Wechselwirkungen und Überlappungen im Selektionseffekt verschiedener Selektionskriterien mit Recht fragen, was die Bestimmung so detaillierter Selektionskriterien überhaupt erbringt. Denn wenn zwei Selektionskriterien Indikatoren für die gleiche zugrundeliegende Eigenschaft sind und in die gleiche Richtung weisen, stellt sich die Frage, ob immer alle Selektionskriterien berücksichtigt werden müssen oder ob ein Selektionskriterium repräsentativ für ein anderes stehen kann. Dem lässt sich zum einen mit Buchbinder auf einer generellen Basis entgegenhalten, dass „die bewußte Zerlegung eines vielseitigen Ganzen [...] zu einem besseren Erkennen des Objekts“ (Buchbinder 1973, S. 307) führen kann. Zum anderen lässt sich die Frage nach

einer möglichen Reduzierung der Selektionskriterien erst beantworten, wenn die konkreten Zusammenhänge im einzelnen klarer sind.

In diesem Sinne möchte ich die Abhängigkeit und Überschneidung, die möglicherweise zwischen zwei Selektionskriterien besteht, als *Wechselwirkung* bezeichnen. Der Effekt, den ein Selektionskriterium auf die endgültige Selektion hat, soll unter dem Begriff des *Wirkungsgrades* des Selektionskriteriums zusammengefasst werden. Beide Aspekte sind in der bisherigen Forschung zum Bereich zentraler Wortschätze nicht genauer diskutiert worden. Sie sollen daher in der nachfolgenden praktischen Untersuchung näher beleuchtet werden (vgl. Abschnitt 9.2).

4.6 Das Problem des Umfangs eines zentralen Wortschatzes

Die Frage danach, ob es einen verbindlichen Umfang für einen bestimmten Typ eines zentralen Wortschatzes geben kann und wenn ja, wie hoch dieser anzusetzen sei, ist eine der grundlegenden Kontroversen in der Grundwortschatzliteratur. Der Grund für diese kontroverse Stellung ist zum einen die Tatsache, dass es keine klaren Ergebnisse zu Umfangsuntersuchungen des Wortschatzes im Allgemeinen gibt, eben weil die Berechnung des Umfangs des Wortschatzes, der ja ein offenes, sich ständig veränderndes System darstellt, nahezu unmöglich ist.

Die Frage nach dem Umfang eines zentralen Wortschatzes ist aber auch deshalb ein Problem, weil es bisher keine klaren Erkenntnisse darüber gibt, welchen Effekt, d.h. welchen Wirkungsgrad einzelne Selektionskriterien auf den Umfang haben. Es fehlen also Antworten auf Fragen wie ‘Wenn man besonders streng mit Selektionskriterium x umgeht, wie wirkt sich das dann auf den Umfang des zentralen Wortschatzes aus?’ etc.¹⁸⁰

Dabei ist die Frage danach, wie wichtig die Festsetzung eines konkreten Umfangs für einen zentralen Wortschatz ist, wie so vieles andere abhängig vom Typus und dem Verwendungszweck des zentralen Wortschatzes. Während die bewußte quantitative Begrenzung bei der Grundwortschatzlexikographie schon immer eine große Rolle gespielt hat, tritt sie beispielsweise bei der Erstellung eines definitiven Vokabulars gegenüber der Nützlichkeit der Zusammenstellung im Sinne des Bereitstellens ausreichend vieler definitiver Termini in den Hintergrund (vgl. Herbst 1986, S. 114).

Verschiedene Entscheidungen bei der Arbeit mit einem zentralen Wortschatz haben dabei einen Einfluss auf den Umfang. Zum einen spielt es eine Rolle, wie genau man mit der Monosemierung von Lexemen umgeht; je genauer die Monosemierung, desto größer wird der Umfang des zentralen Wortschatzes.¹⁸¹ Umgekehrt gilt dann entsprechend, dass die Monosemierung umso weniger detailliert vorgenommen werden kann, je strenger und kleiner man die Grenze eines zentralen Wortschatzes vorab ansetzt (vgl. Kühn 1979, S. 46). Ein

180. Für eine praktische Untersuchung dieser und ähnlicher Fragen vgl. Abschnitt 9.2.

181. Vgl. Krohn 1992, S. 97 und Abschnitt 4.4.1.

anderer Einflussfaktor ist die Zielgruppe (Muttersprachler vs. Fremdsprachenlernende) des zu erstellenden zentralen Wortschatzes, wenn es sich dabei um den Typus des Grundwortschatzes handelt, die Frage danach, wie groß der produktive bzw. rezeptive Wortschatz ist, den die Lernenden bereits mitbringen, und vor allem, wie groß die Synergieeffekte zwischen dem Wortschatz der Muttersprache und der zu lernenden Sprache in Form eines potentiellen Wortschatzes sind (vgl. Krohn 1992, S. 118). Aus einer rein lexikologischen Perspektive heraus spielt es auch eine Rolle, ob man nur einfache Wörter oder auch Komposita zulässt, ob man Ableitungen aufnimmt oder ihre Darstellung in die Grammatik verschiebt.¹⁸²

An dieser Stelle wollen wir uns in einer kurzen Bestandsaufnahme der Frage zuwenden, welche Schätzungen es für den Gesamtwortschatz des Deutschen gibt und mit welchen Größenordnungen für einen Grundwortschatz man bisher gearbeitet hat. Diese Bestandsaufnahme soll die Grundlage bilden für die im praktischen Teil dieser Arbeit vorzunehmende Umfangsanalyse des Wortschatzausschnittes, der dort untersucht wird (vgl. Abschnitt 9.3.2).

Es liegt in der Natur des Wortschatzes als offenem System, dass es für seinen Umfang nur sehr ungefähre und stark schwankende Schätzungen gibt. Wir wollen uns dabei an die bei Glück ((Hg.) 2000, S. 799f.) angegebenen halten. Er gibt den Gesamtwortschatz des Deutschen mit 300 000 bis 500 000 Wörtern an und schätzt den Bestand neuerer einsprachiger Wörterbücher auf 100 000 bis 200 000 Lemma. Den Umfang des Standardwortschatzes schätzt er auf ca. 75 000 Wörter, während er den produktiven Wortschatz eines einzelnen Sprechers des Deutschen auf rund 8 000 bis 10 000 Wörter schätzt.¹⁸³

Für die Grundwortschatzlexikographie hat, abgesehen von einigen wenigen Arbeiten, die mit einem Umfang von unter 1000 Einträgen arbeiten, lange Zeit eine Begrenzung auf 2000 Einheiten als adäquat gegolten, ohne dass für diesen spezifischen Wert eine objektiv nachweisbare Begründung geliefert wurde.¹⁸⁴ Eine Rolle spielte sicher die Beobachtung, dass Grundwortschätze, die nur 500 Einträge haben, zumal solche, die vor allem auf frequentuellen Selektionskriterien beruhen, vor allem Funktionswörter abdecken und daher didaktisch uninteressant sind (vgl. Erk 1963, S. 87). Der „magische[n] 2000-Wort-Grenze“ (vgl. Kühn 1979, S. 46) steht die Schätzung Hendrickx' (1987, S. 2) gegenüber, der Fragen der Mitaufnahme von reihenbildenden Wortbildungen bei seinen Überlegungen mitberücksichtigt und mit zwei Größenordnungen operiert:

182. Vgl. Hendricx 1987, S. 2 und Krohn 1992, S. 118.

183. Für entsprechende Zahlen zum englischen Wortschatz bzw. zur quantitativen Entwicklung des kindlichen Wortschatzes vgl. Carter 1998, S. 236 bzw. Clark 1993, S. 13f. – Für eine grundlegende Diskussion der Frage nach dem Umfang des Wortschatzes eines Sprechers des Deutschen vgl. auch Zimmer 1990.

184. Mit einem Umfang von weniger als 1000 Einheiten arbeiten etwa Ogden 1930, dessen „Basic English“ auf 850 Wörter begrenzt ist, oder Augst (1987, S. 119f.), der für einen Rechtschreibgrundwortschatz 400 Funktionswörter und 200 „Inhaltswörter“ ansetzt. Dabei mag auch die Glottochronologie, die mit kleinen Listen von 200 bzw. 100 Einträgen (vgl. Wang 1994, S. 1445) arbeitet, ebenfalls die Diskussion beeinflussen haben.

„[Or] the basic vocabulary contains about 3.000 lemmata (in which case the morphological rules must be included as well), or it has about 20.000 words. It all depends on one's starting point: if one chooses the native speakers as the target group, the basic vocabulary will be more extended than in the case it is meant for non-native speakers.“ (Hendrickx 1987, S. 2)

Es wird also deutlich, dass es für den Umfang eines zentralen Wortschatzes keine absoluten Werte gibt; das ist auch nicht verwunderlich, denn auch der Umfang eines zentralen Wortschatzes ist, wie andere Aspekte der Arbeit mit zentralen Wortschätzen auch, abhängig von der praktischen Anwendung, für die der Wortschatz erstellt wird. Abschließend wollen wir daher nur grundlegende Tendenzen festhalten:

Es ist offensichtlich, dass ein zentraler Wortschatz gleich welchen Typs in jedem Fall kleiner sein muss als die Gesamtmenge, d.h. der Ausgangswortschatz, dem er entnommen ist.

Zentrale Wortschätze des didaktischen Typs, die sich ja an Sprecher oder Deutschlernende wenden, müssen entsprechend kleiner sein als der angenommene Wortschatz eines Sprechers, also weniger als 10 000 Wörter. Zentralen Wortschätzen des analytischen Typs hingegen liegt theoretisch der Gesamtwortschatz des Deutschen zugrunde, ihr Umfang sollte also unter 75 000 Wörtern liegen.

Bei all diesen Angaben ist wie gesagt zu bedenken, dass diese Angaben nur als sehr grobe Richtlinien zu betrachten sind, gerade weil es so schwierig ist, Wortschatz zu zählen.¹⁸⁵

4.7 Fragen der Darstellung und Aufbereitung des zentralen Wortschatzes

Neben dem Problem der Selektion und den damit verknüpften Fragestellungen ist die Frage nach der richtigen Darstellung und Kodierung des erstellten zentralen Wortschatzes eines der wesentlichen Themen in der (Grundwortschatz-)Diskussion. Dies hat seinen Grund sicherlich zum einen darin, dass diese Frage bei jedem lexikographischen Projekt eine Rolle spielt. Zum anderen hat diese Frage besonders in didaktisch ausgerichteten lexikographischen Projekten eine wichtige Funktion, weil die Art und Weise der Darstellung, d.h. die Anordnung der Wortschatzeinträge und die Gliederung der zu jedem Eintrag zu gebenden Informationen einen wichtigen Einfluss auf die Vermittlung des Wortschatzes hat. Denn sie kann, wenn sie sorgfältig konzipiert ist, beispielsweise „Bedeutungen miterklär[en] und Wortbildungen [...] in einen strukturell-semanticen Kontext“ (Schnörch 2002, S. 372) stellen.

In der Grundwortschatzdiskussion sind dabei Fragen der Darstellung lange Zeit vernachlässigt worden. Erst Kühn 1979 nahm diesen Fragenkomplex explizit in die Reihe der zu lösenden Probleme auf; bis dahin waren Grundwortschätze oft einfache Listen von Wörtern ohne größere zusätzliche Erläuterungen oder gar Bedeutungsangaben, die keine systematische Wortschatzarbeit zuließen (vgl. Schnörch 2002, S. 23).

185. Vgl. die Diskussion in Abschnitt 4.2.

Dabei ist die Antwort auf die Frage nach der Art und Weise der Darstellung – genauso wie die Selektion – stark abhängig vom Typus des anvisierten Wortschatzes (vgl. Kapitel 3). Für die didaktischen Typen stehen vor allem Fragen der semasiologischen vs. onomasiologischen Darstellung und der Organisation der Bedeutungsangaben im Vordergrund, während für analytische Typen zentraler Wortschatze Fragen der Darstellung vor allem um Aspekte des schnellen und eindeutigen Zugriffs auf einzelne Wortschatzeinheiten kreisen. Sowohl die Probleme der lexikographischen Darstellung als auch der datentechnischen Aufbereitung sind im Prinzip allgemeine Probleme der Lexikographie bzw. maschinellen Sprachverarbeitung und damit hinlänglich bekannt; wir wollen uns daher im Folgenden nur auf eine kurze Zusammenfassung der Diskussionsergebnisse im Rahmen der Grundwortschatzlexikographie konzentrieren.

Wie bereits angesprochen, dreht sich bei didaktisch orientierten Wortschatztypen die Diskussion zu Fragen der Darstellung vor allem um zwei Problemkomplexe: das Problem der makrostrukturellen Organisation des Wortschatzes und das Problem der mikrostrukturellen Informationen für jeden Wortschatzeintrag.

Auf der makrostrukturellen Ebene geht es dabei grundsätzlich um die Entscheidung zwischen einer semasiologischen (also bezeichnungsorientierten, d.h. in der Regel alphabetisch angeordneten) und einer onomasiologischen (also begriffsorientierten) Darstellung.

Typischerweise ist dabei die semasiologische Darstellung mit einem Verlust von Informationen über Wortbildungszusammenhänge verbunden, weil Wörter mit dem gleichen Stamm, aber unterschiedlichen Anfangsbuchstaben an verschiedenen Stellen in der Darstellung auftauchen und nicht ohne weiteres als zusammengehörig erkannt werden können. Auch semantische Zusammenhänge können verloren gehen, weil Wörter, die Begriffe bezeichnen, die in bestimmten semantischen Beziehungen zueinander stehen (wie etwa Synonyme, Hypo- und Hyperonyme, Antonymen etc.) nicht zusammen dargestellt werden.

Umgekehrt bringt eine rein onomasiologische Darstellung den gesamten Komplex ungeklärter Fragen rund um das Konzept *Begriff* mit sich (vgl. Abschnitt 4.2). Über das für eine onomasiologische Darstellung nötige System von Begriffen besteht bisher keine Einigkeit und dürfte auch in Zukunft kaum Einigkeit erzielt werden. Die bisherigen Begriffssysteme sind unmittelbar von der (subjektiven) Ordnung der Verfasser abhängig.¹⁸⁶ Zum anderen ist längst nicht klar, was eigentlich ein Begriff ist, insbesondere die Abgrenzung von *Begriff* gegenüber *Wort* und *Begriffsgruppe* und *Thema* ist problematisch, weil das Wort als sprachliches Zeichen, mit dem man einem Begriff einen Namen gibt, zugleich die sprachliche Form ist, mit der der Begriff realisiert ist.¹⁸⁷ Auch der Vergleich verschiedener Sprachen zur ‘Destillierung’ einer Liste von Begriffen, wie sie etwa in der Glottochronologie versucht wurde, führt nicht immer zu einer Liste von durch Wörter repräsentierten Begriff-

186. Vgl. etwa Kühn 1979, S. 116, Kloster 1997, S. 75ff. und die umfassende Darstellung bei Reichmann 1990.

187. Vgl. zum Beispiel Kühn 1979, S. 116 und 130.

fen, denn die Art und Weise, wie ein Begriff sprachlich realisiert ist, ist arbiträr, weshalb es keine Aussagekraft hat, „ob ein Begriff oder Prädikat auf der Wort-, Satz- oder Textebene einer bestimmten natürlichen Sprache dargestellt wird.“ (Kühn 1979, S. 123).¹⁸⁸ Der Vorteil einer onomasiologischen Darstellung wiederum liegt in der Transparenz der Darstellung. Dies gilt sowohl für den Benutzer, der Wortbildungszusammenhänge und semantische Relationen besser erkennen kann, als auch für den Lexikographen, dem es durch die onomasiologische Darstellung möglich wird, lexikalische Über- bzw. Unterrepräsentationen bestimmter Wortfelder aufzudecken (vgl. Krohn 1992, S. 88).¹⁸⁹

Auf der mikrostrukturellen Ebene geht es in der Grundwortschatzdiskussion vor allem um die Frage, welche Bedeutungen zu einem Wortschatzeintrag angegeben werden sollen und wie man sie anordnet. Dabei spielt vor allem die Frage eine Rolle, ob und wie sich feststellen ließe, welche Bedeutung eines polysemen Lexems die Hauptbedeutung ist und damit an erster Stelle genannt werden sollte.¹⁹⁰ Dabei ist die mikrostrukturelle Darstellung stark abhängig von vorab getroffenen Entscheidungen zur Grundeinheit, denn je nach dem, was man als kleinste Einheit betrachtet, ob Lesarten oder Lexeme, werden sich die Fragen der mikrostrukturellen Darstellung unterschiedlich darstellen.

Sowohl Kühn 1979 als auch Krohn 1992 und Schnörch 2002 diskutieren diese beiden Fragenkomplexe ausführlich.¹⁹¹ Alle drei kommen zu dem Schluss, dass für die Darstellung des Grundwortschatzes, also des didaktisch orientierten Typs eines zentralen Wortschatzes, eine Grundausrichtung an der onomasiologischen Darstellung ergänzt durch eine semasiologische ideal wäre, trotz der mit einer onomasiologischen Darstellung verbundenen Probleme.¹⁹² Man kann ergänzend hinzufügen, dass mit den modernen Mitteln der elektronischen Darstellung von Informationen, etwa der Verknüpfung von Informationen durch *hyperlinking* eine solche integrierende Darstellung rein praktisch erleichtert wird, weil so Informationen sowohl alphabetisch als auch aufgrund einer semantischen oder morphologischen Organisation dargestellt werden können.

188. Für eine Zusammenfassung dieser Probleme vgl. etwa Buchbinder 1973, S. 304, Kühn 1979, S. 115-120 oder Schnörch 2002, S. 152-156.

189. Auch die Frage danach, ob ein Wortschatz vor allem produktiv oder hauptsächlich rezeptiv gebraucht wird, spielt bei der Abwägung einer semasiologischen gegenüber einer onomasiologischen Darstellung eine Rolle. Geht es dem Wörterbuchbenutzer um das Produzieren eines Textes, ist ihm mit einer onomasiologischen Darstellung eher gedient; geht es ihm um das Verstehen eines vorhandenen Textes, ist ihm eine semasiologische Darstellung leichter zugänglich (vgl. Herberg 1992, S. 142f.).

190. Vgl. auch die Diskussionen in Abschnitt 4.4.1.

191. Vgl. Kühn 1979, S. 76f. und 98ff., Krohn 1992, S. 84-91 und Schnörch 2002, S. 336-347. – Ein weiteres Beispiel der Verwendung onomasiologischer Darstellungsprinzipien ist die Frequenzerhebung des Wortschatzes von Kindern bei Augst 1985. Für eine kritische Besprechung dieses Verfahrens vgl. Schnörch 2002, S. 152-156 und S. 371f.

192. Vgl. Kühn 1979, S. 76 und Schnörch 2002, S. 339. Ähnlich auch Buchbinder 1973, S. 304 – Für ein Beispiel einer solchen Kombination vgl. Schnörch 2002, S. 307ff.

4.8 Die Arbeit mit zentralen Wortschätzen im Schnittpunkt mehrerer linguistischer Disziplinen

Abschließend wollen wir uns noch einmal einem Problem zuwenden, das sich aus der hier beschriebenen Sammlung von Einzelproblemen ergibt. Denn die Liste der weiter oben beschriebenen spezifischen Fragestellungen zeigt eines ganz deutlich – die Arbeit mit zentralen Wortschätzen bewegt sich im Grenzland verschiedenster linguistischer Disziplinen: Aspekte der Semantik, der Lexikologie, der Morphologie, der Lexikographie, der Sprachstatistik und – im Falle des Grundwortschatzes – der Sprachdidaktik müssen berücksichtigt werden. Zu den einzelnen Fragestellungen, die es zu lösen gilt, kommt also noch die Notwendigkeit, detaillierte Kenntnisse aus diesen verschiedenen Teilbereichen gleichzeitig zu berücksichtigen und sinnvoll miteinander zu kombinieren.

Andererseits kann man diese Tatsache umgekehrt auch als Stärke des Konzeptes betrachten – gerade weil das Konzept zentraler Wortschätze unabhängige Erkenntnisse aus verschiedenen linguistischen Teilbereichen zusammenführt, kann die Beschäftigung damit zu einer gegenseitigen Befruchtung dieser Teilbereiche führen.

Deshalb wollen wir uns am Ende der Diskussion der mit der Arbeit an zentralen Wortschätzen verbundenen Probleme Schnörch anschließen, der eben diesen positiven Aspekt der Beschäftigung mit diesem Konzept herausstellt:

„Nimmt man die Sprachwissenschaft zum Ausgangspunkt, so befassen sich Semantik und Lexikologie ganz allgemein mit 'dem Wortschatz'. In diesen Bereichen gelangt man zu eigenständigen Ergebnissen, und diese lassen sich in begrenzterem Umfang speziell auch für die Grundwortschatzforschung nutzbar machen. Deren Untersuchungen können wiederum unmittelbar einen Beitrag leisten für die Fremdsprachendidaktik, und diese kann wiederum Rückmeldungen geben, beispielsweise an die Lexikografie.“ (Schnörch 2002, S. 38)

Teil B: Theoretisch-methodische Weiterführung

Überblick

In der vorangegangenen Bestandsaufnahme haben wir gesehen, welche Anwendungstypen zentraler Wortschätze es gibt und welche Probleme mit ihrer Erstellung verbunden sind. Dabei ist deutlich geworden, dass diese Probleme an eine Vielzahl komplexer Fragestellungen der allgemeinen Linguistik anknüpfen, zu denen es keine einfache Lösung gibt. Im vorliegenden Teil der Arbeit wollen wir uns daher auf einige ausgewählte Problemstellungen konzentrieren.

Wir wollen zum einen die in den Abschnitten 4.2 und 4.5 aufgeworfenen Fragestellungen zur Grundeinheit eines Wortschatzes und zur Zentralität von Wörtern vertiefen. Dazu werden wir der Frage nachgehen, wie die Begriffe *Lexikon* und *Wortschatz* definiert sind, welche Elemente zum Wortschatz zu rechnen sind, wie sich Wortschatz schichten lässt und ob es so etwas wie ein absolutes Wortschatzzentrum gibt (Kapitel 5).

Zum anderen wollen wir versuchen, das Problem der bisher sehr heterogenen Terminologie zum Thema zentraler Wortschätze ein Stück weit zu lösen, indem wir die bisherigen Überlegungen zur Modellierung von Wortschatz zu einem Arbeitsmodell zusammenführen, das Konzept zentraler Wortschätze definitorisch festlegen und einige terminologische Klärungen vornehmen (Kapitel 6).

Dieser Teil soll damit helfen, die in Abschnitt 1.1 erwähnte theoretisch-methodische Lücke zu schließen und ein Bezugsrahmen dafür zu schaffen, wie man über Wortschatz sprechen und wie man ihn modellieren kann.

Kapitel 5: Faktoren der Wortschatzschichtung

5.1 Einführung

Das hier folgende Kapitel beschäftigt sich mit einigen zentralen Aspekten der Makrostrukturierung des Wortschatzes, genauer: Beschreibungskategorien, die es erlauben, den Wortschatz einer Sprache auf einem der Ebene der individuellen Wörter übergeordneten Niveau zu strukturieren. Ausgangspunkt sind dabei zwei große Fragen: Worüber reden wir eigentlich, wenn wir von Wortschatz sprechen? Und wie lässt sich Wortschatz schichten? In einem Versuch der Annäherung an diese Fragen wollen wir uns unter anderem mit dem Begriff des Lexikons beschäftigen, die Abgrenzung von Gemeinsprache und Sprachvarietäten am Beispiel der Fachsprachen untersuchen und wesentliche Aspekte der bisherigen Diskussion zur makrostrukturellen Wortschatzschichtung zusammenfassen.

Eine solch eingehende Beschäftigung mit lexikologischen Aspekten der Wortschatzschichtung ist wichtig, weil die damit verknüpften Beschreibungskategorien und Abgrenzungsschwierigkeiten bei der Arbeit mit zentralen Wortschätzen eine Rolle spielen. Dabei ist es nicht so, dass Fragen nach diesen makrostrukturellen Dimensionen des Wortschatzes in früheren Arbeiten zum Thema zentraler Wortschätze nicht auftauchten; sie werden durchaus etwa in der Grundwortschatzliteratur angesprochen, allerdings nicht explizit und systematisch dargestellt. Das ist nachvollziehbar, weil die Grundwortschatzlexikographie zunächst ja aus einem praktischen Bedürfnis heraus entstand, das Augenmerk also vornehmlich auf die Lösung praktischer Probleme gerichtet war. Mit der Weiterentwicklung des Konzeptes zentraler Wortschätze etwa bei Kühn 1979 und Schnörch 2002, die grundsätzliche lexikologische Fragen ins Zentrum der Diskussionen rücken, dabei aber Fragen der Makrostruktur des Wortschatzes nur in Teilen behandeln, ist es jedoch an der Zeit, die bisherigen praktischen lexikographischen Versuche mit Erkenntnissen aus der Lexikologie zu verknüpfen, um ihnen so eine theoretische Basis zu geben.

5.2 Der Begriff des Lexikons

5.2.1 Der Terminus *Lexikon* und seine Verwendungsweisen

Zunächst wollen wir uns der Frage zuwenden, was wir betrachten, wenn wir von der Strukturierung des Wortschatzes sprechen, wie also etwa die Termini *Wortschatz* und *Lexikon* zusammenhängen, was sie bedeuten. Zu diesem Zweck wollen wir uns zunächst dem Terminus *Lexikon* zuwenden. Er spielt in verschiedenen Teildisziplinen der Linguistik eine wichtige Rolle, wird aber oft entsprechend vieldeutig, d.h. ohne explizite Beschreibung dessen, worauf man sich mit dem Terminus bezieht, verwendet. Es erscheint daher nötig, den Begriff etwas ausführlicher zu beschreiben, als das bisher in der Literatur zu zentralen Wortschätzen geschehen ist, und ihn zumindest für den Rahmen dieser Arbeit explizit festzulegen.

Traditionelle Definitionen des Terminus *Lexikon* betrachten ihn vor allem im Zusammenhang mit Fragen der Lexikologie und Lexikographie, später kommt noch der Bereich der kognitiven Linguistik hinzu.¹⁹³ Wir wollen uns hier die Definition und Abgrenzung bei Schindler 2002 näher ansehen, weil sie auch den Bereich der maschinellen Sprachverarbeitung miteinbezieht und damit alle Aspekte umfasst, die in dieser Arbeit eine Rolle spielen. Schindler (2002, S. 35) grenzt vier Bedeutungen des Begriffs *Lexikon* ab und gibt ihnen spezifische Namen:

Als „Lexikon“ definiert Schindler die „Speicherkomponente“ einer Sprache, „die mit der Regelelemente (Grammatik bzw. Syntax) interagiert“. Er schließt sich damit dem traditionellen Verständnis von *Lexikon* in der Lexikologie und allgemeinen Sprachtheorie an, nach dem das Lexikon als offenes Teilsystem der Sprache dem geschlossenen Teilsystem der Grammatik gegenübergestellt und oft vereinfachend als *Wortschatz* bezeichnet wird (vgl. Glück (Hg.) 2000, S. 411). Etwas anders als Schindler, aber zum Zwecke einer deutlicheren Terminologie, bei der wir *Lexikon* als Oberbegriff verstehen wollen, wollen wir diese Bedeutung des Begriffs *Lexikon* als *Wortschatz* bezeichnen.¹⁹⁴

Als „mentales Lexikon“ (Schindler 2002, S. 35) bezeichnet Schindler Spracheinheiten im Sinne „mentale[r] Einheiten“, die er „als vielfältig miteinander vernetzte Elemente des Langzeitgedächtnisses“ beschreibt und auf die man sich in der kognitiven Linguistik bezieht. Wir wollen diese Definition noch um eine am Spracherwerb ausgerichtete Definition ergänzen und konkretisieren, bei der das mentale Lexikon definiert wird als...

„...stock of established words speakers can draw on when they speak and have recourse to in understanding what they hear. This stock is stored in memory in such a way that speakers can locate the relevant units to use in both speaking and understanding. To do this, of course, speakers have to be able to identify words either by looking them up in memory (for comprehension) or by retrieving them as appropriate forms for conveying specific meanings (for production).“ (Clark 1993, S. 2)

Für die Lexikographie definiert Schindler den Terminus *Lexikon* als die „Menge der Grundelemente einer natürlichen Sprache, die nach zweckgerichteten und praktischen Gesichtspunkten [...] geordnet ist“ (Schindler 2002, S. 35); er bezeichnet diese Bedeutungsvariante des Terminus *Lexikon* als „Wörterbuch“, einer Bezeichnung, der wir uns anschließen wollen.

Für die maschinelle Sprachverarbeitung schließlich definiert Schindler den Terminus *Lexikon* als „Komponente eines Systems maschineller Sprachverarbeitung“ und spricht dabei präzisierend vom „maschinelle[n] Lexikon“. ¹⁹⁵

193. Für die traditionelle Definition vgl. etwa Bußmann 1990, S. 455. Für die erweiterte Definition vgl. etwa Glück (Hg.) 2000, S. 411.

194. Vgl. auch die terminologischen Klärungen in Verbindung mit der Entwicklung eines Arbeitsmodells in Kapitel 6.

Wir wollen uns dieser Vierteilung grundsätzlich anschließen, jedoch auch einige kleine definitorische und terminologische Präzisierungen vornehmen, die für den Rahmen dieser Diskussion nötig sind:

Zunächst verwenden wir den Terminus *Lexikon* als Oberbegriff für die verschiedenen hier beschriebenen Typen eines Lexikons. Darüber hinaus sprechen wir von *Wortschatz*, wenn wir das Lexikon im lexikologischen Sinne, also die „Speicherkomponente“ (Schindler 2002, S. 35) gegenüber der „Regelkomponente“ und damit die Gesamtheit aller etablierten und potentiellen Wörter einer Sprache meinen. Wir sprechen von *Wörterbuch*, wenn wir die lexikographische Erfassung des Wortschatzes oder eines Wortschatzausschnittes meinen. Wir sprechen von *mentalem Lexikon*, wenn wir die im Langzeitgedächtnis eines Sprechers gespeicherten Spracheinheiten, sein internes Lexikon meinen, und meinen damit insbesondere die Konzentration auf Fragen des Aufbaus und Zugang zu diesem Lexikon, wie sie etwa in der Spracherwerbsforschung im Zentrum stehen. Und schließlich sprechen wir vom *maschinellen Lexikon*, wenn wir den Bestand an Wortinformationen meinen, mit denen ein Computer gespeist wird, um bestimmte maschinelle Verarbeitungsprozesse natürlicher Sprache auszuführen.¹⁹⁶

Mit Bonan-Garrigues wollen wir dabei das mentale Lexikon eines Sprechers, also die kognitive Bedeutung des Terminus *Lexikon*, als Grundlage aller anderen Arten von Lexika auffassen: „L'élaboration du lexique commun à un groupe donné, passe par l'exploration du lexique mental d'un individu.“ (Bonan-Garrigues 1993, S. 7). Sowohl das Lexikon im Sinne eines Wortschatzes einer Sprache als auch das Lexikon im Sinne eines Wörterbuchs, das die etablierten Einheiten dieses Wortschatzes lexikographisch erfasst, und das maschinelle Lexikon, das diese Wortschatzeinheiten für maschinelle Sprachverarbeitungsprozesse bereitstellt, bauen auf diesem mentalen Lexikon auf.¹⁹⁷

5.2.2 Die Abgrenzung von Wortschatz und Grammatik

Wie wir bereits in Abschnitt 4.2 gesehen haben, spielt die Frage der Abgrenzung von Einheiten des Wortschatzes gegenüber den Einheiten der Grammatik bei der Arbeit mit zentralen Wortschatzen eine wesentliche Rolle, weil sie die anzusetzende Grundeinheit und damit sowohl das Selektionsverfahren, den Umfang und die Darstellung eines zentralen Wortschatzes beeinflusst. Wir wollen daher im Folgenden einen genaueren Blick darauf werfen, wie eine solche Grenze zwischen der „Speicherkomponente“ (Schindler 2002, S. 41) und

195. Gegenüber der von Schindler angegebenen englischen Entsprechung – „machine-readable lexicon“ (Schindler 2002, S. 35) wäre allerdings der Terminus *machine-tractable lexicon* vorzuziehen. Damit ließe sich verdeutlichen, dass ein Lexikon, das maschinell lesbar ist, noch nicht unmittelbar zur maschinellen Sprachverarbeitung herangezogen werden kann, sondern erst in ein maschinell bearbeitbares Lexikon umgewandelt werden muss. Dieser terminologische Unterschied spielt im Rahmen der Computerlinguistik eine gewisse Rolle (vgl. zum Beispiel. Bruce / Guthrie / Wilks 1993, S. 251. Ähnlich auch Mehl 1996.) – Für eine Übersicht über die Rolle von maschinellen Lexika in der maschinellen Sprachverarbeitung vgl. Heyer 1993.

der „Regelkomponente“ gezogen werden kann, was also den Wortschatz von der Grammatik unterscheidet.

Die traditionelle Hauptunterscheidung zwischen Wortschatz und Grammatik liegt darin, dass man den Wortschatz als offenes System betrachtet, das im Prinzip quantitativ unbegrenzt ist, während die Grammatik ein in sich geschlossenes System endlicher Regeln darstellt.¹⁹⁸ Eine Ursache für die Offenheit des lexikalischen Systems wird dabei in der Notwendigkeit gesehen, die Sprache immer neuen Phänomenen der zu beschreibenden Welt anzupassen (vgl. Labov 1976, S. 224).

Schindler 2002 differenziert diese grundlegende Unterscheidung und stellt im Wesentlichen zwei verschiedene Modelle der Abgrenzung vor:

Traditionell, also etwa in der generativen Grammatik, betrachtet man den Wortschatz als eine einfache Liste von lexikalischen Einheiten, aus denen einzelne Einheiten zur Bildung einer Äußerung ausgewählt werden. Grammatische und lexikalische Elemente werden dabei unterschieden danach

- worauf sie sich beziehen: Grammatische Elemente operieren nur über Symbolen von Kategorien, d.h. Platzhaltern für lexikalische Elemente, während lexikalische Elemente selbst kategorisierte Elemente sind und die Kategorien füllen
- wie produktiv sie sind: Grammatische Elemente sind uneingeschränkt produktiv, lexikalische Elemente sind nicht produktiv
- inwieweit sie Irregularitäten aufweisen: Grammatische Elemente sind frei von Irregularitäten, für lexikalische Elemente ist es geradezu typisch und konstituierend, dass sie irreguläre Eigenschaften haben bzw. phraseologisch sind.

196. Einen übersichtlichen Vergleich des mentalen Lexikons gegenüber dem maschinellen gibt Handke (1995, S. 312f.). Er sieht die wesentlichen Stärken des mentalen Lexikons gegenüber dem maschinellen Lexikon in den folgenden Punkten:

- in der Fähigkeit, mehrere Informationen parallel zu verarbeiten und ständig miteinander zu verknüpfen
- in der Fähigkeit, semantische Informationen zu verarbeiten und zu größeren semantischen Netzen zusammenzusetzen
- in der Fähigkeit, die im Lexikon gespeicherten Informationen zu verwalten und ständig zu aktualisieren
- schließlich in der Fähigkeit, sich eine Vielzahl von Informationen in flexibler Weise zu Nutzen zu machen

Einem weiteren von Handke aufgeführten Unterschied muss allerdings widersprochen werden: Handke geht von der Silbe als Grundbaustein und grundlegende Zugangseinheit des mentalen Lexikons aus, während maschinenlesbare Lexika höchstens Silbenschemata („cohorts“) verzeichnen könnten. Wie wir in Abschnitt 4.2 gesehen haben und in 5.2.2 noch sehen werden, kann als kleinster Grundbaustein des Lexikons allenfalls das Morphem betrachtet werden.

197. Für einen Überblick über die verschiedenen Aspekte und Erklärungsmodelle zur Organisation des mentalen Lexikons, zu Mechanismen des Zugangs zum gespeicherten lexikalischen Wissen und zum Vergleich von rezeptivem und produktivem Wortschatz vgl. Schindler 2002, S. 41-43.

198. Vgl. etwa Glück (Hg.) 2000, S. 411 oder, für die Grundwortschatzlexikographie, Kühn 1979, S. 19 und 22.

Von dieser strikten Trennung geht man in der kognitiven Linguistik ab. Dort fasst man Wortschatz und Grammatik als zwei Endpunkte eines Kontinuums auf, wobei typische Elemente des Lexikons als „specified symbolic units“ (ebda., S. 40) und typische Elemente der Grammatik als „schematic symbolic units“ definiert sind. Ausgangspunkt für dieses Modell sind Beobachtungen wie die, dass es Einheiten des Wortschatzes gibt, die syntaktische Strukturen aufweisen, wie etwa bestimmte Phraseologismen (*fix und fertig*), dass es umgekehrt aber auch syntaktische Strukturen gibt, die aus festen lexikalischen Einheiten und kategorialen Leerstellen bestehen (*Wenn X wüsst-, dass/w- Y – Wenn sie wüsste, dass er raucht / wie er sein Geld verdient*) (vgl. Schindler 2002, S. 40).

In der maschinellen Sprachverarbeitung wiederum hat es sich immer wieder als nötig herausgestellt, die Trennung zwischen Wortschatz und Grammatik soweit wie möglich aufzuheben. Versuche in dieser Richtung sind zum Beispiel die Lexikongrammatik von Maurice Gross, auf die sich auch Bonan-Garrigues bezieht:

„Composé de dictionnaires électroniques et de tables syntaxiques qui leur sont corrélées, le lexique-grammaire est associé à un certain nombre d'outils logiciels [...] permettant de les manipuler.“ (Bonan-Garrigues 1993, S. 4)¹⁹⁹

Generell spielt in der Lexikographie die Abgrenzung von Grammatik und Wortschatz insofern eine Rolle, als man bei der Erarbeitung eines Wörterbuchs immer der Notwendigkeit der quantitativen Begrenzung und damit dem Zwang zur Auswahl unterliegt. Hier kann die Abgrenzung von Grammatik vs. Lexikon der „Lemmaentlastung“ (Bergenholtz 1994, S. 300) dienen.²⁰⁰ Entsprechend diskutiert Carter 1998 das Problem der Abgrenzung von Wortschatz und Grammatik aus der Sicht der Sprachdidaktik. Hier spielen vor allem Fragen der adäquaten, d.h. benutzerfreundlichen und zugleich effektiven Darstellung eine Rolle, also etwa zum Beispiel Fragen der Anordnung von Derivativen (bei ihrem Stamm oder als separate Einträge?), von Komposita (beim Grund- oder beim Bestimmungsglied?) und von syntaktischen Informationen. Carter kommt dabei zu dem Schluss:

„[It] is difficult to draw a line between sufficiently detailed exemplificatory information and brevity and economy in the entry; it is often the simplest and most common words which contract the most complex syntactic and collocational partnerships. To these can be added the observation that the balance between accessing and interpreting related lexical items is an extremely delicate one within pedagogical lexicography.“ (Carter 1998, S. 158)

5.2.3 Die Heterogenität der Einheiten des Wortschatzes

Schwierigkeiten bereitet der Begriff des Wortschatzes aber nicht nur in seiner Abgrenzung zur Grammatik, sondern auch dadurch, dass der Wortschatz ein System mit untereinander inhomogenen Elementen darstellt.

Schon allein die Frage, was als Morphem, Wort oder Phraseologismus zu rechnen ist und in den Wortschatz aufgenommen werden soll, ist nicht pauschal und konsistent zu beant-

199. Vgl. Gross 1994 und auch Abschnitt 2.2.3.

200. Vgl. auch Abschnitt 4.2.

worten. Es ergeben sich vielerlei Abgrenzungsprobleme, sowohl formaler als auch inhaltlicher Art.²⁰¹ Aber, so wird bei Schindler (2002, S. 37) deutlich, auch die sprachlichen Einheiten, die sich eindeutig dem Lexikon zuordnen lassen, also etwa so 'typische' Wörter wie autosemantische Substantive, Adjektive und Verben, bereiten Schwierigkeiten, wenn man den Wortschatz global betrachten und strukturieren will. Denn nicht alle Eigenschaften, mit denen sich Wörter beschreiben lassen, lassen sich auch bei allen Wörtern beobachten. So lassen sich manche Wörter etwa hinsichtlich bestimmter Modalitäten beschreiben (*klettern* als ein motorisches Verb), andere sind hinsichtlich einer Modalität gar nicht gekennzeichnet (*Zufall*, *verwalten*), sondern haben abstrakten Charakter; manche Lexeme lassen sich einer Darstellungsfunktion zuordnen (*Haus*), andere Lexeme lassen sich hinsichtlich einer Darstellungsfunktion gar nicht beschreiben, sondern haben etwa vornehmlich Appellfunktion (*potzblitz*); Inhaltswörter werden kognitiv anders verarbeitet als Funktionswörter etc. Die Gruppe derjenigen Eigenschaften, nach denen sich alle Einheiten des Wortschatzes beschreiben lassen, ist also eher begrenzt.

Umgekehrt lassen sich aber durchaus Beziehungen zwischen Wörtern und semantische Ordnungsmuster erkennen. So etwa lassen sich Wörter danach gruppieren, ob sie in einer synonymischen oder antonymischen Beziehung zueinander stehen oder ob sie sich in eine Hyponym-Hyperonym-Hierarchie einordnen lassen, ob sie Wortfelder bilden oder zu Wortfamilien zusammengefasst werden können.²⁰² Da diese Ordnungsprinzipien aber nur eine beschränkte Gruppierung der Einheiten des Wortschatzes zulassen und sozusagen nur die Beschreibung und Abgrenzung kleiner 'Inseln' innerhalb der Gesamtmenge des Wortschatzes erlauben, nicht aber die Schichtung des Wortschatzes in seiner Gesamtheit, wollen wir sie als Beschreibungsmuster unterhalb der makrostrukturellen Ebene bezeichnen. Sie spielen bei der Arbeit mit zentralen Wortschatzen durchaus eine wichtige Rolle, weil sie unmittelbar an Probleme der lexikographischen Praxis anknüpfen.²⁰³ Sie sind aber wegen ihrer begrenzten Reichweite keine Hilfe, wenn es darum geht, herauszufinden, wie sich die Gesamtheit des Wortschatzes in Schichten zentraler und weniger zentraler Einheiten einteilen lässt. Wir wollen stattdessen versuchen, mit Hilfe eines Blicks auf die Varietäten der Sprache eine Vorstellung davon zu erhalten, wie sich Wortschatz schichten lassen könnte.

201. Vgl. die Diskussion bei Schindler 2002, S. 36-39. Beispiele für die bei Schindler angeführten Abgrenzungsprobleme können sein: Sind Interjektionen oder Zeichen der Gebärdensprache Teile des Lexikons (S. 36)? Sind Äußerungen wie *tra la la* Teile des Lexikons (S. 37)? Wie geht man mit sprachlichen Einheiten um, die eine primär grammatische Bedeutung und Funktion haben (S. 36f.)? Sind Fugenelemente noch als Teile des Lexikons zu rechnen (S. 38)?

202. Für einen detaillierten Überblick über semantische Beziehungen zwischen Wörtern (Sinnrelationen) vgl. etwa Carter 1998, S. 19-23 und Kühn 1979, S. 83-95 oder Schnörch 2002, S. 257-262 und 347-366 im spezifischen Kontext der Grundwortschatzlexikographie. – Für eine ausführliche Darstellung des Ordnungsprinzips der Wortfelder vgl. etwa Lutzeier 1981 und Lutzeier (Hg.) 1993. Für eine Diskussion des Ordnungsprinzips der Wortfamilie in der lexikographischen Praxis vgl. etwa Augst 1997.

203. Vgl. etwa die Diskussionen in den Abschnitten 4.4.1 und 4.4.5.

5.3 Makrostrukturelle Dimensionen der Wortschatzschichtung

5.3.1 Sprachvarietäten: die Abgrenzung von Gemeinsprache und Sondersprachen am Beispiel der Fachsprachen

Der Gebrauch der Sprache variiert auf vielfältige und gruppenspezifische Weise. Dementsprechend werden von der sog. Gemeinsprache eine Vielzahl von Sprachvarietäten unterschieden.²⁰⁴ Zu diesen Sprachvarietäten werden beispielsweise die Fachsprachen, die Dialekte oder auch die Jugendsprache gerechnet. Da diese Varietäten nur von einem Teil der Sprachgemeinschaft verwendet werden, liegt die These nahe, dass Wortschatzelemente, die einer spezifischen Sprachvarietät zuzurechnen sind, nicht Teil zentraler Wortschätze sein können. Ob diese These stimmt oder ob sich nicht doch die Beachtung dieser Wortschatzdimension für die Frage nach der Zusammensetzung zentraler Wortschätze nutzbar machen lässt, ist Thema der folgenden Abschnitte.

Bei allen Sprachvarietäten steht die Frage nach ihrer Abgrenzung zu einem Standard, also etwa zur Gemeinsprache im Mittelpunkt. Ein typisches Beispiel einer solchen Sprachvarietät und der dazugehörigen Diskussion ist die Fachsprache. Wir wollen an ihrem Beispiel diskutieren, wie sich die Abgrenzung von Gemeinsprache und Sprachvarietät gestaltet.

Mit Glück ((Hg.) 2000) wollen wir Fachsprache definieren als...

„...[entweder] die sprachl[ichen] Spezifika oder die Gesamtheit der sprachl[ichen] Mittel, die in einem Fachgebiet verwendet werden. [...] Den Kern einer F[achsprache] bildet in der Regel ihre [...] Terminologie, in der sich die Fachkenntnisse spiegeln. F[achsprachen] können jedoch auch syntakt[ische] Besonderheiten (vgl. z. B. die aussagenlog[ischen] oder mathemat[ischen] Konnektoren) und spezif[ische] Textformen (Versuchsbeschreibung usw.) aufweisen.“ (Glück (Hg.) 2000, S. 203)

Von den Fachsprachen abgegrenzt werden manchmal auch die Berufssprachen im Sinne einer...

„...Spr[ache] einer bestimmten Berufsgruppe, vor allem ihre [...] Fachsprache. Den Kern einer B[erufssprache] bildet ihre [...] Terminologie; sie kann jedoch auch in informellen Gruppenprozessen entstandene Besonderheiten umfassen ([...] Phraseologismen des speziellen [...] Jargons) sowie besondere Textsorten.“ (Glück (Hg.) 2000, S. 106)

In Anlehnung an Glück (Hg.) 2000, S. 203 wollen wir aus Gründen der terminologischen Vereinfachung die beiden Termini *Berufssprache* und *Fachsprache* als identisch betrachten – trotz kleiner terminologischer Unterschiede – und uns auf den Gebrauch des Terminus *Fachsprache* beschränken.²⁰⁵

204. Mit dem Begriff *Sprachvarietät* beziehen wir uns hier auf die Definition bei Glück ((Hg.) 2000, S. 771, nach der eine Varietät ein „Subsystem [bzw.] Teil einer ganzen Spr[ache] [darstellt], die in aller Regel eine größere Zahl von V[arietäten] umfaßt, z. B. [...] Dialekte, eine [...] Standardvarietät“. – Verwendet wird auch der Begriff der *Existenzformen der Sprache* (vgl. etwa Besch 1983, S. 962).

205. Wie die Definition bei Glück zeigt, bestehen Fachsprachlichkeit und Gemeinsprachlichkeit genau genommen aus mehr als nur einem fachsprachlichen bzw. gemeinsprachlichen Vokabular. Eigenheiten auf anderen sprachlichen Ebenen, etwa der Syntax oder Morphologie, tragen auch zur Fachsprachlichkeit eines Textes bei (vgl. etwa auch Kalverkämper 1990, S. 119). Wir sprechen hier vereinfachend von *Fachsprache*, wenn wir uns allein auf den fachsprachlichen Wortschatz beziehen.

Dabei spielt die Frage danach, ob eine Wortschatzeinheit als Teil einer Fachsprache oder als Teil der Standardsprache zu betrachten ist, bei der Arbeit mit zentralen Wortschätzen und der hier geführten Diskussion in dreierlei Hinsicht eine Rolle:

Wenn es um die Erstellung eines zentralen Wortschatzes geht, dessen Ausgangspunkt der Allgemeinwortschatz ist, dann ist eine als fachsprachlich erkannte lexikalische Einheit der Peripherie zuzurechnen und eher nicht Kandidat für den zentralen Wortschatz. Geht es um die Erarbeitung eines zentralen Wortschatzes innerhalb einer Fachsprache, ist die Eigenschaft einer lexikalischen Einheit, dieser Fachsprache anzugehören, gerade ein konstituierendes Merkmal. In beiden Fällen ist es wichtig, beurteilen zu können, ob eine lexikalische Einheit als fachsprachlich oder als allgemeinsprachlich einzuordnen ist. Und schließlich sind Fragen der Abgrenzung von fachsprachlicher Lexik und allgemeinsprachlicher Lexik Teil der grundlegenden Frage danach, ob man einen generischen zentralen Kern des Wortschatzes annehmen soll oder nicht.²⁰⁶

Wie sich hier bereits zeigt, ist die Abgrenzung der Fachsprache von der Gemeinsprache konstituierendes Merkmal von Fachsprachen bzw. fachsprachlicher Wortschätze und zugleich Hauptstreitpunkt der Diskussionen zum Thema Fachsprache.²⁰⁷ Die Trennung von Fachsprache und Gemeinsprache, ja sogar die Annahme der Existenz einer Gemeinsprache, ist in der Fachsprachenforschung kontrovers diskutiert worden (vgl. Nuopponen 2002, S. 857). Bei dieser Kontroverse geht es vornehmlich um die Frage danach, welchem von zwei grundlegenden Ansätzen recht zu geben sei:

Die Vertreter des ersten Ansatzes sehen Fachsprache und Gemeinsprache bzw. Fachwortschatz und Gemeinwortschatz als zwei komplementäre Mengen, wobei der Gemeinwortschatz den Ausgangswortschatz darstellt, zu dem der Fachwortschatz ergänzend hinzutritt.²⁰⁸

Einer der Hauptkritiker der Annahme einer strikten Opposition von Fachsprachlichkeit gegenüber Gemeinsprachlichkeit und damit Vertreter des zweiten Ansatzes ist Hartwig Kalverkämper (1990). Versuche, die Dichotomie von Fachsprache vs. Gemeinsprache mit Hilfe des Merkmals der Verständlichkeit aufzulösen,²⁰⁹ stellt Kalverkämper mit dem Ar-

206. Vgl. die Diskussion in Abschnitt 5.5.

207. Manche Ansätze vertreten auch die These, dass es zusätzlich zu den Fachsprachen und der Gemeinsprache noch so etwas wie eine allgemeine Fachsprache gibt, also eine Fachsprache, die mehreren Fachsprachen gemeinsam ist, sich aber von der Gemeinsprache abhebt. Dieser Ansatz spiegelt sich beispielsweise in den Arbeiten von Erk (1972, 1975 und 1982) wider, der versucht, die allgemeine Wissenschaftssprache, wie sie den von ihm untersuchten Texten von insgesamt 34 Fachgebieten gemeinsam ist (vgl. Erk 1972, S. 27), frequentiell zu erfassen.

208. Vgl. z. B. Wichter 1994 oder Nuopponen 2002.

209. Vgl. etwa Heller 1981 und seine Matrix von Fachbezogenheit und Verständlichkeit. Heller kommt dabei zu einer Einteilung des Wortschatzes in unter anderem Allgemeinwortschatz, nicht fachbezogene Fremdwörter, allgemein verständliche Fachlexik und nicht allgemein verständliche Fachlexik. Er schließt dabei auch Soziolekte in seine Ontologie mit ein. Für eine detailliertere Darstellung vgl. Nuopponen 2002, S. 858.

gument in Frage, die Eigenschaft der Verständlichkeit sei eine „grundsätzlich relationale Qualität“ (Kalverkämper 1990, S. 103):

„Gemeinsprache oder – wie es ja traditionellerweise praktisch immer gemeint ist – der gemeinsprachliche Wortschatz kann als solcher also gar nicht die absolute Qualität einer 'allgemeinen Verständlichkeit' innehaben – schon diese beiden kombinierten ganz relationalen Begriffe belegen die Unmöglichkeit -, und so kann sie erst recht nicht – als ein besonderes Charakteristikum – verständlicher sein als die Fachsprachen. 'Verständlichkeit' ist eine Textqualität, und als solche hat sie Gültigkeit für alle Arten von Texten und ist prinzipiell abhängig von den Beteiligten jeder funktionierenden Kommunikation; von Sprecher/Schreiber, Hörer/Leser, von der Sprache (als Texte oder in Texten) und von der jeweiligen Situation. Somit ist auch die vorangenommene 'Allgemeinheit' eine relationale Größe, die sich sowohl auf die sozialen Gruppen, Strukturen und Rollen bezieht, als auch auf eine wie auch immer geartete (geschweige erarbeitete) Situationen-Typologie. In der konkreten Kommunikation wird diese Relativität festgelegt, die 'Allgemeinheit' wird in der jeweiligen Situation mit den jeweils beteiligten Partnern zur Individualität, reduziert eben von Faktoren wie Alter, Geschlecht, Herkunft, Lebenslauf, soziokulturellem Hintergrund, Intelligenz, Lerngeschichte und Bildung oder Vorkenntnissen.“ (Kalverkämper 1990, S. 104)

Er stellt gerade diese Sichtweise, d.h. die Einteilung des Wortschatzes in eine „als homogen ausgewiesene“ (ebda., S. 101) und allgemeinverständliche Gemeinsprache und verschiedene daran angegliederte Fachsprachen stark in Frage und plädiert für ein graduelles Modell, bei dem Texte hinsichtlich ihrer Fachlichkeit und Fachsprachlichkeit auf einer Skala von „(extrem) merkmalarreich“ (ebda., S. 112) bis „(extrem) merkmalararm“ angeordnet werden. Seine Argumente für diese Aufhebung sind vielfältig:

Zum einen bestehe auch das sogenannte Alltagsleben und -handeln, wo man die sogenannte Gemeinsprache verwendet, aus einer Vielzahl spezifischer Handlungen, die alle irgendwie etwas mit einem Fachbereich zu tun haben:

„Was allgemein und üblicherweise als Prototyp des unspezifischen Handelns, der normalen Tätigkeiten, des interesselosen grauen Einerleis, der Routine ohne notwendige Kennerschaft gilt und entsprechend auch in Befragungen eingeschätzt wird [...], entpuppt sich, empirisch verfolgt, als eine dichte Überschneidung und Abfolge von – gerade gegenteilig – speziellen Arbeiten in den verschiedensten Bereichen und mit unterschiedlichen Zielsetzungen.“ (Kalverkämper 1990, S. 91)

Die Schwierigkeiten der sprachlichen Abgrenzung spiegeln sich also in den Schwierigkeiten der faktischen Abgrenzung wider.

Darüberhinaus gebe es keine klaren Anhaltspunkte dafür, welche Fächer bzw. Fachsprachen überhaupt annehmen soll, im Gegenteil, es gebe vielmehr eine Vielzahl verschiedenartiger Fächereinteilungen (vgl. ebda., S. 93 und 100). Damit spricht Kalverkämper ein grundlegendes Problem der Fachsprachenforschung an – wie lassen sich überhaupt Fachbereiche voneinander abgrenzen?²¹⁰

Desweiteren sei der Begriff der *Gemeinsprache* an sich problematisch, weil sie keine so homogene Varietät sei, wie es in der Fachliteratur postuliert werde, sondern eine „ausschließlich relativ bestimmte Größe“ (ebda., S. 105), und zwar anhand der (relativen) Merkmale

210. Einen praktischen Versuch einer Lösung dieses Problems stellt das bei Bergenholtz/Pedersen 1994 beschriebene Fachkorpus der Pumpentechnologie dar.

der Verständlichkeit und dem Kontext des Laien, ein Konzept, das in sich wiederum relativ ist, nämlich im Kontrast zum Kontext des Experten. Daraus folgt:

„Eine faßbare sprachliche [...] Kategorie 'Gemeinsprache' kann es somit nicht geben, jedenfalls nicht in einer Seinsweise oder in einer Vorkommensart, die es gestattet, sie im Umfang zu begrenzen, sie inhaltlich zu präzisieren und sie in ihren Funktionen zu erkennen.“ (Kalverkämper 1990, S. 106)

Und schließlich seien die Grenzen zwischen Gemeinsprache und Fachsprache nicht nur fließend, sondern oft auch nur sehr schwer argumentativ zu stützen:

„Ist [...] das Wort *Hammer* gemeinsprachlich oder fachsprachlich? Oder das Wort [...] *Fünfyylinder, Satellitenübertragung, Sweat Shirt*? Oder ist [...] das Gespräch einer Hausfrau mit einer Nachbarin über die angemessene Wahl von Pflegemitteln für die Terrasse mit Marmorboden gemeinsprachlich? Oder fachsprachlich? Ist es fachlich (sachbezogen), aber nicht fachsprachlich?“ (Kalverkämper 1990, S. 108)

Dabei lässt Kalverkämper auch die Alternative, statt von Fachsprache vs. Gemeinsprache von fachsprachlich vs. nicht-fachsprachlich zu sprechen, nicht gelten, mit der Begründung, dies sei nur eine versteckte Form der Fachsprachen-Gemeinsprachen-Dichotomie (vgl. ebda., S. 117).

An dieser Stelle der Argumentation führt Kalverkämper sein Modell der graduellen Darstellung als die adäquatere Sichtweise ein:

„Das tertium comparationis ist damit [auf die] 'Fachsprachlichkeit' festgelegt; von ihr aus bestimmt sich jene Sprache-in-Texten, die bislang intuitiv als 'Gemeinsprache' eingeschätzt wurde, als (merkmal-)arm, als (merkmal-)reduziert, also als Verringerung derjenigen Merkmale, die in Fachsprachen reich, vielfältig, komplex, auffällig vorhanden sind. Die Perspektive bei der Besetzung der Skala mit Merkmalen verläuft, da es sich um (gestufte) Merkmale der 'Fachsprachlichkeit' handelt, somit in einer Richtung: von den Fachsprachen als Orientierungsbereich hin zu der sogenannten 'Gemeinsprache'.“ (Kalverkämper 1990, S. 112)

Kalverkämper kann dabei sein Modell mit einer Reihe von Argumenten stützen (vgl. Kalverkämper 1990, S. 112-116). Unter anderem führt er an, dass sich der Status als Laie aus der Negation des Experten definiere – ein Laie ist ein Nicht-Experte. Des weiteren gebe es zwar Begriffe für das Phänomen, wenn ein Terminus aus der Fachsprache von Laien verwendet wird (z.B. *popularisieren*), aber für den umgekehrten Weg, die 'Verfachlichung' eines Ausdrucks, gebe es kaum solche direkten Bezeichnungen, die 'Verfachlichung' eines gemeinsprachlichen Wortes sei also nicht markiert.

Ähnlich, so argumentiert Kalverkämper weiter, spreche man auch vom Wissensgefälle und vom Wissenstransfer, die beide eine Richtung vom Fachlichen hin zum Laienhaften implizieren. Darüberhinaus sei es inzwischen eine anerkannte wissenschaftliche Erkenntnis,...

„...daß die Fachsprachen und die sogenannte 'Gemeinsprache' nicht, wie ehemals angenommen, eigenständige Charakteristika aufweisen, vielmehr die Fachsprachen ihre Mittel durch Abwahl, durch Selektion aus dem 'Gesamtbestand der Sprache' (Hoffmann 1985, S. 47) gewinnen und für ihre speziellen Zwecke in besonderer Weise bereithalten und einsetzen.“ (Kalverkämper 1990, S. 114)²¹¹

Zu vergessen sei schließlich auch nicht, dass es kaum Texte gibt, die man ohne weiteres als der Gemeinsprache zugehörig einordnen könne.

Daraus zieht Kalverkämper den Schluss:

„Demnach wäre alles Kommunizieren über die Gegenstände, Sachverhalte und Handlungszusammenhänge – kurz: über 'die Welt' – als fachsprachlich anzusehen, und zwar [...] in gestufter [...] Merkmalhaftigkeit. Streng fachliches Kommunizieren zwischen Fachleuten wäre somit äußerst merkmalsreich, ein Gespräch von Hausfrau zu Hausfrau über die angemessene Waschmitteldosierung oder eine Nachfrage und Antwort zum richtigen Zeitabstand beim Füttern von Jodperlen für den Wellensittich wären merkmalarms.“ (Kalverkämper 1990, S. 112)

Für die Skala der Fachsprachlichkeit, die Kalverkämper vorschlägt, gilt dann auch, dass sie eine Skala der gleitenden Übergänge ist und damit die traditionelle Dichotomie von Fachsprache vs. Gemeinsprache durch ein Kontinuum der beiden Pole merkmalsreich vs. merkmalarms hinsichtlich des Merkmals der Fachsprachlichkeit ablöst (vgl. Kalverkämper 1990, S. 117-119). Damit seien auch Problemfälle wie Wörter, die aus einem Fachbereich in die Gemeinsprache wandern, angemessen zu erklären.

Das Modell von Kalverkämper führt also zu einer Reduzierung der ursprünglichen vierdimensionalen Einteilung nach Laie vs. Experte und Gemeinsprache vs. Fachsprache. Die erste Dichotomie wird reduziert zu einer Skala der Fachlichkeit, die zweite zu einer Skala der Fachsprachlichkeit (vgl. Kalverkämper 1990, S. 123). Dabei sind die Merkmale für Fachsprachlichkeit an den Text gebunden (ebda., S. 122). Damit geht die „Zuweisung eines Wortes im traditionell gesehenen Spannungsfeld zwischen der sogenannten 'Gemeinsprache' und den Fachsprachen [...] nicht mehr intuitiv“ (Kalverkämper 1990, S. 123) vor sich, „sondern nunmehr angemessen, weil über die textuell verankerten sprachlichen Merkmale beschreibbar, aus den Texten-in-Funktion ableitbar.“ (Kalverkämper 1990, S. 123).

Zur Erarbeitung der Merkmale der Skala der Fachsprachlichkeit fordert Kalverkämper schließlich ausgedehnte induktive, d.h. korpusgestützte Untersuchungen, und zwar nicht allein im Hinblick auf die Lexik, sondern auf den gesamten Text (vgl. Kalverkämper 1990, S. 105 und 119). Insbesondere eine „Hierarchie der Merkmale“ (ebda., S. 125) stehe noch aus, also die Beantwortung von Fragen wie:

„Wieviele [...] und welche Merkmale [...] müssen im Text gegeben sein [...], um ihn als fachsprachlich merkmalsreich, als einen Fachtext, erkennbar werden zu lassen? Wie sind die Merkmale im Text verteilt? Lassen sich Korrelationen zwischen Quantität [...], Qualität [...], Rekurrenz [...] und Distribution [der Merkmale] erkennen und als Hierarchie auswerten?“ (Kalverkämper 1990, S. 125)

Mit diesem Modell und der grundlegenden Kritik an der herkömmlichen Dichotomie von Fachsprache(n) vs. Gemeinsprache weicht Kalverkämper von der bis dahin allgemein akzeptierten Einschätzung ab, und auch in neueren Arbeiten scheint sein Vorschlag nicht übernommen worden zu sein.²¹²

211. Vgl. ähnlich auch Rosengren (1972, S. XII), die von der Fachsprache (in ihrem Fall der Zeitungssprache, auch wenn man diese nicht unmittelbar als Fachsprache bezeichnen würde) als einem Phänomen der *parole*, also der Performanz spricht, da sie nur eine Auswahl aus dem System der Möglichkeiten darstelle. – Ähnlich auch Nuopponen 2002, S. 858.

Dabei scheint sein Modell einer skalenhaften Fachsprachlichkeit den praktischen lexikographischen Gegebenheiten besser zu entsprechen. Denn wie so oft, wenn es darum geht, sprachliche Erscheinungen der einen oder anderen Kategorie zuzuordnen, sieht man sich auch bei der Einordnung eines Wortes hinsichtlich seiner Fachsprachlichkeit einer Reihe von Zweifelsfällen gegenüber:

Zum einen ist ganz allgemein die Zuordnung eines Wortes zu einem bestimmten Fachbereich ohne irgendwelche Kontextangaben nahezu unmöglich.²¹³ Aber auch wenn Wörter in einen Kontext eingebettet sind, kann es schwierig sein, ihre Fachsprachlichkeit zu beurteilen, etwa wenn es sich dabei um Wörter handelt, die zunächst eindeutig fachsprachlichen Charakter haben, dann aber durch einen weiteren Benutzerkreis und größere Bekanntheit des von ihnen Bezeichneten (Verbreitung der Sache) aus der Fachsprache in die Gemeinsprache übernommen werden.²¹⁴ Beispiele hierfür sind etwa die Wörter *Karies* (vgl. Kühn 1979, S. 7), *AIDS* oder auch Termini aus dem Bereich der Autotechnik. Anzumerken ist, dass beim Prozess der Übernahme in die Gemeinsprache das jeweilige Lexem oft die semantische Präzision verliert und eine verschwommenere Bedeutung erhält (vgl. Nuoppo 2002, S. 859).²¹⁵ Zum anderen gibt es auch Fälle, in denen ein Lexem eine eher gemeinsprachliche und eine eher fachsprachliche Bedeutung hat, was zu erheblichen Zuordnungsschwierigkeiten führt.²¹⁶

Vor dem Hintergrund dieser Beispiele von Zweifelsfällen, die zeigen, wie schwierig es in manchen Fällen tatsächlich ist, ein konkretes Wort eindeutig der Gemeinsprache oder einer spezifischen Fachsprache zuzuordnen, etwa weil ein und dasselbe Wort sowohl in gemeinsamen als auch fachsprachlichen Zusammenhängen vorkommt oder etwa auch aus rein fachsprachlichen Kontexten in gemeinsprachliche Kontexte wandern kann, erscheint hier ein Skalenmodell, das gleitende Zugehörigkeiten und Übergangszonen erlaubt, angemessener

212. So arbeitet etwa Sigurd Wichter (1994) mit dem Begriff der „vertikale[n] Wortschatzvariation“ (Wichter 1994, S. 9), d.h. der „Vertikalität“, definiert als „Wissensvertikalität [...], d.h. [als] ein bestimmtes Verhältnis zwischen zwei oder mehr Wissenskomplexen“, und geht von der horizontalen Anordnung verschiedener Fachbereiche zueinander aus. Diese Fachbereiche wiederum sind in sich vertikal (Experte – Laie) strukturiert. Er arbeitet also mit einer Hierarchie, der die Dichotomie von Experte vs. Laie und damit die Dichotomie von Fachsprache und Gemeinsprache zugrundeliegt. Zugleich kritisiert Wichter den Ansatz von Kalverkämper 1990, weil dieser durch die Betonung der Fachsprachlichkeit als der Ausgangsdimension den „fachfreien Kontext“ (Wichter 1994, S. 25) zu sehr vernachlässige, und ignoriert damit das Hauptargument Kalverkämpers, dass es bei genauem Hinsehen praktisch keine allgemeinsprachliche, fachfreie Kommunikation gebe, jedenfalls keine, die klar von einer fachlichen abzugrenzen sei (vgl. z. B. Kalverkämper 1990, S. 91).

213. Vgl. z.B. Kalverkämper 1990, S. 121. – Ähnliche Erfahrungen macht auch Bonan-Garrigues, wenn es um die Beurteilung der Plausibilität der Einträge des LADL geht, die ja ohne Kontext angegeben sind (vgl. Bonan-Garrigues 1993, S. 48f.). Auf demselben Phänomen beruht auch die Kritik Schnörchs an Krohns Verfahren, aus den Hapaxlegomena der von ihm untersuchten Korpora thematische Profile zu erarbeiten (vgl. Schnörch 2002, S. 219-223).

214. Zu einer kritischen Beurteilung dieser Metapher des Wanderns von einem Fachbereich zum anderen vgl. Kalverkämper 1990, S. 111.

215. Vgl. etwa den Terminus *AIDS*, der in der Medizin das Krankheitsbild, das durch eine HIV-Infektion ausgelöst wird, bezeichnet, während *AIDS* und *HIV* im allgemeinsprachlichen Kontext oft auch synonym verwendet werden.

als ein Modell zweier oppositioneller Mengen. Mit seinen beiden grundlegenden Skalen von merkmalsreich bis merkmalarms hinsichtlich Fachlichkeit (der Sprecher) und Fachsprachlichkeit (der Texte) erlaubt ein solches Skalenmodell eine differenziertere Darstellung von gemeinsprachlichen bzw. fachsprachlichen Phänomenen als die traditionelle Dichotomie und sorgt zugleich für mehr Transparenz bei der Textsortenanalyse.²¹⁷ Dies gilt allgemein; darüber hinaus aber wiegt noch schwerer, dass die Frage nach der Fachsprachlichkeit von Lexemen vor allem bei der Frage nach einem zentralen Wortschatz für eine automatische, d.h. computergestützte Klassifikation von Texten nach Themen bzw. Fachbereichen eine Rolle spielt. Gerade hier ist eine Einordnung eines Textes, die auf vielfachen Parametern beruht, wie sie Kalverkämper (1990, S. 25) beispielhaft anführt, möglich und vor allem nötig, denn die Beurteilung wird von Computern durchgeführt und ein manuelles Eingreifen ist nicht möglich. Je differenzierter die Beurteilungsmöglichkeiten in einem solchen Fall, desto besser.

Unabhängig davon, welchen der beiden hier dargelegten Ansätze man vertritt, lassen sich in jedem Fall jedoch gewisse typische Züge fachsprachlich merkmalsreicher Lexeme ausmachen.²¹⁸ Sie sollen im Folgenden kurz zusammengefasst werden.

- Zunächst ist ein fachsprachlich geprägter Wortschatz häufig das Ergebnis bewusster Festlegungen:

„[The] general vocabulary of a language is used to comprehend and organise the world intuitively, while special vocabularies, particularly terminologies, are based on more conscious cognitive processes and more or less explicit agreements on the relationship between the linguistic form and the concept [in Anlehnung an Albrecht 1992, S. 63; V. H.]“ (Nuopponen 2002, S. 857)

Fachsprachlich merkmalsreiche Lexeme sind also semantisch stärker festgelegt als allgemeinsprachliche Wörter, d.h. sie weisen weniger Polysemie auf, im Gegenteil, in fachsprachlichen Nomenklaturen sind sie oft genau definiert.²¹⁹ Dieses Phänomen wird auch als „Terminologisierung“ (Kalverkämper 1990, S. 120) bezeichnet.

- Als fachsprachlich markiert gelten auch Häufungen von Komposita, von Derivationen, von Funktionsverbgefügen und Nominalisierungen (vgl. Kalverkämper 1990, S.

216. Vgl. etwa das Beispiel der Lexeme von *Lampe*, *Leuchte* und *Birne* bei Seebold 1993

„Die technische Bezeichnung für die *Glühbirne* ist *Glühlampe* oder einfach *Lampe*; während das Gerät, in das diese Lampe eingeschraubt wird, als *Leuchte* zu bezeichnen ist. In der Gemeinsprache [...] spricht man aber allgemein von *Glühbirne* oder einfach *Birne*, während man den Beleuchtungskörper eher als *Lampe* bezeichnet. Der Fachmann schraubt also eine Lampe in die Leuchte, während der Laie eine Birne in die Lampe schraubt.“ (Seebold 1993, S. 216).

217. Vgl. das bei Kalverkämper 1990, S. 124 angeführte Beispiel.

218. Für eine knappe Übersicht über alle, nicht nur die lexikalischen, Merkmale der Fachsprachlichkeit vgl. Kalverkämper 1990, S. 120.

219. Vgl. auch die Gegenüberstellung eines allgemeinsprachlichen und eines medizinischen Korpus bei Fabre / Habert / Labbé 1997. Sie können anhand einer Kollokationsanalyse nachweisen, dass allgemeinsprachliche Texte nicht in gleichem Maße semantisch organisiert sind wie fachsprachliche Texte. Vgl. z. B. Fabre / Habert / Labbé 1997, S. 28 – Mit der starken semantischen Festlegung fachsprachlicher Wörter werden auch Fragen der Abgrenzung von sprachlicher und enzyklopädischer Bedeutung in die Diskussion eingeführt, ein Thema, auf das hier aus Platzgründen nicht eingegangen werden kann.

120).²²⁰ Typisch für die allgemeine Wissenschaftssprache als ein Typus der Fachsprache ist darüberhinaus, dass sie viele sogenannte Internationalismen enthält; besonders grundlegende Gegensatzpaare wie *aktiv – passiv*, *individuell – kollektiv*, *objektiv – subjektiv* sind durch Internationalismen vertreten (vgl. Braun 1990, S. 24). Auch die Namen vieler fachwissenschaftlicher Disziplinen sind, so Braun, Internationalismen.

- Oft drücken Fachtermini auch ihre Zugehörigkeit zu einer Hierarchie von Begriffen aus, wie etwa das bei Nuopponen (2002, S. 863) angeführte Beispiel der verschiedenen Lokomotivtypen (*steam locomotive – diesel locomotive – electric locomotive*) zeigt.
- Außerdem weisen Fachtermini einen geringen Grad an stilistischer Färbung und Konnotation auf:

„A further characteristic of terminologies is said to be their neutrality. In subject fields there is a need to objectively describe specific knowledge. Thus terminologies try to avoid all possible emotive and negative connotations. It cannot be helped that sometimes subject field specific terms develop a negative connotation when taken over to the general language.“ (Nuopponen 2002, S. 863)

Wenn man sich nun rückblickend die Überlegungen zur makrostrukturellen Wortschatzdimension der Fachsprachlichkeit ansieht, dann wird deutlich, dass sie bei zentralen Wortschatze dann eine Rolle spielt, wenn es um Wortschatze zur maschinengestützten Textanalyse und Textklassifikation geht, bei der Texte unter anderem auch an dem ihnen eigenen Vokabular klassifiziert und nach Themen gruppiert werden. Dabei ist eine Einteilung des Wortschatzes im Sinne von ‘Lexem x ist in der Regel ein Fachbegriff der Medizin, Lexem y dagegen ein Fachbegriff der Biologie’ und eine daraus folgende Zuordnung eines Textes zu einem der beiden Fachbereiche illusorisch. Was erreichbar ist, ist ein Erkennen von Lexemen, die in der Regel (und das auch nur als Faustregel) keinem spezifischen Fachbereich angehören, sondern fach- bzw. themenneutral sind. Diese Themenneutralität ließe sich dann kombinieren mit dem Festlegen von Schlüsselwörtern, die einerseits selten den Fachbereich wechseln und andererseits sehr typisch für einen Fachbereich sind. Zusätzlich müsste man noch typische Kollokationen dazu induktiv, d.h. durch Korpusanalysen zu identifizieren.²²¹

Zugleich ist in dieser Zusammenstellung deutlich geworden, dass sich die Opposition von Fachsprache und Gemeinsprache auch als Kontinuum beschreiben lässt, bei der man Wörter danach betrachtet, ob sie hinsichtlich der Fachsprachlichkeit merkmalsreich oder merkmalsarm sind. Wir wollen uns dieser Sicht anschließen und sie von der Fachsprache auch

220. Für eine ausführliche Übersicht der verschiedenen Mechanismen der Generierung neuer Terme vgl. Nuopponen 2002, S. 863f.

221. Eine wichtige Aufgabe erfüllen Fachsprachen auch in der maschinellen Sprachverarbeitung allgemein. Dort können sie im Sinne eines sprachlichen ‘Mikrokosmos’ einen wichtigen Beitrag zur Erforschung der Gemeinsprache leisten:

„In this way sublanguages play the same role for computational linguistics as fruitflies (*drosophilae*) play for the study of genetics. In sublanguages one can study language systems which are microcosms (in many if not all respects) of the whole language. Just as *drosophilae* facilitate the study of genetics for biologists, sublanguages make natural language's information-carrying mechanisms more transparent to linguists and computer scientists.“ (Kittredge 2003, S. 437)

auf andere Sprachvarietäten übertragen. Demnach lassen sich Wörter beispielsweise danach beschreiben, ob sie hinsichtlich eines regional begrenzten Gebrauchs merkmalsreich sind – dann ist ihr Gebrauch dialektal geprägt – oder ob sie hinsichtlich regionaler Beschränkungen merkmalsarm sind – dann ist ihr Gebrauch nicht dialektal geprägt. Bei der Erarbeitung eines Wortschatzmodells (vgl. Kapitel 6) werden wir dieses Skalenmodell teilweise übernehmen.

5.3.2 Frequentielle Phänomene der Wortschatzschichtung

Neben den bis jetzt beschriebenen lexikologischen Dimensionen, anhand derer man Wortschatz strukturieren kann, lassen sich auch einige allgemeine frequentielle Phänomene der Wortschatzschichtung beobachten, die in der Diskussion zu zentralen Wortschätzen eine Rolle spielen und immer wiederkehren.

Eines dieser beobachteten Phänomene ist das der sogenannten „degressive[n] Struktur des Wortschatzes“ (Augst 1987, S. 117). Damit bezieht man sich auf ein Phänomen der Frequenzverteilung in Texten: Es läßt sich beobachten, dass ein Großteil der laufenden Wortformen eines Textes auf eine verhältnismäßig geringe Zahl an Lexemen zurückzuführen ist, dass also, umgekehrt gesprochen, eine geringe Anzahl an Lexemen hinter dem Großteil der laufenden Wortformen eines Textes steht.²²² Je stärker ausgeprägt diese Degressivität ist, desto geringer ist die Anzahl der Wörter, die den Großteil der Wörter eines Textes abdecken. Dabei beruht dieses Phänomen vor allem auf dem Verhalten der Funktionswörter, also Präpositionen, Partikel, Artikel und Pronomina, die, weil sie semantisch vergleichsweise arm sind und zugleich eine wichtige syntaktische Funktion übernehmen, in der Regel eine hohe Frequenz in Texten aufweisen.

Allerdings muss man einschränkend hinzufügen, dass man sich mit dem Phänomen der degressiven Struktur des Wortschatzes vor allem auf allgemeinsprachliche Texte und Häufigkeitsuntersuchungen bezieht. Bei fachsprachlichen Texten scheint die Streuung, d.h. die Anzahl von Wörtern, die mit nur geringer Frequenz vorkommen, weitaus geringer zu sein, die degressive Form der Häufigkeitsverteilungskurve also wesentlich weniger steil zu sein (vgl. Erk 1972, S. 27). Das bedeute, so Erk, dass es bei fachsprachlichen Texten einen stärkeren ‘Mittelbau’ an mittelfrequenten Wörtern und eben nicht nur ganz wenige hochfrequente und sehr viele niedrigfrequente Wörter.²²³

Neben diesen allgemeinen Beobachtungen zur Worthäufigkeitsverteilung in Texten tritt auch die Beobachtung, dass sich Wortarten quantitativ unterschiedlich verhalten. So stellt

222. Laufende Wortformen werden zum Teil auch als „Textwörter“ (vgl. etwa Bergenholtz / Pedersen 1994, S. 165) oder „tokens“ (Fox 1992, S. 103) bezeichnet, und auch der Terminus „grammatical words“ (Carter (1998, S. 92) scheint, zumindest in Verbindung mit lexikalischer Dichte eines Textes (vgl. ebda.) so verwendet zu werden.

223. Vgl. auch Kühn 1979, S. 57, Kühn 1990, S. 1360 und Carter 1998, S. 234.

etwa Ruoff bei seiner Untersuchung der quantitativen Verhältnisse innerhalb eines Sets mündlicher Sprachdaten fest:

„60% der Wörterbucheinträge sind Substantive, die aber nur gut 10% der Belege stellen. Anders gesagt: jedes zehnte Wort der gesprochenen Sprache ist ein Substantiv, aber sehr viel häufiger immer wieder ein anderes Wort als bei anderen Wortarten. Das Verhältnis Belege zu Lemmata ist beim Substantiv ungefähr 5 : 1, beim Adjektiv 10 : 1, beim Verb 25 : 1, beim Adverb schon 250 : 1, bei den anderen Wortarten natürlich noch größer. Substantive und Adverbien haben fast gleich viele Belege, der Fundus an Substantiven ist aber ungefähr der fünfzigfache.“ (Ruoff 1983, S. 21)

5.3.3 Weitere makrostrukturelle Dimensionen der Wortschatzschichtung

Neben den weiter oben diskutierten Sprachvarietäten lässt sich Wortschatz auch hinsichtlich anderer Dimensionen bzw. Dichotomien strukturieren. So lässt sich Wortschatz zum Beispiel anhand der Abgrenzung von Standardsprache und Umgangssprache betrachten, und auch Fragen der schriftlichen und mündlichen Kommunikation und des produktiven vs. rezeptiven Wortschatzes können helfen, Wortschatz in Schichten einzuteilen. Aber im Vergleich zur Abgrenzung von Sprachvarietäten zur Gemeinsprache spielen diese Abgrenzungen eine für die Arbeit an zentralen Wortschätzen weniger wesentliche Rolle, weshalb sie hier nicht ausführlich behandelt werden.

5.4 Das Kern-Peripherie-Modell des Prager Strukturalismus

In der bisherigen Forschungsliteratur zum Thema zentraler Wortschätze wird immer wieder am Rande erwähnt, dass die Arbeit an diesem Konzept nachhaltig vom Kern-Peripherie-Modell des Prager Strukturalismus geprägt sei. So baut zum Beispiel Krohn seine Arbeit auf einem Schichtenmodell von Synsemantika, themenunspezifischen Autosemantika und themenspezifischen Autosemantika auf, die gemeinsam ein System von Schichten bilden. Die Grenzen dieser Schichten seien „[in] Anlehnung an das wissenschaftstheoretisch flexible Kategoriedenken des Prager Strukturalismus [...] nur relativ definierbar“ und „[die] Unschärfe der Grenzziehung demnach ein konstitutives Charakteristikum dieser Differenzierung“ (Krohn 1992, S. 114). Schnörch greift dieses Vorgehen auf und kritisiert zwar die Unschärfe der Grenzen, nicht aber das generelle Vorgehen (vgl. Schnörch 2002, S. 47).²²⁴ Weil dieses Modell also eine gewisse Rolle in der bisherigen Diskussion gespielt hat, wollen wir es uns als eine Möglichkeit der Wortschatzschichtung und Modell für die Arbeit mit zentralen Wortschätzen näher ansehen.

Die Vertreter der Prager Schule des Strukturalismus beschäftigen sich vornehmlich mit der Phonologie. In diesem Zusammenhang haben sie ein Modell der Phänomene sprachlicher Ausdrücke (Laute) entwickelt, das die Phänomene danach gruppiert, ob sie zum „solid central core“ (Vachek 1966, S. 27) gehören, der kaum Veränderungen unterworfen ist, oder ob sie zur Peripherie des Systems gehören, in der Phänomene 'kommen und gehen', die also

224. Ähnliche Rückbezüge finden sich auch bei Kühn 1979, S. 21 und Bonan-Garrigues 1993, S. 6 und S. 48.

sowohl solche Phänomene enthält, die gerade entstanden sind, als auch solche, die gerade auf dem Weg sind, ganz aus dem System zu verschwinden.

Dieses Modell ist eng verknüpft mit der Diskussion um die Frage der diachronen vs. synchronen Sprachwissenschaft. Nachdem die Junggrammatiker sich ausschließlich auf die diachrone Seite der Untersuchung von sprachlichen Phänomenen konzentriert hatten – wobei es zu ihrer Zeit die Dichotomie von diachroner und synchroner Betrachtungsweise noch gar nicht gab – rücken die Prager Strukturalisten die synchrone Sichtweise in das Zentrum linguistischer Untersuchungen (vgl. ebda., S. 19ff.). Dabei versuchen sie, der naheliegenden Annahme, ein synchroner Blick auf sprachliche Phänomene sei eine statische Sache, zu begegnen, indem sie das sprachliche System als ständig in Bewegung befindlich betrachten. Und an dieser Stelle wird das Kern-Peripherie-Modell wichtig: Denn es erklärt, wie ein synchron betrachtetes System dennoch seinen dynamischen Charakter bewahren kann, indem es den Phänomenen stabilere und weniger stabile Eigenschaften zuordnet:

„On the contrary, at any moment in its development a language system displays items that are on the point of disappearing, and – conversely – items that are only just being born into it. [...] In other words, any language system has, besides its solid central core, its periphery, which need not be in complete accordance with the laws and tendencies governing its central core.“ (ebda., S. 27)

Darüberhinaus spielt dieses Konzept bei den Überlegungen zum „therapeutic character“ (ebda., 21) der Veränderungen im Sprachsystem eine Rolle. Das Sprachsystem strebt nach einem Gleichgewicht, d.h. nach einem geordneten Verhältnis der Teile des sprachlichen Systems zueinander, das verhindert, dass die Teile des Systems – im konkreten Fall die Phoneme – nicht miteinander vermischt werden (ebda., S. 21). Dem gegenüber steht allerdings das Prinzip der Sprachökonomie, das als „Ursache bzw. Anlaß für die Tendenz, mit einem Minimum an sprachlichem Aufwand ein Maximum an sprachlicher Effektivität zu erzielen“ (Bußmann 1990, S. 711) Forderungen nach Veränderungen am System stellt, dadurch das Sprachsystem ständig in Bewegung hält und damit das Gleichgewicht gefährdet (vgl. Vachek 1966, S. 24). Diesen Veränderungen begegnet das System durch sogenannte ‘therapeutische Veränderungen’ (vgl. ebda., S. 21f.). Die Annahme der Prager Strukturalisten ist nun die, dass alle Typen von Veränderungen, ob Veränderungen, die auf sprachökonomische Tendenzen zurückgehen, oder Veränderungen, die auf therapeutische Tendenzen zurückgehen, sich in der Peripherie des Systems abspielen.

Zu erwähnen ist allerdings auch, dass dieses Modell grundsätzlich vor dem Hintergrund der Betrachtung eines phonologischen Systems entstanden ist und dass es daher nicht unmittelbar auf andere Systeme der Sprache wie zum Beispiel das Lexikon übertragen werden kann, sondern einen bewussten Umgang erfordert. Entsprechendes räumen die Prager Strukturalisten auch selbst ein:

„The complexity of the system is also reflected in the fact that within the system taken as a whole one can distinguish a number of sub-systems or levels, each of which has its own particular structure and, consequently, its own specific structural problems.“ (ebda., S. 28)

Glück ((Hg.) 2000, S. 518) konkretisiert die Beschreibung der Peripherie gegenüber dem Kern des Systems dadurch, dass er den Unterschied im Merkmalsumfang hervorhebt: Phänomene der Peripherie verfügen über weniger Merkmale als Phänomene des Zentrums, die alle Merkmale aufweisen. Übertragen auf die lexikalische Semantik rückt Glück damit das Zentrum-Peripherie-Modell in die Nähe der Prototypentheorie. Auch hebt er den fließenden Charakter der Schichtgrenzen hervor: In den Grenzbereichen der einzelnen Schichten des Systems können Überschneidungen und Mischmerkmale vorkommen, die eine eindeutige Einordnung der betroffenen Phänomene in die eine oder andere Schicht nicht zulassen. Und er betont, dass das Prager Kern-Peripherie-Modell der

“Inhomogenität sprachl[icher] Klassen bzw. Kategorien und ihren fließenden Grenzen Rechnung [trägt]. Während die Elemente aus dem Zentrum über alle klassenspezif[ischen] Eigenschaften verfügen, ist die P[eripherie] durch einen unvollständigen Merkmalsatz gekennzeichnet, dessen Defektivität mit zunehmender Entfernung vom Zentrum ansteigt.“
(Glück (Hg.) 2000, S. 518)

Wenn wir nun die bisherigen Ergebnisse zusammenfassen, dann sieht das Prager Modell eines sprachlichen Systems folgendermaßen aus:

Ein sprachliches System hat einen Kern und einen oder mehrere periphere Bereiche. Der Kern besteht aus stabilen, d.h. wenig veränderlichen sprachlichen Phänomenen. In der Peripherie sind die veränderlichen, noch nicht oder nicht mehr stabilen Phänomene angesiedelt. Die jeweiligen Schichten dieses Modells, also der Kern und der periphere oder die peripheren Bereiche, sind in sich selbst Kontinua. Darüber hinaus sind die Grenzen zwischen Kern und Peripherie und zwischen den einzelnen peripheren Schichten fließend, d.h. es gibt sprachliche Phänomene, die sich nicht der einen oder der anderen Schicht eindeutig zuordnen lassen.

Für die weitere Diskussion rund um das Konzept zentraler Wortschätze besonders wichtig sind dabei folgende Eigenschaften des Modells:

- die Stabilität der Mitglieder des Kerns gegenüber dem nicht gefestigten Status der Mitglieder der Peripherie,
- die fließenden Grenzen zwischen den Schichten und schließlich
- der Unterschied im Merkmalsumfang: Die Mitglieder des Kerns besitzen alle Merkmale, die ein Element des Systems haben kann, während Mitglieder der Peripherie nur Teile dieses Merkmalssets besitzen.

Diese Eigenschaften werden später bei der Erarbeitung eines Arbeitsmodells eine wichtige Rolle spielen (vgl. Abschnitt 6.2).

5.5 Die Frage nach dem lexikologischen Zentrum einer Sprache gegenüber der Relativität zentraler Wortschätze

In der bisherigen Diskussion ist vielfach die Rede von der Relativität zentraler Wortschätze, davon, wie sehr ihre Erarbeitung und Ausgestaltung abhängt von dem Zweck, zu dem sie erstellt werden.²²⁵ Zugleich wird, wie wir in Abschnitt 4.1 gesehen haben, die Bezeichnung *Grundwortschatz* oder *basic vocabulary* als ein konkreter anwendungsorientierter Wortschatzausschnitt immer wieder in vager Weise mit dem „Kerninventar“, „core vocabulary“ oder „lexikalischen Kern“ einer Sprache in Verbindung gesetzt, das dem Gesamtwortschatz einer Sprache zugrundeliege. Darin scheint zunächst ein Widerspruch zu liegen – wenn ein Grundwortschatz in seinem Umfang, seinem Inhalt und seiner Darstellung abhängig ist von dem Zweck, zu dem er erstellt wird, wie kann er dann zugleich das Zentrum des Wortschatzes einer Sprache abbilden? Ein näherer Blick auf die Frage danach, wie ein solches Wortschatzzentrum aussehen könnte und wie es greifbar gemacht werden kann, soll helfen, diesen Widerspruch aufzulösen.

5.5.1 *coreness* und die Frage nach dem lexikologischen Zentrum einer Sprache

Um der Frage nach einem möglichen Zentrum des Wortschatzes einer Sprache nachzugehen, wollen wir uns den bei Carter 1998 besprochenen Begriff der *coreness*, also der allgemeinen lexikologisch feststellbaren Zentralität bestimmter Wörter im Verhältnis zu anderen Wörtern, ansehen. Zunächst soll es dabei um die Frage gehen, welche Indizien dafür sprechen, dass ein solches Wortschatzzentrum existiert.

Carter 1998 führt hierfür mehrere Argumente an. Zum einen ist ein solches Wortschatzminimum aus praktischen Gründen nötig, wenn es darum geht, in Situationen, bei denen eine vereinfachte Kommunikation nötig ist – etwa mit Kindern oder fremdsprachigen Erwachsenen (vgl. Carter 1998, S. 34f.) – ein geeignetes Set an Wortschatzelementen zur Verfügung zu haben.

Zum anderen unterstützen Ergebnisse der Spracherwerbsforschung eine solche Annahme, wenn man davon ausgeht, dass die Reihenfolge, in der Kleinkinder lexikalisches Wissen erwerben, ein Indiz für die Zentralität bestimmter Teile des Wortschatzes sein kann.

So geschieht etwa der Erwerb lexikalischen Wissens stufenweise: Die Anzahl der Wortfelder, aus denen Kinder ihre ersten Wörter lernen, ist zunächst sehr beschränkt (vgl. Clark 1993, S. 21). Zudem überwiegen unter den erlernten Wörtern Bezeichnungen für Objekte, während Bezeichnungen für Handlungen und Zustände in der Minderzahl sind (ebda., S. 29). Und vor allem bei Verben zeigt sich, dass Kinder zunächst sogenannte „general-pur-

225. Vgl. etwa Kühn 1979, S. 60f., Schnörch 2002, S. 7, 24, 38, 81 oder 373; Bauer (2005, S. 248), der die Einführung pragmatischer Selektionskriterien als wichtigen Grund für die Notwendigkeit einer Orientierung am Zweck eines zentralen Wortschatzes anführt.

pose verbs like *do, make, get, and go*“ (ebda., S. 30) erlernen und sie auch auf Fälle anwenden, in denen ein erwachsener Sprecher ein spezifischeres Verb einsetzen würde. Auch tendieren Kleinkinder dazu, zunächst eher über eigene Handlungen zu sprechen und erst später über Handlungen anderer Personen; dabei bevorzugen sie Verben, die eine konkrete Bewegung bezeichnen (*walk*) als eine Zustandsänderung (*break*) (ebda., S. 30).

Vor dem Hintergrund der Diskussion um die Rolle der lexikalischen Stabilität als Zeichen für die Zentralität eines Wortschatzelements (vgl. Abschnitt 4.5.2.3) ist auch die folgende Beobachtung besonders interessant: Die Gruppe der ersten Wörter, die Kinder erlernen, gehört zu den ungefähr gleichen Kategorien von Wörtern wie die Wörter, die Kinder vor fünfzig oder hundert Jahren als erste Wörter gelernt haben (vgl. Clark 1993, S. 28).²²⁶

Während ein Kind seinen Wortschatz immer mehr erweitert, erweitern und verfeinern sich auch die Kategorien, nach denen es die lexikalischen Informationen ordnet. Dabei kann es zu sogenannten Überextensionen und Unterextensionen (vgl. Clark 1993, S. 33f.) kommen, wenn also etwa ein Kind die Bezeichnung für einen bestimmten Typus eines Gegenstandes – etwa das Wort *Ball* – für alle möglichen runden Gegenstände verwendet (Überextension) oder wenn es die Bezeichnung für eine bestimmte Gruppe von Tieren – etwa das Wort *Hund* – nicht für alle zur Gruppe der Hunde gehörenden Tiere verwendet (Unterextension). Interessant, wenn auch nicht überraschend, ist hier vor allem die Beobachtung, dass es eher zu Unterextensionen kommt, wenn es sich bei dem zu Bezeichnenden um einen eher peripheren Vertreter der Kategorie handelt: „[A] child might readily use *dog* for any kinds of dogs that are central or prototypical (most dog-like), but not for a Chihuahua or a Pekingese [...]“ (ebda., S. 34). Überextensionen scheinen eher an spezifische Kategorien von Bezeichnungen als an die Prototypenhaftigkeit einzelner Vertreter der Kategorie gebunden zu sein: Bestimmte Bezeichnungen für Fahrzeuge, Kleidungsstücke und Tiere sind eher von Überextensionen berührt als andere Wortfelder. Besonders hervorzuheben ist dabei, dass Überextensionen zwar im produktiven Wortschatz stattfinden, nicht aber im rezeptiven Wortschatz:

„Children who over-extend words in production typically do not over-extend them in comprehension. Instead, they restrict themselves to identifying an appropriate referent [...]. That is, children understand the words they over-extend in production. this would be consistent with over-extension being a communicative strategy: use the nearest word that seems appropriate.“ (Clark 1993, S. 35)

Auch die Frage danach, welche Rolle die Markiertheit vs. Unmarkiertheit von Wortschatzelementen im Spracherwerb spielt, ist, wie wir später noch sehen werden, ein wichtiges Indiz dafür, dass einige Wörter grundlegender sind als andere. Es zeigt sich zum Beispiel, dass Kleinkinder beim Erwerb dimensionaler Adjektive (*big/small, tall/short, high/low*) das positive, d.h. unmarkierte Wort (*big – tall – high*) zuerst aktiv gebrauchen und das negative

226. Wir setzen hier vereinfachend voraus, dass die bei Clark 1993 referierten Untersuchungsergebnisse, die sich auf das Englische beziehen, sich unmittelbar auf andere Sprachen übertragen lassen.

Wort erst zeitverzögert in Gebrauch nehmen, und dies sogar unabhängig davon, mit welcher Häufigkeit ihnen die jeweiligen Wörter begegnen (vgl. ebda., S. 32).

Neben der Spracherwerbsforschung stützen beispielsweise auch die Beobachtungen syntaktischer Umschreibungen von Tabuwörtern, die Dixon 1971 in Bezug auf zwei Sprachvarietäten der Aborigines-Sprache Dyirbal macht, die Annahme von der Zentralität bestimmter Wörter. Die beiden bei Dixon untersuchten Sprachvarietäten werden in unterschiedlichen Kommunikationssituationen verwendet; die eine wird in der alltäglichen Kommunikation gebraucht, die andere ist eine „special (so called ‘mother-in-law’ [...]) language used in the presence of certain ‘taboo’ relatives“ (Dixon 1971, S. 436). Dabei müssen bei der Verwendung dieser zweiten Sprachvarietät bestimmte Tabuwörter, darunter auch Verben, umschrieben werden. Dies, so beobachtete Dixon, geschieht mithilfe der Substitution dieser Verben durch eine Konstruktion aus einem grundlegenden, oberbegriffshaften Verb, dem sogenannten ‘nuklearen Verb’, das das ‘non-nukleare’ Verb ersetzt und umschreibt.²²⁷ Dabei setzen die nuklearen Verben sich aus einfachen semantischen Eigenschaften zusammen, die oft mehreren nuklearen Verben gemeinsam sind, während non-nukleare Verben komplexe semantische Eigenschaften haben (vgl. ebda., S. 441).²²⁸ Zudem sind diese nuklearen Verben in der Regel häufiger als die ihnen verwandten non-nuklearen Verben (vgl. ebda., S. 441).

Wie kann man aber nun konkret herausfinden, welches Wort zentraler ist als andere? Carter führt dazu eine Reihe von Tests an. Sie sollen hier kurz vorgestellt werden.

(Syntaktische) Ersetzbarkeit

Dies ist ein auf Sprecherbefragungen gestützter Test, bei dem es darum geht, herauszufinden, welches Wort aus einer Reihe von Kohyponymen am ehesten dazu dienen kann, die anderen Kohyponyme zu definieren bzw. umschreibend zu ersetzen, ohne dass sich die Bedeutung des Satzes, in dem sie verwendet werden, ändert. Dabei wird angenommen, dass das Wort, das die Kohyponyme umschreibend ersetzen kann, das grundlegende und damit zentrale Wort der untersuchten Kohyponymenreihe darstellt:

„Thus in the lexical set *gobble, dine, devour, eat, stuff, gormandize* each of the words could be defined using ‘eat’ as a basic semantic feature, but it would be inaccurate to define *eat* by reference to any other of the words in the set (i.e. *dine* entails *eat* but *eat* does not entail *dine*).“ (Carter 1998, S. 37)

Je weniger sich ein Wort durch ein Kohyponym umschreiben lässt, d.h. je grundlegender oder übergeordneter es ist, desto zentraler ist dieses Wort. Je leichter es durch ein semantisch neutraleres Wort der Kohyponymenreihe ersetzbar ist, d.h. je stärker untergeordnet das Wort ist, desto weniger zentral ist seine Rolle.²²⁹ Dieser Test zieht also auf die Grund-

227. Vgl. das bei Dixon 1971, S. 441 aufgeführte Beispiel der Verben *look* (nuklear) und *stare* (non-nuklear).

228. Dabei stellt Dixon auch die Hypothese auf, dass diese Unterscheidung nicht arbiträr ist, sondern natürlichsprachlich, und weist das anhand einer Vielzahl von Gegenüberstellungen von lexikalischen Entsprechungen der beiden Sprachvarietäten nach (vgl. Dixon 1971, S. 443-451)

begrifflichkeit bzw. semantische ‘Neutralität’ eines Wortes innerhalb einer Reihe von Kohyponymen ab; implizit liegt er damit auch dem Verfahren zur Erstellung eines kontrollierten definitorischen Vokabulars, wie es Herbst 1986 beschreibt (vgl. Abschnitt 3.2) zugrunde (vgl. auch Carter 1998, S. 36).

Antonymie

Der zweite Test auf „coreness“ (Carter 1998, S. 38), den Carter anführt, geht davon aus, dass...

„...[the] less ‘core’ a word is, the more difficult it is to find an antonym for it. So, while *fat-thin* and *laugh-cry* are antonyms, it is problematic to locate precise antonyms for items occupying different points in the semantic space created by these words. Thus *emaciated*, *corpulent*, *obese*, *guffaw*, *sob* are by this definition less core items.“ (Carter 1998, S. 38)

Der Test zielt also darauf ab, diejenigen Wörter herauszuarbeiten, die die beiden semantischen Pole eines Spektrums von verwandten Bedeutungen besetzen, weil erst durch die lexikalische Besetzung der Pole das semantische Zwischenspektrum möglich wird.²³⁰

Kollokationsfähigkeit („ collocability“)

Der dritte Test, den Carter anführt, beruht auf der Annahme, dass...

„...the more core a lexical item is, the more partnerships it will contract with other lexical items. That is, if the words *bright*, *radiant*, *shiny*, *gaudy* are hypothesized as operating along a continuum from core to non-core, then one measure of the relative coreness of *bright* is that we have: *bright sun / light / sky / idea / colours / red, green / future / prospects / child*. *Radiant* collocates less widely. Since there are few true synonyms, it contracts its own partnerships but it is not generally able to collocate within the same range as *bright*. For example: *radiant light / smile / sun / flame? / *green / *idea / *prospects / future? / *child [...]*“ (Carter 1998, S. 38f.)

Was die Nützlichkeit dieses Tests einschränken kann, ist die Schwierigkeit, genau zu bestimmen, welche Kollokationen akzeptabel sind und welche nicht, weshalb korpusbasierte Frequenzuntersuchungen zu den jeweiligen Kollokationen nötig sind (vgl. ebda., S. 39). Einwenden ließe sich auch, dass man es hier vielleicht einfach nur mit verschiedenen Bedeutungsschattierungen zu tun habe.

„So, *shiny*, for example, creates its own collocations (e.g. *shiny nose*, *shiny coin*, *shiny car*) because it has a particular meaning which is different from *bright*, *radiant* or *gaudy*. Core words are then, according to this argument, simply polysemous words dressed up by collocation theory.“ (Carter 1998, S. 39)

229. Dabei räumt Carter ein, dass syntaktische und semantische Kriterien dafür, wann eine Ersetzung noch akzeptabel ist, nicht von einander zu trennen sind. Vgl. die Beispiele zu den Verben *rob* und *steal* bei Carter 1998, S. 48.

230. Dabei muss man hinzufügen, dass dieser Test vor allem für die bei Glück ((Hg.) 2000, S. 48) von anderen Typen von Antonymen unterschiedenen sogenannten „kontradiktorische[n] A[ntonymien]“ gilt, denn nur sie sind graduierbar und eröffnen ein semantisches Kontinuum von Übergangsbedeutungen zwischen sich, im Gegensatz etwa zu den „komplementäre[n] A[ntonymien]“, wie etwa *tot – lebendig*, *männlich – weiblich*, die keine Zwischenbedeutungen zulassen.

Carter verteidigt dennoch diesen Test, mit dem Argument, dass ein solcher Test wertvolle Informationen über stilistische Eigenschaften einzelner Wörter enthalten könne, die man nicht ignorieren dürfe.

Ergänzen kann man Carters Argumentation auch mit der Erklärung, dass die Wörter, die aufgrund eines solchen Tests auf Kollokationsfähigkeit als zentral beschrieben werden, Wörter sind, die im Zentrum verschiedener Achsen von Bedeutungsnuancen stehen. So würde etwa *shiny* eine Bedeutungsachse besetzen, die das buchstäblich Glänzende von konkreten Gegenständen beschreibt, während *radiant* eine Bedeutungsachse belegt, die den Aspekt des Strahlens hervorhebt. *Bright* hingegen wäre dann als so generisch in seiner Bedeutung aufzufassen, dass es die Bedeutungsachsen von sowohl *shiny* als auch *radiant* repräsentieren kann, also irgendwo da anzusiedeln ist, wo diese beiden Bedeutungsachsen sich treffen. Bei der Bestimmung des zentralen Wortes einer Reihe von Kohyponymen ginge es also darum, herauszufinden, welches Mitglied dieser Kohyponymenreihe die in ihr repräsentierten Bedeutungsachsen komplett repräsentiert.

Das Beispiel zeigt aber auch, dass systemimmanente Zentralität nicht unmittelbar mit der Zentralität für eine spezifische Anwendung zu tun hat. Denn die systemimmanente Zentralität von *bright* betrifft nur seinen Status innerhalb eines Kontinuums von Kohyponymen. Sie hilft, wenn es darum geht, zu entscheiden, welches der gegebenen Kohyponyme man auswählen soll; unbeantwortet bleibt aber die für einen konkreten zentralen Wortschatz oftmals viel wichtigere Frage, ob man gerade das durch dieses Kohyponym besetzte Bedeutungsspektrum im zentralen Wortschatz angeben soll oder ob nicht etwa ein ganz anderes Bedeutungsspektrum, zum Beispiel das von *hot* und seinen Kohyponymen, wichtiger ist.

Extension

Ein weiterer Test ist der bereits bei Stubbs 1986 beschriebene Test auf die „property of extension“ (Carter 1998, S. 39). Dieser Test beruht auf der Annahme, dass die Stellung eines Wortes im Wortschatz umso zentraler ist, je mehr Einträge dieses Wort in einem Wörterbuch auf sich vereint. Dabei gelten als Extensionen sowohl Unterbedeutungen als auch verschiedene Wortbildungen und Syntagmen wie etwa Komposita, Phraseologismen oder zusammengesetzte Verben (vgl. ebda., S. 39f.). Die Beispiele, die Carter hier nennt sind etwa Einträge zum Stichwort *run* wie *run-of-the-mill*, *run about*, *run up* (für Schulden), *run down* (im Sinne von ‘kritisieren’), *on the run*, *run in a car* oder – für das Stichwort *bright* – *bright and breezy*, *brighten up*, *look on the bright side*, *the brighter moments in one’s life*. Damit rückt dieser Test in die Nähe des traditionellen Selektionskriteriums der morphologischen Produktivität und des semantischen Expansionspotentials, wie sie vielfältig in der Grundwortschatzliteratur diskutiert werden (vgl. Abschnitt 4.5); außerdem ist er eng mit dem Test auf Kollokationsfähigkeit verwandt und macht damit die wechselseitigen Beziehungen zwischen den einzelnen Tests deutlich (vgl. Carter 1998, S. 40).

Gegen diesen Test ließe sich einwenden, dass er eigentlich nur die lexikographischen Entscheidungen der Wörterbuchverfasser widerspiegelt und damit kein rein systeminternes Testwerkzeug mehr darstellt, sondern auf der (subjektiven) Beurteilung von Lexikographen beruht. Hinzu kommt, dass unterschiedliche Notationen dazu führen können, dass in manchen Wörterbüchern Wortbildungsableitungen und Phraseologismen zu einem Stichwort nicht unmittelbar untereinander stehen, während sie das in anderen tun, und dass dies sogar innerhalb eines Wörterbuchs unterschiedlich gehandhabt werden kann. Daher sind solchermaßen ermittelte Werte nicht immer zuverlässig. Andererseits spiegelt die Anordnung und Unterteilung von Bedeutungen im Wörterbuch, wenn sie auf gründlicher lexikographischer Vorarbeit beruht, zumindest eine intersubjektiv nachvollziehbare Aufgliederung wieder, die zudem oftmals die einzige Möglichkeit ist, das Bedeutungsspektrum eines Wortes systematisch erfassen zu können. Behält man also die oben genannten Beschränkungen im Auge, kann dieser Test, vor allem in Kombination mit anderen Tests, wertvolle Resultate liefern.²³¹

Überordnung („superordinateness“)

Auch dieser Test beruht, wie schon die Tests auf Ersetzbarkeit und Kollokationsfähigkeit (vgl. S. 116), auf der Annahme, dass zentrale Wörter typischerweise grundlegende, d.h. unspezifische semantische Eigenschaften haben, die sie dazu befähigen, übergeordnete Bezeichnungen, d.h. Hyperonyme zu Gruppen anderer Wörter darzustellen. Hier nimmt man also an, dass die Bezeichnungen für Klassen von Objekten – *Blume* als übergeordnete Bezeichnung, die sowohl *Rose* als auch *Tulpe* umfasst; *Fahrzeug* als übergeordnete Bezeichnung, die sowohl Wörter wie *Auto*, *Bus*, *Lastwagen* umfasst – tendenziell zentrale Wörter darstellen. Damit sei aber nicht gesagt, räumt Carter ein,...

„...that all superordinates will be core words or that superordinateness itself is as easy to specify as is implied by the neat but limited examples cited above or in the literature on structural semantics (e.g. do *caravan*, *bulldozer*, *bicycle* all come under *vehicle*?). In fact, [...] as soon as such groupings of words are explored in use in communicative discourse, it is the case that superordinates are sometimes distinctly marked rather than unmarked. [...] [It] may be more productive in certain cases to consider items which are [...] one stage lower in the hierarchy of generality as being more core or unmarked [...] A schema for *furniture* might be represented [...] in such a way as to suggest that items such as *bed* or *chair* may be more appropriately superordinate. However, the existence of a further word *seat* which is more widely usable in so far as it can be used in more contexts than either *chair* or *furniture* further problematizes the relations between coreness and superordinate terms.“ (Carter 1998, S. 40f.)

Die von Carter hier geführte Diskussion macht deutlich, dass bei der Frage nach der Zentralität eines Wortes die Position eines Begriffs in einer vertikalen Hierarchie von Unter- und Überbegriffen mit der Prototypenhaftigkeit und Repräsentativität eines Begriffs für seine Kohyponyme konkurriert: Ein Möbelstück ist zwar ein generischer Sammelbegriff und könnte daher als zentral gelten, andererseits sind *Stuhl* oder *Tisch* Bezeichnungen für sehr

231. In Abschnitt 9.3.3 werden wir die Frage nach dem Zusammenhang zwischen der Extension eines Wortes und seiner Zentralität anhand der praktischen Analyse genauer betrachten.

typische Möbelstücke, die möglicherweise bei Disponibilitätstests viel eher genannt würden als *Möbelstück*. Diese Vermutung wiederum knüpft an die Beobachtung an, dass es für manche Gruppen von Wörtern (*Buntstift, Federhalter, Bleistift, Filzschreiber, Kugelschreiber*) keine natürlichen Hyperonyme gibt, so dass diese Lücke zwar durch die Bildung eines neuen Wortes, etwa eine Wortbildungskonstruktion (*Schreibgerät*), geschlossen werden kann, dieses Hyperonym aber eine okkasionelle Bildung bleibt und kaum wirklich von den Sprechern verwendet wird (und zum Beispiel auch nicht in Wörterbüchern auftaucht).

Hier zeigt sich wieder, dass Zentralität im Sinne der Zugehörigkeit eines Wortes zum Kern des lexikologischen Systems einer Sprache nicht gleichzusetzen ist mit Zentralität im Sinne seiner Zugehörigkeit zu einem konkreten, anwendungsorientierten zentralen Wortschatz. Sowohl bei didaktischen als auch analytischen zentralen Wortschatzen spielt die tatsächliche Nützlichkeit (für die didaktischen zentralen Wortschatze) bzw. das tatsächliche oder wahrscheinliche Vorkommen (für die analytischen zentralen Wortschatze) eine größere Rolle als die systembedingte Zentralität eines Wortes.

Kulturelle Neutralität („culture-free“)

Ein weiterer Test geht von der Annahme aus, dass...

„...[the] more core a word is the less likely it is to be restricted to culture-specific uses. Stubbs (1986a) [= Stubbs 1986, V. H.] discusses the extent to which it is rare for words in certain categories to be borrowed from one language into another. Thus, in English, while certain cooking terms or furniture terms (e.g. *pouf, chaise-longue*) are borrowed from French, words for basic bodily functions, natural physical phenomena, dimensions of size and shape, words for pronouns etc. will form core components in a language. [...] The notion itself must, however, be relative to particular geographical areas. For example, the dimensions of human shape conveyed by the word *thin* would be 'neutral' in most Western cultures but would be marked for certain African cultures; also, related words with positive and sex-specific associations in Western culture such as *slim* and *slender* may be reversed elsewhere. By this test *round* would be more core as a word than *thin*.“ (Carter 1998, S. 41f.)

Abgesehen von der geographischen Relativität der 'kulturellen Neutralität', die Carter diskutiert, ist es auch interessant festzuhalten, dass die *basic word list* der Glottochronologie auf einer ähnlichen Argumentation beruht: Gerade weil man annimmt, dass die zentralsten Begriffe am wenigsten Phänomenen des Sprachwandels wie Entlehnungen unterworfen sind, verwendet man sie, um den Trennungszeitpunkt zweier Sprachen zu bestimmen. Was die Glottochronologen dabei allerdings ignorierten, war die bei Carter erwähnte beschränkte Gültigkeit des Kulturraumes, dass eben kulturell neutrale Begriffe nur innerhalb eines abgegrenzten Raumes einander ähnlicher Kulturen neutral sind.

Eignung zur Zusammenfassung („summary“)

Bei diesem Test handelt es sich um eine Informantenbefragung. Ihr liegt die Beobachtung zugrunde, dass Informanten, die gebeten werden, eine Handlung oder ein Ereignis zusammenzufassen, Wörter verwenden, von denen man annimmt, sie haben eine zentrale Stellung

im Wortschatz, weil sie Informationen in kompakter, zusammengefasster und vor allem wertneutraler Form vermitteln:

„Thus, in summaries of Hemingway's short story *Cat in the Rain* informants unanimously preferred the term *cat* to alternatives available within and outside the story such as *kitty*, *pussy*, *feline*, *mog* [...]. This seems to suggest that summaries are a genre in which it is perceived that the propositions conveyed should be represented without stylistic, rhetorical or evaluative overlay. It explains why for literary critics there will always be a 'heresy of paraphrase'. Further tests could be designed here by asking informants to summarize texts constructed from sets of lexically related items.“ (Carter 1998, S. 42)

Aus der Diskussion bei Carter wird deutlich, dass Wörter, die sich zur Zusammenfassung von Informationen eignen, oft zugleich auch stilistisch neutrale Wörter darstellen. Ähnlich argumentiert der folgende Test.

Konnotativität („associationism“)²³²

Dieser Test schließt sich an die Arbeit von C.E. Osgood an (vgl. Osgood / Suci / Tannenbaum 1957) und geht von folgender Informantenbefragung aus:

„[I]nformants were asked to rate words along scales. The scales were those revealed by Osgood as being most relevant to semantic space analysis – an evaluation scale and a potency scale and to these were added a formality scale. The results of the tests are interesting in themselves for word definition but for our purposes here it is not insignificant that certain words from each lexical set gravitated towards a central or neutral point in each scale. In the case of a lexical set involving *emaciated*, *skinny*, *lean*, *slender*, *slim*, *thin*, *weedy*, the word *thin* came out along formality, potency and evaluation scales around the mid-point of a 10-point cline. [...] Other words from the set converged less consistently, thus carrying greater associations and being less neutral.“ (Carter 1998, S. 42)

Damit zielt dieser Test auf die Ermittlung derjenigen Wörter ab, die innerhalb einer Kohyponymenreihe die stilistisch neutralen Wörter darstellen.

Themenneutralität („neutral field of discourse“)

Dieser Test beruht auf der Beobachtung, dass zentrale Wörter oft keinen Hinweis darauf geben, um welches spezifische Thema es bei einem Text oder bei einem Gespräch geht:

„Thus the words *galley*, *port* and *starboard*, *fore* and *aft*, *knots* immediately recall nautical or aeronautical contexts while corresponding items like *kitchen*, *left* and *right*, *miles per hour* etc. do not.“ (Carter 1998, S. 43)

Auch hier wird wieder deutlich, dass systemimmanente Zentralität und anwendungsspezifische Zentralität unterschiedlich sind, denn: „This is of course not to say that for specific purposes such nautical lexis is not core in a quite crucial sense.“ (ebda., S. 43). Zugleich bestätigt sich hier auch die in Abschnitt 3.3 angesprochene Annahme, dass etwa eine Stopworlliste für das *Information Retrieval* oder der Neutralwortschatz der automatischen Text-

232. Auch wenn Carter sich hier explizit auf Assoziationen bzw. das Fehlen von Assoziationen bezieht, wähle ich hier den Begriff der *Konnotation*, weil die semantischen 'Beiklänge', auf die sich der Test von Osgood / Suci / Tannenbaum 1957 bezieht, typische Beispiele einer konnotativen Bedeutung darstellen (vgl. Glück (Hg.) 2000, S. 367, der die in diesem Test abgefragten Konnotationen als Beispiel für den Subtypus der „emotive[n] K[onnotation]“ anführt).

klassifikation keine absoluten Größen sind, sondern relativ zu dem konkreten Textbestand, auf den sie angewandt werden sollen.

Stilistische Neutralität („neutral tenor of discourse“)

Der letzte der von Carter angeführten Tests auf die Zentralität von Wörtern zielt in dieselbe Richtung wie der Konnotativitätstest (vgl. S. 121):

„Here core words will be those which emerge as neutral in formality tests. For example, of *podgy*, *corpulent* and *fat*, *fat* will in most tests emerge for most informants as the one word expressing greatest neutrality in respect of tenor of discourse.“ (Carter 1998, S. 44)

Eingeschränkt werde dieser Test jedoch dadurch, dass sich nicht alle Gruppen von semantisch identischen, pragmatisch aber verschiedenen Wörtern auf diese Weise testen ließen. Dies sei durch die Schwierigkeiten bedingt, die das Abfragen von (kontextlos präsentierten) Gruppen von Wörtern für die Messung von Bedeutungen mit sich bringe. Vor allem sei...

„...formality in language [...] not a wholly lexical matter. A good illustration of these kinds of difficulties would be the words *whine*, *moan*, *grumble*, *complain*, *bellyache*, *whinge*, *lodge a complaint*. Certain items here carry associative values which one is usually happier to attribute to others rather than to oneself and no word from this particular set could be said to be culture- or value-free.“ (Carter 1998, S. 44)

Zugleich illustriert dieses Beispiel, dass...

„...any communicative discourse in core vocabulary may also be initially restricted in terms of the semantic notions involved and would probably be a peculiarly ‘detached’ noncommittal activity and, for better or for worse, be as ‘unliterary, unaesthetic and unemotional’ as Quirk (1982) claims.“ (Carter 1998, S. 44)

Neben den zehn hier beschriebenen Zentralitätstests referiert Carter auch noch einige bei Stubbs 1986 angeführte Merkmale, die man als Kennzeichen zentraler Wörter bezeichnen kann. Einige davon gelten spezifisch für das Englische.²³³ Hervorheben möchte ich aber die Beobachtung, dass zentrale Wörter häufig morphologisch unregelmäßig seien, also etwa unregelmäßig flektiert werden, „perhaps reflecting that such words have sufficient centrality to resist regularization over a period of time“ (Carter 1998, S. 45).²³⁴ Allerdings muss diese letzte Aussage insofern präzisiert werden, als es nicht unbedingt die Zentralität dieser Wörter ist, die direkt dafür sorgt, dass diese Wörter ihre Irregularitäten behalten können, sondern dass die Zentralität eines Wortes möglicherweise zu einer höheren Frequenz führt und diese wiederum dafür sorgt, dass sich irreguläre Formen im Gedächtnis der Sprecher halten können.²³⁵

233. Vgl. etwa die Angabe, dass Adjektive, die auf -y enden, weniger zentral sind, weil sie nicht stilistisch neutral sind, oder die Angabe, dass Wörter mit angelsächsischer Wurzel eher zentral sind als Wörter mit romanischer Wurzel (vgl. Carter 1998, S. 45).

234. Dies lässt sich auch im Deutsche beobachten, wo starke, d.h. unregelmäßig konjugierte Verben häufig solche sind, die grundlegende Aktivitäten bezeichnen (*sitzen*, *stehen*, *gehen*, *essen*)

235. Vgl. etwa auch die Rolle des Frequenzbegriffs in der kognitiven Linguistik, wie sie beispielsweise in den Darstellungen bei Glück (Hg.) 2000, S. 197 oder Schindler 2002, S. 42 deutlich wird.

Wichtig für die weitere Diskussion sind darüberhinaus zwei wesentliche Annahmen des hier entworfenen Bildes eines Wortschatzzentrums:

Grundeinheit eines solchen Wortschatzzentrums ist das (wie auch immer vage definierte) Wort, nicht die einzelne Lesart bzw. das Semem. Diese Annahme ist fast konstituierend, da oft erst die Mehrdeutigkeit von Wörtern ihre Zentralität erkennen lässt.

Zudem sind – hält man sich streng an Carter – nur lexikalische Wörter, also Autosemantika, Teil dieses Wortschatzzentrums, auch wenn Synsemantika, also Funktionswörter, auch zentral für das Sprachsystem sind (vgl. Carter 1998, S. 36). Sie bilden jedoch eine gesonderte, geschlossene Klasse und brauchen durch die hier beschriebenen Zentralitätstests nicht ermittelt zu werden. Damit unterscheidet sich dieses Wortschatzzentrum deutlich von verschiedenen Beispielen der Grundwortschatzdiskussion, bei denen die Aufnahme von Funktionswörtern und die Forderung nach der Orientierung an Semem als Grundeinheit zu den wesentlichen Anforderungen an einen zentralen Wortschatz gehören.

Wie die kurze Diskussion einzelner Tests gezeigt hat, ist die Arbeit mit diesen Tests nicht unproblematisch. Zunächst bleibt festzuhalten, dass zwischen den einzelnen Tests ein beträchtlicher Teil an Überschneidungen besteht, d.h. sie testen auf verschiedene Weise oft ähnliche oder gleiche Eigenschaften von Wörtern. Als Beispiel einer solchen Überschneidung führt Carter (ebda., S. 44) das Kohyponymenset *odour, stench, aroma, smell, fragrance, stink, scent* und *bouquet* an und macht deutlich, dass Wörter, die bei dem Test auf Konnotativität einen geringen Grad an Formalität aufweisen (*stench, stink*) zugleich auch im Test auf stilistische Neutralität als nicht neutral ausgewiesen werden; oder dass die Wörter *smell* und *scent*, die häufig als syntaktische Ersetzungen für andere Wörter desselben Kohyponymensets verwendet werden, ebenso regelmäßig für die Zusammenfassung einer Beschreibung oder Handlung, bei der dieses Set von Kohyponymen vorkommt, verwendet werden. Insgesamt, so Carter, hätten Tests an anderen Gruppen von Wörtern bestätigt, dass es eine gewisse Überschneidung zwischen dem Konnotativitätstest und dem Test auf stilistische Neutralität bzw. zwischen dem Test auf syntaktische Ersetzung und dem Test auf die Eignung zur Zusammenfassung gibt (vgl. ebda.).

Das von Carter angeführte Beispielsset zeigt auch ein weiteres Problem, nämlich dass es Fälle gibt, in denen zwei Wörter eines Sets von Kohyponymen – im vorliegenden Fall *smell* und *scent* – beide in gleicher Weise zentral erscheinen, woraus für Carter die Frage erwächst: „[Are] all the tests of equal relevance in determining coreness and should there be then a hierarchical ranking of the relative criteriality of the individual tests?“ (Carter 1998, S. 45). Dem von Carter im Zusammenhang mit der Überschneidung mehrerer Tests ange deuteten Vorschlag, man könne eventuell einige Tests aus dem Repertoire streichen, kann man daher bisher nicht zustimmen. Bedenkt man die Unsicherheiten, die mit jedem der Tests verbunden sind, kann es durchaus sinnvoll sein, zunächst alle von Carter vorgeschlagenen Tests beizubehalten, um unsichere Ergebnisse eines Tests durch die Ergebnisse eines

anderen Tests abstützen oder widerlegen zu können. Eine ähnliche Schlussfolgerung zieht auch Carter, wenn er die akkumulative Wirkung des Testrepertoires hervorhebt:

„It is, however, important to underline at this point that no single test will on its own be a sufficiently systematic measure and that core vocabulary itself has no unambiguously clear boundaries. That is why it is more accurate to speak of clines and gradients and of degrees of coreness in words and why as a general rule it is claimed that, once the tests have been satisfactorily tested, the more tests a word passes the greater the degree of coreness it will have.“ (Carter 1998, S. 45)

Was die hier angeführten Tests, die ja ausschließlich linguistische Tests sind, ebenfalls nicht abfangen, ist die subjektiv empfundene Zentralität einzelner Wörter, wie sie etwa Disponibilitäts- und Vertrautheitstests zutage fördern können (vgl. Carter 1998, S. 237).²³⁶ Vor allem aber reflektieren die hier beschriebenen Zentralitätstests möglicherweise nur eine statische Augenblicksaufnahme eines mentalen Lexikons eines oder mehrerer individueller Sprecher, die wiederum nur mit kontextfreien Listen von Wörtern konfrontiert sind. Diese Begrenzungen tragen unter anderem dazu bei, dass ein auf diese Weise abgegrenztes Wortschatzzentrum nicht unmittelbar auf eine konkrete Anwendungssituation, etwa einen didaktisch ausgerichteten zentralen Wortschatz übertragen werden kann (vgl. Carter 1998, S. 237f.).

5.5.2 Lexikologisches Wortschatzzentrum und anwendungsspezifische zentrale Wortschätze – eine konzeptuelle Trennung

Viele Indizien und Argumente sprechen also für die Existenz eines Zentrums des Wortschatzes einer Sprache. Was die Diskussion allerdings erschwert, ist das unklare Verhältnis zwischen diesem Zentrum und den Anwendungen des Konzeptes zentraler Wortschätze, insbesondere der des Grundwortschatzes. Der Grund für dieses unklare Verhältnis ist sicher darin zu suchen, dass viele der in Abschnitt 4.5 beschriebenen traditionellen Selektionskriterien den von Carter angeführten Zentralitätstests ähneln, vor allem aber auch, dass, wie wir in Abschnitt 4.1 gesehen haben, in der Diskussion keine klare terminologische Trennung zwischen zentralen Wortschätzen und Wortschatzzentrum vollzogen wird. Wir wollen daher eine terminologische Klärung vornehmen:

Zentrale Wortschätze wie etwa der Grundwortschatz sind zu unterscheiden vom lexikologischen Zentrum einer Sprache. Zentrale Wortschätze sind anwendungsorientierte Wortschatzausschnitte, die, will man eine solche Zuordnung vornehmen, eher im Bereich der *parole* angesiedelt und damit Teil der von Kühn für den Grundwortschatz geforderten „funktionalen Sprachauffassung“ (Kühn 1979, S. 20) sind. Sie sind in ihrem jeweiligen Teilbereich der angewandten Linguistik zuhause und werden nach anwendungsspezifischen Kriterien erarbeitet. Dabei ist neben dem Auswahlverfahren auch die Frage nach der anzunehmenden Grundeinheit (Lexem, Semem) von der Anwendung abhängig. Das Bün-

236. Deshalb fordert Carter auch eine Ergänzung der Testserie durch psycholinguistisch orientierte Tests (vgl. Carter 1998, S. 46).

del von Eigenschaften, durch das sich die einem zentralen Wortschatz zugehörigen Wortschatzelemente auszeichnen, bezeichnen wir als *anwendungsspezifische Zentralität*.

Das lexikologische Zentrum einer Sprache dagegen ist ein Teil des Sprachsystems, also der *langue*, und damit Teil der von Kühn als „systemisch“ bezeichneten Sichtweise von Sprache.²³⁷ Dabei bezeichnet der Terminus *lexikologisches Zentrum einer Sprache* den Kern des lexikalischen Bestandes einer Sprache, der nach systemlinguistischen Kriterien und nicht etwa anhand von spezifischen Kommunikationsbedürfnissen erarbeitet wird. Grundeinheit des lexikologischen Zentrums ist das Wort bzw. der Begriff. In seiner Systemorientiertheit ist es den Häufigkeitswörterbüchern verwandt (aber nicht mit ihnen gleichzusetzen), indem es einen Teilbereich der Sprache als System zu beschreiben bzw. zu charakterisieren versucht, unabhängig von einer konkreten Anwendung. Die Eigenschaften, durch die sich die diesem Zentrum zugehörigen Wortschatzelemente auszeichnen, bezeichnen wir als *systemimmanente Zentralität*.²³⁸

Dabei gehen wir davon aus, dass diese beiden Typen von Wortschatzausschnitten grundsätzlich nicht deckungsgleich sind. Das heißt, ein konkreter zentraler Wortschatz für eine spezifische Anwendung kann im Prinzip ganz andere Wortschatzelemente enthalten als diejenigen Elemente, die sich aufgrund systemlinguistischer Kriterien als zentral für das Wortschatzsystem einer Sprache an sich ergeben hat. Damit folgen wir Carter 1998, der die potentielle Diskretheit dieser zwei Wortschatzausschnitte am Beispiel didaktischer zentraler Wortschätze, bei denen beispielsweise die Erlernbarkeit von Wortschatzelementen wichtiger ist als systemlinguistische Zentralität, illustriert (vgl. Carter 1998, S. 236).²³⁹

Im Extremfall können lexikologisches Wortschatzzentrum und ein zentraler Wortschatz zwei völlig diskrete Wortschatzmengen darstellen.²⁴⁰ Das muss aber nicht zwangsläufig so sein. Denkbar sind zentrale Wortschätze, die viele Wortschatzelemente mit dem lexikologischen Wortschatzzentrum gemeinsam haben und damit in seine Nähe rücken. Das ist etwa der Fall bei Grundwortschätzen, die eine thematisch möglichst breite Basis des Wortschatzes einer Sprache einem breiten, wenig spezifizierten Benutzerkreis didaktisch vermitteln sollen. Umgekehrt ist auch vorstellbar, dass die systemimmanente Strukturierung des Wortschatzsystems einer Sprache in Wörterbüchern in Form einer entsprechenden Annotation zentraler Wortschatzelemente zum Ausdruck kommt, als eine Art linguistischer Information für Fachleute. Es sind wohl gerade diese Wechselbeziehungen zwischen den beiden

237. Für eine detaillierte Gegenüberstellung dieser beiden Auffassungen von Sprache im Zusammenhang mit der Grundwortschatzlexikographie vgl. Kühn 1979, S. 19f.

238. Engl. *coreness*, in Anlehnung an Carter 1998. – Das lexikologische Wortschatzzentrum hat trotz seiner Eigenschaft der ‘coreness’ nichts zu tun mit dem in der Korpuslinguistik verwendeten Begriff des *core corpus*. Das ist ein Textkorpus, das auf einer thematisch und texttypisch breiten Basis aufbaut und als neutrale Vergleichsgrundlage bei komparativen Korpusstudien herangezogen wird (vgl. Kennedy 1998, S. 20).

239. Ähnlich auch Carter 1998, S. 46.

240. Vgl. etwa das Beispiel zentraler Wortschatzeinheiten in der Kommunikation zwischen Eltern und Kleinkind bei Carter 1998, S. 35f.

Konzepten, die dazu beigetragen haben, dass die Diskussionen zum Grundwortschatz und anderen zentralen Wortschätzen sie terminologisch vermengen. Erschwerend kommt hinzu, dass man auch beim lexikologischen Zentrum einer Sprache von einer Gruppe von Varianten sprechen kann, die durch die Unterschiede zwischen Sprechern und sogar durch die Unterschiede bei nur einem Sprecher aber variierender Gesprächssituation bedingt sein können (vgl. Carter 1998, S. 34f.). Dennoch bzw. gerade deshalb ist eine terminologische Unterscheidung der beiden Konzepte, die eine Wechselbeziehung zwischen beiden durchaus zulässt, sie aber als zwei eigenständige Konzepte mit einer gewissen internen Schwankungsbreite betrachtet, wichtig.

Zum Schluss muss noch eine Gemeinsamkeit der beiden Konzepte hervorgehoben werden: Bei beiden handelt es sich nicht um Grenzziehungen in Form von Ja-Nein-Unterscheidungen, sondern um graduelle, kontinuierliche Übergänge.²⁴¹

5.5.3 Zusammenfassender Rückblick

Mit der bisher geführten Diskussion im Blick halten wir also fest:

Es gibt klare Indizien dafür, dass es einen Bereich des Gesamtwortschatzes einer Sprache gibt, den man als dessen Zentrum bezeichnen könnte, weil die ihm zugehörigen Elemente besondere Eigenschaften aufweisen, die sie als (kognitiv) besonders wichtig auszeichnen.

Den anfangs genannten Widerspruch zwischen der Abhängigkeit zentraler Wortschätze von der intendierten Anwendung im Allgemeinen und dem Grundwortschatz im Besonderen gegenüber einem eventuellen festen Zentrum des Wortschatzes einer Sprache wollen wir auflösen, indem wir eine terminologische Abgrenzung vornehmen. *Grundwortschatz* oder *basic vocabulary* sind Bezeichnungen für zentrale Wortschätze, also anwendungsorientierte Wortschatzausschnitte, während wir die Gruppe der für das Sprachsystem als solches zentralen Wortschatzelemente als das *lexikologische Zentrum einer Sprache* bezeichnen.

Die terminologische Trennung zwischen zentralen Wortschätzen und lexikologischem Zentrum soll aber auch Raum für Wechselbeziehungen zwischen den beiden Konzepten geben. Die Eigenschaften, die nach systemimmanenten Kriterien ein Wort als zentral einstufen, können auch im Anwendungszusammenhang eine Rolle spielen. So könnte etwa die Erkenntnis aus der Spracherwerbsforschung, dass Überextensionen von Kindern aktiv als eine kommunikative Strategie benutzt werden, um Dinge zu bezeichnen, für die sie noch keine Wörter besitzen (vgl. S. 115), in der Grundwortschatzlexikographie zu einer verstärkten Berücksichtigung von stark polysemen Wörtern führen, die eine flexible Anwendung zulassen.

241. Vgl. auch Abschnitt 7.3.5 und Carter 1998, S. 238.

Kapitel 6: Entwurf eines Arbeitsmodells

6.1 Einführung

Nachdem wir bisher in einer Art Bestandsaufnahme gesehen haben, in welchen Dimensionen sich der Wortschatz betrachten läßt, welche Typen von zentralen Wortschätzen es in der linguistischen Praxis bereits gibt, welche Probleme mit ihrer Erarbeitung und Darstellung verbunden sind und auf welchen Wegen man bisher versucht hat, solche zentralen Wortschätze zu erarbeiten, ist es nun nötig, die so gesammelten Informationen zu systematisieren und aus ihnen ein Arbeitsmodell zu erarbeiten.

Das hier zu entwerfende Arbeitsmodell hat dabei zwei Ziele:

Zum einen sollen mit seiner Hilfe die bisherigen Ergebnisse der forschungsgeschichtlichen Analyse zusammengefasst und terminologisch festgelegt werden. Wie sich aus den Diskussionen in Kapitel 4 ergeben hat, ist es dabei erforderlich, sowohl ein grundlegendes Wortschatzmodell als auch eine Definition des Konzeptes zentraler Wortschätze zu erarbeiten.

Zum anderen geht es aber auch darum, einen theoretisch-methodischen Rahmen aufzubauen, auf den sich die nachfolgende praktische Untersuchung beziehen kann.

Das Arbeitsmodell besteht aus zwei Teilen:

- einem generischen Wortschatzmodell: Grundlage ist ein generisches Modell der Makrostruktur von Wortschätzen, ein lexikologisches Modell, das die in Kapitel 5 besprochenen Dimensionen des Wortschatzes berücksichtigt und beispielhaft auf den Gesamtwortschatz des Deutschen anwendet.
- einer Definition des Konzeptes zentraler Wortschätze: Vor dem Hintergrund dieses generischen Wortschatzmodells wird dann, sowohl deduktiv (durch die Betrachtung des Wortschatzmodells) als auch induktiv (durch die Betrachtung der in Kapitel 3 beschriebenen Anwendungstypen des Konzeptes zentraler Wortschätze) eine Definition des lexikographischen Konzeptes zentraler Wortschätze entwickelt.

Hervorzuheben ist dabei, dass das Wortschatzmodell und das Konzept zentraler Wortschätze, auch wenn sie inhaltlich eng miteinander verwoben sind, sachlich und damit terminologisch deutlich voneinander zu unterscheiden sind: Während das Wortschatzmodell ein lexikologisches Konstrukt ist, ist das Konzept zentraler Wortschätze ein lexikographisches. Dabei nimmt das Konzept gegenüber dem Wortschatzmodell die Rolle eines Werkzeugs ein, mit dessen Hilfe man den Kernbereich eines Wortschatzes erarbeiten und annähernd abgrenzen kann, während das Modell eine Art 'Landkarte' über den zu untersuchenden Wortschatz darstellt.

6.2 Das generische Wortschatzmodell

Bei der Entwicklung des generischen Wortschatzmodells wollen wir an die in Kapitel 5 besprochenen Dimensionen des Wortschatzes und insbesondere dem in Abschnitt 5.4 dargestellten Kern-Peripherie-Modell des Prager Strukturalismus anknüpfen.

Wir nehmen an, dass sich jeder Wortschatz – der Wortschatz des Deutschen insgesamt, aber auch der Wortschatz einer Sondersprache oder eines Dialektes – nicht nur in semantische oder grammatikalische Klassen zerlegen lässt, sondern auch nach der Zentralität der einzelnen Wortschatzeinheiten strukturiert werden kann und ein Kontinuum von zentralen hin zu peripheren Einheiten darstellt. Um dem Kontinuum-Gedanken gerecht zu werden, aber zugleich auch eine gewisse Schichtung dieser Wortschatzstruktur vornehmen zu können, wollen wir statt von einer einfachen Unterscheidung zwischen Kern und Peripherie von einem Drei-Schichten-Modell ausgehen (vgl. Abb. 6-1).

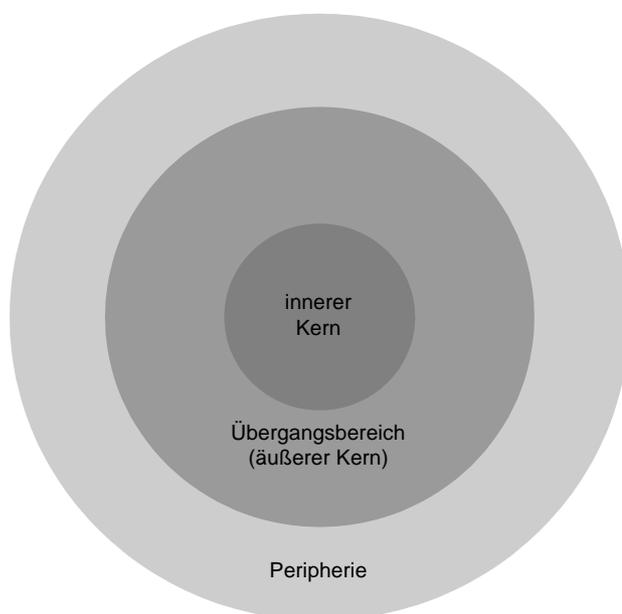


Abb. 6-1: Das generische Wortschatzmodell

Ausgehend von den Eigenschaften, die nach dem Kern-Peripherie-Modell die Mitglieder der jeweiligen Schichten prägen – der Grad an Stabilität des jeweiligen Elements sowie der Umfang an gegebenen Merkmalen, die das Element mit anderen Elementen teilt – und in Anlehnung an das Skalenmodell von Kalverkämper (vgl. Abschnitt 5.3.1), lassen sich die jeweiligen Schichten folgendermaßen beschreiben:

Mitglieder des inneren Kerns sind stabile Bestandteile des Wortschatzes, die alle Merkmale eines gegebenen Sets derjenigen Eigenschaften tragen, die für den zu gliedernden Wortschatz als wichtig und spezifisch betrachtet werden. Dabei knüpfen wir indirekt an Kalverkämper an, indem wir das Skalenhafte seines Modells der Fachsprachlichkeit auch für die die hier betrachteten Eigenschaften übernehmen. Darüber hinaus wollen wir diejenigen Wortschatzeinheiten als stabil bezeichnen, die im Wortschatz wohletabliert und gebräuchlich sind.

In der Peripherie befinden sich Wörter, die entweder vergleichsweise neu sind oder auf dem Weg sind zu veralten, also ihren Status der Wohletabliertheit noch gewinnen müssen oder bereits verlieren. Vor allem aber bewegen sich Wörter, die nicht alle Merkmale des zugrundeliegenden Merkmalssets haben, in der Peripherie des Wortschatzes.

Wörter, die weder eindeutig dem inneren Kern noch der Peripherie zuzuordnen sind, bewegen sich im Übergangsbereich, der hier als äußerer Kern bezeichnet werden soll.

Auf den ersten Blick scheint die unmittelbare Übertragung dieser ursprünglich auf die Phonetik zugeschnittenen Modellierung auf das System des Wortschatzes einer Sprache unproblematisch, auch wenn die Frage nach der Stabilität bzw. Wohletabliertheit eines Wortschatzelements nicht immer auf Anhieb zu beantworten ist.²⁴²

Auf den zweiten Blick zeigt sich jedoch, dass gar nicht so klar ist, welche Eigenschaften ein Element des Wortschatzes in ausgeprägtem Maße besitzen muss, um der einen oder anderen Schicht zugeordnet werden zu können. Das Problem beginnt bereits damit, dass man entscheiden muss, woraus das Set von Eigenschaften, an dem man das jeweils zu beurteilende Wortschatzelement messen soll, besteht. Gehören zu einem solchen Merkmalsset etwa 'typische' Worteeigenschaften wie Flektiertheit, Austauschbarkeit im Satz, die Eigenschaft, eine "minimale freie Form" (Glück 2000, S. 792) zu sein? Oder gehört dazu etwa die Fähigkeit zur Referenz? Gehört dazu die Eigenschaft, nicht fachsprachlich markiert zu sein, spielt die regionale Festlegung eines Wortes eine Rolle?

Sieht man sich aber einmal an, welchem Zweck der Abgleich eines Wortschatzelements mit einem gegebenen Set von Eigenschaften dient, dann wird Folgendes deutlich: Mit Hilfe der Frage, wie sehr ein gegebenes Wortschatzelement ein gegebenes Set von Eigenschaften aufweist, soll es möglich sein, zu beurteilen, wie 'gut' das Element die Gesamtmenge repräsentiert und wie zentral dieses Element für die Gesamtmenge ist. Wie sich diese Zentralität bestimmen lässt, ist jedoch wiederum abhängig von den Eigenschaften der Menge, die strukturiert werden soll, und von der Anwendung eines zentralen Wortschatzes, für die der Wortschatz strukturiert werden soll. Daher lässt sich im Rahmen des generischen Modells der Makrostruktur eines Wortschatzes eine konkrete Menge an Eigenschaften nicht angeben. Die konkrete Beschreibung dieses Merkmalssets ist Aufgabe der Umsetzung des Modells auf den jeweiligen Anwendungsfall.

So wird etwa für den konkreten Fall eines fremdsprachendidaktischen Grundwortschatzes das Set an Eigenschaften, anhand derer die Zentralität eines Wortes bestimmt werden soll, aus den Merkmalen der Stabilität, der Gebräuchlichkeit bzw. Frequenz, der Neutralität, der Vertrautheit und der kommunikativ-thematischen Relevanz eines Wortes bestehen, während das Merkmalsset zur Bestimmung des Neutralwortschatzes für die automatische Text-

242. Vgl. Abschnitt 8.6 zur Diskussion der Operationalisierbarkeit des Merkmals der Stabilität in der hier vorgenommenen praktischen Untersuchung.

klassifikation vermutlich aus den Eigenschaften der Distribution, der Stabilität und der Neutralität bestehen wird.

6.3 Das Konzept zentraler Wortschätze und seine Definition

Betrachtet man die in Kapitel 3 beschriebenen Typen von Wortschätzen, die das Konzept eines zentralen Wortschatzes anwenden, dann wird deutlich, dass ihnen allen folgende drei Charakteristika gemeinsam sind:

In allen Fällen wird versucht, der unstrukturierten Menge eines Gesamtwortschatzes einer Sprache, Sprachvarietät oder Textsammlung eine auf einer bestimmten Rangfolge beruhende Struktur zu geben. In der Regel geht es dabei um das Prinzip der Effektivisierung bzw. Präzisierung – dadurch, dass man sich auf den zentralen Bereich eines Wortschatzes konzentriert, hofft man beispielsweise, das Erlernen einer Fremdsprache oder die Klassifikation von Texten anhand gegebener thematischer Kategorien effektiver bzw. präziser zu gestalten. Und schließlich zielen alle Ansätze auf eine quantitative Begrenzung des Gesamtwortschatzes einer Sprache, Sprachvarietät oder Textsammlung ab.

Darauf aufbauend kann man nun das Konzept eines zentralen Wortschatzes folgendermaßen definieren:

Ein zentraler Wortschatz ist die anwendungsorientierte, bewusst quantitativ begrenzte und wohldefinierte Submenge eines Wortschatzes einer Sprache, Sprachvarietät oder Textsammlung, die nach bestimmten Kriterien durch Selektion gewonnen wurde.

Das Selektionsverfahren ist dabei direkt abhängig von der Anwendung, für die der zentrale Wortschatz entwickelt werden soll, und zielt darauf ab, die für diese Anwendung zentralen Wortschatzelemente von den nicht-zentralen zu trennen.

Was die Zentralität eines Wortschatzelementes bestimmt, aus welchen konkreten Eigenschaften sich die Zentralität zusammensetzt, ist dabei abhängig von dem konkreten Ausgangswortschatz und der spezifischen sprachdidaktischen oder computerlinguistischen Anwendung, für die der zentrale Wortschatz erstellt werden soll.

Der Wortschatz einer Sprache, Sprachvarietät oder Textsammlung stellt dabei den Ausgangswortschatz dar, zu dem der zentrale Wortschatz eine Teilmenge bildet.

Die Definition der Grundeinheit eines zentralen Wortschatzes wiederum ist abhängig davon, wie die Grundeinheit des Ausgangswortschatzes definiert ist, prinzipiell können aber der Begriff, das Semem oder das Lexem die Grundeinheit bilden.

Im Detail betrachtet stützt sich diese Definition auf folgende Kriterien:

- Anwendungsorientiertheit und bewusste quantitative Begrenzung:

Diese beiden Elemente der Definition stellen die konstituierenden Merkmale eines zentralen Wortschatzes dar. Eine Submenge eines Wortschatzes ist dann ein zentraler Wortschatz, wenn sie einem bestimmten sprachdidaktischen oder computerlinguistischen Zweck dient und bewusst quantitativ begrenzt wird.

Damit schließt diese Definition Wortschatzsammlungen aus, die die Anzahl ihrer Mitglieder nicht nach deren Zentralität im Verhältnis zu anderen Wortschatzelementen, sondern nach anderen, nicht-hierarchisierenden Kriterien auswählen. Vor allem schließt diese Definition Wortschatzsammlungen aus, die zwar den Wortschatz von einem bestimmten Blickwinkel aus darstellen, ihn aber nicht aus der Orientierung an einem bestimmten linguistischen Zweck heraus, sondern allein aus praktischen Gründen quantitativ begrenzen, wie das etwa Valenzwörterbücher oder rückläufige Wörterbücher tun oder auch Wörterbücher, die Wörter nach der Länge ihrer Grundform sortieren. Damit sind aber auch reine Häufigkeitswörterbücher von der Definition als zentrale Wortschätze ausgeschlossen, sodass sich die hier vorgenommene Definition zentraler Wortschätze an die in der bisherigen Forschung bereits mehrfach angesprochene und in Abschnitt 3.4.1 verdeutlichte Unterscheidung von zentralen Wortschätzen auf der einen und Häufigkeitswörterbüchern auf der anderen Seite anschließt.

- Wohldefiniertheit:

Ein zentraler Wortschatz ist dann eine wohldefinierte Submenge eines Wortschatzes, wenn er durch Angaben zum Ausgangswortschatz, zum Verwendungszweck, zu den herangezogenen Auswahlkriterien und zu den Kennzeichen dessen, was als unbestritten zum Ausgangswortschatz gehörig zu zählen ist, gekennzeichnet ist.

- Zentralität und die dazugehörigen Eigenschaften:

Zentralität wird hier als ein Bündel von Eigenschaften und nicht als ein monolithisches Merkmal der Elemente eines zentralen Wortschatzes betrachtet. Woraus sich dieses Bündel von Eigenschaften zusammensetzt, ist abhängig von den Anforderungen des konkreten zentralen Wortschatzes und seinem Ausgangswortschatz.

- Ausgangswortschatz einer Sprache, Sprachvarietät oder Textsammlung:

Mit dieser Formulierung soll deutlich gemacht werden, dass es je nach Ausgangswortschatz und dessen Zugehörigkeit zu einer Sprache, Sprachvarietät oder Textsammlung unterschiedliche zentrale Wortschätze geben kann, dass sich also ein zentraler Wortschatz nicht ausschließlich auf den Gesamtwortschatz einer Sprache beziehen muss.²⁴³

243. Ähnliches führt bereits Kühn an: „[Es] lassen sich auch Grundwortschätze erstellen, die auf dem symptomfunktionalen lexikalischen Potential aufbauen und somit dem Prinzip der stilistischen Neutralität nicht genügen würden. Man ist also gezwungen, den Grundwortschatzbegriff dahingehend zu differenzieren, daß man vom Grundwortschatz eines Dialektes, eines Soziolektes, einer Generationalsprache, [...] einer Standardsprache, [...] einer Umgangssprache, einer Sondersprache, einer Fachsprache oder einer Literatursprache sprechen müßte.“ (Kühn 1979, S. 27)

Mit diesen Definitionsmerkmalen schließt sich die hier vorgenommene Definition zu einem gewissen Grad den impliziten Definitionen von Kühn und Schnörch an, die ihre Definitionen allerdings ausschließlich auf den (didaktisch orientierten) Grundwortschatz und seine Rolle für die Kommunikation beziehen. Kühn bezeichnet einen solchen zentralen Wortschatz als „funktional ausgewählte Materialgrundlage“ (Kühn 1980a, S. 57) und hebt damit den Einfluss der Funktion, also des Zwecks eines zentralen Wortschatzes, hervor. Schnörch geht noch weiter, indem er die Bedeutung der intendierten Anwendung betont und feststellt:

„Der Begriff 'Grundwortschatz' ist [...] nur als relative Größe zu verstehen: Die Sprache ist kein mathematisches Gebilde, dessen Mittelpunkt sich genau errechnen lässt. Infolgedessen kann es **den** allgemein verbindlichen Grundwortschatz auch nicht geben, sondern lediglich einen Grundwortschatz (von mehreren möglichen), der nur für ein ganz bestimmtes, vorab definiertes Anwendungsgebiet Gültigkeit besitzt.“ (Schnörch 2002, S. 24)

Abschließend ist noch einmal zu betonen, dass dieses Konzept die gemeinsame begriffliche Grundlage der verschiedenen in Kapitel 3 beschriebenen Anwendungstypen darstellt, die, wie wir gesehen haben, dieses Konzept in ganz unterschiedlicher Weise realisieren.

6.4 Terminologische Klärungen

Wie wir in Abschnitt 4.1 gesehen haben, ist eines der Probleme der bisherigen Forschung rund um das Thema zentraler Wortschätze die terminologische Unbestimmtheit und Inkonsistenz, die herrscht, wenn es um die Benennung verschiedener Typen und Anwendungen von zentralen Wortschätzen geht. Daher ist es gleichermaßen notwendig wie sinnvoll, den Begriffsumfang und -inhalt relevanter Begriffe, die in der Literatur vielfältig verwendet werden, ausgehend von den bisherigen Erkenntnissen aus Teil A und den hier vorgenommenen Definitionen terminologisch festzulegen und um weitere Begriffe zu ergänzen.

Wortschatz – Wortschätze

In den bisherigen Ausführungen war sowohl von *Wortschatz* als auch von *Wortschätzen* die Rede. Dabei handelt es sich um eine terminologische Vereinfachung: Anstatt von Teilwortschätzen zu sprechen, wenn bestimmte Bereiche des Gesamtwortschatzes gemeint sind, zum Beispiel anderssprachliche oder fachsprachliche Wortschatzbereiche, wird vereinfachend von Wortschätzen gesprochen. Dies hat vor allem den Grund, dass auch diese Wortschatzbereiche eine interne Makrostruktur haben, in zentrale und periphere Bereiche eingeteilt werden können und sich damit ähnlich wie der Gesamtwortschatz verhalten. Der Einfachheit halber sprechen wir daher auch in diesen Fällen von einem Wortschatz oder mehreren Wortschätzen. Wenn jedoch tatsächlich der gesamte Wortschatz des Deutschen gemeint ist, der in anderen Terminologien einfach als ‘der deutsche Wortschatz’ bezeichnet wird und sowohl Gemeinsprache als auch die verschiedenen Sprachvarietäten umfasst, dann wird vom *Gesamtwortschatz* des Deutschen gesprochen.²⁴⁴

Ausgangswortschatz – zentraler Wortschatz

Eine wichtige terminologische Unterscheidung bilden auch die beiden Termini *Ausgangswortschatz* und *zentraler Wortschatz*. Mit *Ausgangswortschatz* ist damit der Wortschatz gemeint, aus dem eine Teilmenge im Sinne eines zentralen Wortschatzes gebildet werden soll, während der Terminus *zentraler Wortschatz* die Teilmenge selbst bezeichnet.

Zentralität

Wie wir bei der Entwicklung des Wortschatzmodells und der Definition des Konzepts zentraler Wortschatze gesehen haben, ist Zentralität kein einheitliches, festes und eindeutiges Merkmal, sondern setzt sich aus einem Bündel von Eigenschaften zusammen, die die für die jeweilige Anwendung charakteristische Zentralität beschreiben. Das gilt für die lexikographische Zentralität. Daneben gibt es noch, wie wir in Abschnitt 5.5 gesehen haben, eine lexikologische Zentralität, die sich durch die beschriebenen *coreness*-Tests erfassen lässt. Sie kann sich in Teilaspekten mit der lexikographischen Zentralität überschneiden, ist grundsätzlich aber von ihr zu unterscheiden (vgl. Abschnitt 5.5.2).

Grundwortschatz

Wie wir in Abschnitt 4.1 gesehen haben, wurde der Terminus Grundwortschatz bisher in vielfältiger Weise verwendet, wobei die sprachdidaktische Bedeutung im Vordergrund stand. Wir wollen hier den Terminus *Grundwortschatz* explizit auf diese Bedeutung beschränken und ihn auf die didaktisch orientierten zentralen Wortschatze hin festlegen. Ähnlich verwendete Begriffe wie *kommunikatives Minimum* oder *Grunddeutsch* sind als synonym zu *Grundwortschatz* zu verstehen, der damit Hauptterm bleibt.

244. Im speziellen Fall der hier vorzunehmenden praktischen Untersuchung betrachte ich vereinfachend die im DUW erfassten Wörter als eine gute Repräsentation dieses Gesamtwortschatzes (vgl. Abschnitt 7.3.2).

Teil C: Praktische Untersuchung

Überblick

In Teil B haben wir uns einigen der in Kapitel 4 beschriebenen Probleme von einem lexikologischen Winkel aus genähert. Im vorliegenden Teil wollen wir nun zu einigen praktischen Problemen eine Antwort suchen.

Konkret wird es dabei um drei Problembereiche gehen:

Zum einen wollen wir ein Verfahren skizzieren, das es erlaubt, die Selektion für einen zentralen Wortschatz in transparenter und individuell justierbarer Weise vorzunehmen.

Zum anderen wird es darum gehen, den Wirkungsgrad und die gegenseitige Wechselwirkung bestimmter Selektionskriterien zu untersuchen.

Und schließlich wollen wir uns mit der Frage beschäftigen, wie sich die einzelnen Selektionskriterien auf den Umfang eines zentralen Wortschatzes auswirken.

Kapitel 7: Die praktische Untersuchung eines Lexikonausschnittes – Überblick über die Vorgehensweise

7.1 Ziel der praktischen Untersuchung

Die hier vorgestellte praktische Untersuchung hat das Ziel, die Überlegungen rund um das Thema zentraler Wortschätze um einige Beobachtungen zu erweitern, diese mit den in Kapitel 5 und 6 zusammengetragenen Einsichten um Aspekte der Wortschatzschichtung zu verknüpfen und damit die Diskussion voranzutreiben.

Zu diesem Zweck habe ich die Lexeme des Buchstaben O des *Duden. Deutsches Universalwörterbuch*²⁴⁵ auf einige ausgewählte Eigenschaften, die für die Selektion zu einem zentralen Wortschatz eine Rolle spielen können, hin untersucht und mich dabei auf die folgenden drei Fragestellungen konzentriert:

Zum einen wird es um die Frage nach dem Wirkungsgrad und der Wechselwirkung bestimmter Selektionskriterien gehen:²⁴⁶ Wie differenzierend und unterscheidend wirken die verschiedenen Selektionskriterien eigentlich? In welcher Wechselwirkung stehen sie zueinander? Führen zwei Selektionskriterien zu identischen Ergebnissen, und sind sie daher vielleicht etwa Symptome desselben sprachlichen Phänomens und daher gegeneinander austauschbar? Lässt sich das zugrundeliegende Phänomen vielleicht direkt ermitteln, oder ist eines der beiden Selektionskriterien vielleicht in der Praxis leichter zu ermitteln?

Zum anderen wird es um Fragen der Quantität gehen: Mit welchen Größenordnungen hat man es eigentlich zu tun, wenn man zum Beispiel versucht, den Wortschatz des Deutschen anhand ausgewählter Selektionskriterien auf zentrale und weniger zentrale Elemente hin zu gliedern?

Zum dritten soll es auch um Fragen der Methodik gehen: Wie könnte ein Selektionsverfahren aussehen, das den in den jüngsten Arbeiten verstärkt geäußerten Forderungen nach vermehrter Transparenz der Vorgehensweise entgegenkommt? Hier soll es darum gehen, einige der bisherigen Literatur angeführten Überlegungen zu Selektionskriterien transparenter zu machen, indem sie unabhängig von einer unmittelbaren Verwendung für einen spezifischen zentralen Wortschatz als Eigenschaften von Wörtern betrachtet werden. Diese Eigenschaften können eine mehr oder weniger wichtige Rolle spielen, wenn es darum geht zu bestimmen, ob ein Wort in einen konkreten zentralen Wortschatz aufgenommen werden soll oder nicht. Wie dieses 'Mehr oder Weniger' praktisch nutzbar gemacht werden kann, soll in der praktischen Untersuchung näher betrachtet werden.

245. Im Folgenden abgekürzt durch DUW.

246. Vgl. auch die Diskussionen in Abschnitt 4.5.3.

7.2 Besonderheiten dieser praktischen Untersuchung

Bevor es im folgenden um die konkreten Entscheidungen und unternommenen Schritte dieser Untersuchung geht, ist es wichtig, einige Besonderheiten dieser Untersuchung hervorzuheben, weil sie Einfluss auf das Ergebnis haben können und daher bei der Interpretation der Ergebnisse mitberücksichtigt werden müssen.

Ein Faktor, der bei der Interpretation der Ergebnisse eine Rolle spielt, besteht darin, dass die praktische Untersuchung auf die Stichwörter des Buchstaben O des DUW beschränkt ist. Diese Entscheidung war vor allem praktisch begründet:

Es ging einerseits darum, einen Ausschnitt des deutschen Wortschatzes zu finden, der sowohl leicht (d.h. öffentlich und elektronisch) zugänglich war – damit kamen gedruckte Wörterbücher und nicht öffentlich zugängliche Korpora nicht in Betracht – als auch einen gewissen Grundbestand an linguistischen, wohlaufbereiteten Informationen zu den Wortschatzeinheiten umfasste (damit waren nicht annotierte Korpora ausgeschlossen) und eine Art zufälliger (d.h. nicht von mir getroffener) Vorauswahl von Lexemen bereits zur Verfügung stellte (damit fielen öffentliche annotierte Korpora, bei denen ich zunächst selbst eine Auswahl hätte treffen müssen, weg).

Zum anderen spielte die Anzahl der zu bearbeitenden Wortschatzeinheiten eine Rolle: Einerseits war es wichtig, eine ausreichend große Menge an Lexemen zur Verfügung zu haben, um möglichst viele Problembereiche und Aspekte der Arbeit an einem zentralen Wortschatz abzudecken. Daher kam zum Beispiel der Buchstabe X mit den wenigen darunter aufgelisteten Stichwörtern nicht in Frage. Umgekehrt war es aus praktischen Gründen nötig, den Umfang der zu untersuchenden Anzahl von Lexemen zu begrenzen, weshalb sich zum Beispiel die Stichwörter des Buchstaben B oder S, die im DUW viele Seiten einnehmen, auch nicht für eine Untersuchung eigneten. Die Anzahl der Stichwörter des Buchstaben O dagegen war mit 1695 Einheiten gerade angemessen.²⁴⁷

Zu bedenken ist auch, dass man es bei der hier vorliegenden Untersuchung mit einem anderen Ausgangsmaterial zu tun hat als bei einer klassischen Korpusanalyse, bei der man das Wortmaterial direkt aus laufenden Texten gewinnt, wie wir bereits in Abschnitt 4.3 gesehen haben.

Neben diesen materialbedingten Besonderheiten spielen auch die in Abschnitt 7.1 besprochenen Zielsetzungen bei der weiteren Vorgehensweise eine Rolle: Eben weil es um den grundsätzlichen Effekt spezifischer Selektionskriterien geht, müssen andere zentrale Fragen der Lexikographie, etwa die nach der adäquaten Darstellung der Einheiten eines zentralen Wortschatzes (vgl. Abschnitt 4.7) oder die Frage nach der adäquaten Aufgliederung polysemer Lexeme hintangestellt werden. Soweit möglich sollen sie jedoch bei der Analyse der Untersuchungsergebnisse berücksichtigt werden, etwa wenn es darum geht, zu beurtei-

247. Zur Diskussion der mit dieser willkürlichen Auswahl verbundenen Begrenzungen bzw. ihrem Einfluss auf das Ergebnis siehe Abschnitt 9.4.4.

len, ob die Bedeutungsunterteilungen, die die Verfasser des DUW vorgenommen haben, hinsichtlich der Arbeit mit zentralen Wortschätzen hilfreich oder eher störend sind.

Darüber hinaus wollen wir die Beschreibung der Wortschatzeinheiten hinsichtlich der Selektionseigenschaften von ihrer Bewertung trennen. Das bedeutet unter anderem, dass wir bei der Frage danach, ob eine Einheit der Untersuchungsmenge ein typisches Wort darstellt oder nicht und deshalb ein Kandidat für einen zentralen Wortschatz sein kann oder nicht, zunächst hintanstellen. Damit wollen wir einer impliziten und verfrühten Selektion vorbeugen. Statt also zu Beginn der Untersuchung ‘untypische’ Wörter aus der Untersuchungsmenge herauszunehmen, wollen wir die Tatsache, dass einzelne Einheiten der Untersuchungsmenge eher dem Status eines Wortes entsprechen als andere, nutzbar machen, indem wir die Worthaftigkeit als eine weitere Eigenschaft in den Beschreibungsapparat aufnehmen und differenzieren und später bei der Gewichtung als einen weiteren Bewertungsfaktor in das Verfahren mitaufnehmen.²⁴⁸

7.3 Ausgangshypothesen und Grundsatzentscheidungen

Die Diskussion der mit der Arbeit an zentralen Wortschätzen verbundenen Probleme (vgl. Kapitel 4) macht deutlich, dass einige dieser Probleme eine grundsätzliche Vorabentscheidung erfordern. Weil für diese Vorabentscheidungen zumeist eine objektive Entscheidungsgrundlage fehlt und man höchstens auf intersubjektive Nachvollziehbarkeit hoffen kann, müssen diese Entscheidungen zum Teil willkürlich getroffen werden. Das ist hinnehmbar, solange gewährleistet ist, dass diese Entscheidungen transparent sind und konsistent durchgehalten werden.²⁴⁹ In diesem Sinne soll es im Folgenden um die dieser Untersuchung zugrundeliegenden Vorabentscheidungen gehen.

7.3.1 Die allgemeine Ausrichtung der praktischen Untersuchung gegenüber der Frage nach dem Zweck eines zentralen Wortschatzes

Wie aus der Definition des Konzeptes zentraler Wortschätze in Abschnitt 6.3 hervorgeht, gehört zu einem wohldefinierten zentralen Wortschatz die Angabe des konkreten Verwendungszwecks, für den er erstellt wird. Die vorliegende Untersuchung behandelt jedoch allgemeine Fragen der Selektion und ist nicht an einem konkreten zentralen Wortschatz ausgerichtet. Deshalb fehlt die Angabe eines konkreten Zwecks an dieser Stelle.

7.3.2 Der Bestand des DUW als ‘gute’ Repräsentation des Wortschatzes des Deutschen

Zu den hier beschriebenen Grundsatzentscheidungen gehört, dass wir vereinfachend annehmen, dass der Bestand des DUW eine ‘gute’ Repräsentation des Gesamtwortschatzes des Deutschen darstellt, auch wenn das DUW mit seiner Größe als einbändiges Wörterbuch si-

248. Vgl. auch Abschnitt 7.3.4.

249. Vgl. auch Schnörch 2002, 102f.

cher nicht zu den umfassendsten Darstellungen des deutschen Wortschatzes gehört. Mit ‘gut’ meine ich hier die Annahme, dass der Bestand des DUW den Wortschatz des Deutschen zu großen Teilen abdeckt und dass die Angaben, die der DUW zu jedem Eintrag bezüglich seiner Bedeutung, grammatischen Eigenschaften und Sondermarkierungen macht, umfassend und korrekt sind.²⁵⁰

7.3.3 Die betrachteten Eigenschaften

Bei der praktischen Untersuchung der Lexeme des Buchstaben O im DUW konzentriere ich mich auf die Beschreibung von sechs Eigenschaften, von denen vier in der bisherigen (Grundwortschatz-)Literatur wiederholt zur Selektion von Lexemen für einen zentralen Wortschatz verwendet und beschrieben wurden:

Zum einen werde ich die Elemente der Wortschatzmenge hinsichtlich ihrer Frequenz, Stabilität und Neutralität beschreiben – diese drei Eigenschaften stellen klassische Selektionskriterien dar, weshalb die Untersuchung ihres Wirkungsgrades und ihrer Wechselwirkung besonders wichtig ist.

Zum anderen werde ich die Wortschatzelemente mit Hilfe der transparenten Introspektion (vgl. Abschnitt 4.5.2.2), d.h. hinsichtlich ihrer Repräsentiertheit und Plausibilität beschreiben. Es erschien mir wichtig, die von Bonan-Garrigues (1993) erstmals systematisch beschriebene und eingeführte Methode der differenzierten Introspektion in die Untersuchung miteinzubeziehen. Einerseits ist die von ihr vorgeschlagene Vorgehensweise ein deutlicher Fortschritt zu dem sonst eher diffus beschriebenen Verfahren der lexikographischen Introspektion; andererseits kann der Vergleich der Ergebnisse der introspektiven Beschreibung der Lexeme mit der anderer Eigenschaften zu interessanten Ergebnissen führen; und schließlich ist die Bestimmung der Eigenschaften von Lexemen und deren Gewichtung bis zu einem gewissen Grad niemals völlig objektiv, sodass ein explizites Miteinbeziehen der Introspektion in Verbindung mit einem transparenten Vorgehen nur eine implizite Tatsache sichtbar und damit nachvollziehbar macht.

Und zum dritten schließlich soll jedes Element der zu untersuchenden Wortschatzmenge danach beschrieben werden, wie sehr es ein typischer Vertreter des Wortschatzes ist. Ich werde also den Grad der Worthaftigkeit der Wortschatzelemente als sechste Eigenschaft in die Beschreibung und Bewertung mit einbeziehen.

Alle diese sechs Eigenschaften sollen dann in verschiedenen Bewertungsversuchen hinsichtlich ihrer Aussagekraft untersucht werden. Dabei sind sie als Indikatoren der Zentralität einer Wortschatzeinheit zu verstehen. Je ausgeprägter eine Wortschatzeinheit eine bestimmte Eigenschaft aufweist, desto zentraler ist sie hinsichtlich dieser Eigenschaft. In diesem Zusammenhang spreche ich daher von Zentralitätsgraden: Eine Wortschatzeinheit, die die Eigenschaft der Stabilität in hohem Maße aufweist, hat einen Zentralitätsgrad von 1

250. Vgl. auch Abschnitt 7.3.4 und 8.3.

hinsichtlich der Eigenschaft der Stabilität; eine Wortschatzeinheit, die die Eigenschaft der Neutralität in hohem Maße aufweist, hat einen Zentralitätsgrad von 1 hinsichtlich der Eigenschaft der Neutralität etc. Die Summe der Zentralitätsgrade, die eine Wortschatzeinheit hinsichtlich aller sechs Eigenschaften besitzt, soll als *kumulative Zentralität* bezeichnet werden.

7.3.4 Die Frage nach der Grundeinheit und das Problem der impliziten Selektion

Wie wir in Abschnitt 4.2 gesehen haben, stellt die lexikalische Einheit, also die Einheit von genau einer Grundform und genau einer Bedeutung eine gute Grundeinheit eines zentralen Wortschatzes dar. Dieser Regel folgt auch die vorliegende praktische Analyse. Konkret ist damit gemeint, dass die Stichwörter des DUW im Laufe der praktischen Analyse monosemiert werden müssen, d.h. dass ihre kompakte Darstellung, bei der unter einem Hauptstichwort verschiedene Bedeutungen und grammatische Varianten angegeben sind, in 1:1-Paare von (grammatisch eindeutiger) Wortform und semantisch eindeutiger Bedeutung aufgelöst werden.²⁵¹

Wir wollen also die lexikalische Einheit als Grundeinheit der praktischen Analyse betrachten.²⁵² Allerdings ist hier eine terminologische Verdeutlichung nötig: Wenn wir von einer Wortschatzeinheit als ‘lexikalischer Einheit’ sprechen, soll damit nichts über ihren Status als Wort ausgesagt werden. Wir wollen – bis auf wenige Ausnahmen (vgl. Abschnitt 8.3) – alle Teilbedeutungen und monosemen Stichwörter des DUW als lexikalische Einheiten betrachten. Dass sie einen unterschiedlichen Wortstatus haben, wird erst bei der Beschreibung ihrer Worthaftigkeit eine Rolle spielen.

Die Entscheidung, die Wahl der Grundeinheit von der Frage nach dem Wortstatus gedanklich zu trennen, hat zwei Gründe: Zum einen soll damit eine verfrühte, implizite Selektion vermieden werden, wie sie ein Aussortieren nicht-worthafter lexikalischer Einheiten zu Beginn des Beschreibungsverfahrens bedeutet hätte. Zum anderen soll das hier vorgestellte Verfahren eine allgemeine methodische Grundlage für die Erstellung aller Arten von zentralen Wortschätzen zur Verfügung stellen; und da spielt es gerade bei der Erarbeitung zentraler Wortschätze des analytischen Typs, also derjenigen, die maschinenlesbar sein sollen und mit sehr kleinen Elementen eines Textes, etwa Morphemen, operieren können müssen, eine wichtige Rolle, dass auch ‘untypische’ lexikalische Einheiten berücksichtigt werden, auch wenn ein solches Verfahren für didaktisch orientierte zentrale Wortschätze zunächst eher kontra-intuitiv wirken kann.

Zur Entscheidung, die Stichwörter des DUW zu monosemieren, gehört auch, dass diese Stichwörter mit den Beschreibungen, die der DUW gibt, als ‘gutes’ Korpus betrachtet wer-

251. Vgl. die Diskussion in Abschnitt 4.2 und die entsprechende praktische Umsetzung in Abschnitt 8.3.

252. In der Darstellung der Daten in Auszugstabellen wird der Terminus *lexikalische Einheit* mit *LE* verkürzt.

den. Damit ist die Annahme gemeint, dass die Angaben, die das DUW zu den einzelnen Stichwörtern macht, grundsätzlich richtig und umfassend sind (vgl. Abschnitt 7.3.2). Konkret bedeutet das, dass ich von Stichwörtern, zu denen im DUW nur eine Bedeutung aufgelistet ist, vereinfachend annehme, sie bildeten bereits lexikalische Einheiten, ohne dass ich nachprüfe, ob es zu diesen einfachen Stichwörtern noch weitere Unterbedeutungen gibt. Mit 'gutem' Korpus meine ich auch, dass ich die Sondermarkierungen und grammatischen Angaben zu den einzelnen Stichwörtern als zutreffend und korrekt betrachte und zum Beispiel die Zuordnung zu Wortarten oder das System der Sondermarkierungen direkt übernehme. Der Grund für dieses Vorgehen liegt in der thematischen Ausrichtung dieser Arbeit: Zwar gehören Fragen der sprachlichen Sondermarkierung oder der semantischen Aufgliederung von Wortbedeutungen mit zur Arbeit an zentralen Wortschätzen, allerdings sind sie nur ein Mittel zum Zweck und nicht der eigentliche Gegenstand solcher Arbeiten. Dies ist der Grund dafür, dass ich die Angaben im DUW vereinfachend übernehme, ohne sie per se in Frage zu stellen.

7.3.5 Zur Trennung von Beschreibung und Gewichtung

In der Literatur werden einzelne Selektionskriterien zumeist wie Ja-Nein-Kriterien besprochen. Die Zugehörigkeit einer lexikalischen Einheit zu einem zentralen Wortschatz, ihre Zentralität, ist jedoch das Ergebnis einer Funktion über mehrere Eigenschaften oder Faktoren. Diese Eigenschaften an sich sind inhärente Merkmale der lexikalischen Einheit und zunächst 'wertfrei'. Erst wenn man sie mit dem Typus des anvisierten zentralen Wortschatzes verknüpft und entsprechend bewertet, bekommen diese Eigenschaften selektierenden Charakter. Aber auch hier ist es nicht die einzelne Eigenschaft, die ausschlaggebend ist für die Zugehörigkeit einer lexikalischen Einheit zu einem zentralen Wortschatz, sondern erst die Summe der verschiedenen untersuchten Eigenschaften.

Somit ist die Frage nach der Zugehörigkeit einer lexikalischen Einheit zu einem zentralen Wortschatz eher eine Mehr-oder-Weniger- als eine Ja-Nein-Entscheidung.²⁵³ An dieser Stelle setzt nun die Trennung von Beschreibung und Gewichtung an: Sie erlaubt es, zum einen die Eigenschaften der lexikalischen Einheiten zunächst wertfrei zu bestimmen und sie dann zum anderen einzeln zu gewichten. Praktisch bedeutet das, dass man die Auswahl der Einheiten für einen zentralen Wortschatz in zwei Schritte zerlegen muss: In einem ersten Schritt geht es um die qualitative Beschreibung der Wörter, unter denen ausgewählt werden soll, um eine Bestandsaufnahme sozusagen, bei der man untersucht und feststellt, inwieweit bestimmte Eigenschaften auf die Wörter, unter denen ausgewählt werden soll, zutreffen; in einem zweiten, separaten Schritt geht es dann um die Gewichtung dieser Eigenschaften, bei der man jeder Eigenschaft ein bestimmtes Gewicht beimisst, das relativ ist zum Gewicht anderer Eigenschaften.

253. Vgl. auch Krohn 1992, S. 213.

Konkret bedeutet die Trennung von Beschreibung und Gewichtung, dass wir grundsätzlich zunächst erst einmal alle Einträge des DUW-Ausschnittes als Elemente des Wortschatzes betrachten und in die Beschreibung mit einbeziehen und erst bei der Bewertung des Grades der Worthaftigkeit zwischen ‘typischen’ und ‘weniger typischen’ lexikalischen Einheiten unterscheiden (vgl. Abschnitt 7.3.4).

Diese Zerlegung des Auswahlverfahrens in zwei Schritte erlaubt eine detaillierte Gewichtung im Sinne einer Mehr-oder-Weniger-Zugehörigkeit anstatt einer wenig flexiblen Ja-Nein-Selektion. Darüberhinaus wird durch diese Trennung das gesamte Auswahlverfahren transparenter. Außerdem erlaubt diese Vorgehensweise eine genauere Feinsteuerung der Ergebnisse und ein kontrolliertes repetitives Verfahren zur Erstellung eines zentralen Wortschatzes, weil die Transparenz der einzelnen Schritte es möglich macht, im Nachhinein Selektionsschritte rückgängig zu machen, abzuändern und die Ergebnisse miteinander zu vergleichen.

Damit wird es möglich, zwei Forderungen, die bei Schnörch 2002 gestellt werden, zu erfüllen: Transparenz des Vorgehens als Ersatz für eine nicht erreichbare Objektivität (vgl. Schnörch 2002, S. 102f.) und differenzierte Betrachtung der Mittel, die zur Verfügung stehen (vgl. Schnörch 2002, S. 23).²⁵⁴

7.3.6 Die Reihenfolge der Bearbeitung

In welcher Reihenfolge die Eigenschaften beschrieben und bewertet werden, ist im Prinzip gleichgültig. Eine Ausnahme gibt es allerdings: Die Besonderheiten der Beschreibung der intuitiven Eigenschaft der Repräsentiertheit eines Lexems erfordern, dass man, wenn man diese Eigenschaft beschreibt, durch möglichst wenig linguistisches Detailwissen zu den jeweiligen Lexemen in seiner Urteilsbildung beeinflusst sein sollte.²⁵⁵ Daher ist es essentiell, dass die Beschreibung der Stichwörter des DUW anhand dieser Eigenschaft zuerst stattfindet, noch bevor Informationen über verschiedene Bedeutungen, Sondermarkierungen etc. zu den einzelnen Lexemen eingeholt werden. Dagegen sind Eigenschaften wie Stabilität oder Frequenz Merkmale, die unabhängig von der Intuition des zu Beurteilenden sind und die vor allem ein linguistisches Detailwissen zu den einzelnen Lexemen voraussetzen. Sie können also erst beschrieben werden, wenn die linguistischen Details zu den Lexemen erfasst sind.

254. Eine darüber hinausgehende Frage ist diejenige danach, wieweit sich diese Trennung in der Praxis konsequent durchführen lässt. Diese Frage wird in Abschnitt 9.4.1 diskutiert.

255. Vgl. auch Bonan-Garrigues (1993, S. 28), die betont, dass die Kontextlosigkeit der hinsichtlich ihrer Repräsentiertheit zu beschreibenden Wörter gerade ein konstituierendes Element dieser Beschreibung ist. Im Gegensatz dazu ist es allerdings, so Bonan-Garrigues (1993, S. 52) weiter, wichtig, dass der Bearbeitende imstande ist, Sprache als Objekt zu betrachten.

7.4 Die praktischen Schritte im Überblick

An dieser Stelle möchte ich noch einmal betonen, dass es mir nicht um die Bestimmung eines konkreten zentralen Wortschatzes geht, sondern um die Erarbeitung einer Vorgehensweise, mit deren Hilfe sich solche Wortschätze in Zukunft besser und transparenter erarbeiten lassen. Kern meines Vorschlages ist dabei die deutliche Trennung zwischen der Beschreibung lexikalischer Eigenschaften, die mehr oder minder zur Zentralität eines Lexems beitragen, und der Gewichtung dieser Eigenschaften. Um diesen schrittweisen Charakter meiner Vorgehensweise zu verdeutlichen, habe ich die einzelnen Schritte in Abbildung 7-2 zusammengefasst.

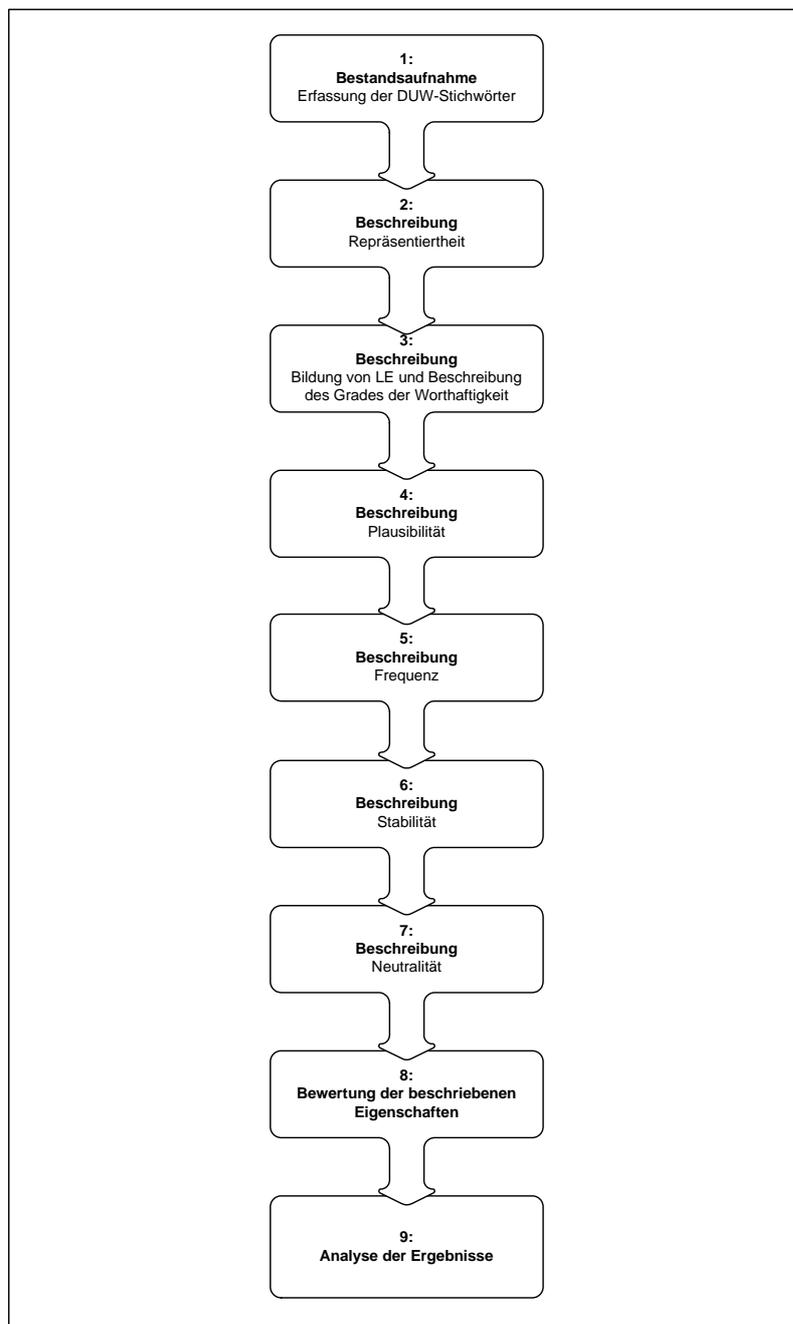


Abb. 7-2: Überblick über die Arbeitsschritte der praktischen Untersuchung

Kapitel 8: Die Schritte der praktischen Untersuchung im Einzelnen

8.1 Die Erfassung der Hauptstichwörter des DUW

In einem ersten Schritt habe ich die Hauptstichwörter des Buchstaben O des DUW von der CD-ROM-Ausgabe direkt in eine elektronische Tabelle übernommen. Dabei habe ich jedem Hauptstichwort, d.h. jedem fettgedruckten übergeordneten Lemma im DUW, einen eigenen Listeneintrag zugeordnet und mit einer laufenden DUW-Eintragsnummer versehen. Wenn ein Stichwort mehrere untergeordnete Stichwörter (Unterbedeutungen) hatte, habe ich diese zunächst ignoriert und erst im dritten Untersuchungsschritt separat erfasst.

Detaillierte Informationen über den jeweiligen Listeneintrag habe ich unberücksichtigt gelassen, um allzuviel Vorwissen über die einzelnen Einträge zu umgehen (vgl. Abschnitt 7.3.6). Dies war auch der Grund dafür, dass ich bei der Erfassung der DUW-Stichwörter und ihrer Überführung in die Datenbank die CD-ROM-Ausgabe benutzt habe: Sie erlaubte es mir, die Stichwörter in ihrer graphischen Form in die Datenbank zu übernehmen, ohne zugleich die Informationen über das jeweilige Stichwort zufällig ‘mitzulesen’.

Insgesamt enthält das DUW unter dem Buchstaben O 1695 Stichwörter; damit enthielt auch die erste Tabelle der Datenbank nach dem ersten Schritt 1695 Einträge. Tabelle A-1 in Appendix A gibt in einem Ausschnitt dieser Datenbanktabelle einige Beispieleinträge wieder.²⁵⁶

8.2 Die Beschreibung der Einträge nach ihrer Repräsentiertheit

In einem nächsten Schritt habe ich alle Einträge danach beschrieben, wie sehr sie in meinem internen Lexikon repräsentiert sind, beruhend auf dem bei Bonan-Garrigues (1993) dargelegten Verfahren (vgl. Abschnitt 4.5.2.2.). Damit wechsle ich von der reinen Bestandsaufnahme zur Beschreibung der Daten. Der Grund dafür, dass ich die Beschreibung der Repräsentiertheit vor allen anderen Beschreibungsschritten vornehme, liegt in der in Abschnitt 7.3.6 beschriebenen Notwendigkeit, möglichst wenig linguistisches Detailwissen zu den zu beschreibenden Lexemen zu besitzen, wenn es um die Beschreibung der Repräsentiertheit eines Lexems geht.

Wie in Abschnitt 4.5.2.2 und bei Bonan-Garrigues (1993, S. 40ff.) beschrieben, beruht das Verfahren von Bonan-Garrigues darauf, dass man Wörter in ihrer graphischen Erscheinung innerhalb einer einfachen, kontextlosen Wortliste danach einteilt, ob sie einen Wiedererkennungseffekt hervorrufen und damit im mentalen Lexikon des Beurteilenden repräsentiert sind oder nicht. Dabei geht Bonan-Garrigues von einer binären Definition der Repräsentiertheit aus, bei der ein Wort repräsentiert ist, wenn es auch nur ein schwaches

256. Alle hier angegebenen Datentabellen sind in ihrem gesamten Umfang auf der CD-ROM, die der gedruckten Ausgabe der Dissertation beiliegt, individuell abgespeichert. Leser der elektronischen Ausgabe der Dissertation können die entsprechenden Dateien unter dissertationsdaten.haderlein@gmail.com anfordern.

mentales Bild hervorruft (vgl. Bonan-Garrigues 1993, S. 40). Wörter, die eine solche Reaktion nicht hervorrufen, gelten nach ihrem Verfahren als nicht repräsentiert.

In einem ersten Versuch, ihr Verfahren auf die hier untersuchte Wortschatzmenge anzuwenden, habe ich alle Datenbankeinträge, so wie sie nach dem ersten Bearbeitungsschritt in Ausgangstabelle vorlagen, beschrieben. Im Laufe der Bearbeitung zeigte sich jedoch, dass ich das von Bonan-Garrigues vorgeschlagene Kriterienset ein wenig modifizieren musste, um überhaupt einen differenzierenden Effekt zu erzielen. Ich musste nicht nur fragen, ob ein Wort ein mentales Bild erzeugt oder nicht, sondern, wie deutlich dieses mentale Bild ist, und vor allem, was der Grund dafür ist, dass das mentale Bild zwar vorhanden, aber nicht sehr deutlich ist. Anlass für diese differenziertere Vorgehensweise waren zwei Beobachtungen während dieser ersten Beschreibungsrunde:

Zum einen zeigte sich, dass mir fast alle Stichwörter bzw. Datenbankeinträge in irgendeiner der oben genannten Formen bekannt waren. Nach diesem ersten Durchgang waren von den insgesamt 1695 Stichwörtern nur knapp 70 als mir völlig unbekannt markiert. Ein Grund für dieses Phänomen besteht wohl darin, dass das Wortmaterial, das ich untersuche, bereits eine lexikographisch bearbeitete Wortliste ist, die bereits eine Auswahl des Gesamtwortschatzes des Deutschen darstellt und kein direktes, nicht-bereinigtes Korpusmaterial.²⁵⁷ Es enthält somit vor allem lexikalisierte oder zumindest konventionalisierte Elemente des Wortschatzes und kaum, wie der von Bonan-Garrigues untersuchte Wortbestand, okkasionelle Bildungen. Ebenso spielt sicher auch der persönliche Hintergrund desjenigen, der die Repräsentiertheit eines Wortes beschreiben soll, eine Rolle. Idealerweise sollte daher dieser Schritt der Beschreibung, ebenso wie derjenige der Beschreibung der Plausibilität (vgl. auch Abschnitt 8.4), von mehreren Sprechern des Deutschen durchgeführt werden, um so eine breitere Beurteilungsbasis zu erhalten.²⁵⁸

Zum anderen war die Einteilung eines Wortes nach den von Bonan-Garrigues angeführten Kriterien, die ja die geringste Reminiszenz, die ein Wort hervorruft, bereits als hinreichend für seine Einstufung als repräsentiert auszeichnen, insgesamt schwierig. Denn es zeigte sich, dass fast alle Stichwörter des DUW Eigenschaften besitzen, die sie dem Beurteilenden bekannt vorkommen lassen können, allerdings in unterschiedlich starkem Maß. Manche Wörter (*Obdachlose*, *örtlich*, *Orthographie*), sind so stark repräsentiert, dass es spontan möglich ist, eine Bedeutung und verschiedene Anwendungsbeispiele zu ihnen anzugeben; andere Wörter (*Obi*, *Ordal*, *oblong*) sind so schwach repräsentiert, dass es abgesehen von der Groß- und Kleinschreibung kaum möglich war, ihnen auch nur eine Wortart zuzuordnen. Gäbe es nur diese beiden Gruppen von Wörtern, wäre eine Einteilung nach Bonan-Garrigues in repräsentierte und nicht repräsentierte Wörter möglich. Doch zeigt sich, dass es eine ganze Reihe von Wörtern gibt, bei denen zwar eine vollständige Bedeutungsangabe

257. Vgl. die Abschnitte 4.3 und 7.2.

258. Aus diesem Grund baut Bonan-Garrigues in ihre Untersuchung auch eine Sprecherbefragung ein. Vgl. Bonan-Garrigues 1993, S. 50ff.

kaum möglich ist, dennoch aber Teile oder Teilaspekte der Wörter wie zum Beispiel einzelne Teile einer Wortbildungskonstruktion repräsentiert sind. Hier zeigt sich, was auch Bonan-Garrigues (1993, S. 40) andeutet, nämlich dass es sich bei der Frage nach der Repräsentiertheit eines Wortes um ein Kontinuum handelt, zwischen dessen Endpolen sich viele Zwischentypen befinden. Das von Bonan-Garrigues entwickelte Verfahren eignet sich gut für einen Wortbestand, der aus einem großen Textkorpus gewonnen wurde und daher Wörter mit verschiedenen Graden an Lexikalisiertheit enthält; für den hier vorliegenden Fall der Einteilung eines Wortbestandes, der aus einem Wörterbuch, das ja bereits einen vorab bearbeiteten Wortbestand darstellt, gewonnen wurde, führt das streng binäre Verfahren von Bonan-Garrigues allerdings dazu, dass fast alle Einheiten mehr oder minder repräsentiert sind. Daher war es im konkret vorliegenden Fall nötig, differenzierter vorzugehen und all die Fälle, die zwischen eindeutiger Repräsentiertheit und eindeutiger Nicht-Repräsentiertheit liegen, zu sammeln und zu systematisieren.

Im Sinne der Ausgangsregel von der Trennung von Beschreibung und Gewichtung habe ich also zunächst einmal alle DUW-Stichwörter danach beschrieben, ob sie voll repräsentiert sind, ob sie zur Gänze nicht repräsentiert sind, oder ob irgendeine Eigenschaft eine teilweise Repräsentiertheit hervorruft und ihre Beschreibung hinsichtlich ihrer Repräsentiertheit unklar ist.²⁵⁹ Dabei habe ich für die Gruppe von Stichwörtern mit unklarer Repräsentiertheit genau vermerkt, was zu ihrer teilweisen Repräsentiertheit beiträgt. Der Grund für dieses detaillierte Verfahren ist der, dass es bei der späteren Bewertung des Repräsentiertheitsgrades möglich sein sollte, ähnliche Zweifelsfälle gleich zu behandeln, damit eine Konsistenz bei der Bewertung gewahrt bleibt. Hervorzuheben ist dabei, dass ich mit dieser differenzierten Vorgehensweise vom Verfahren und der Argumentation bei Bonan-Garrigues abweiche; Bonan-Garrigues argumentiert, dass, gerade weil es so schwierig sei, zu bestimmen, wann einem Sprecher ein Wort bekannt sei, die einzige pragmatische Lösung eine binäre Einteilung sei.²⁶⁰ Wie weiter oben erwähnt gilt diese Forderung sicherlich für die konkrete Situation der Beurteilung eines Korpusbestandes, der viele okkasionelle Bildungen enthält; für den vorliegenden Fall eines lexikographisch bereits bearbeiteten Wortbestandes scheint es allerdings nötig und auch legitim zu sein, von der strengen Einteilung bei Bonan-Garrigues abzuweichen. Um dem Unterschied im Verfahren allerdings auch in der Terminologie gerecht zu werden, ist es genau genommen nötig, nicht von Repräsentiertheit schlechthin, sondern von einer graduellen Repräsentiertheit zu sprechen. Es ist diese graduelle Repräsentiertheit, auf die ich mich im Folgenden aus Gründen der Einfachheit mit dem Terminus *Repräsentiertheit* beziehe.

Die Fälle unklarer Repräsentiertheit lassen sich in die folgenden Gruppen einteilen:

259. Die entsprechenden Kommentare zu diesen Datenbankeinträgen beschreiben den jeweiligen Typus von Zweifelsfall.

260. Vgl. die Diskussion bei Bonan-Garrigues 1993, S. 25-28 und die daraus gezogenen praktischen Konsequenzen (ebda., S. 40f.).

- Komposita und Ableitungen, bei denen Teile der Konstruktion bekannt sind, andere nicht: Beispiele dieses Typs sind *Obstruktionspolitik* oder *OAU-Staaten*, bei denen zwar das jeweilige Grundglied (*Politik*, *Staaten*) repräsentiert ist, nicht aber das jeweilige Bestimmungswort (*Obstruktion*, *OAU*). Nach dem Verfahren von Bonan-Garrigues müsste man sie als repräsentiert einordnen, weil ein Teil der Konstruktion repräsentiert ist. Zugleich sind diese Wörter nicht in gleicher Weise repräsentiert wie etwa *Onkel*, *örtlich* oder *Orthographie*.
- Ableitungen mit gebundenen Lexemen, wie etwa *Ophthalmologie*, *Ostrazismus* oder *obstruktiv*: Hier sind zwar die Wortbildungskonstruktion und das darin verwendete gebundene Lexem erkennbar, die Gesamtbedeutung aber bleibt dunkel, weil der entscheidende Teil der Konstruktion, der Wortstamm (*Ophthalm-*, *Ostraz-*, *obstrukt-*), nicht repräsentiert ist. Auch hier müssten, ginge man streng nach Bonan-Garrigues' Verfahren vor, die jeweiligen Stichwörter als repräsentiert eingeordnet werden, dennoch ist ihre Repräsentiertheit eine andere als die der eindeutig repräsentierten Stichwörter.
- Ähnliches gilt für Wörter anderer Sprachen, bei denen zwar die sprachliche Herkunft erkennbar ist, die Bedeutung sich aber nicht näher benennen lässt (Beispiel: *Obiter Dictum* oder *Open Shop*): Hier ließe sich argumentieren, dass allein die Tatsache, dass sich die Wörter als lateinisch bzw. englisch erkennen lassen, ein Zeichen für ihre Repräsentiertheit im internen Lexikon ist, da sie durchaus einen Wiedererkennungseffekt hervorrufen. Im Vergleich zu anderen Stichwörtern (Beispiel: *Office*), bei denen sich eine unmittelbare Bedeutung angeben lässt, ist der Wiedererkennungseffekt allerdings gering.
- Symbole und Einzelbuchstaben: Der DUW enthält nicht nur voll ausgebildete Wörter, sondern auch Symbole und Einzelbuchstaben wie etwa *ö*, *Ö* oder *Ω*. Richtet man sich streng nach den Vorgaben von Bonan-Garrigues, so müssen diese als repräsentiert eingeordnet werden, da sie in der Regel in irgendeiner Form bekannt sind. Nur ist es schwierig, ihnen irgendeine Art der Bedeutung zuzuordnen, weshalb es nicht gerechtfertigt erscheint, sie auf eine Ebene mit voll repräsentierten Wörtern wie *Onkel* zu stellen. Zugleich stellen sie eine noch geringere semantische Repräsentation dar als etwa die Fälle der Komposita, bei denen das Grundglied bekannt ist, die Gesamtbedeutung aber nicht, weshalb sie auch nicht mit diesem Typus auf eine Stufe gestellt werden können.
- Abkürzungen: Abkürzungen wie etwa *OAU*, *ÖAMTC*, *ÖBB* oder *OSB* stellen ebenfalls einen Problemfall dar, wenn man sich streng an die von Bonan-Garrigues vorgegebenen Regeln hält, denn zumeist sind sie lautlich repräsentiert, ohne dass notwendigerweise eine Bedeutung in irgendeiner Weise mit ihnen verknüpft werden kann. Was sie in ihrer Repräsentiertheit voneinander unterscheidet, ist die Frage, ob sie mit einer bereits bekannten Vollform oder einer Bedeutungsumschreibung verknüpft werden können.

Quantitativ betrachtet handelt es sich bei diesen Zweifelsfällen um eine kleine Gruppe: Von den insgesamt 1695 DUW-Stichwörtern weisen 134 eine unklare Repräsentiertheit auf.²⁶¹ Tabelle A-2 in Appendix A gibt für die verschiedenen hier angegebenen Typen von Repräsentiertheit Beispiele wieder.

8.3 Die detaillierte Erfassung der Stichwortangaben des DUW und die Bildung von lexikalischen Einheiten

In einem nächsten Schritt habe ich die DUW-Stichwörter, die ich zunächst ja nur ganz ohne zusätzliche Informationen in die Datenbank aufgenommen hatte, mit den im DUW angegebenen detaillierten Informationen (Bedeutungsbeschreibung, Wortart, Genus, Sondermarkierungen und sonstige formale, zumeist syntaktische oder morphologische Besonderheiten) versehen und dabei aus dem DUW-Stichwort lexikalische Einheiten gebildet. Ziel dieses Schrittes ist es, aus der undifferenzierten Liste der Wortformen der DUW-Stichwörter eine Liste von lexikalischen Einheiten im weiter oben definierten Sinne (vgl. Abschnitt 7.3.4) zu bilden, die schließlich die Grundlage für die eigentliche Analyse darstellen.

Die Darstellungsweise im DUW enthält vier Typen von Stichwörtern:

- monoseme Stichwörter (wie etwa das in den Beispieltabellen wiedergegebene Stichwort *opportun*);
- polyseme Stichwörter, d.h. Stichwörter, die Platzhalter bzw. Haupteintrag für mehrfache Lesarten sind (wie etwa das in den Beispieltabellen wiedergegebene Stichwort *obenauf*);
- Stichwörter, bei denen die Formseite in verschiedenen Varianten existiert (wie die in den Beispieltabellen wiedergegebenen Stichwörter *Obstbau*, *Obstanbau*, *Obdachlose*, das sowohl das maskuline als auch das feminine Genus repräsentiert und *Orthographie*, *Orthografie*);
- und schließlich Stichwörter, die sowohl auf der Bedeutungsseite als auch auf der Formseite Variationen aufweisen (wie etwa das in den Beispieltabellen wiedergegebene Stichwort *Output*, das mit drei verschiedenen Lesarten angegeben wird und zudem sowohl das maskuline als auch das neutrale Genus annehmen kann).

Alle Typen von Stichwörtern habe ich nun in einzelne lexikalische Einheiten, d.h. in Einheiten von einer (Stamm-)Form und einer Bedeutung umgeformt. Das bedeutete, dass ich monoseme Stichwörter direkt in die Liste der lexikalischen Einheiten aufnehmen konnte, polyseme und multifforme Stichwörter aber in ihre jeweiligen Grundeinheiten auflösen musste. Zugleich habe ich dabei die sie bis dahin vertretenden Platzhalter aus der weiteren Bearbeitung herausgenommen.

261. Während 1494 Stichwörter eine eindeutige Repräsentiertheit aufweisen und 67 eindeutig nicht repräsentiert sind.

Bei der Bildung der lexikalischen Einheiten bin ich vereinfachend vorgegangen, d.h. ich habe die im DUW angeführten Daten zu den jeweiligen Stichwörtern als gegeben angenommen und sie direkt als Angaben der lexikalischen Einheiten übernommen.²⁶² So habe ich etwa die Bedeutungsbeschreibung, die im DUW angegeben war, direkt als Bedeutungsbeschreibung der jeweiligen lexikalischen Einheit übernommen, ebenso die Sondermarkierungen und sonstigen Angaben zu grammatischem Verhalten und anderen Besonderheiten. Außerdem habe ich diejenigen lexikalischen Einheiten gesondert gekennzeichnet, bei denen es sich offensichtlich um Abkürzungen, Eigennamen, mehrgliedrige Formen, Symbole oder Einzelbuchstaben, ein gebundenes Lexem, ein reihenbildendes Kompositum mit einem Eigennamen als Bestimmungsglied oder einfach um eine andere Wortart als die Hauptwortarten N, V, Adj und Adv handelte.

Grundsätzlich habe ich dabei nur die Bedeutungen berücksichtigt, die das DUW auflistet; andere mögliche Bedeutungen, die im DUW nicht genannt werden, habe ich unberücksichtigt gelassen, weil ich davon ausgehe, dass das DUW 'die Welt komplett abbildet'.²⁶³ Der Grund dafür liegt darin, dass ich mich für den hiesigen Zweck auf diejenigen Fragestellungen konzentrieren wollte, die unmittelbar mit der Erarbeitung zentraler Wortschätze zusammenhängen. Fragen danach, ob eine Bedeutungsumschreibung oder die Darstellung der Unterbedeutungen eines Polysems komplett sind oder nicht, hätten den Rahmen dieser Arbeit gesprengt. Zudem habe ich zunächst einmal alle Bedeutungen, die zu einem polysemen Stichwort angegeben waren, als eigenständige lexikalische Einheit übernommen. Das muss nicht so sein, denn in vielen Fällen ist die Bedeutungsunterteilung, die der DUW vornimmt,

262. Dabei habe ich in einigen wenigen Fällen Korrekturen vorgenommen:

1. Unter dem Stichwort *orientalisieren* führt das DUW zwei Bedeutungen an:

„a) *orientalische Einflüsse aufnehmen*: orientalisierende Kunst;

b) *einer Gegend o.Ä. ein orientalisches Gepräge geben*“

Hier ist es die erste Bedeutung, die Probleme bereitet. Denn hier handelt es sich nicht um eine Bedeutung des Verbes *orientalisieren*, sondern um die Bedeutung des partizipialen Adjektivs *orientalisierend*, wie der Beispielsatz zeigt. Eine Suche im Mannheimer COSMAS-Korpus zeigt, dass die Belegstellen, die sich zu den Suchwörtern *orientalisieren*, *orientalisiert* und *orientalisierten* ergeben, aus verbalen Konstruktionen im Sinne der Bedeutung b) des DUW-Stichwortes bestehen, in denen das Verb entweder die Rolle des Prädikats übernimmt oder in Form eines Partizip Perfekt das Ergebnis der Handlung im Sinne von b) wiedergibt („der orientalisierten deutschen Nationalhymne“, „dem orientalisierten Klischee“), nicht aber in einer aktiv verbalen Verwendung der Bedeutung a). Ich nehme daher an, dass es sich bei der Bedeutung a) um eine fälschliche Zuordnung des Adjektivs *orientalisierend* zum transitiven Verb *orientalisieren* handelt und habe daher die im DUW angegebene Unterbedeutung a) als adjektivi-sche lexikalische Einheit *orientalisierend* (DUW-ID 1301.1.-) erfasst.

2. Ähnliches gilt für das Stichwort *organisieren* (DUW-ID 1258.2.a): „*in einer Organisation, einem Verband o.Ä. od. zu einem bestimmten Zweck zusammenschließen*: gewerkschaftlich organisierte Arbeiter“. Hier handelt es sich nicht um die Bedeutung des transitiven Verbes *organisieren*, wie ursprünglich im DUW angegeben, sondern um die Bedeutung des partizipialen Adjektivs *organisiert*. Die Form der zugehörigen lexikalischen Einheit ist dementsprechend als *organisiert* vermerkt.

3. Unter dem Stichwort *Oberste* führt das DUW an, dass es sich hier um ein Neutrum handelt, und verweist ansonsten auf das Stichwort *ober...*. Daraus wird nicht ersichtlich, weshalb gerade das deadjektivische Neutrum *das Oberste* als eigenes Stichwort angeführt wird, nicht aber *der Oberste* oder *die Oberste*. Wenn man diese Möglichkeiten, aus einem Adjektiv Substantive zu bilden, verzeichnen möchte, dann muss man das für alle Genera tun und für alle im Wörterbuch angeführten Adjektive. Dies tut das DUW aber nicht. Daher habe ich den Eintrag *Oberste* nicht als lexikalische Einheit weiter übernommen.

diskutabel. Andererseits hängt die Frage, ob eine Lesart eines Lexems als eigenständige lexikalische Einheit zu werten ist oder nicht, zu einem Großteil auch von den Anforderungen der Anwendung ab, für die man lexikalische Einheiten erstellt. Daher wollte ich die Frage, welcher Grad an Bedeutungsunterteilungen praktikabel ist, aus der praktischen Analyse ausklammern und erst vor dem Hintergrund der gesammelten Erfahrungen diskutieren.

Im Zuge der Bildung von lexikalischen Einheiten habe ich auch der jeweiligen Einheit, soweit sie aus einem polysemen Stichwort abgeleitet ist, die Bedeutungsnummer und ggf. den Bedeutungsbuchstaben zugeordnet, unter der sie als Lesart im DUW angeführt ist, um so ein Zurückverfolgen der Lesart zu ihrer ursprünglichen Angabe im DUW zu gewährleisten. Dabei habe ich die Markierung „Ü“, die im DUW zwar nicht zu einem gesondert eingerückten Eintrag für eine Lesart führt, auch als eine Art Bedeutungsbuchstabe behandelt und der jeweiligen Lesart zugeordnet.

Bei der detaillierten Erfassung nicht berücksichtigt habe ich die fettgedruckten Beispiele, die die Verwendung eines Stichwortes in einem idiomatischen Ausdruck darstellen und mit „*“ (zur Kennzeichnung von Verwendungsweisen in Form fester Phraseologismen), „R“ (zur Kennzeichnung von redensartigen Verwendungsweisen des Stichwortes) oder „Spr“ (zur Kennzeichnung von Verwendungsweisen des Stichwortes in einem Sprichwort) markiert sind. Zwar berührt die Frage nach dem Status von idiomatischen Ausdrücken als lexikalische Einheiten auch die Arbeit mit zentralen Wortschätzen, und im Sinne der in Abschnitt 7.1 besprochenen Forderung, die Bestimmung des Wortstatus offen zu lassen und vielmehr die Informationen über den Wortstatus als eine weitere Eigenschaft eines Wortschatzelementes anzusehen, sollte man diese Verwendungsweisen auch nicht ohne weiteres ausschließen. Aber diese Ausdrücke stellen im DUW eine offene Gruppe dar, die man, wenn es darum geht, den Umfang der zu untersuchenden lexikalischen Einheiten aus praktischen Gründen zu begrenzen, auch weglassen kann.

Eine Ausnahme stellen allerdings diejenigen Einträge dar, die allein als Platzhalter für einen Phraseologismus stehen (vgl. *Olim*) als Platzhalter für den Phraseologismus *wie zu Olims Zeiten*). Sie sind als eigene Stichwörter im DUW eingetragen, daher behandle ich sie als Wortschatzelemente und versuche, sie mit allen Eigenschaftsbeschreibungen zu versehen, die ich für alle anderen Wortschatzelemente auch ermittle.

Tabelle A-3 in Appendix A gibt einige der hier genannten Beispiel-Stichwörter und ihre Umwandlung in lexikalische Einheiten wieder.

Die so entstandene Liste von lexikalischen Einheiten enthielt neben einer Mehrzahl von ‘unproblematischen’ Fällen auch einige Einheiten, bei denen der Mehrwert der weiteren Bearbeitung gering schien. Es handelt sich dabei um zwei Gruppen von Einheiten.

263. Wie bereits in Abschnitt 7.3.2 angesprochen.

Die eine besteht aus Einträgen, die eindeutig nur deshalb angeführt waren, weil sie unregelmäßige Formen eines Paradigmas darstellen (vgl. *Omnien* als Plural von *Omnium* oder *Organa* als Plural von *Organum*). Es handelt sich dabei um insgesamt nur 12 Einheiten; da sie eher einen Teil der Grammatik darstellen als einen Teil des Wortschatzes, habe ich sie aus der weiteren Bearbeitung herausgenommen.

Die andere Gruppe von Einheiten, deren Mehrwert bei der weiteren Bearbeitung gering erschien, sind diejenigen, die eine rein formale Variante zu einer anderen Einheiten darstellen und bei der die Art der Variation keine semantische oder syntaktische Bedeutung hat (im Gegensatz etwa zur Variation zwischen zwei verschiedenen Genera). Hierbei handelt es sich um die folgenden Fälle:

- regelhafte orthographische Varianten (etwa *Orthographie* und *Orthografie*) – insgesamt 7 lexikalische Einheiten werden aufgrund dieser Entscheidung aus der weiteren Bearbeitung herausgenommen): Hier nehme ich nur die *-f*-Variante auf, weil die Alternation mit *-ph-* durch Regeln erklärbar ist.²⁶⁴
- regelhafter Wechsel zwischen Klein- und Großschreibung von Abkürzungen, Einzelbuchstaben bzw. Symbolen und gebundenen Lexemen, je nach Stellung im Satz (etwa *o. Z.*, *O. Z.*, *o*, *O* oder ω , Ω , *orth-*, *Orth-* – insgesamt 18 lexikalische Einheiten): Hier nehme ich nur die Kleinschreibung auf, weil die Alternation durch die Großschreibung durch Regeln erklärbar ist.²⁶⁵
- regelhafte Kongruenz, je nach Verwendung des bestimmten Artikels (etwa *Oberdeutsch* – *das Oberdeutsche* oder *Okzitanisch* – *das Okzitanische* – insgesamt drei lexikalische Einheiten gehören zu dieser Gruppe): Hier nehme ich nur die Variante ohne *-e*-Endung auf, weil die Alternation durch Regeln erklärbar ist.
- konventionalisierter Wechsel zwischen *i* und *y* (etwa in *Oxidation* – *Oxydation* oder *oxidisch* – *oxydisch* – insgesamt 5 lexikalische Einheiten gehören zu dieser Gruppe): Hier nehme ich nur die Variante mit *y* auf, weil die Alternation durch Regeln erklärbar ist.²⁶⁶
- konventionalisierter Wechsel zwischen *ö* und *oe* (*Oesophagus* – *Ösophagus* oder *Örsted* – *Oersted* – insgesamt zwei lexikalische Einheiten): Hier nehme ich nur die Variante mit *oe-* auf, weil die Alternation durch Regeln erklärbar ist.²⁶⁷
- konventionalisierter Wechsel zwischen Schreibung mit und ohne Punkt (*OSB* – *O.S.B.* oder *OvD* – *O.V.D.* – insgesamt zwei lexikalische Einheiten): Hier nehme ich nur die Variante mit dem Punkt auf, weil die Alternation durch Regeln erklärbar ist.²⁶⁸

264. Allerdings habe ich bei der Beschreibung der Stabilität (vgl. Abschnitt 8.6) die entsprechende lexikalische Einheit als im jeweiligen Wörterbuch vorhanden zugelassen, wenn sie nur in der anderen, hier ausgeschiedenen graphischen Variante (also etwa in der Schreibung mit *-ph-*) vorkam war.

265. Vgl. Fußnote 264.

266. Vgl. Fußnote 264.

267. Vgl. Fußnote 264.

268. Vgl. Fußnote 264.

- regelhafte Verwendung von Fugenelementen (etwa in *ot-* – *oto-*, *orangerfarben* – *orangefarben* – insgesamt 10 Einheiten): Hier nehme ich nur die Variante ohne Fugenelement auf, weil die Alternation durch Regeln erklärbar ist.
- Normunsicherheit bei Getrennt- und Zusammenschreibung von Fremdwörtern (etwa bei *Oral History* – *Oralhistory* – insgesamt vier Einheiten): Hier nehme ich nur die zusammengeschrriebene Variante auf, weil die Alternation durch Regeln erklärbar ist.²⁶⁹

Drei Typen von Mehrfachformen habe ich allerdings komplett aufgenommen:

- Alternierendes Genus aufgrund von Normunsicherheit (etwa *overkill*, das mit neutralem und maskulinem Genus angegeben ist, oder *Ort2*, das mit maskulinem und neutralem Genus angegeben ist): Hier nehme ich beide Varianten auf, weil das Genus eine Information ist, die bei der weiteren Beschreibung eine Rolle spielen kann und weil die Genusalternation nicht durch eine Regel oder Konvention erklärt werden kann.
- Idiosynkratische Varianten, d.h. Varianten, die nicht unmittelbar durch eine Regel oder Konvention erklärt werden können (etwa *Oxygen* – *Oxygenium*, *Ophthalmiatrie* – *Ophthalmiatrik*, *Ordonnanz* vs. *Ordonanz*, *orangerfarben* – *orangerfarbig*, *Obstanbau* vs. *Obstbau* oder *Obödienz* – *Obedienz* – insgesamt 19 DUW-Hauptstichwörter gehören in diese Gruppe): Hier nehme ich beide Varianten auf, weil streng betrachtet erst die Beschreibung der Selektionseigenschaften die Entscheidungsgrundlage liefern soll dafür, welche der beiden Alternativen aufgenommen werden soll.
- Zugehörigkeit zu zwei Wortarten (*och*, das als Interjektion und Partikel angegeben ist): Hier nehme ich beide Varianten auf, weil unterschiedliche Wortarten einen gewissen semantischen Eigenwert und einen syntaktischen Einfluss haben.

Nun ließe sich argumentieren, diese Vorfilterung verstieße gegen das Prinzip, eine verfrühte oder implizite Selektion nach Möglichkeit zu umgehen, und natürlich stellt diese Vorfilterung auch eine Art Selektion dar. Allerdings gehe ich dabei sehr konservativ vor, d.h. ich versuche, so wenige lexikalische Einheiten wie nur möglich von der weiteren Bearbeitung auszuschließen. Und vor allem betrifft diese Vorfilterung allein formale Aspekte der Variation von lexikalischen Einheiten, und auch hier nur diejenigen, die durch Regeln erklärbar sind. Varianten, die auf idiosynkratischen Gründen beruhen, habe ich zwar als solche markiert, aber in die weitere Bearbeitung aufgenommen.

Die Anzahl der von diesen Schritt der Vorfilterung von der weiteren Bearbeitung ausgeschlossenen Einheiten beläuft sich auf insgesamt 64, verteilt auf 56 DUW-Hauptstichwörter.²⁷⁰

269. Die Variation von *-farben* gegenüber *-farbig* (vgl. *orangefarben* vs. *orangefarbig*, ebenso *olivfarben* vs. *olivfarbig*) habe ich dagegen beibehalten, weil sie potentiell semantisches Gewicht haben kann.

270. Von der weiteren Bearbeitung ausgeschlossen habe ich auch die Stichwörter *Obödienz*, das im DUW bereits unter dem alphabetisch früher rangierenden Stichwort *Obedienz*, *Obödienz* erfasst war und nur aufgrund der Darstellungsregeln des DUW noch einmal separat aufgelistet war, und *Oxyd*, das im DUW bereits unter *Oxid*, *Oxyd* erfasst war.

Tabelle A-4 in Appendix A zeigt anhand einiger der hier genannten Beispiele, wie diese Vorfilterung in der Praxis vor sich geht.

Nach dem hier beschriebenen Schritt der Bildung von lexikalischen Einheiten aus den ursprünglichen DUW-Angaben haben wir es mit einer Liste von 2294 Einheiten zu tun. Diese Liste wollen wir als *Ausgangswortschatz* bezeichnen. Sie ist in der gedruckten Ausgabe beiliegenden Datei *Ausgangswortschatz.xls* erfasst, zusammen mit den detaillierten Angaben zu Bedeutung, Wortart etc. jeder einzelnen Einheit.

8.4 Die Beschreibung der Plausibilität

Bei der weiteren Beschreibung der Eigenschaften der lexikalischen Einheiten soll es zunächst um die Beschreibung der bei Bonan-Garrigues 1993 beschriebenen introspektiven Eigenschaft der Plausibilität im Sinne einer intuitiven Beurteilung der Möglichkeit, dass ein bestimmtes Wort im Ganzen der schriftlichen und mündlichen Äusserungen, denen wir gewöhnlich ausgesetzt sind, verwendet wird.²⁷¹ Im Gegensatz zur Repräsentiertheit, die allein die Erscheinungsform von Wörtern berücksichtigt, erfordert die Beschreibung der Plausibilität, dass man mit Einheiten von Form und Bedeutung arbeitet, also nicht nur mit Wortformen. Da die Beschreibung der Plausibilität ein iterativer Prozess ist, der in der Relativität des Begriffs der Plausibilität begründet ist (vgl. Bonan-Garrigues 1993, S. 48), bin ich mit einem Abstand von mehreren Wochen die Zuordnung der Plausibilität noch einmal durchgegangen und habe sie vor dem Hintergrund der Erfahrungen aus der ersten Beurteilung noch einmal überarbeitet. Erst als dieser Schritt abgeschlossen war, bin ich zur detaillierten Betrachtung der Stichwörter übergegangen.

Im Unterschied zum bei Bonan-Garrigues beschriebenen Verfahren habe ich dabei allerdings nicht unmittelbar Zahlenwerte eingetragen, sondern eher versucht, die Plausibilität der jeweiligen Einheit verbal zu beschreiben, um so die Trennung von Beschreibung und Gewichtung auch äußerlich deutlich zu vollziehen. Erst in einem späteren Schritt (vgl. Abschnitt 8.8.4) habe ich die verbalen Beschreibungen in ein Bewertungssystem umgesetzt.

Bei der Beschreibung der Plausibilität habe ich versucht, die Plausibilität für jede Einheit individuell zu bestimmen, auch wenn sie ihre Form und eine mögliche Grundbedeutung mit anderen Einheiten teilte. Zu berücksichtigen war, in welcher spezifischen Bedeutung die zu beurteilende lexikalische Einheit in Texten und Gesprächen begegnet, also etwa ob das Wort *Obstler* eher in seiner Bedeutung ‘Obstwasser’ (DUW-ID 301.1.-) oder ‘Obsthändler’ (DUW-ID 301.2.-) plausibler erscheint.

Umgekehrt hatte ich aber auch zu bestimmen, mit welcher Plausibilität mir eine Einheit begegnen wird, die sich die Bedeutung mit einer anderen Einheit teilt, aber eine andere Wortform hat, wie zum Beispiel bei den beiden Einheiten *Obstbau* und *Obstanbau*. Die erste

271. Vgl. Abschnitt 4.5.2.2.

habe ich als gar nicht plausibel beschrieben, die zweite als wenig plausibel, weil mir das Wort *Obstanbau* vertrauter ist und ich glaube, dass es mir eher begegnen wird als *Obstbau*.

Eine interessante Beobachtung im Zuge der Beurteilung der Plausibilität war die Tatsache, dass ich innerhalb einer Wortfamilie oftmals das Adjektiv (*ökologisch*) als eher plausibel eingestuft habe als das Substantiv (*Ökologie*). Zugleich erwies sich manchmal eine differenzierte Bestimmung als schwierig, wenn es sich um sehr feine Bedeutungsunterscheidungen handelte, etwa bei den beiden Einheiten des DUW-Stichwortes *Oxydation*, bei der es nicht möglich gewesen wäre, den beiden Bedeutungen ‘Reaktion, Verbindung eines chemischen Elements od. einer chemischen Verbindung mit Sauerstoff’ (DUW-ID 1660.1.-) und ‘Vorgang, bei dem ein chemisches Element od. eine chemische Verbindung Elektronen abgibt, die von einer anderen Substanz aufgenommen werden’ (DUW-ID 1660.2.-) eventuell unterschiedliche Plausibilitäten zuzuordnen.

Wie schon bei der Beschreibung der Repräsentiertheit habe ich auch hier die Plausibilität zunächst verbal beschrieben, um sie in einem späteren Bewertungsschritt konsistent in ein quantitatives System umsetzen zu können. Dabei habe ich mich allerdings vereinfachend auf drei verschiedene Beschreibungsformen beschränkt:

- ‘wenig plausibel’ (etwa *offen* im Sinne von ‘nicht genügend auf Deckung achtend und so dem Gegner die Möglichkeit zum erfolgreichen Gegenangriff gebend’ (DUW-ID 432.6.-))
- ‘relativ plausibel’ (etwa *Ozonschicht*)
- ‘sehr plausibel’ (etwa *Oktober*).

Im quantitativen Überblick zeigte sich anschließend, dass sich hier im Vergleich zur Repräsentiertheit die Mehrheitsverhältnisse verschoben: Die Mehrzahl der Einheiten (1361 Einheiten) waren als wenig plausibel eingestuft, eine kleinere Gruppe (803 Einheiten) als relativ plausibel, und eine Minderheit (130 Einheiten) als sehr plausibel.

In Abschnitt 8.8.4 werden wir eine Bewertung dieser Eigenschaft vornehmen.

Die der gedruckten Ausgabe beigelegte Datei *Beschreibung_Plausibilität.xls* listet zu allen Einheiten die Plausibilitätsbeschreibungen auf.

8.5 Die Beschreibung der Frequenz

In einem weiteren Teilschritt der Beschreibung habe ich nun die exemplarische Frequenz ermittelt, die die Wortschatzerhebung des sogenannten *Projekt Deutscher Wortschatz* an der Universität Leipzig über das Internet zur Verfügung stellt.²⁷² Der dabei erhobene Frequenzwert ist kein absoluter Wert, sondern gibt an, um wieviel häufiger das im Wortschatz-Projekt absolut häufigste Wort *der* im Vergleich zu dem jeweils abgefragten Wort ist. Ent-

272. Vgl. <http://wortschatz.uni-leipzig.de/>. – Im Folgenden wird das Projekt *Deutscher Wortschatz* verkürzt als *Wortschatz-Projekt* bezeichnet.

sprechend ihrer relativen Häufigkeit werden die Wörter dann in Frequenzklassen von 0 bis 24 gruppiert.²⁷³

8.5.1 Das Projekt *Deutscher Wortschatz* der Universität Leipzig

Das Projekt *Deutscher Wortschatz* stellt ein Vollformenlexikon dar, das 1995 begonnen wurde und seither kontinuierlich erweitert wird. Es enthält zur Zeit ca. 750 Millionen laufende Wortformen und beruht vor allem aus den Volltexten von Tageszeitungen; aber auch Fachlexika, Fachzeitschriften und Monographien zu verschiedenen Wissensgebieten, soweit sie in elektronischer Form von Verlagen und Autoren zur Verfügung gestellt werden, fließen in den Bestand mit ein.²⁷⁴ Zeitweise war es auch für individuelle Nutzer möglich, mit Hilfe einer CD-ROM, auf der der gesamte Wortbestand des Wortschatzprojektes zusammen mit dem Softwareprogramm *Neuwort* gespeichert war, eigene Texte auf Wörter zu untersuchen, die noch nicht im Wortschatzbestand des *Wortschatz*-Projektes enthalten sind, und sie an das Leipziger Institut für Informatik zu schicken.²⁷⁵ Insgesamt beruht also der Wortbestand des Wortschatz-Projektes auf schriftlichen, relativ formalisierten Quellen. Quellen. Material mündlicher Kommunikation oder weniger formale Texte sind nicht verarbeitet. Das ist in der vorliegenden Untersuchung kein allzu gravierendes Problem, als es sich auch bei dem zu untersuchenden Wortschatzausschnitt um einen hauptsächlich schriftsprachlich-konventionalisierten Wortschatz handelt.²⁷⁶

8.5.2 Probleme bei der Bestimmung der Frequenz

Das *Wortschatz*-Projekt an der Universität Leipzig stellt also ein Vollformenlexikon dar. Dabei ist die Erhebung der Frequenz einzelner Wörter nur ein nachrangiges Ziel des Projektes; im Vordergrund stehen vielmehr statistisch basierte Kollokationsanalysen. Die Orientierung an der Vollform, d.h. an der laufenden Form von Wörtern bringt einige Besonderheiten mit sich, die bei der Auswertung der Frequenzangaben für unsere Zwecke von Bedeutung sind.

273. Dabei ist die Ziffer, die die Frequenzklasse bezeichnet, umgekehrt proportional zur Häufigkeit der darin enthaltenen Wörter: Im Vergleich zu Wörtern der Frequenzklasse 1 etwa ist *der* 2¹ häufiger, zu Wörtern der Frequenzklasse 2 ist *der* 2² häufiger, zu Wörtern der Frequenzklasse 3 ist *der* 2³ häufiger etc. – Der Vorteil dieses Verfahrens liegt darin, dass sich die Zugehörigkeit eines Wortes zu einer Frequenzklasse bei Erweiterung des Korpus praktisch nicht ändert, während die absolute Häufigkeit eines Wortes mit der Korpuserweiterung variiert (vgl. Quasthoff 1998, S. 5).

274. Zum Vergleich: Das dem *Français Fondamental* (Gougenheim 1958) zugrundeliegende Korpus enthält 312 000 laufende Wörter und ca. 8 000 Lexeme (vgl. Pfeffer 1975, S. 9), die dem Frequenzwörterbuch von Wängler (Wängler 1963) zugrundeliegenden Korpora 160 000 Einträge. – Für eine ausführliche Darstellung des *Wortschatz*-Projektes und aktueller Daten vgl. Quasthoff / Richter 2005. Für eine Übersicht über die benutzten Quellen vgl. Richter 2004, S. 103f.

275. Vgl. Quasthoff 1997.

276. Ginge es bei der Untersuchung um die konkrete Erfassung und Erarbeitung eines Grundwortschatzes, also eines didaktischen Typs eines zentralen Wortschatzes, der allgemeine, d.h. auch mündliche kommunikative Fähigkeiten vermitteln soll, wäre ein solches Korpus nicht adäquat, weil es den Aspekt der Mündlichkeit vernachlässigt.

Zunächst ist die Angabe von Frequenzklassen, mit der im *Wortschatz*-Projekt gearbeitet wird, gegenüber der Angabe von absoluten Frequenzen zu beachten. Die Angabe der Frequenzklasse, zu der ein Wort gehört, führt mit sich, dass keine Unterscheidung zwischen Wörtern einer Klasse möglich ist. Darüberhinaus besteht die Tendenz, dass niedrige Frequenzklassen überproportional groß werden, weil die meisten Wörter sehr niedrige Frequenzen aufweisen (vgl. Abschnitt 5.3.2). Die direkte Angabe der absoluten Frequenz eines Wortes wäre für unsere Zwecke also der Angabe von Frequenzklassen vorzuziehen. Dem steht die Tatsache gegenüber, dass das Korpus des *Wortschatz*-Projektes ständig erweitert wird und sich dadurch die absoluten Frequenzwerte der einzelnen Wörter laufend ändern. Zudem ist die Angabe der absoluten Frequenz eines Wortes nur dann wirklich sinnvoll, wenn eine genaue Zuordnung von Frequenz und betrachteter lexikalischer Einheit (mitsamt ihrer spezifischen Einzelbedeutung) möglich wäre. Nachdem die Art der Datenerhebung des *Wortschatz*-Projektes eine so detaillierte Zuordnung ohnehin nicht zulässt (s. u.), muss die Angabe von Frequenzklassen genügen.

Gravierender als das Problem der pauschalen Angabe von Frequenzklassen ist die Tatsache, dass im *Wortschatz*-Projekt wie die meisten anderen großangelegten Häufigkeitsuntersuchungen keine unmittelbare Unterscheidung von Unterbedeutungen polysemer Wörter zur Verfügung stellt. Allenfalls eine genaue Untersuchung der zu jeder Wortform angegebenen Kollokationen ließe Rückschlüsse auf die durch sie repräsentierten Unterbedeutungen zu. Da sich die Angabe der Frequenzklasse aber ohnehin auf die Wortform an sich bezieht und nicht auf deren Unterbedeutungen, lässt sie sich nur als pauschale Angabe interpretieren.

Das Problem mit den potentiell weitaus größten Konsequenzen ist schließlich die Tatsache, dass im *Wortschatz*-Projekt nur sehr begrenzt zwischen Wortarten unterschieden wird (etwa anhand der Groß- und Kleinschreibung). Zwar erhält man bei der Suche nach einer bestimmten Wortform durchaus alle Wortarten, denen die entsprechende Wortform zugeordnet werden kann; die zugehörige Frequenzangabe aber ist ein akkumulierter Wert für alle möglichen Wortarten dieser Wortform. Beispiele hierfür ist etwa die Frequenzangabe zum Wort *öden*, die sowohl die Frequenz des Infinitivs *öden* als auch die paradigmatische Adjektivform *öden* umfasst. Sucht man nach dem Eintrag *offenbaren*, so erhält man die Angabe Frequenzklasse 13. Sieht man sich die im Ergebnis aufgelisteten Formen an, dann wird deutlich, dass dort auch Wortformen angegeben sind, die zugleich zum Paradigma des Adjektivs *offenbar* gehören, so zum Beispiel *offenbare*. Das Problem betrifft also die ambigen Wortformen. Bei anderen Einträgen, wie zum Beispiel dem Wort *Offenbarung*, die durch ihre Form eindeutig als Substantiv erkennbar sind, fällt es weg. Dieses Problem tritt übrigens nicht auf, wenn es sich um eine Kombination von Verb und Substantiv oder von Adjektiv und Substantiv handelt, weil das Frequenzberechnungsprogramm offensichtlich anhand der Großschreibung der Substantive die Wortarten unterscheiden kann (vgl. *Offensive* vs. *offensiv*). Ähnlich ambig sind die Frequenzangaben von Wörtern, bei denen eine Unterscheidung von Appellativum und Eigenname wichtig sein kann; der Ein-

trag *Ofenloch* zum Beispiel ist allein als Eigenname (Nachname) verzeichnet, während aber eine Vielzahl konkreter Belege zeigt, dass das Wort als auch als Appellativum verwendet wird.

All diese Einschränkungen sind keine spezifischen Schwächen des *Wortschatz*-Projektes, sondern, wie wir in Abschnitt 4.5.2.1 gesehen haben, Begrenzungen der Frequenzerhebung, die auch mit den heutigen Mitteln noch nicht in großem Umfang zwischen Einzelbedeutungen unterscheiden kann. Dennoch stehen wir vor einem Dilemma: Einerseits ist Frequenz eine Eigenschaft, anhand derer wir Wörter beschreiben können und wollen, weil wir vermuten, dass sie grundsätzlich ein gutes Kriterium zur generellen Erfassung der Gebräuchlichkeit von Wörtern darstellt. Andererseits sind die Daten des *Wortschatz*-Projektes, das derzeit die einzige frei zugängliche, aktuelle und umfangreiche Frequenzerhebung des Deutschen darstellt, nicht in vollem Ausmaß verwendbar. Es gilt also, die darin zugänglichen Informationen soweit wie möglich zu nutzen und zugleich die inhärenten Begrenzungen im weiteren Verfahren zu berücksichtigen.

8.5.3 Konsequenzen

Was bedeuten die gerade beschriebenen Einschränkungen des *Wortschatz*-Projektes für die weitere Bearbeitung unseres Wortbestandes?

Betroffen ist zunächst die hier vorzunehmende Erfassung der Frequenz der Einheiten unserer Untersuchungsmenge. Während die Frequenzangaben des *Wortschatz*-Projektes sich pauschal auf die Wortform beziehen und keinen Unterschied zwischen polysemen bzw. homonymen Teilbedeutungen machen, beziehen sich die Einheiten der Untersuchungsmenge jeweils auf genau eine Bedeutung. Damit können aber die feinen Bedeutungsunterscheidungen, die im DUW vorgenommen werden und die wir auf die Untersuchungseinheiten übertragen haben, nicht in den Frequenzangaben reflektiert werden. Wir legen also vereinfachend fest, dass allen Einheiten, die sich eine Wortform teilen, die gleiche Frequenzklasse zugeteilt wird, d.h. diejenige Frequenzklasse, die die Wortform im *Wortschatz*-Projekt hat. Damit werden zum Beispiel alle lexikalischen Einheiten, die aus dem DUW-Eintrag *offen* abgeleitet sind, der Frequenzklasse 8 zugeordnet.²⁷⁷

Wichtiger noch ist es jedoch, die Begrenzungen der Frequenzbeschreibung bei der Bewertung dieser Eigenschaft zu berücksichtigen (vgl. Abschnitt 8.8.5). Konkret bedeutet das, dass wir uns bei der Bewertung der Frequenz auf die niedrigfrequenten Frequenzklassen konzentrieren und sie dazu heranziehen, in sehr cursorischer Weise niedrigfrequente von hochfrequenten Wörtern zu unterscheiden, ohne dabei eine Rangordnung aufstellen zu wollen. Denn eine geringe Frequenz ist in der Regel darauf zurückzuführen, dass dahinter nur eine Bedeutung bzw. ein Wort einer Wortart steckt, sodass man mit der alleinigen Konzen-

277. Dies gilt hingegen nicht für Einheiten, die zwar ursprünglich von einem gemeinsamen DUW-Stichwort abstammen, wie etwa *Obstanbau* und *Obstbau* (beide abgeleitet aus dem DUW-Eintrag *Obstanbau*, *Obstbau* (DUW-ID 283)). In diesen Fällen wurden die Frequenzen für jede Form gesondert ermittelt.

tration auf Wörter der niedrigfrequenten Klassen das Problem hochfrequenter Wörter, deren Frequenzangaben auf homonymen Wortformen verschiedener Grundformen beruhen, weitgehend minimieren kann.

Unabhängig davon habe ich bei der Übernahme der Frequenzangaben versucht, soweit wie möglich die jeweiligen Problemfälle zu kommentieren. In der Praxis bedeutet das, dass nach diesem Arbeitsschritt jede Untersuchungseinheit mit einer der jeweiligen Frequenzklasse entsprechenden Frequenzziffer zwischen 1 und 24 und – soweit es sich bei der Frequenzangabe um einen der hier besprochenen Problembereiche handelt – einem beschreibenden Kommentar versehen ist. Zusätzlich habe ich den Einheiten, die nicht im *Wortschatz*-Projekt zu finden waren, die Frequenzziffer 25 zugeordnet.²⁷⁸

Tabelle A-6 in Appendix A gibt die in früheren Beispieltabellen aufgeführten Beispielseinheiten mit ihren Frequenzangaben wieder. Der gesamte Ausgangswortschatz, also alle zu untersuchenden lexikalischen Einheiten, ist mit den jeweiligen Frequenzangaben in der gedruckten Ausgabe beiliegenden Datei *Beschreibung_Frequenz.xls* elektronisch abrufbar.

8.6 Die Beschreibung der zeitlichen Stabilität

Wie wir in den Abschnitten 4.4.3 und 4.5.2.3 gesehen haben, hat die Bestimmung der Stabilität eines Wortes zum Zweck, Wörter, die über einen bestimmten Zeitraum hinweg fester Bestandteil des zu untersuchenden Wortschatzes sind, von Wörtern, die noch nicht im Wortschatz etabliert sind, zu unterscheiden. Während die Frequenz die Zentralität eines Wortes ‘im Augenblick’ betrachtet, sozusagen in einer Art horizontalen Momentaufnahme des Wortschatzes, ist die Untersuchung der Stabilität eine ‘vertikale’ Abbildung der Zentralität, d.h. über mehrere Zeiträume hinweg.

Wie aber findet man heraus, ob ein Wort bzw. eine lexikalische Einheit stabil ist? Auf den ersten Blick scheint es sich dabei um eine einfache Frage zu handeln. Man entscheidet sich für einen bestimmten Zeitraum und schlägt das Wort in verschiedenen Wörterbüchern, die diesen Zeitraum abdecken, nach. Wenn sich das Wort erst in den jüngsten untersuchten Wörterbüchern zeigt, dann ist es nicht stabil, wenn es sich sowohl in den ältesten als auch in den jüngsten Wörterbüchern zeigt, dann ist es stabil. Diese vereinfachte Herangehensweise birgt jedoch einige Risiken. Sie sollen im folgenden kurz diskutiert werden.

278. Die Frequenzklassenangaben 24 bezieht sich dabei auf einen Sonderfall: Es gibt Wörter, die zwar zu einem Suchergebnis im *Wortschatz*-Projekt führen, aber mit einer absoluten Beleganzahl von 0 angegeben sind. Vgl. etwa die Wörter *Oblation* oder *oktaedrisch*. Nach Angaben von Mitarbeitern des *Wortschatz*-Projektes hat dies historische Gründe. Die entsprechenden Wörter waren in einer frühen Version der Frequenzerhebung registriert, sind aber im aktuellen Korpus nicht belegt.

8.6.1 Zur Eignung des Vergleichs von Wörterbüchern zur Bestimmung der Stabilität lexikalischer Einheiten

Eine grundsätzliche Überlegung, die beim Vergleich von Wörterbüchern eine Rolle spielt, betrifft die Frage nach dem Zeitraum, auf den die Untersuchung sich beziehen soll. Dabei müssen Fragen berücksichtigt werden wie: Soll ein möglichst breiter Zeitraum abgedeckt werden, sollen möglichst frühe Quellen herangezogen werden? Und wie verhält es sich mit der zeitlichen Lücke zwischen der Erarbeitung eines Wörterbuchs durch einen Lexikographen und seiner Veröffentlichung? Auf welchen zeitlichen Querschnitt bezieht sich das Wörterbuch?²⁷⁹

Zum anderen sagt ein einfacher Vergleich zweier einsprachiger Wörterbücher nicht unbedingt direkt etwas darüber aus, wie lange ein Wort sich bereits im Wortschatz befindet. Vielmehr reflektiert das Nichtvorhandensein eines Wortes in einem der untersuchten Wörterbücher eher Unterschiede in der lexikographischen Herangehensweise, lexikographische ‘Moden’ sozusagen. Dies hat Matthias Schulz in einer kritischen Untersuchung dreier Beispiele eines Wörterbuchvergleichs deutlich gemacht (vgl. Schulz 2000): Zum einen dürfe man nicht den Inhalt eines spezifischen Wörterbuchs mit dem Lexikon einer Sprache gleichsetzen. Denn das Vorhandensein oder Nichtvorhandensein eines Lexems in einem spezifischen Wörterbuch sage etwas über die Zusammensetzung dieses Wörterbuchs aus, aber nicht unmittelbar etwas über die Zusammensetzung des Wortschatzes der Zeit, in der das Wörterbuch entstanden ist (vgl. Schulz 2000, S. 64). Zum anderen dürfe man das, was an Daten über das jeweilige Lexem lexikographisch im untersuchten Wörterbuch vorhanden ist, nicht gleichsetzen mit dem tatsächlichen Gebrauch des Lexems in der Sprache. Angaben des Lexikographen, das Wörterbuch würde die Gesamtheit des Lexikons einer Sprache abbilden, seien gerade bei älteren Wörterbüchern aufgrund der Materiallage nicht als wahr zu begreifen (vgl. ebda.). Vor allem das Nichtvorhandensein eines Lexems dürfe nicht als Aussage über die Sprache verstanden werden, sondern “höchstens als Aussage über die Buchungsprinzipien des Lexikographen“ (Schulz 2000, S. 64). Und schließlich sei die Gegenüberstellung zweier Wörterbücher dadurch erschwert, dass sie selten nach den gleichen Aufnahmeprinzipien erstellt und in ihrem Umfang in der Regel nicht identisch seien (Schulz 2000, S. 64). Schulz konkludiert damit, dass Textkorpora für eine zuverlässige und aussagekräftige Untersuchung des Wortschatzwandels eine geeignetere Grundlage darstellten (vgl. ebda., S. 73f.).

Mit diesen Argumenten und der ausführlichen Analyse der drei Beispieluntersuchungen macht Matthias Schulz eine Grundannahme der Lexikologie und Lexikographie konkret

279. Eine Alternative zu dem hier vorgeschlagenen Verfahren könnte auch sein, in einem ausführlichen etymologischen Wörterbuch nachzusehen, wann das zu beschreibende Lexem in die Sprache gekommen ist. Damit wird man allerdings die wenigsten Lexeme abdecken können, weil etymologische Wörterbücher oft die ‘außergewöhnlichen’ oder auch verdunkelten Lexeme beschreiben, nicht aber zum Beispiel Ableitungen oder Zusammensetzungen. Ich werde daher auf den Vergleich verschiedener zweisprachiger Wörterbücher des Deutschen zurückgreifen.

fassbar, dass nämlich der Vergleich von Wörterbüchern mit dem Ziel, Veränderungen im Wortschatz aufzudecken, durchaus mit Vorsicht zu genießen ist. Dem liegt ein grundsätzliches Problem zugrunde: Wörterbücher sind Kodifikationen des Wortschatzes und damit Abstraktionen des Lexikons einer Sprache. Mit ihrer Erarbeitung führt man unweigerlich subjektive Elemente ein. Damit aber enthalten zwei Wörterbücher, die ja zwei verschiedene Abstraktionen des Wortschatzes darstellen, auch unterschiedliche subjektive Elemente und sind damit nur bedingt vergleichbar.

Diesem Problem steht wiederum der Bedarf an Material zur Bestimmung von Veränderungen im Wortschatz gegenüber. Gäbe es die von Schulz als Vergleichsgrundlage geforderten Korpora in der notwendigen Form, dann wäre ein Vergleich dieser Korpora sehr wohl einem Vergleich von Wörterbüchern vorzuziehen. Aber in der Regel sind die zugänglichen Korpora einfache Listen von Wörtern oder Volltextkorpora ohne semantisch-lexikologische Angaben. Schlägt man in diesen Textkorpora nach, um das Vorhandensein eines Lexems in verschiedenen Sprachstadien zu untersuchen, kann man in der Regel nur Wortformen untersuchen, nicht aber Sememe bzw. lexikalische Einheiten. Oder man muss sich die Sememe erst aus den unterschiedlichen Belegen zusammensetzen. Für eine Untersuchung einer Vielzahl von lexikalischen Einheiten ist das eine große praktische Hürde. Solange es keine Textkorpora gibt, die in einer Art und Weise aufbereitet sind, die es erlaubt, direkt auf einzelne Einheiten zuzugreifen, und die zugleich für eine Vielzahl von Sprachstadien vorhanden sind, muss man sich mit der komprimierten Darstellung von semantisch-lexikologischen Angaben in Wörterbüchern behelfen.

Damit stehen wir vor einem ähnlichen Dilemma wie bei der Diskussion der Verwendbarkeit der Frequenzangaben des *Wortschatz*-Projektes der Universität Leipzig (vgl. Abschnitt 8.5) – wir sind auf die Erhebung bestimmter Informationen angewiesen, die zugänglichen Daten haben aber gewisse Begrenzungen. Und auch hier gilt es, diesem Dilemma durch eine differenzierte Verwendung der Informationen zu begegnen. Konkret bedeutet das: Zum einen müssen die herangezogenen Wörterbücher sehr bewusst gewählt werden. Zum anderen müssen die Begrenzungen der zugänglichen Daten bei der Bewertung der Eigenschaft der Stabilität entsprechend berücksichtigt werden. Das bedeutet wiederum, dass der Wörterbuchvergleich zunächst vor allem dazu dienen soll, vermutete Neologismen als solche zu bestätigen, dass aber auch das Nichtvorhandensein einer Untersuchungseinheit in einem der beiden Wörterbücher nicht überbewertet wird. Und schließlich bedeutet dies auch, dass die durch den Wörterbuchvergleich erarbeitete Eigenschaft der Stabilität nicht die einzige Eigenschaft sein darf, anhand derer eine Einheit als zentral oder nicht zentral bewertet wird, gerade weil den Ergebnissen eines Wörterbuchvergleichs nur in Maßen vertraut werden kann.

Die Frage nach den herangezogenen Wörterbüchern soll im folgenden diskutiert werden; die Frage nach der adäquaten Berücksichtigung der Datenlage soll bei der Bewertung der Stabilität diskutiert werden (vgl. Abschnitt 8.8.6).

8.6.2 Die herangezogenen Wörterbücher

Um so weit wie möglich auszuschließen, dass allein lexikographische Eigenheiten der herangezogenen Wörterbücher und nicht tatsächliche Veränderungen im Wortschatz den Ausschlag geben für das Fehlen oder Vorhandensein einer lexikalischen Einheit, war es wichtig, Wörterbücher zu wählen, die sich ihrer lexikographischen Anlage möglichst ähnlich waren. Dies war mit den mir zur Verfügung stehenden einsprachigen Wörterbüchern nicht möglich, weil sie in der Regel nicht detailliert genug sind. Ich habe mich daher entschieden, auf zweisprachige Wörterbücher zurückzugreifen und zur Beurteilung der Stabilität zwei deutsch-englische Wörterbücher herangezogen: Wildhagen 1972 und Muret-Sanders 1899.²⁸⁰

Wildhagen 1972 ist die überarbeitete Neuauflage eines 1953 zum erstenmal veröffentlichten zweibändigen englisch-deutschen / deutsch-englischen Wörterbuch. Es ist...

„...als eine *enzyklopädische, streng wissenschaftliche Darstellung* des deutschen Wortschatzes der Neuzeit und Gegenwart mit besonderer Berücksichtigung der *Syntax*, des *Stils* und des *idiomatischen Brauchs*“ (Wildhagen 1972, S. XXVI)

konzipiert. Die Stichwörter bei Wildhagen sind nicht so reichhaltig mit grammatischen und etymologischen Angaben versehen wie die Stichwörter des DUW, auch fehlt oft eine Bedeutungsumschreibung. Diese wird aber durch die Angabe der englischen Entsprechung kompensiert, und auch die reichhaltige Angabe von Beispielsätzen unterstützt die Identifizierung der lexikalischen Einheiten. Leider enthält der einleitende Teil des Wörterbuchs keine Angaben über den Stichwortumfang, daher ist es kaum möglich, Wildhagen und DUW vom Umfang her miteinander zu vergleichen, abgesehen von der Tatsache, dass beide einbändige Wörterbücher sind.

Während es bei Wildhagen um die möglichst kompakte Darstellung großer Mengen von Informationen geht, steht bei Muret-Sanders die möglichst deckende Erfassung des deutschen Wortschatzes im Vordergrund.²⁸¹

Nachdem die Beschreibung der Stabilität und die Auswahl der Wörterbücher in gewisser Weise arbiträr und zugleich eine optimale Datengrundlage nicht möglich ist, ist es um so wichtiger, die Beurteilungsgrundlage, d.h. die Wörterbücher, auf denen sich die Stabilität gründen soll, genau in ihren Eigenheiten und Begrenzungen zu kennen. Dies gilt zumindest in dem Maße, in dem diese einen Einfluss auf das Ergebnis haben. Für die hier vorzunehmende Untersuchung – nämlich die Beurteilung der Stabilität eines Wortes auf der Grundlage dessen, ob es in einem älteren Wörterbuch bereits erwähnt ist oder nicht – sind daher alle Aspekte, die mit dem Vorhandensein bzw. der Auffindbarkeit eines Wortes im jeweiligen Wörterbuch zu tun haben, wichtig. Denn es geht darum, herauszufinden, was die

280. In den Auszugstabellen der Datenbank abgekürzt mit WH bzw. MUS.

281. Vgl. das Vorwort zur ersten Hälfte des deutsch-englischen Bandes (Muret-Sanders 1899a).

Gründe dafür sind, dass ein Wort in einem Wörterbuch nicht verzeichnet ist. Damit aber geht es konkret um folgende Fragen:

- Wortschatzselektion:

Unter welchen Gesichtspunkten haben die Autoren des jeweiligen Wörterbuchs die Stichwörter ausgewählt? Im Bereich der Wortschatzselektion versucht Wildhagen möglichst breit vorzugehen; sowohl dialektale als auch stilistische Varianten und verschiedene Sachgebiete sollen berücksichtigt sein. Ebenso verfahren Muret-Sanders, denen ganz explizit an einer möglichst umfassenden Erfassung des Wortschatzes gelegen ist. Dabei sind allerdings folgende Aspekte zu berücksichtigen: Anglizismen und Amerikanismen, „die in die deutsche Sprache Eingang gefunden haben und als solche noch empfunden werden“ (Wildhagen 1972, S. XVII) sind bei Wildhagen nicht im deutsch-englischen Band verzeichnet, sondern im englisch-deutschen Band. Muret-Sanders wiederum ist deutlich am Bildungsstandard des Universalgelehrten orientiert; das wird deutlich an den Quellbeispielen, die im Vorwort genannt werden: Das Wörterbuch soll den Wortschatz von Martin Luther, Albrecht Dürer, Paracelsus und anderen wiedergeben (vgl. Muret-Sanders 1899a, S. VII) und auch dafür sorgen, dass Wörter abgedeckt sind, die zwar eigentlich nicht zum unmittelbaren Wortschatz gehören, aber bei wichtigen Autoren vorkommen.²⁸² Dieser Aspekt ist jedoch weniger problematisch, denn ob ein Wort aufgeführt ist, das nicht in den von mir untersuchten Wortschatzausschnitt fällt, besagt nichts über die Stabilität der zu untersuchenden Wörter. Zentral ist die Frage, ob ein Wort des DUW-Wortschatzausschnittes aufgeführt ist oder nicht.

- Wortschatzanordnung:

Welche Anordnungsprinzipien haben die Autoren angewandt? Bei der Untersuchung, ob ein gegebenes Wort in einem Wörterbuch aufgeführt ist oder nicht, spielt auch die Anordnung der Stichwörter im herangezogenen Wörterbuch eine Rolle. Denn wenn ein Wort nicht an der erwarteten Stelle aufzufinden ist, bedeutet das noch nicht, dass es nicht im Wörterbuch aufgelistet ist, sondern es kann auch an anderer Stelle zu finden sein. Das hat sich bei beiden Wörterbüchern als etwas schwierig erwiesen: Wildhagen scheint zwar einem alphabetischen Ordnungsprinzip zu folgen, das aber von dem im DUW abweicht. Bei Muret-Sanders werden die Stichwörter, sofern sie Zusammensetzungen sind, teilweise bei ihrem Grundwort angesiedelt. Das bedeutet, dass man bei der Untersuchung, ob ein bestimmtes Wort bei Muret-Sanders aufgeführt ist, in größerem Maße nachschlagen muss als bei einem Wörterbuch, das streng einer rein alphabetischen Anordnung folgt.

Vor dem Hintergrund dieser Eigenheiten der herangezogenen Wörterbücher geht es nun darum, festzulegen, wann wir davon ausgehen wollen, dass eine Einheit der Untersu-

282. Vgl. die Anmerkung auf der selben Seite, dass auch schottische Wörter aufgenommen sind, soweit sie bei Walter Scott vorkommen.

chungsmenge im jeweiligen Wörterbuch aufgeführt ist, und wann nicht. Dazu gehört eine praktikable Definition der lexikalischen Identität; denn wenn man bestimmen will, wann ein Lexem bereits in einem Wörterbuch verzeichnet ist, muss man auch wissen, wann man ein Stichwort in einem Wörterbuch als lexikalisch identisch mit der Einheit, deren Stabilität bestimmt werden soll, betrachtet.

8.6.3 Die Frage nach der lexikalischen Identität

Wir gehen davon aus, dass die lexikalische Identität zweier zu vergleichender Wörter von drei Faktoren bestimmt wird – von der graphischen Form, von der angegebenen grammatischen Form und von der Bedeutungsangabe.²⁸³ Betrachtet man die untersuchten lexikalischen Einheiten und die Stichwörter bei Wildhagen und Muret-Sanders unter diesen Aspekten, dann ergeben sich verschiedene Typen bzw. Grade von lexikalischer Identität, die sich auf einer Skala anordnen lassen:

Am einen Ende der Skala haben wir den Fall, dass zum einen die Wortform und die grammatischen Angaben der Einheiten des DUW-Bestandes mit der Wortform bzw. den grammatischen Angaben des jeweiligen Wörterbucheintrages bei Wildhagen bzw. Muret-Sanders übereinstimmen, und dass zum anderen sowohl lexikalische Einheit als auch Stichwort eine einzige, übereinstimmende Bedeutung angeben und man daher annehmen muss, dass sie in ihrer Bedeutung identisch sind. Am anderen Ende der Skala stehen die Fälle, in denen der lexikalischen Einheit bei Wildhagen bzw. Muret-Sanders kein Stichwort gegenübersteht, wie zum Beispiel *ostensibel* oder *Ostermarsch* bei Wildhagen. Dazwischen befinden sich alle Mischfälle.

Um eine konsistente Beschreibung der lexikalischen Identität zwischen einer Einheit der Untersuchungsmenge und dem jeweiligen Stichwort bei Wildhagen bzw. Muret-Sanders zu gewährleisten, habe ich den Begriff der lexikalischen Identität in drei verschiedene Teilidentitäten zerlegt und jede Einheit der Untersuchungsmenge dahingehend mit dem Wortbestand bei Wildhagen und Muret-Sanders abgeglichen. Die zu untersuchenden Teilidentitäten waren:

- graphische Identität, d.h. die Identität der (orthographischen) Form der lexikalischen Einheit mit dem jeweiligen Wörterbucheintrag
- grammatische Identität, d.h. die Identität der grammatischen Angaben (Genus, Valenzangaben) der lexikalischen Einheit mit denen des jeweiligen Wörterbucheintrages
- Bedeutungsidentität, d.h. der Grad der semantischen Übereinstimmung zwischen lexikalischer Einheit und dem jeweiligen Wörterbucheintrag

Während sich der Abgleich der Form bzw. grammatischen Angaben relativ problemlos durchführen ließ, war bei der Bestimmung der Bedeutungsidentität ein differenziertes Vor-

283. Vgl. auch die drei Eigenschaftsbündel, die die Worthaftigkeit eines Wortschatzelementes ausmachen, wie bei Schindler 2002 beschrieben und in Abschnitt 8.8.3 referiert.

gehen nötig.²⁸⁴ Ausgangspunkt war die Annahme, dass die englische Übersetzung des jeweiligen Wörterbucheintrages mitsamt allen Beispielsätzen und Untereinträgen als Bedeutungsangabe gilt; darauf aufbauend habe ich drei Grade der Bedeutungsidentität unterschieden:

1) direkte Bedeutungsidentität:

Wenn es zur untersuchten lexikalischen Einheit einen monosemen Wörterbucheintrag gibt, dessen englische Übersetzung mit der Bedeutung der lexikalischen Einheit übereinstimmt, dann habe ich diese Identität als direkte Bedeutungsidentität klassifiziert (vgl. etwa *Obstkorb* oder *ornamentieren* bei Wildhagen).

2) abgeleitete Bedeutungsidentität

Wenn der der lexikalischen Einheit zugehörige Wörterbucheintrag ein polysemes Lexem darstellte, dessen Bedeutungsangaben einen Anhaltspunkt dafür gaben, dass die Bedeutung der Einheit in diesem Wörterbucheintrag enthalten ist, habe ich diese Bedeutungsidentität als abgeleitet bezeichnet (vgl. etwa *Oase* (DUW-ID 15) bei Wildhagen). Dabei gilt als Anhaltspunkt eine der Bedeutung der Untersuchungseinheit entsprechende englische Bedeutungsangabe oder ein entsprechender Beispielsatz.²⁸⁵ Auch wenn es sich bei der Einheit der Untersuchungsmenge bzw. dem Wörterbucheintrag um einen Internationalismus handelt, der direkt ins Englische übersetzt wurde ((vgl. *obedience* bei Wildhagen 1972), gilt in diesem Fall die Bedeutung der Untersuchungseinheit als aus den Bedeutungsangaben des Wörterbucheintrages ableitbar. Die Bedeutung der Untersuchungseinheit gilt allerdings nicht als ableitbar, wenn sie aufgrund ihrer Sondermarkierung an einen speziellen Kontext geknüpft ist, der in den Angaben zum jeweiligen Wörterbucheintrag fehlt, wie das etwa bei medizinischen Termini der Fall ist.²⁸⁶ Die Angabe des spezifischen Kontextes ‘Medizin’ durch die

284. Natürlich ließen sich auch bei der Bestimmung der graphischen Identität einige wenige Grenzfälle nicht vermeiden. Wenn man etwa graphische Varianten zuläßt, wann spricht man dann etwa vom Vorhandensein einer Einheit in dem untersuchten Wörterbuch? Vgl. etwa die Einheit *Oktagon*, die bei Wildhagen nicht in genau der Form auftaucht, sondern nur als *Oktogon*. In diesem speziellen Fall, bei dem die entsprechende Variante – *Oktogon* – auch eine lexikalische Einheit darstellt, habe ich die Wortschatzeinheit *Oktagon* als nicht bei Wildhagen repräsentiert betrachtet, um eine Doppelzählung zu vermeiden.

Die Frage nach dem Vorhandensein oder Fehlen bestimmter Einheiten stellt sich auch im Falle von Komposita. Beispielsweise findet sich die Einheit *Ölwechsel* nicht als solche bei Wildhagen, aber durchaus in der komplexeren Form *Ölwechselzeit*. Man könnte in diesem Fall also annehmen, die Einheit *Ölwechsel* sei durchaus auch indirekt bei Wildhagen vertreten. Diese feinen Unterschiede habe ich allerdings aus Gründen der Arbeitsökonomie nicht registriert. Ähnliches gilt auch für andere reihenbildende Formen, wie etwa *Opportunismus* (nicht bei Wildhagen vorhanden) gegenüber *Opportunist* (das bei Wildhagen vorhanden ist).

285. Dabei richten wir uns vereinfachend nach Haß-Zumkehr (2001, S. 25f. und S. 32), für die zwei Bedeutungen dann als identisch gelten, wenn sie denselben Bezugsbereich haben, wenn also etwa die Bedeutungsangaben miteinander verknüpfte Synonyme oder Hyperonyme verwenden.

286. Vgl. etwa *sich organisieren* in der Bedeutung ‘selbsttätig in gesundes Gewebe umwandeln’ (DUW-ID 1258.4.-), das im DUW als medizinischer Fachbegriff markiert ist, bei Wildhagen und Muret-Sanders aber nicht gesondert markiert wird. Ähnlich auch: *Ordensstern* (‘Stapelia’, DUW-ID 1187.2.-), *ordinär* (in seinen abwertenden Bedeutungen, DUW-ID 1197.1.n), *Ordination* (‘ärztliche Praxis’, DUW-ID 1203.2.c), *Ost* (‘Ostmark, Ostgeld’, DUW-ID 1483.1.c)

Sondermarkierung bildet einen wichtigen Bestandteil der Bedeutung der lexikalischen Einheit, weshalb ihr Fehlen im jeweiligen Wörterbuch die Bedeutungsidentität beschränkt.

Ähnliches gilt für den umgekehrten Fall, wenn also ein Wörterbucheintrag bei Wildhagen oder Muret-Sanders sich explizit auf eine bestimmte Verwendungsweise des Stichwortes bezieht.²⁸⁷ Auch hier kann man nicht unmittelbar davon ausgehen, die Bedeutung der Einheit unserer Untersuchungsmenge sei durch das jeweilige Stichwort bei Wildhagen oder Muret-Sanders abgedeckt oder daraus ableitbar. Entsprechend wurden Untersuchungseinheiten, die nur in einer spezifischen Bedeutung bei Wildhagen oder Muret-Sanders angegeben waren, als in ihrer Bedeutung nicht ableitbar eingestuft.

3) Fälle, in denen eine Bedeutungsidentität nicht ableitbar war:

Wenn es sich bei dem einer Untersuchungseinheit zugehörigen Wörterbucheintrag um ein polysemes Lexem handelt, dessen Bedeutungsauffächerung keinen Anhaltspunkt dafür gibt, dass die spezifische Bedeutung der Untersuchungseinheit abgedeckt ist, habe ich diese Bedeutungsidentität als nicht identifizierbar eingestuft (vgl. *Operette* (DUW-ID 981 bei Wildhagen)). Das gilt auch umgekehrt für den Fall, wenn die Untersuchungseinheit ursprünglich Teil eines polysemen Lexems im DUW war, der Wörterbucheintrag bei Wildhagen oder Muret-Sanders aber nur ein Monosem ist; auch hier läßt sich die spezifische Bedeutung der Untersuchungseinheit nicht unmittelbar im Wörterbucheintrag identifizieren und gilt daher als nicht identifizierbar (vgl. *Obedienz* (DUW-ID 38.1.- und 38.2.-) bei Wildhagen).

Die so gewonnenen Einzelidentitäten (Identität der graphischen Form, Identität der grammatischen Angaben, Bedeutungsidentität) habe ich in einem abschließenden Schritt für jede untersuchte lexikalische Einheit zu einer Gesamtidentität zusammengefasst. Dabei habe ich die verschiedenen Einzelidentitäten unterschiedlich gewichtet. So wurde etwa eine schwache Bedeutungsidentität (abgeleitete oder nicht ableitbare Bedeutungen) gegenüber fehlender graphischer oder grammatischer Identität geringer gewichtet, weil eine vermeintlich fehlende Bedeutung immer auch durch eine allgemeine Grundbedeutung abgedeckt sein kann.²⁸⁸ Insgesamt entstand daraus eine Dreiteilung:

- Als identisch werden demnach Untersuchungseinheit und Wörterbucheintrag betrachtet, wenn sie in graphischer und grammatischer Form übereinstimmen und zudem eine direkte oder abgeleitete Bedeutungsidentität aufweisen. Sie gelten auch dann als iden-

287. Vgl. etwa *Operateur* bei Muret-Sanders, das auf die Verwendung 'Chirurg' beschränkt ist.

288. Mit Abweichung in der grammatischen Form meine ich vor allem abweichendes Genus, abweichende Valenz und Abweichungen bei Singulariatantum bzw. Pluraliatantum (wie zum Beispiel bei der Untersuchungseinheiten *Originalaufnahme* vs. dem Stichwort *Originalaufnahmen* bei Wildhagen). Mit Abweichung in der graphischen Form meine ich lautgleiche Varianten in der Schreibung (*Occultismus* vs. *Okkultismus*, *Oxidation* vs. *Oxydation*) und Unterschiede in der Gestaltung der Fugenelemente (*olivfarben* vs. *olivenfarben*, *Omelette* vs. *Omelett*).

tisch, wenn sie nicht in graphischer Form übereinstimmen, aber eine direkte oder ableitbare Bedeutungsidentität aufweisen.

- Als teilweise identisch gelten Untersuchungseinheit und Wörterbucheintrag, wenn sie entweder keine ableitbare Bedeutungsidentität aufweisen oder in den grammatischen Angaben voneinander abweichen.
- Als nicht identisch schließlich gelten solche Paare von Untersuchungseinheiten und Wörterbucheinträgen, bei denen sowohl die graphische und grammatische Form voneinander abweichen als auch eine klare (direkte oder abgeleitete) Bedeutungsidentität fehlt oder wo es gar keinen der Untersuchungseinheit entsprechenden Wörterbucheintrag gibt.

Tabelle 8-1 listet die spezifischen Kombinationen von Einzelidentitäten und ihre Zusammenführung zu Gesamtidentitäten detailliert auf:

Tabelle 8-1 Typen lexikalischer Identität

Gesamtidentität	LE als Wörterbucheintrag vorhanden?	Orthographische Identität	Grammatische Identität	Bedeutungsidentität	Begründung für den Grad der Gesamtidentität	Beispiel
identisch:	ja	identisch	identisch	direkt: LE und Wörterbucheintrag sind Monoseme	alle drei Teilidentitäten sind voll ausgeprägt	Obstgarten bei WH
		identisch	identisch	ableitbar: Wörterbucheintrag ist ein Polysem mit einer der LE entsprechenden Unterbedeutung oder einem entsprechenden Beispielsatz	alle drei Teilidentitäten sind voll ausgeprägt	Original (DUW-ID 1323.3.-) bei WH
		abweichend	identisch	direkt: LE und Wörterbucheintrag sind Monoseme	Abweichungen in der orthographischen Form soll ein geringes Gewicht beigemessen werden, solange die Bedeutung klar identifizierbar ist.	Okkultismus (DUW-ID 617) bei MUS
		abweichend	identisch	ableitbar: Wörterbucheintrag ist ein Polysem mit einer der LE entsprechenden Unterbedeutung oder einem entsprechenden Beispielsatz	Abweichungen in der orthographischen Form soll ein geringes Gewicht beigemessen werden, solange die Bedeutung klar identifizierbar ist.	okzidentalisch (DUW-ID 694.n.-n) bei MUS
teilweise identisch:	ja	identisch	identisch	Identität nicht ableitbar: Entweder ist der Wörterbucheintrag ein Polysem ohne eine direkt erkennbare, der LE entsprechende Unterbedeutung. Oder die LE ist Teil eines polysemen DUW-Eintrages und der Wörterbucheintrag ist ein Monosem ohne entsprechende polyseme Auffächerung.	Es läßt sich nicht ausschließen, dass die in der LE angegebene Bedeutung durch die Grundbedeutung oder eine der Unterbedeutungen des Wörterbucheintrages abgedeckt ist.	Operette (DUW-ID 981.-b) bei WH
		abweichend	identisch	Identität nicht ableitbar: Entweder ist der Wörterbucheintrag ein Polysem ohne eine direkt erkennbare, der LE entsprechende Unterbedeutung. Oder die LE ist Teil eines polysemen DUW-Eintrages und der Wörterbucheintrag ist ein Monosem ohne entsprechende polyseme Auffächerung.	Es läßt sich nicht ausschließen, dass die in der LE angegebene Bedeutung durch die Grundbedeutung oder eine der Unterbedeutungen des Wörterbucheintrages abgedeckt ist.	Opazität (DUW-ID 945.n.n) bei MUS
		identisch	abweichend	direkt: LE und Wörterbucheintrag sind Monoseme	Die grammatische Form ist nicht ganz identisch	Oman (DUW-ID 837.-) bei WH
		identisch	abweichend	ableitbar: Wörterbucheintrag ist ein Polysem mit einer der LE entsprechenden Unterbedeutung oder einem entsprechenden Beispielsatz	Die grammatische Form ist nicht ganz identisch	-
		identisch	abweichend	Identität nicht ableitbar: Entweder ist der Wörterbucheintrag ein Polysem ohne eine direkt erkennbare, der LE entsprechende Unterbedeutung. Oder die LE ist Teil eines polysemen DUW-Eintrages und der Wörterbucheintrag ist ein Monosem ohne entsprechende polyseme Auffächerung.	Es läßt sich nicht ausschließen, dass die in der LE angegebene Bedeutung durch die Grundbedeutung oder eine der Unterbedeutungen des Wörterbucheintrages abgedeckt ist.	Orchis (DUW-ID 1170) bei WH
		abweichend	abweichend	direkt: LE und Wörterbucheintrag sind Monoseme	Zwar ist die Bedeutung identisch, aber orthographische und grammatische Form sind nicht ganz identisch	Opera buffa (DUW-ID 959) bei MUS
		abweichend	abweichend	ableitbar: Wörterbucheintrag ist ein Polysem mit einer der LE entsprechenden Unterbedeutung oder einem entsprechenden Beispielsatz	Zwar ist die Bedeutung identisch, aber orthographische und grammatische Form sind nicht ganz identisch	-
nicht identisch:	ja	abweichend	abweichend	Identität nicht ableitbar: Entweder ist der Wörterbucheintrag ein Polysem ohne eine direkt erkennbare, der LE entsprechende Unterbedeutung. Oder die LE ist Teil eines polysemen DUW-Eintrages und der Wörterbucheintrag ist ein Monosem ohne entsprechende polyseme Auffächerung.	Weder orthographische noch grammatische Form sind identisch, zusätzlich ist die Bedeutung nicht mit Sicherheit beim Wörterbucheintrag zu finden	Optant (DUW-ID 1077) bei MUS
	nein	LE nicht als Wörterbucheintrag vorhanden			LE nicht als Wörterbucheintrag vorhanden	Ombudsleute bei MUS

Mit dieser Dreiteilung soll erreicht werden, dass das Fehlen einer Untersuchungseinheit in einem der benutzten Wörterbücher nicht übermäßig bewertet wird: Nur diejenigen Untersuchungseinheiten, die hinsichtlich aller drei Aspekte der lexikalischen Identität (graphische und grammatische Form, Bedeutungsidentität) nicht identisch sind, und solche Einheiten, die im untersuchten Wörterbuch überhaupt nicht auftauchen, sollen als nicht identisch gewertet werden.

Darüberhinaus soll eine Überbetonung graphischer Abweichungen gegenüber grammatischen Abweichungen vermieden werden. Fälle, bei denen Untersuchungseinheit und Wörterbucheintrag graphisch voneinander abweichen, sollen als identisch gelten, solange sich die Bedeutung der Einheit im Wörterbucheintrag identifizieren lässt und die grammatische Form übereinstimmt. Demgegenüber wird in Fällen, bei denen es Abweichungen gibt zwischen der grammatischen Beschreibung bzw. Wortartenzuordnung der Untersuchungseinheit und der des Wörterbucheintrages, die lexikalische Identität als stärker beeinträchtigt beurteilt; solche Paare von Untersuchungseinheiten und Wörterbucheinträgen gelten dementsprechend als nur teilweise identisch. Wichtigster Anwendungsfall sind hier Einheiten, bei denen noch eine Normunsicherheit bezüglich des Genus besteht.²⁸⁹

Die Anwendung dieses Systems ist, wie sich leicht denken lässt, sehr arbeitsintensiv. Alternativ wäre auch die Anwendung einer vereinfachenden Daumenregel denkbar: Allein die Tatsache, dass die Wortform der Untersuchungseinheit bei Wildhagen bzw. Muret-Sanders aufgeführt ist, kann als Zeichen für das Vorhandensein der Einheit im jeweiligen Wörterbuch gewertet werden. Damit gäbe man aber das Konzept der lexikalischen Einheit als Grundeinheit weitgehend auf. Eklatante Fälle wie *Objekt* in 214.2 a-d, die eindeutig bei Muret-Sanders nicht erwähnt werden, kommen dann nicht zum Tragen.

Betrachtet man, wieviele der Einheiten der Untersuchungsmenge bei Wildhagen bzw. Muret-Sanders identifizierbar, teilweise identifizierbar oder nicht identifizierbar sind, dann ergibt sich folgendes Bild (Tabelle 8-2):²⁹⁰

Tabelle 8-2: Das Vorkommen der lexikalischen Einheiten der Untersuchungsmenge in den beiden Wörterbüchern

	Wildhagen 1972		Muret-Sanders 1899	
Anzahl der LE mit Identität = ja	680	30%	801	35%
Anzahl der LE mit Identität = teilweise	382	16%	381	17%
Anzahl der LE mit Identität = nein	1232	54%	1112	48%

Dieses Bild passt zu dem allgemeinen Unterschied zwischen den beiden Wörterbüchern: Muret-Sander 1899 macht deutlich reichhaltigere Angaben zu den Bedeutungen seiner

289. Vgl. etwa *Outback* oder *Overkill*, die im DUW mit neutralem oder maskulinem Genus angegeben sind. Ein ähnlicher Anwendungsfall, bei dem eine Unterscheidung nach dem Genus wichtig ist, sind etwa die deadjektivischen Personenbezeichnungen *Obdachlose* bzw. *Oberkommandierende*, bei denen der DUW das maskuline und das feminine Genus unterscheidet.

290. (Alle Prozentangaben in den folgenden quantitativen Analysen sind gerundet).

Stichwörter und enthält auch mehr Stichwörter als Wildhagen 1972. Die reichhaltigeren Angaben erleichtern die Zuordnung der lexikalischen Einheiten der Untersuchungsmenge zu den jeweiligen Stichwörtern. Daher war es bei Muret-Sanders in mehreren Fällen möglich, eine eindeutige Ja- bzw. Nein-Zuordnung herzustellen als das bei Wildhagen der Fall war. Daraus ließe sich der Einwand ableiten, dass man Wildhagen bei der Bewertung der Stabilität auch hätte weglassen können. Dem läßt sich entgegenhalten, dass die Menge der Einheiten, die bei Wildhagen zu finden ist, nur zu einem Teil (518 von insgesamt 676 Untersuchungseinheiten) auch bei Muret-Sanders zu finden ist und Wildhagen damit bei der Gruppe der identischen Einheiten einen deutlich unterscheidenden Beitrag leistet. Was man allerdings deutlich sieht, ist, dass der Vergleich der aus dem DUW entnommenen lexikalischen Einheiten mit dem Wortbestand bei Muret-Sanders und Wildhagen keiner chronologischen Entwicklung folgt, denn dann müssten die Übereinstimmungen zwischen Wildhagen – der ja das jüngere und damit dem DUW zeitlich näher stehende Wörterbuch darstellt – und den DUW-Einheiten größer sein als die Überschneidungen zwischen Muret-Sanders und DUW. Man kann also nicht von einer generellen zeitlichen Stabilität sprechen, sondern eher davon, wie sehr und wie lange eine lexikalische Einheit lexikographisch konventionalisiert ist.

Die hohen Prozentzahlen für die Anzahl der Untersuchungseinheiten, die nicht bei Wildhagen bzw. Muret-Sanders identifiziert werden konnten, dürfen allerdings nicht missverstanden werden: Die jeweils betroffenen Mengen sind nicht identisch, d.h. eine Einheit, die nicht bei Muret-Sanders identifiziert werden konnte, kann durchaus bei Wildhagen identifizierbar sein und umgekehrt. Die Anzahl der Einheiten, die weder bei Wildhagen noch bei Muret-Sanders identifizierbar waren, liegt bei 42 %, ist allerdings damit immer noch relativ hoch. Wenn man sich diese Gruppe genauer ansieht, dann wird deutlich, dass sich darunter zwar viele durchsichtige Wortbildungen wie Komposita und Movierungen (*Oberflächenstruktur, Orgelmusik; Oberbürgermeisterin, Oppositionsführer*) befinden, deren Aufnahme in ein Wörterbuch nicht immer üblich ist. Andererseits sind darunter auch viele Einheiten, die sich auf tendenziell 'junge' Sachverhalte beziehen (*Organspende, Ökobilanz, Ölpest, online, Organigramm*), so dass durchaus der zeitliche Unterschied in der Entstehung des DUW gegenüber der Entstehung von Wildhagen und Muret-Sanders am Bestand der Untersuchungseinheiten erkennbar ist.

Tabelle A-7 in Appendix A listet einige Beispiel-Beschreibungen auf; auf der der gedruckten Ausgabe beiliegenden Datei *Beschreibung_Stabilität.xls* finden sich alle lexikalischen Einheiten der Untersuchungsmenge mitsamt ihren Stabilitätsangaben.

Das so entstandene System von nicht, teilweise oder schon bei Muret-Sanders oder Wildhagen nachweisbaren Stichwörtern kann nun unterschiedlich gewichtet werden. In Abschnitt 8.8.6 wird ein solches Bewertungssystem entworfen.

8.7 Die Beschreibung der Neutralität

Bei der Beschreibung der stilistischen Neutralität habe ich vereinfachend auf die Markierungen zurückgegriffen, die bereits den jeweiligen DUW-Stichwörtern zugeordnet waren.²⁹¹ Unter Neutralität verstehe ich dabei eine graduelle Eigenschaft, die von Unmarkiertheit, d.h. vom Fehlen jeglicher Markierung hinsichtlich einer Sondersprache, über einen gewissen, noch zu definierenden Grad an Markiertheit, der durch bestimmte Typen von Sondermarkierungen hervorgerufen wird, bis hin zu einer deutlichen Markiertheit hinsichtlich bestimmter Typen von Sondermarkierungen reicht.

Konkret habe ich die Sondermarkierungen, die zu den jeweiligen DUW-Stichwörtern gegeben sind, auf die aus dem jeweiligen DUW-Stichwort abgeleiteten lexikalischen Einheiten übertragen.²⁹² Dabei habe ich mehrgliedrige Markierungen, die deutlich gleichwertige Markierungen nebeneinander stellen, wie zum Beispiel „Med., Biol.“ oder „schweiz., sonst veraltet“, als getrennte Sondermarkierungen erfasst, so dass die entsprechende Einheit nicht nur eine Sondermarkierung (wie etwa *Ohrenbeichte*, das mit „kath. Kirche“ markiert ist), sondern zwei (vgl. *Offside*, das mit „Ballspiele“ und „schweiz.“ markiert ist) oder sogar drei Sondermarkierungen (vgl. *Onstimme*, das mit „Film“, „Ferns.“ und „Theater“ markiert ist) haben kann. Diese genaue Erfassung und Auffächerung der Markierungen (vgl. Abschnitt 8.8.7) war nötig, damit ich beim späteren Schritt der Gewichtung der Neutralität der lexikalischen Einheiten der Untersuchungsmenge entsprechend differenzierend vorgehen kann. Darüberhinaus habe ich einige wenige Markierungen zusammengelegt, um das weitere Verfahren zu vereinfachen.

Die der gedruckten Ausgabe beiliegende Datei *Beschreibung_Neutralität.xls* enthält alle Einheiten, die eine Sondermarkierung haben. Tabelle A-8 in Appendix A zeigt einige Beispielbeschreibungen.

291. Das bedeutet nicht, dass die Zuordnungen, die die DUW-Redaktion vorgenommen hat, völlig unanfechtbar seien; gerade der Überblick über das Thema Sondersprachen in Abschnitt 5.3.1 zeigt, wie schwierig und arbiträr eine solche Zuordnung fast zwangsläufig ist. Doch im Sinne einer Konzentration auf die wesentlichen Fragen dieser Arbeit gehe ich, wie bereits in Kapitel 7 mehrfach angesprochen, vom DUW als einem 'guten' Korpus und damit von der Richtigkeit der darin gemachten Angaben aus. – Für eine kritische Beurteilung von Fächerzuordnungen im allgemeinsprachlichen Wörterbuch vgl. ansonsten auch Bergenholtz 1994.

292. Inkonsistenzen in der Markierung, wie sie zum Beispiel bei den femininen Varianten zu maskulinen Substantiven vorkommen (wie zum Beispiel *Oberbundesanwalt* (DUW-ID 80), das mit „Bundesrepublik Deutschland“ markiert ist, gegenüber *Oberbundesanwältin* (DUW-ID 81), das keine Markierung trägt), habe ich, soweit es keinen ersichtlichen Grund für den Unterschied in der Markiertheit gab, entsprechend ergänzt. Darüberhinaus habe ich Inkonsistenzen in der Schreibweise, wie zum Beispiel „geh.“ vs. „gehoben“ vereinheitlicht, um so eine konsistente Datenauswertung zu ermöglichen. Weitere Vereinheitlichungen im Sinne der Konsistenz: „Bundesrepublik Deutschland“ -> „BRD“, „chem. Fachspr.“ -> Chemie; „Fachspr. Anat.“ -> „Anat.“; „Fachspr., bes. Math., Sprachw.“ -> „Math.“, „Sprachw.“; „Papierdt.“ -> „Amtsspr.“; „Bergbau“ -> „Bergmannsspr.“; „bayr.“ -> „südd.“; „bes. Fachspr.“ -> „Fachspr.“; „meist Fachspr.“ -> „Fachspr.“; „meist scherzh.“ -> „scherzh.“; „seltener“ -> „selten“; „emotional verstärkend“ -> „verstärkend“; „leicht abwertend“ -> „abwertend“.

Insgesamt enthält der hier untersuchte Wortschatzausschnitt 101 verschiedene Markierungen. Sie sind in Tabelle A-9 in Appendix A zusammen mit der Anzahl der Untersuchungseinheiten, die sie markieren, aufgelistet. Von den 2294 hier untersuchten lexikalischen Einheiten sind 1479, also eine deutliche Mehrheit, unmarkiert, während 815 Einheiten in irgendeiner Weise markiert sind. Dabei haben die meisten (d.h. 688 Einheiten) der markierten Untersuchungseinheiten nur eine Sondermarkierung; eine kleinere Menge (d.h. 115 Einheiten) hat zwei Markierungen, und ein Bruchteil (12 Einheiten) hat drei Markierungen.

8.8 Die Gewichtung der Eigenschaften

8.8.1 Vorüberlegungen

Die hier vorgenommene Gewichtung und Bewertung der sechs betrachteten Eigenschaften einer lexikalischen Einheit ist eine exemplarische Gewichtung, mit deren Hilfe spätere Vergleiche möglich sein sollen. Zugleich soll die Diskussion der Bewertung auch zeigen, dass andere Gewichtungen durchaus möglich und denkbar sind, denn gerade das ist ja der Sinn dieser strikten Trennung von Beschreibung und Gewichtung: eine größere Transparenz des Verfahrens und vor allem mehr Flexibilität und Kontrolle.

Um das, was mit dieser Arbeit zu zeigen ist, möglichst einfach zu halten, habe ich versucht, mich bei der Gewichtung der Eigenschaften an eine Dreiteilung zu halten, d.h., ich habe versucht, jede Eigenschaft so zu quantifizieren und zu bewerten, dass die Untersuchungseinheiten anhand dieses Quantifizierungssystems in jeweils drei Gruppen eingeteilt werden können: Die erste Gruppe umfasst diejenigen Einheiten, die die jeweilige Eigenschaft am stärksten aufweisen, die zweite Gruppe umfasst diejenigen Einheiten, die die jeweilige Eigenschaft in mittlerem Maße aufweisen, und die dritte Gruppe umfasst diejenigen Einheiten der Untersuchungsmenge, die die jeweilige Eigenschaft in geringstem Maße aufweisen. Dabei gehe ich davon aus, dass alle Eigenschaften Indikatoren für die Zentralität eines Wortes sind und dass damit diejenigen lexikalischen Einheiten, die alle Eigenschaften in stärkstem Maße ausweisen, zu den zentralen Einheiten der untersuchten Menge gehören.

8.8.2 Die Bewertung der Repräsentiertheit

In einem ersten Schritt der Bewertung habe ich nun die Beschreibungen der Repräsentiertheit in ein dreigeteiltes Punktesystem umgesetzt. Dabei habe ich denjenigen lexikalischen Einheiten, die voll repräsentiert sind, einen Repräsentiertheitswert 1 zugeordnet und denjenigen Einheiten, die eindeutig nicht repräsentiert waren, einen Repräsentiertheitswert von 3. Diejenigen Einheiten, die eine unklare Repräsentiertheit aufweisen, habe ich wie folgt bewertet:

- Komposita und Ableitungen, bei denen Teile der Konstruktion bekannt sind, andere nicht (*Obstruktionspolitik* (DUW-ID 310), *OAU-Staaten* (DUW-ID 17)). Hier gehe ich von einer strengeren Beurteilung aus als diejenige, die Bonan-Garrigues vornimmt. Ich

ordne diesen Fällen einen Repräsentationsgrad von 2 zu: Wenn es nicht möglich ist, die gesamte Konstruktion mit einer Bedeutungsumschreibung zu verknüpfen, die nicht nur eine direkte Umformung der Wortbildungskonstruktion in ein Prädikat darstellt ('eine Politik der Obstruktion'), sondern alle Bedeutungsbestandteile umschreibt ('eine Politik, bei der man durch eine Flut von Anträgen das Parlament lahmzulegen versucht'), dann ist das entsprechende Wort mit einem Repräsentiertheitsgrad von 2 (= teilweise repräsentiert) belegt.²⁹³

- Noch strenger verfare ich bei Ableitungen mit gebundenen Lexemen (*Ophthalmologie*, *Ostrazismus* oder *obstruktiv*): In diesen Fällen habe ich dem jeweiligen Wort den geringsten Repräsentiertheitsgrad (3) zugeordnet. Zwar ist die Zuordnung der Wortart bereits mehr als eine völliges Nichtrepräsentiertsein, aber der semantische Mehrwert, den man durch das Erkennen der Wortart gewinnt, ist geringer als der, den man beispielsweise durch das Verstehen des Grundgliedes eines Kompositums (s. o.) hat.
- Wörter anderer Sprachen, bei denen zwar die sprachliche Herkunft erkennbar ist, die Bedeutung sich aber nicht näher benennen lässt (*Obiter Dictum* oder *Open Shop*): Hier ließe sich argumentieren, dass allein die Tatsache, dass sich die Wörter als lateinisch bzw. englisch erkennen lassen, ein Zeichen für ihre Repräsentiertheit im internen Lexikon ist, da sie durchaus einen Wiedererkennungseffekt hervorrufen. Im Vergleich zu anderen Stichwörtern (Beispiel: *Office*), bei denen sich eine unmittelbare Bedeutung angeben lässt, ist der Wiedererkennungseffekt allerdings gering. Diese Gruppe von Wörtern, also Wörter, bei denen ich die Herkunft aus einer anderen Sprache erkennen kann, bei denen es mir aber nicht möglich war, eine auch nur ungefähr zutreffende Bedeutung anzugeben, wurden mit einer Repräsentiertheit von 3 (= nicht oder nur geringfügig repräsentiert) belegt.
- Symbole und Einzelbuchstaben (*ö*, *Ö* oder *Ω*): Ihnen habe ich einen Repräsentiertheitsgrad von 3 (= nicht oder nur geringfügig repräsentiert) zugeordnet.
- Abkürzungen (*OAU*, *ÖAMTC*, *ÖBB*, *OSB*): Abkürzungen, bei denen es mir möglich war, sowohl eine Vollform anzugeben (auch wenn die Vollform, mit der man die Abkürzung verknüpft, nicht identisch ist mit der im DUW angegebenen) als auch eine Bedeutung, (vgl. *ÖBB*), habe ich mit dem Repräsentiertheitsgrad 1 belegt. Abkürzungen, denen ich keine Vollform, wohl aber eine Bedeutung zuordnen konnte (vgl. *ÖAMTC*), habe ich mit dem Repräsentiertheitsgrad 2 belegt. Abkürzungen, denen ich weder eine

293. Die hier beschriebene Vorgehensweise beim Umgang mit Wortbildungskonstruktionen kann allerdings auch zu vermeintlichen Inkonsequenzen führen, wenn die einzelnen Bestandteile einer solchen Konstruktion auch separate Stichwörter des DUW sind und möglicherweise einen anderen Repräsentationswert haben als die Gesamtkonstruktion. Hierbei handelt es sich aber eben nur vermeintlich um eine Inkonsistenz, denn die Angaben zur Repräsentiertheit eines Stichwortes sind grundsätzlich unabhängig von der Repräsentiertheit anderer Stichwörter. Ist eine Wortbildungskonstruktion nach den oben genannten Regeln als nicht repräsentiert zu betrachten, so muss sie auch so eingestuft werden, unabhängig davon, ob ihre Bestandteile einzeln betrachtet repräsentiert sind oder nicht.

Vollform noch eine Bedeutung zuordnen konnte (vgl. *OAU*, *OSB*), habe ich mit dem Repräsentiertheitsgrad 3 belegt.

Betrachtet man nun die Gesamtheit der nach ihrer Repräsentiertheit zu beschreibenden DUW-Stichwörter ergibt sich folgendes quantitative Bild (vgl. Tabelle 8-3):

Tabelle 8-3: Die Schichtung der lexikalischen Einheiten der Untersuchungsmenge in Bezug auf die Repräsentiertheit

Repräsentiertheitsgrad	Anzahl LE	
1 (voll repräsentiert)	2060	90%
2 (teilweise repräsentiert)	14	1%
3 (nicht oder gering repräsentiert)	220	9%

Damit ergibt sich – wie schon zu Beginn der Diskussion zu erwarten – ein anderes Bild als bei Bonan-Garrigues: Zwar gibt sie keine genauen Zahlen zur Verteilung der von ihr untersuchten Wörter über die beiden Repräsentiertheitsgrade von 0 (nicht repräsentiert) und 1 (repräsentiert) an, aber aus ihrer zusammenfassenden Darstellung der Verteilung der Wörter über die durch Repräsentiertheit und Plausibilität bestimmten Schichten C1, C2 und C3 (vgl. Bonan-Garrigues 1993, S. 57) ergibt sich, dass bei ihrer Einteilung die große Mehrheit der Wörter (52%) der Schicht C3 angehören und damit zugleich auch nicht repräsentiert sind²⁹⁴. Wie schon weiter oben erwähnt, ist diese Diskrepanz dadurch zu erklären, dass Bonan-Garrigues 'Rohmaterial' bearbeitet hat, d.h. Wörter, die aus einem Textkorpus extrahiert und nur geringfügig lexikographisch bearbeitet (etwa auf ihre Grundform reduziert) wurden; dem gegenüber stellt das Material des DUW Wortschatzelemente bereit, die bereits intensiv lexikographisch bearbeitet wurden und den etablierten Teil des Wortschatzes darstellen sollen, so dass etwa okkasionelle Bildungen, die eher einen geringeren Repräsentiertheitswert aufweisen als lexikalisierte Wörter, gar nicht erst im Bestand auftauchen.

8.8.3 Die Bewertung des Grades der Worthaftigkeit

Bei der Erstellung von lexikalischen Einheiten (vgl. Abschnitt 8.3) haben wir genau notiert, wie sich die einzelnen Einheiten syntaktisch verhalten, ob sie morphologische, phonetische oder syntaktische Besonderheiten aufweisen. Dabei haben wir festgestellt, dass der hier untersuchte Lexikonausschnitt unterschiedlich 'typische' Vertreter von lexikalischen Einheiten enthält. Wir wollen diese Unterschiede zu erfassen versuchen und sammeln diese Unterschiedlichkeit unter dem Begriff der Worthaftigkeit.

Ziel ist es, typische, d.h. formal, semantisch und von ihrer Lexikalisiertheit her typische Lexeme von weniger typischen Lexemen zu unterscheiden. Als Grundlage der Beurteilung dient uns dabei die Diskussion der Worthaftigkeit bei Schindler 2002, der den Begriff Wort definiert als...

294. Alle Wörter, die einen Repräsentiertheitswert von 0 haben, ordnet Bonan-Garrigues automatisch der Schicht C3 zu, vgl. Bonan-Garrigues 1993, S. 45.

„...die konventionalisierte Verbindung einer phonologisch organisierten, vom menschlichen Stimmapparat hervorgebrachten Lautfolge mit einer begrifflichen Bedeutung. Diese Bestimmung wird inzwischen um die kombinatorischen Eigenschaften erweitert [...], sodass das Wort phonologische, syntaktische bzw. allgemein kombinatorische Eigenschaften [...] und konzeptuelle Informationen aufeinander bezieht.“ (Schindler 2002, S. 36)

In Anlehnung an Jackendoff 1995 beschreibt Schindler ein typisches Wort als eine Einheit, die „phonologische (phonological structure, kurz: PS), syntaktische (syntactic structure, SS) und semantische Informationen (als conceptual structure bzw. CS) zu einem Triplet <PS, SS, CS>“ (ebda., S. 36) vereint. Als „[b]esonders typisch“ gelte dabei „das autosemantische Wort (N, V, Adj, Adv), dem im Unterschied zur grammatischen Bedeutung sog. 'Funktionswörter' (Det, Konj, Präp) eine sog. 'lexikalisch-begriffliche Bedeutung' zugesprochen wird.“ (Schindler 2002, S. 36)

Die Mehrzahl der Einheiten unseres Untersuchungsmaterials stellt ‘unproblematische’, d.h. typische Wörter dar im Sinne der von Schindler referierten autosemantischen Wörter. Ein kleinerer Teil besteht jedoch aus Einheiten, die man intuitiv nicht unmittelbar als ‘typische’ Wörter betrachten möchte. Um diese Gruppen von Einheiten soll es im Folgenden gehen.

Mehrgliedrige lexikalische Einheiten und Phraseologismen

Das DUW führt neben Lexemen mit einer geschlossenen Wortform auch lexikalische Einheiten mit einer mehrgliedrigen Form auf. Dazu gehören getrennt geschriebene Verben und Adjektive wie *offen legen* (DUW-ID 447) oder *oben stehend* (DUW-ID 59), mehrgliedrige Substantive wie *Open Air* (DUW-ID 949) oder *Opera buffa* (DUW-ID 959), aber auch feste Wendungen wie *oh, la la* (DUW-ID 540), *ora et labora* (DUW-ID 1109) oder *out of area* (DUW-ID 1625). Nun haben, intuitiv betrachtet, getrennt geschriebene Verben und Adjektive zwar einen deutlicheren Status als Wort als etwa feste Wendungen, weil sie beispielsweise konjugierbar bzw. deklinierbar sind. Andererseits ist eine eindeutige Abgrenzung und Unterteilung dieser heterogenen Gruppe von Einheiten des Wortschatzes eine komplexe Fragestellung.²⁹⁵ Wir wollen daher vereinfachend alle Einheiten, die eine mehrgliedrige Form aufweisen (und nicht zu einem der anderen Typen von Sonderfällen gehören), zu einer Gruppe zusammenfassen und ihnen einen Grad der Worthaftigkeit von 2 zuordnen.

Insgesamt handelt es sich bei den mehrgliedrigen Untersuchungseinheiten um 46 Einheiten. Die der gedruckten Ausgabe beigelegte Datei *Mehrgliedrige_Einheiten.xls* führt sie komplett auf.

Atypische Wortarten

In Anlehnung an Schindler, der die Wortarten Nomen, Verb, Adjektiv und Adverb als die ‘typischen’ Wortarten betrachtet, müssten eigentlich die lexikalischen Einheiten, die anderen Wortarten angehören, also Interjektionen, Konjunktionen und Partikel, als nicht den typischen lexikalischen Einheiten zugehörig betrachten. Weil sie aber durchaus phonetische,

295. Vgl. etwa die Diskussion bei Schindler 2002, S. 38f.

syntaktische und semantische oder zumindest pragmatische Eigenschaften aufweisen und weil wir das Ergebnis einer Beispielselektion auch auf ihre Zusammensetzung hinsichtlich der Wortarten betrachten wollen (vgl. Abschnitt 9.3.4), habe ich ihnen einen Grad der Worthaftigkeit von 1 zugeordnet.²⁹⁶

Insgesamt handelt es sich bei dieser Gruppe um 35 Einheiten der Untersuchungsmenge. Sie sind in der der gedruckten Ausgabe beigelegten Datei *Atypische_Wortarten.xls* vollständig aufgeführt.

Reihenbildende Wortbildungen mit Eigennamen

Im Rahmen der Bestimmung des Grades der Worthaftigkeit der Einträge wollen wir auch den Grad der Lexikalisiertheit gegenüber dem Grad der Okkasionalität als Beschreibung mit aufnehmen. Dies hat seinen Grund darin, dass bei der Erstellung eines zentralen Wortschatzes die quantitative Begrenzung bzw. Abgrenzung eine Rolle spielt, während zugleich Fragen der Lückenhaftigkeit bzw. Komplettheit (*linguistic closure*) von reihenbildenden Komposita etwa oft diskutiert werden (vgl. Abschnitt 4.4.5).

Dieses Problem stellte sich, als ich die Wortschatzmenge auf das Vorkommen von Eigennamen hin untersuchte und dabei entdeckte, dass sie einige Wortbildungen mit Eigennamen enthält, wie zum Beispiel *Oberitalien* oder *Oberammergau*. Dass es zu diesen Wörtern komplementäre Zusammensetzungen gibt (*Unteritalien*, *Unterammergau*), verleiht diesen Wörtern den Status eines reihenbildenden Kompositums, bei denen sich die Frage stellt, ob nicht andere Mitglieder der Reihe in den Wortschatz aufzunehmen wären. Warum etwa sollte man nur *Oberpfalz* aufführen, nicht aber *Oberbayern*?

Wie wir jedoch in den Diskussionen zur Behandlung von Wortfeldern in Abschnitt 4.4.5 gesehen haben, ist Vollständigkeit bei solchen Reihenbildungen ohnehin nicht zu erreichen. Zugleich ist die quantitative Begrenzung eines zentralen Wortschatzes ein wichtiger Aspekt. Daher habe ich entsprechend der in den Abschnitten 4.2 und 5.2.2 diskutierten Abgrenzungen von Wortschatz und Grammatik Komposita und andere Wortbildungen, die in reihenhafter Weise mit Eigennamen Verbindungen eingehen, mit einem Grad der Worthaftigkeit von 2 belegt.

Insgesamt handelt es sich dabei um 35 Einheiten. Sie sind in der der gedruckten Ausgabe beigelegten Datei *Reihenbildende_Wortbildungen_mit_Eigennamen.xls* vollständig aufgeführt.

296. In einem ersten Versuch der Bildung und Beschreibung einer Referenzselektion (vgl. Abschnitt 9.3) wurden Interjektionen, Konjunktionen und Partikel noch der Worthaftigkeitsgrad 2 zugeordnet, was dazu führte, dass Einheiten, die diesen Wortarten angehörten, gar nicht in die Referenzselektion gelangten, so dass auch keine Analyse der Referenzselektion möglich war. Durch die deutliche Trennung von Beschreibung und Bewertung aber war es möglich, rückwirkend die Beurteilung von Einheiten dieser Wortarten ganz gezielt hinsichtlich der Worthaftigkeit zu ändern, ohne dabei die Beurteilung aller anderen Untersuchungseinheiten revidieren zu müssen.

Abkürzungen und Kurzwörter

Zwischen den beiden Polen ‘typisches Wort’ und ‘atypisches Wort’ bewegen sich Abkürzungen und Kurzwörter. Glück ((Hg.) 2000, S. 392) klassifiziert diese Wortbildungsphänomene folgendermaßen:

„A[bkürzungen] sind Kurzformen von Ausdrücken, die mit Bezug auf die geschriebene Sprachform im Gegensatz zu Kurzwörtern mit Bezug auf die gesprochene Sprachform gebildet werden.“ (Glück (Hg.) 2000, S. 2).

Dabei unterscheidet Glück verschiedene Grade der Abkürzung: Die grundlegende Form der Abkürzung „ist die (durch einen Punkt abgeschlossene) Verkürzung eines Ausdrucks [...] oder eine Auswahl charakterist[ischer] Buchstaben, die i.d.R. das Konsonantengerüst oder Teile davon repräsentieren“. Darüberhinaus gebe es die Form der...

„...Akronyme, meist für nominale Ausdrücke, in denen die Anfangsbuchstaben der Bestandteile eines Syntagmas oder eines Kompositums für den ganzen Ausdruck stehen. Geschrieben werden sie teilweise in Versalien, dann in der Regel ohne Punkt [...], teilweise gemischt entsprechend der Schreibweise des abgekürzten Ausdrucks [...], teilweise (vorzugsweise bei Syntagmen) mit Punkt hinter jedem Bestandteil“ (Glück (Hg.) 2000, S. 2).

Nachdem aber Abkürzungen zunächst einmal „völlig synonym mit dem Ausdruck [sind], aus dem sie gebildet“ sind, ist die Frage naheliegend, wie sehr Abkürzungen überhaupt lexikalische Einheiten darstellen. Dabei scheint es mir richtig, die Abkürzung und ihren ursprünglichen Ausdruck als zwei Formvarianten ein und derselben Bedeutung zu betrachten. Zwei unterschiedliche Formen führen aber dazu, dass auch zwei unterschiedliche lexikalische Einheiten angesetzt werden müssen, und hier setzt nun die Frage an, wie sehr eine Abkürzung als typische lexikalische Einheit betrachtet werden kann. Ein Ansatz könnte sein, das Verhalten der Abkürzung daraufhin zu untersuchen, ob es dem eines Wortes ähnelt, beispielsweise, ob die Abkürzung mit einem Artikel verwendet oder dekliniert wird. Ein anderer Aspekt kann sein, zu fragen, wie sich die Abkürzung phonetisch verhält.

Ich schlage daher vor, Abkürzungen anhand der folgenden Regeln zu behandeln:

- Regel 1: Grundsätzlich ist jede Abkürzung als eine synonymische Formvariante des ausgeschriebenen Lexems oder Syntagmas zu betrachten und daher als eigenständige Einheit zu behandeln. Dabei stellt sie jedoch aufgrund des Entstehungsweges – die Vollform existiert vor der Kurzform – im Vergleich zu ihrer Vollform grundsätzlich eine weniger typische lexikalische Einheit dar.
- Regel 2: Innerhalb der Gruppe der Abkürzungen wiederum sind diejenigen Abkürzungen eher typische lexikalische Einheiten, deren Form phonetisch gesehen Wortstatus hat, d.h. dass die Form phonetisch zusammenhängend ausgesprochen werden kann und auch phonetisch gebraucht wird (vgl. etwa *OPEC* oder *ÖBB*). Wird sie das nicht, wie das etwa bei Abkürzungen, die mit einem Punkt versehen sind (*obd.*, *od.* oder *Ob.-Insp.*) oder nur in der geschriebenen Sprache existieren (*O*), der Fall ist, dann gehören diese Abkürzungen in die Gruppe der lexikalischen Einheiten, die am wenigsten ‘wort-

typisch' sind. Darüberhinaus sollen Abkürzungen mit einem Wortstatus 2. Grades auch mit einem Artikel kombinierbar sein, soweit sie Substantive darstellen.

Ausgehend von diesen Regeln habe ich alle im Ausgangswortschatz enthaltenen Abkürzungen – insgesamt 54 Einheiten – entsprechend ihres phonetischen und syntaktischen Verhaltens einer der beiden Gruppen weniger typischer lexikalischer Einheiten zugeteilt und mit entsprechenden Kommentaren versehen. Dabei haben 25 Einheiten in Form von Abkürzungen einen Wortstatus 2. Grades, d.h. sie sind phonetisch repräsentiert und deklinierbar bzw. konjugierbar, und 29 Untersuchungseinheiten haben einen Wortstatus 3. Grades.

Die der gedruckten Ausgabe beigelegte Datei *Abkürzungen.xls* enthält alle diese 54 Einheiten der Untersuchungsmenge.

Symbole und Einzelbuchstaben

Der DUW verzeichnet auch Symbole und Einzelbuchstaben wie etwa ω , Ω . Da diesen syntaktische Eigenschaften fehlen, habe ich sie mit einem Worthaftigkeitsgrad von 3 bewertet.

Insgesamt handelt es sich bei dieser Gruppe um fünf Einheiten. Sie sind in der der gedruckten Ausgabe beigelegten Datei *Symbole_und_Einzelbuchstaben.xls* aufgeführt.

Gebundene Lexeme

Gebundene Lexeme stellen einen sehr wortnahen Typus untypischer lexikalischer Einheiten dar. Sie haben sowohl in phonetischer als auch in semantischer und morphologischer Hinsicht klare Strukturen vorzuweisen. Was sie von typischen lexikalischen Einheiten unterscheidet, ist, dass sie keine eigenständigen Einheiten im Satz darstellen können. Ich habe daher den gebundenen Lexemen, die der DUW auflistet (*ober-*, *öko-*, *-oid*) den Worthaftigkeitsgrad 3 zugeordnet.

Insgesamt handelt es sich bei diesem Typus um 27 Einheiten. Sie sind in der der gedruckten Ausgabe beigelegten Datei *Gebundene_Lexeme.xls* aufgeführt.

Eigennamen

Der Status von Eigennamen als lexikalischen Einheiten ist schwierig zu bestimmen, weil es auch hier ein Kontinuum an Eigenschaften gibt. Ich habe bei der hier vorliegenden praktischen Untersuchung alle Einheiten, die Eigennamen von Personen (*Osiris*), Orten (*Osnabrück*) oder Markennamen (*Orlon*) darstellen, einen Grad der Worthaftigkeit von 3 zugeordnet; dies aus dem einfachen Grund, dass die Aufnahme von einzelnen Eigennamen unmittelbar die Frage danach aufwirft, warum nicht auch andere Eigennamen in das Lexikon aufgenommen sind.

Insgesamt sind 35 Eigennamen in der untersuchten Wortschatzmenge enthalten.

Die der gedruckten Ausgabe beigelegte Datei *Eigennamen.xls* enthält die gesamte Liste der Eigennamen.

Betrachtet man nun, wie sich die Gesamtheit der Einheiten der Ausgangsmenge über die verschiedenen Grade der Worthaftigkeit verteilen, dann ergibt sich ein ähnliches Bild wie bei der Verteilung der Repräsentiertheit (vgl. Tabelle 8-4):

Tabelle 8-4: Die Schichtung der lexikalischen Einheiten der Untersuchungsmenge in Bezug auf den Grad der Worthaftigkeit

Grad der Worthaftigkeit	Anzahl LE	
1 (vollwertiges Wort)	2092	91%
2 (wortnah)	106	5%
3 (geringer Grad an Worthaftigkeit)	96	4%

Eine große Mehrzahl an Einheiten weist das ‘typische’ Verhalten eines Wortes auf, wie wir es zu Beginn dieses Abschnittes definiert haben, eine Minderheit entspricht nur in geringem Maße den Kriterien für ein vollwertiges Wort. Auch diese Verteilungsverhältnisse sind wohl mit der lexikographischen Vorab-Bearbeitung des Materials zu erklären.

8.8.4 Die Bewertung der Plausibilität

Im Gegensatz zu den bisher betrachteten Eigenschaften bin ich bei der Plausibilität bereits beim Schritt der Beschreibung von einer Dreiteilung (sehr plausibel – relativ plausibel – wenig plausibel) ausgegangen (vgl. Abschnitt 8.4), weil Plausibilität eine Eigenschaft ist, die nicht auf greifbaren Kriterien, die in der lexikalischen Einheit selbst zu finden sind, beruht und eine differenziertere Einteilung daher schwierig war. Diese Dreiteilung will ich bei der Bewertung beibehalten, indem ich Einheiten, die ich als sehr plausibel beschrieben habe, mit dem Wert 1 belege, Einheiten, die als relativ plausibel beschrieben sind, mit dem Wert 2 und Einheiten, die als wenig plausibel beschrieben sind, mit dem Wert 3.

Damit ist die quantitative Verteilung der Einheiten der Untersuchungsmenge entsprechend ihres Plausibilitätswertes die gleiche wie noch bei der Beschreibung (vgl. Tabelle 8-5):

Tabelle 8-5: Die Schichtung der lexikalischen Einheiten der Untersuchungsmenge in Bezug auf die Plausibilität

Grad der Plausibilität	Anzahl LE	
1 (sehr plausibel)	130	6%
2 (relativ plausibel)	803	35%
3 (wenig plausibel)	1361	59%

Hier kehrt sich zum ersten Mal das Bild um: Diejenigen Einheiten, deren Grad an Zentralität hinsichtlich der untersuchten Eigenschaft am stärksten ist, sind in der Minderzahl.

8.8.5 Die Bewertung der Frequenz

Sieht man sich an, welche der Frequenzklassen von 1 bis 25 der Frequenzerhebung des Wortschatzprojektes vertreten sind und wieviele Einheiten sie jeweils auf sich vereinen (vgl. Tabelle 8-6), dann wird die „degressive Struktur des Wortgebrauchs“ (Augst 1987, S.

115) des Wortschatzes ganz deutlich: Wenige Wörter haben eine hohe Frequenz, viele Wörter haben eine sehr geringe Frequenz.

Tabelle 8-6: Die Verteilung der lexikalischen Einheiten über die verschiedenen Frequenzklassen hinweg

Frequenzklasse lt. Wortschatz-Projekt	Anzahl der LE je Frequenzklasse	
4	7	0.30%
6	12	0.50%
7	6	0.30%
8	40	1.70%
9	43	1.90%
10	62	2.70%
11	87	3.80%
12	95	4.10%
13	119	5.20%
14	124	5.40%
15	166	7.20%
16	197	8.60%
17	181	7.90%
18	187	8.10%
19	194	8.50%
20	152	6.60%
21	135	5.90%
22	84	3.70%
23	90	3.90%
24	163	7.10%
25	151	6.60%

Wie wir bei der Beschreibung der Frequenz gesehen haben (vgl. Abschnitt 8.5), sind die Frequenzangaben, die sich aus dem Leipziger Wortschatz-Projekt gewinnen lassen, nur bedingt verwendbar, weil die Orientierung an der laufenden Wortform und die mangelnde Unterscheidung von Wortarten zu künstlich hohe Frequenzen führen können, so dass man hohe Frequenzangaben nicht direkt als Indikator für eine echte hohe Frequenz der jeweiligen lexikalischen Einheit interpretieren kann. Die Daten lassen sich aber durchaus nutzen, um sehr niedrigfrequente Einheiten entsprechend niedrig zu bewerten. Das bedeutet, dass ich diejenigen Einheiten der Untersuchungsmenge als niedrigfrequent bezeichne, die zu den Frequenzklassen 25 (die Einheit ist nicht in der Datenbank des Wortschatz-Projektes verzeichnet und ergibt daher keinen Treffer), 24 (die Einheit ist in der Datenbank des Wortschatz-Projektes verzeichnet, aber mit einer Frequenz von 0 angegeben) und zu den Klassen 23, 22, 21 und 20, in denen nur Wörter mit einer absoluten Frequenz von weniger als 6 Vorkommen (vgl. Quasthoff 1998, S. 6), gehören. Alle anderen Einheiten werden ganz allgemein als nicht niedrigfrequent bezeichnet. Damit erhalten wir folgende quantitative Verteilung (vgl. Tabelle 8-7):

Tabelle 8-7: Anzahl der niedrigfrequenten gegenüber den nicht niedrigfrequenten lexikalischen Einheiten der Untersuchungsmenge

Frequenzklassen	Anzahl zugehöriger LE	
niedrigfrequente Klassen (25, 24, 23, 22, 21, 20)	774	34%
nicht niedrigfrequente Klassen (alle anderen Klassen)	1520	66%

8.8.6 Die Bewertung der Stabilität

Die Beschreibung der Stabilität (vgl. Abschnitt 8.6) hat dazu geführt, dass wir jede Einheit der Untersuchungsmenge dahin gehend einstufen können, ob und inwieweit sie bei Wildhagen 1972 bzw. Muret-Sanders 1899 als Wörterbucheintrag identifizierbar ist (vgl. auch Tabelle 8-1). In der hier vorzunehmenden Bewertung geht es nun darum, die Angaben zum Vorhandensein in den beiden verglichenen Wörterbüchern miteinander zu einem Stabilitätswert zu kombinieren. Dabei habe ich folgendes Regelset verwendet (vgl. Tabelle 8-8):

Tabelle 8-8: Die Quantifizierung der Stabilitätseigenschaften

Bei WH identifizierbar?	Bei MUS identifizierbar?	Stabilitätsgrad
ja	ja	1
ja	teilweise	1
teilweise	ja	1
teilweise	teilweise	2
ja	nein	2
nein	ja	2
teilweise	nein	2
nein	teilweise	2
nein	nein	3

Die Zuordnung zu drei verschiedenen Stabilitätsgraden beruht dabei auf dem Bemühen, dem Fehlen einer Untersuchungseinheit in einem der beiden herangezogenen Wörterbücher nicht zuviel Gewicht beizumessen, indem ich nur die Fälle, in denen die fragliche Einheit weder in dem einen noch in dem anderen Wörterbuch vorkommt, mit dem geringsten Stabilitätsgrad belege. Denn wie in Abschnitt 8.6 bereits diskutiert, bedeutet die Tatsache, dass ein Stichwort nicht in einem Wörterbuch zu finden ist, nicht generell, dass es dieses Wort zur Zeit der Entstehung des Wörterbuchs noch nicht gegeben hat. Es gilt also, das Fehlen nicht überzubewerten und stattdessen dem Vorhandensein eines Wortes das entsprechende Gewicht beizumessen.

Problematisch bei der Beurteilung des Vorhandenseins oder Fehlens einer Untersuchungseinheit waren dabei vor allem folgende Fälle:

- Internationalismen: Wörter zumeist lateinischen Ursprungs (*ominös*) sind in den beiden deutsch-englischen Wörterbüchern oft nur direkt mit der entsprechenden englischen Schreibweise angegeben, ohne genauere Angaben zu den Unterbedeutungen. In solchen Fällen habe ich das jeweilige Wort im englischen Teil des PONS 1999 nachgeschlagen und überprüft, ob es die im DUW angegebene Unterbedeutung dort gibt.

Wenn dies der Fall war, habe ich angenommen, dass die spezifische Untersuchungseinheit mit dieser Wortform und der im DUW angegebenen Bedeutung auch in Wildhagen bzw. Muret-Sanders vorhanden und eben nur durch einen nicht weiter spezifizierten Internationalismus repräsentiert ist.

- Starke semantische Differenzierung im DUW: So zum Beispiel bei Titeln bzw. Berufsbezeichnungen (*Oberlehrer, Oberleutnant*). Dort wird zwischen dem Titel und dem Träger des Titels unterschieden, während bei Wildhagen und Muret-Sanders oft nur das entsprechende Stichwort danebensteht, ohne eine weitere, differenzierende Bedeutungserklärung. In solchen Fällen habe ich beiden Bedeutungen den Stabilitätsgrad ‘teilweise’ zugeordnet.
- Problematisch war auch die Frage, wie streng man mit Abweichungen der Form umgehen soll. Muret-Sanders führt *Objektssatz* als Stichwort auf, nicht aber *Objektsatz*. Ich habe diesen und ähnliche Fälle (*Obligationenrecht, Orthografie*) als graphisch nicht identisch und damit als teilweise vorhanden eingestuft.

Zusammengefasst bedeutet das, dass die Beurteilung der Stabilität der einzelnen Untersuchungseinheiten auf folgenden Grundregeln beruht: Zum einen bedeutet ein fehlendes Stichwort in Muret-Sanders nicht unmittelbar, dass der entsprechende Eintrag der Datenbank nicht abgedeckt ist. Es kann auch sein, dass der Eintrag unter einem anderen Stichwort zu finden ist. Das Fehlen eines Wortes in einem Wörterbuch hat geringeres Gewicht als das tatsächliche Vorhandensein eines Wortes in einem Wörterbuch.

Beruhend auf der oben beschriebenen Quantifizierung der Stabilitätsbeschreibung ergibt sich folgende Verteilung der Untersuchungseinheiten auf die unterschiedlichen Klassen:

Tabelle 8-9: Die Schichtung der lexikalischen Einheiten der Untersuchungsmenge in Bezug auf die Stabilität

Stabilitätsgrad	Anzahl LE	
1 (stabil)	666	29%
2 (teilweise stabil)	652	28%
3 (vermutlich nicht stabil)	976	43%

Im Vergleich zu den anderen bisher betrachteten Eigenschaften führt die hier angenommene Bewertung der Eigenschaft der Stabilität zu einer gleichmäßigeren Verteilung der Einheiten der Untersuchungsmenge über die verschiedenen Grade der Stabilität.

8.8.7 Die Bewertung der Neutralität

8.8.7.1 Die Gruppierung der Sondermarkierungen

Die in dem hier untersuchten Wortmaterial vorkommenden Sondermarkierungen (vgl. Tabelle A-9 in Appendix A) habe ich nun in verschiedene Stufen der Markiertheit eingeteilt. Dabei habe ich die durch das DUW-Material vorgegebenen Sondermarkierungen anhand

der in Abschnitt 5.3.1 besprochenen Aspekte der Wortschatzschichtung in bezug auf Sondersprachen zu den folgenden Gruppen zusammengefasst:

- regionale Markierungen: Markierungen, die vor allem eine regionale lexikalische Variation auszeichnen (wie etwa „schweiz.“ oder „österr.“)
- sozialsprachliche Markierungen: Markierungen, die eine lexikalische Einheit beschreiben, die neben ihrer denotativen Bedeutung vor allem auch die Zugehörigkeit zu einer bestimmten sozialen Gruppierung oder spezifischen sozialen Situation bezeichnet (wie etwa „Jugendspr.“, „Kinderspr.“, aber auch „Amtsspr.“)
- fachsprachliche Markierungen: Markierungen, die eine Einheit beschreiben, die einen engen Bezug zu einem bestimmten Thema und oft eine deutlich eingegrenzte Bedeutung hat (etwa „Geol.“ oder „griech. Myth.“)
- Markierungen der emotionalen Konnotation: Markierungen, die eine Einheit beschreiben, die neben ihrer denotativen Bedeutung eine deutlich emotionale Konnotation trägt bzw. eine ausgeprägte pragmatische Funktion hat (wie etwa „verstärkend“, „Schimpfwort“)
- linguistische Markierungen: Markierungen, die eine Einheit hinsichtlich bestimmter linguistischer Aspekte beschreiben, d.h. vor allem Markierungen zum diachronischen Status der Einheit („veraltet“, „veralternd“), aber auch die Markierung „selten“, die einen lexikographischen Kommentar zum Status der Gebräuchlichkeit der Einheit darstellt.

Tabelle A-10 in Appendix A gibt die Einteilung der Sondermarkierungen in Gruppen im Detail wieder.

Bei diesem Vorgehen ist zu hervorzuheben, dass die Abgrenzungen zwischen den Gruppen nicht immer ganz klar sind, speziell bei der Abgrenzung zwischen sozialsprachlichen und fachsprachlichen Markierungen sind oft keine objektiv eindeutigen Einteilungen möglich; die Markierung ‘Seemannssprache’ etwa, die im hier vorgestellten Verfahren als sozialsprachliche Markierung eingeordnet ist, ließe sich auch als fachsprachliche Markierung führen. Hier spielt die in Abschnitt 5.3.1 diskutierte und vor allem von Kalverkämper 1990 gründlich problematisierte Grenzziehung zwischen Fachsprache, Gemeinsprache und Berufssprache hinein, die deutlich macht, dass Fachsprachen immer auch Berufssprachen sind, und Berufssprachen immer auch sozial bedingte Sondersprachen darstellen.

8.8.7.2 Die Gewichtung der Gruppen

Nachdem wir die Liste der Sondermarkierungen durch die Einteilung in Gruppen von Markierungen auf fünf Typen von Markierungen vereinfacht haben, geht es nun darum, diesen Typen ein Gewicht beizumessen. Dabei steht die Frage im Zentrum, wieviel der jeweilige Typus von Sondermarkierungen über die Neutralität der von ihnen bezeichneten lexikalischen Einheiten aussagt. Dabei soll Neutralität verstanden werden als ein möglichst gerin-

ger Grad an Markiertheit. Der Grad an Markiertheit soll dabei daran bemessen werden, wie sehr die Markiertheit den Gebrauch eines Wortes in einer sowohl pragmatisch wie auch thematisch neutralen Situation einschränkt. Folgende Überlegungen spielen dabei bei den jeweiligen Typen von Sondermarkierungen eine Rolle:

- regionale Sondermarkierungen: Regional markierte Wörter tragen im allgemeinen wenig zu einer emotionalen oder thematischen Färbung einer Äußerung bei, abgesehen davon, dass sie in der Regel einem umgangssprachlichen Kontext verwendet werden und erkennen lassen, welcher regionalen Sprachgemeinschaft der Sprecher oder Verfasser eines Textes angehört. Manchmal stellt ein regional beschränkt gebrauchtes Wort auch ein Synonym zu einem anderen, regional nicht beschränkten Wort dar. Insbesondere aber ist ein regional beschränkt gebrauchtes Wort vor allem nur dann als nicht zum Gemeinsprachlichen gehörig markiert, wenn es außerhalb der regional beschränkten Sprechergemeinschaft verwendet wird. Zugleich ist die Menge der Sprecher, die ein regional beschränktes Wort in für sie unmarkierter Weise verwendet, im Vergleich etwa zu den Vertretern einer Gruppensprache, relativ groß. Berücksichtigt man alle diese Überlegungen – die Sonderstellung eines regional markierten Wortes ist nur durch seine lokale Beschränktheit bedingt; die Anzahl der Sprecher, für die dieses Wort unmarkiert ist, weil sie der regionalen Sprechergemeinschaft angehören, ist vergleichsweise hoch – dann ist der Beitrag eines regional markierten Wortes zur Markiertheit einer Aussage insgesamt und vor allem im Vergleich zu den anderen Sondermarkierungen gering.
- sozialsprachlich bedingte Markierungen: Sozialsprachlich markierte Wörter sind Wörter, die nur von einer geringen Anzahl von Sprechern oder nur in einem sehr spezifischen sozialen Kontext – etwa im Gespräch mit Kleinkindern – verwendet werden. Ihr Gebrauch signalisiert eine deutliche Abgrenzung von einer sozial neutralen Kommunikationssituation. Ihr Beitrag zur Markiertheit einer Aussage ist daher relativ hoch. Zugleich können sie nicht in der Mehrzahl aller Kommunikationssituationen verwendet werden, ihre Verwendung ist eingeschränkt. Damit gehören sie eher zum peripheren Bereich des Wortschatzes.
- fachsprachliche Markierungen: Ähnlich wie sozialsprachlich markierte Wörter signalisieren fachsprachlich bzw. thematisch markierte Wörter eine bestimmte Beschränkung, in diesem Fall den spezifischen Bezug auf ein Thema bzw. Fach. Auch sie können in einem allgemeinen Kontext nicht verwendet werden, ohne dass sie ein gewisses Maß an Markiertheit erhalten.
- Markierungen, die eine emotionale Konnotation ausdrücken: Wörter, die als emotional konnotiert markiert sind, prägen die pragmatische Färbung einer Aussage; sie sind in diesem Sinne sehr typische Markierungen.
- linguistische Markierungen: In der hier untersuchten Wortschatzmenge beziehen sich diese Markierungen auf Eigenschaften, die nicht direkt etwas mit der Verwendung in

einer Kommunikationssituation zu tun haben, sondern aus dem Sprachsystem abgeleitet sind und etwas darüber aussagen, welche Rolle die damit markierte Einheit im System des Wortschatzes spielt, ob sie etwa als veraltet eingestuft wird. Diese Eigenschaften können zwar Folgen für die Verwendbarkeit der so markierten Einheit haben, etwa weil die Verwendung eines veralteten Wortes in bestimmten Kommunikationssituationen auffällig sein kann, ihre Markiertheit ist aber nicht aus der Kommunikationssituation heraus abzuleiten, sondern aus ihrer Position im System.

Vor dem Hintergrund dieser Überlegungen sollen die verschiedenen Typen von Markierungen nun zwei Graden der Markiertheit zugeordnet werden, so dass sich zusammen mit der Unmarkiertheit der Einheiten, die keine Sondermarkierung tragen, folgendes Wertesystem hinsichtlich der Neutralität ergibt (vgl. Tabelle 8-10):

Tabelle 8-10: Grade der Neutralität

Grad der Neutralität	Zugehörige Markierungen
1 (= unmarkiert)	-
2 (= geringfügig markiert)	regionale Markierungen
	linguistische Markierungen
3 (= deutlich markiert)	sozialsprachliche Markierungen
	fachsprachliche Markierungen
	Markierungen der emotionalen Konnotation

Entsprechend dieser Einteilung werden alle Einheiten der Untersuchungsmenge einem der drei Grade der Neutralität zugeordnet. Hat eine Einheit mehr als eine Sondermarkierung, also zwei oder drei Sondermarkierungen, dann erhält sie den geringsten Neutralitätsgrad, der sich aus den verschiedenen Sondermarkierungen ableiten lässt.

Quantitativ betrachtet ergibt sich folgendes Bild der Schichtung der Einheiten der Untersuchungsmenge bezüglich der Eigenschaft der Neutralität (vgl. Tabelle 8-11):

Tabelle 8-11: Die Schichtung der lexikalischen Einheiten der Untersuchungsmenge in Bezug auf die Neutralität

Grad der Neutralität	Anzahl LE	
1 (umarkiert)	1479	65%
2 (geringfügig markiert)	100	4%
3 (deutlich markiert)	715	31%

8.9 Das Bewertungssystem im Überblick

Bevor wir uns nun der Auswertung der Ergebnisse, zu denen das hier entworfene Bewertungssystem führt, zuwenden, soll ein kurzer Rückblick die verschiedenen Dimensionen des Bewertungssystems zusammenfassen:

Wir gehen von sechs Bewertungseigenschaften und damit Bewertungsdimensionen aus – Repräsentiertheit, Grad der Worthaftigkeit, Plausibilität, Frequenz, Stabilität und Neutralität. Fast alle diese Bewertungsdimensionen sind in eine dreistufige Skala von 1 bis 3 einge-

teilt: Weist eine Einheit hinsichtlich einer Bewertungsdimension, etwa hinsichtlich der Stabilität den Grad 1 auf, dann besitzt sie die in dieser Dimension betrachtete Eigenschaft in ausgeprägtem Maße, ist also etwa besonders stabil. Weist sie hingegen den Grad 3 auf, dann besitzt sie die betrachtete Eigenschaft nur in geringem Maß. Allein bei der Eigenschaft der Frequenz war es aufgrund der Datenlage nötig, von dieser Dreiteilung abzuweichen; hier gibt es nur die beiden Werte niedrigfrequent und frequent.

Im nächsten Kapitel wird es nun darum gehen, die so für jede Bewertungsdimension gewonnenen Schichtungen der untersuchten Wortschatzmenge darauf hin zu untersuchen, wie sehr sie sich in ihrem Wirkungsgrad in Bezug auf eine Selektion voneinander unterscheiden, d.h. wie sehr sich die Schichtungen ähneln, ob es etwa Überschneidungen gibt, ob es Eigenschaften gibt, die zu einer anderen Schichtung führen als andere etc. Außerdem wird es darum gehen, eine Referenzselektion vorzunehmen, anhand derer die Frage nach dem Umfang eines zentralen Wortschatzes des gesamten DUW-Wortbestandes in einer quantitativen Annäherung besprochen werden soll.

Kapitel 9: Ergebnisse und Beobachtungen

9.1 Einführung

Ziel dieses Kapitels ist es, die Ergebnisse der praktischen Untersuchung auszuwerten und zugleich die Beobachtungen, die sich auf dem Weg zu den Ergebnissen ergeben haben, zusammenzufassen.

9.2 Der Wirkungsgrad und die Wechselwirkung der verschiedenen Selektionseigenschaften im Vergleich

Bei der Bewertung der sechs Eigenschaften – Worthaftigkeit, Repräsentiertheit, Plausibilität, Frequenz, Stabilität und Neutralität – habe ich versucht, für jede der Eigenschaften einen dreistufigen Einteilung zu erreichen.²⁹⁷ Wie bereits im Abschnitt 8.8 diskutiert, ist diese Dreiteilung zunächst willkürlich gewählt und unabhängig von einem konkreten zentralen Wortschatz. Die Details einer solchen Gewichtung müssen für jeden konkreten zentralen Wortschatz gesondert entschieden werden. Hier soll es einfach darum gehen, zu sehen, ob die jeweiligen Selektionseigenschaften zu ähnlichen Ergebnissen führen, oder ob sie signifikant unterschiedlich sind und damit individuell etwas zum Selektionsprozess beitragen. Dazu wollen wir die sich aus der Anwendung der verschiedenen Eigenschaften ergebende Schichtung der Untersuchungseinheiten danach, wie sehr diese die jeweilige Eigenschaft aufweisen, einer Kookkurrenzanalyse unterziehen und ihre Ergebnisse diskutieren.

Zum Zweck dieser Analyse bilden wir für jede Eigenschaft diejenige Gruppe von Untersuchungseinheiten, die diese Eigenschaft am ausgeprägtesten aufweist, also hinsichtlich dieser Eigenschaft einen Zentralitätsgrad von 1 besitzt. In einem zweiten Schritt betrachten wir, wie hoch der Anteil der Einheiten aus dieser Gruppe ist, die zugleich auch eine bestimmte zweite Eigenschaft zu einem Zentralitätsgrad von 1 aufweist. Die Höhe dieses prozentualen Anteils sagt etwas darüber aus, wie wahrscheinlich es ist, dass zwischen den beiden betrachteten Eigenschaften eine Kookkurrenz besteht – je höher der prozentuale Anteil, desto höher ist der Anteil derjenigen Einheiten, die für beide Eigenschaften einen Zentralitätsgrad von 1 aufweisen. Ob sich diese Kookkurrenz aber eindeutig auf einen ähnlichen Selektionseffekt bezieht oder andere Gründe hat, muss dann in einem weiteren Schritt geklärt werden.

Tabelle 9-12 gibt für jedes geordnete Paar von Eigenschaften den jeweiligen Prozentsatz an gemeinsamen Untersuchungseinheiten, den wir als *Kookkurrenzwert* bezeichnen wollen, wieder.²⁹⁸

297. In den der gedruckten Ausgabe beiliegenden Dateien *Gewichtete_Repräsentiertheit.xls*, *Gewichtete_Plausibilität.xls*, *Gewichtete_Worthaftigkeit.xls*, *Gewichtete_Frequenz.xls*, *Gewichtete_Stabilität.xls* und *Gewichtete_Neutralität.xls* sind alle Untersuchungseinheiten entsprechend ihrer Zuordnung zu einem der drei Zentralitätsgrade für die jeweilige Eigenschaft gespeichert.

Tabelle 9-12: Kookkurrenzanalyse der Selektionseigenschaften bezüglich des Zentralitätsgrades 1

	1	2	3	4	5	6
	Repräsentiertheit	Worthaftigkeit	Plausibilität	Frequenz	Stabilität	Neutralität
Umfang in Anzahl LE	2060	2092	130	1520	666	1479
1 Repräsentiertheit	100%	91%	99%	96%	97%	92%
2 Worthaftigkeit	92%	100%	88%	93%	97%	88%
3 Plausibilität	6%	5%	100%	8%	12%	7%
4 Frequenz	71%	68%	95%	100%	85%	68%
5 Stabilität	31%	31%	61%	37%	100%	32%
6 Neutralität	66%	62%	83%	66%	71%	100%

Wir gehen nun vereinfachend davon aus, dass ein Anteil von 80% und mehr auf eine starke Kookkurrenz zwischen den beiden betrachteten Eigenschaften hindeutet. In der Tabelle sind diese Werte mit roter Farbe hinterlegt. Umgekehrt deutet ein Anteil von 20% und weniger auf eine geringe Kookkurrenz zwischen den beiden betrachteten Eigenschaften hin; diese Werte sind in der Tabelle mit grüner Farbe hinterlegt. Werte, die zwischen 20% und 80% liegen, sind zunächst unklare Fälle.

Um nun beurteilen zu können, wie sehr eine Eigenschaft mit jeweils einer der anderen hier verglichenen Eigenschaften in einem Kookkurrenzverhältnis steht, muss man sich Tabelle 9-12 spaltenweise von links nach rechts ansehen und dabei jede Eigenschaft getrennt betrachten.

Dabei wird schnell deutlich, dass die Eigenschaft der Repräsentiertheit zu jeder der anderen untersuchten Eigenschaften in einem Kookkurrenzverhältnis steht, dass also beispielsweise die Einheiten, die hinsichtlich der Frequenz einen Zentralitätsgrad von 1 aufweisen (weil sie zu den nicht niedrig frequenten Einheiten der Untersuchungsmenge gehören) in 96 von 100 Fällen auch hinsichtlich der Repräsentiertheit einen Zentralitätsgrad von 1 aufweisen; (vgl. Spalte 4 und Zeile 1 in Tabelle 9-12) oder dass etwa diejenigen Einheiten, die hinsichtlich der Neutralität einen Zentralitätsgrad von 1 aufweisen, in 92 von 100 Fällen auch hinsichtlich der Repräsentiertheit einen Zentralitätsgrad von 1 aufweisen (vgl. Spalte 6 und Zeile 1 in Tabelle 9-12).

Für die Eigenschaft der Worthaftigkeit gilt Ähnliches: Auch sie steht zu jeder der anderen Eigenschaften in einem Kookkurrenzverhältnis (vgl. Spalte 2 in Tabelle 9-12). Hervorzuheben ist dabei, dass Worthaftigkeit und Repräsentiertheit in einem nahezu gegenseitigen Verhältnis der Kookkurrenz zueinander stehen, denn gleichgültig, ob man zunächst die Menge derjenigen Einheiten betrachtet, die hinsichtlich der Repräsentiertheit einen Zentralitätsgrad von 1 aufweisen, und daraus dann die Einheiten auswählt, die hinsichtlich der Worthaftigkeit einen Zentralitätsgrad von 1 aufweisen (vgl. Spalte 1 und Zeile 2 in Tabelle

298. Dabei ist die Reihenfolge der Untersuchung ist hier wichtig – zuerst werden Untersuchungseinheiten anhand der ersten zu vergleichenden Eigenschaft ausgewählt, und aus der so entstandenen Selektion werden diejenigen Einheiten ausgewählt, die auch hinsichtlich der zweiten zu untersuchenden Eigenschaften einen Zentralitätsgrad von 1 aufweisen.

9-12), oder ob man umgekehrt vorgeht (vgl. Spalte 2 und Zeile 1 in Tabelle 9-12), ergibt sich ein fast gleicher Kookkurrenzwert von 91 bzw. 92 %.

Umgekehrt stellt sich das für die Eigenschaft der Plausibilität dar: Sie steht mit keiner der anderen Eigenschaften in einem Kookkurrenzverhältnis, denn die Kookkurrenzwerte, die sie erzeugt, wenn man sie zur Selektion anhand einer der anderen Eigenschaften hinzufügt, liegen allesamt unter 20%.

Unklar ist das Ergebnis bei den Eigenschaften der Frequenz, der Stabilität und der Neutralität. Abgesehen von einigen wenigen Ausnahmen liegen ihre Kookkurrenzwerte zwischen 20 und 80 %. Wir werden sie noch weiter unten genauer betrachten.

Nun haben wir uns also zunächst die reinen Kookkurrenzwerte der verschiedenen Kombinationen von Eigenschaften angesehen. Diese Werte können aber nicht direkt als Indikatoren für den Wirkungsgrad, d.h. die selektive Kraft der einzelnen Eigenschaften betrachtet werden. Denn die Kookkurrenzen an sich sind zunächst einfach nur Werte, eine andere Sache ist die Frage danach, was die Ursachen dieser Werte sind, wie diese Werte also zu interpretieren sind. Bei der Frage nach der Interpretation der Kookkurrenzwerte spielt im vorliegenden Fall die Frage des Umfangs der jeweiligen Selektionen eine Rolle. Es zeigt sich, dass in allen Fällen, in denen sich deutliche Kookkurrenzen (Repräsentiertheit und Worthaftigkeit) bzw. deutliche Unabhängigkeiten (Plausibilität) abzeichnen, der Umfang der jeweiligen Selektion entweder extrem groß oder extrem klein ist. Von den 2294 Einheiten der Gesamtmenge, aus der wir eine Auswahl vornehmen wollen, zeigen 2060 bzw. 2092 Untersuchungseinheiten, also knapp 90 %, einen Zentralitätsgrad von 1 hinsichtlich der Eigenschaften der Repräsentiertheit bzw. Worthaftigkeit. In der Praxis bedeutet das, dass fast jede Einheit der Gesamtmenge hinsichtlich der Repräsentiertheit einen Zentralitätsgrad von 1 aufweist und dass fast jede Einheit auch hinsichtlich der Worthaftigkeit einen Zentralitätsgrad von 1 aufweist. Dieses Ergebnis überrascht nicht, wenn man bedenkt, dass das untersuchte Material bereits lexikographisch aufbereitet war und daher eher den konventionalisierten, d.h. bekannten Teil des Gesamtwortschatzes des Deutschen abbildet als etwa unbearbeitetes Korpusmaterial.

Umgekehrt lässt sich über die Eigenschaft der Plausibilität sagen, dass die große Mehrzahl der 2294 Einheiten der Untersuchungsmenge hinsichtlich der Plausibilität nicht einen Zentralitätsgrad von 1 aufweisen, sondern einen Zentralitätsgrad von 2 oder 3, weil nur 130 Einheiten hinsichtlich der Plausibilität einen Zentralitätsgrad von 1 aufweisen. Daraus erklären sich die geringen Kookkurrenzwerte für die Fälle, in denen Plausibilität als zweites Kriterium der Selektion hinzugenommen wird – die 130 Untersuchungseinheiten stellen insgesamt nur knapp 6 % der Gesamtmenge dar und werden dadurch immer kleine Kookkurrenzwerte erzeugen.

Damit birgt der Einsatz dieser drei Eigenschaften jedoch nur wenige Überraschungen. Denn dadurch, dass diese drei Eigenschaften so polarisierte Mengen abbilden – entweder extrem

große oder extrem kleine Teilmengen der Gesamtmenge -, lässt sich über fast jedes beliebige Wort der Gesamtmenge sagen, dass es hinsichtlich der Eigenschaften der Repräsentiertheit und Worthaftigkeit einen Zentralitätsgrad von 1 aufweist und hinsichtlich der Plausibilität einen Zentralitätsgrad von 2 oder 3.

Was bedeutet das aber nun für die Anwendung dieser Eigenschaften als Selektionskriterien?

Zum einen bedeuten die eben besprochenen Ergebnisse, dass Repräsentiertheit und Worthaftigkeit bezüglich der hier betrachteten Datengrundlage – dem Wortbestand eines einsprachigen Wörterbuchs – keinen oder nur einen sehr geringen selektierenden Effekt haben. Das haben wir bereits bei der Beschreibung dieser Eigenschaften (vgl. Abschnitte 8.8.2 und 8.8.3) vermutet und sehen diese Vermutung nun hier bestätigt. Das bedeutet aber nicht, dass diese beiden Eigenschaften grundsätzlich nicht als Selektionskriterien in Frage kommen, denn die Ergebnisse ihrer Anwendung können ganz anders aussehen, wenn sie auf Daten angewendet werden, die aus Korpusuntersuchungen entnommen sind.²⁹⁹

Zum anderen wird deutlich, dass die Eigenschaft der Plausibilität wiederum einen stark selektierenden Charakter hat; allerdings ist das auch hier wieder vor dem Hintergrund der Ausgangslage zu bewerten: Die Eigenschaft der Plausibilität ist allein auf der Beschreibung eines beurteilenden Sprechers entstanden; die Ergebnisse können anders ausfallen, wenn mehrere Sprecher eine solche Beschreibung vornehmen. Darüberhinaus deckt das DUW eine Vielzahl sehr spezifischer Wörter und Termini ab, die kaum in der Alltagssprache eines Sprechers auftauchen, was die geringe Anzahl der Untersuchungseinheiten erklärt, die eine Plausibilität von 1 aufweisen. Wäre ein Textkorpus der Ausgangspunkt gewesen, wäre, je nach Zusammensetzung des Korpus, der Anteil ‘alltäglicher’ und damit plausibler Wörter vermutlich höher ausgefallen.

Nun wollen wir uns der Auswertung der weniger klaren Eigenschaften zuwenden, d.h. derjenigen Eigenschaften, die zum einen über alle Kombinationen hinweg mittlere Prozentwerte aufweisen und deren Anzahl an Untersuchungseinheiten hinsichtlich des Zentralitätsgrades 1 weder besonders hoch noch besonders niedrig ist; also die Eigenschaften der Frequenz, der Stabilität und der Neutralität. Weil die Daten, die wir aus der Betrachtung des Zentralitätsgrades von 1 gewinnen können, keine eindeutige Aussage über ihre Wirkung zulassen, wir also nicht imstande sind, hinsichtlich dieser Eigenschaften für jede beliebige Einheit der Gesamtmenge so klare ‘Vorhersagen’ zu treffen wie das hinsichtlich der Eigenschaften der Repräsentiertheit, Worthaftigkeit und Plausibilität möglich war, können Frequenz, Stabilität und Neutralität tatsächlich eine selektive Wirkung haben und voneinander unabhängig sein.

299. Vgl. dazu auch Abschnitt 4.3.

Um Indizien für diese Vermutung zu finden, wollen wir uns für die Eigenschaften der Stabilität und der Neutralität die jeweiligen Kookkurrenzen, die sie miteinander über alle Zentralitätsgrade hinweg bilden, ansehen:

Tabelle 9-13: Kookkurrenzanalyse der Eigenschaft der Stabilität im Verhältnis zur Neutralität

	Neutralität: Zentralitätsgrad 1	Neutralität: Zentralitätsgrad 2	Neutralität: Zentralitätsgrad 3
Anzahl der LE je Zentralitätsgrad	1479	100	715
Stabilität: Zentralitätsgrad 1	32%	31%	22%
Stabilität: Zentralitätsgrad 2	26%	45%	31%
Stabilität: Zentralitätsgrad 3	42%	24%	47%

Tabelle 9-13 macht deutlich, dass es zwischen der Eigenschaft der Stabilität und der der Neutralität keine eindeutigen Verbindungen gibt. Wählt man eine beliebige Einheit aus der Gruppe derjenigen, die hinsichtlich der Neutralität einen Zentralitätsgrad von 1 aufweisen, dann kann diese Einheit hinsichtlich der Stabilität einen beliebigen Zentralitätsgrad aufweisen, weil keine der drei Zentralitätsgrade hinsichtlich der Stabilität die Selektion dominiert. Ähnliches gilt für die Zentralitätsgrade 2 und 3 hinsichtlich der Neutralität.

Kehrt man das Bild um und betrachtet die Gruppe von Untersuchungseinheiten, die hinsichtlich der Eigenschaft der Stabilität einen Zentralitätsgrad von 1 aufweisen (vgl. Tabelle 9-14), verschiebt sich das Bild ein wenig, bleibt aber im Grunde gleich: Wählt man eine beliebige Einheit aus derjenigen Gruppe, die hinsichtlich der Stabilität einen Zentralitätsgrad von 1 aufweisen, dann weist diese Einheit hinsichtlich der Neutralität tendenziell einen Neutralitätsgrad von 1 oder 3 auf.

Tabelle 9-14: Kookkurrenzanalyse der Eigenschaft der Neutralität im Verhältnis zur Stabilität bezüglich aller Zentralitätsgrade

	Stabilität: Zentralitätsgrad 1	Stabilität: Zentralitätsgrad 2	Stabilität: Zentralitätsgrad 3
Anzahl der LE je Zentralitätsgrad	666	652	976
Neutralität: Zentralitätsgrad 1	71%	60%	63%
Neutralität: Zentralitätsgrad 2	5%	7%	3%
Neutralität: Zentralitätsgrad 3	24%	33%	34%

Was bedeutet das aber nun für die Anwendung der Eigenschaften der Neutralität und der Stabilität als Selektionskriterien? Die geringen Kookkurrenzwerte und vor allem ihre fast gleichmäßige Verteilung über alle Zentralitätsgrade hinweg, gleichgültig, welche der beiden Eigenschaften als erste Selektionseigenschaft betrachtet wird, bedeutet, dass die Anwendung dieser Eigenschaften zu Ergebnissen führt, die nicht bereits aus den Eigenschaften selbst abzulesen sind. Sie bringen also ein Element der ‘Überraschung’ in den Selektionsprozess mit ein und tragen damit individuell zu einem Selektionsergebnis bei, im Gegensatz

etwa zu den Eigenschaften der Repräsentiertheit und der Worthaftigkeit, bei denen wir schon wissen, dass fast alle Einheiten der Untersuchungsmenge hinsichtlich dieser Eigenschaften einen Zentralitätsgrad von 1 aufweisen.

Schließlich wollen wir noch einen kurzen Blick auf die Eigenschaft der Frequenz werfen. Die genauere Betrachtung dieser Eigenschaft macht deutlich, dass hier die Tatsache, dass wir hinsichtlich der Frequenz nur zwei Zentralitätsgrade festgelegt haben (vgl. Abschnitt 8.8.5), von denen der Zentralitätsgrad 1 auf rund 70 % aller Einheiten zutrifft, zu einer ähnlich polarisierenden Wirkung beiträgt, wie sie bei den Eigenschaften der Repräsentiertheit bzw. der Worthaftigkeit eintritt – dadurch dass die Mehrzahl der untersuchten Einheiten hinsichtlich der Frequenz zum Zentralitätsgrad 1 gehören, kann man in der Mehrzahl der Fälle auch davon ausgehen, dass eine beliebige Einheit der Untersuchungsmenge wahrscheinlich zur ersten Frequenzgruppe gehört. Wir wollen deshalb auf eine eingehende Ko-
 okkurrenzanalyse verzichten.

Allerdings wollen wir uns ansehen, ob für die Annahme der Vertreter des frequenzorientierten Ansatzes zentraler Wortschätze, Frequenz sei ein guter Indikator für die Zentralität eines Wortes, weil sie verschiedene Eigenschaften, wie etwa die Gebräuchlichkeit oder Neutralität eines Wortes, die zu seiner Wichtigkeit beitragen, widerspiegeln, in der hier untersuchten Datenmenge Anhaltspunkte zu finden sind. Die Grundlage dafür soll ein exemplarischer Blick auf die verschiedenen Zentralitätsgrade der Eigenschaften der Repräsentiertheit, der Plausibilität, der Stabilität und der Neutralität und die Verteilung der beiden Frequenzgruppen auf diese Teilmengen liefern:

Tabelle 9-15: Die Verteilung der Frequenzen über die verschiedenen Grade der Repräsentiertheit

	Repräsentiertheit: Zentralitätsgrad 1	Repräsentiertheit: Zentralitätsgrad 2	Repräsentiertheit: Zentralitätsgrad 3
	2060	14	220
Frequenz: Zentralitätsgrad 1	71%	21%	24%
Frequenz: Zentralitätsgrad 2	29%	79%	76%

Tabelle 9-16: Die Verteilung der Frequenzen über die verschiedenen Grade der Plausibilität

	Plausibilität: Zentralitätsgrad 1	Plausibilität: Zentralitätsgrad 2	Plausibilität: Zentralitätsgrad 3
	130	803	1361
Frequenz: Zentralitätsgrad 1	95%	81%	55%
Frequenz: Zentralitätsgrad 2	5%	19%	45%

Tabelle 9-17: Die Verteilung der Frequenzen über die verschiedenen Grade der Stabilität

	Stabilität: Zentralitätsgrad 1	Stabilität: Zentralitätsgrad 2	Stabilität: Zentralitätsgrad 3
	666	652	976
Frequenz: Zentralitätsgrad 1	85%	75%	47%
Frequenz: Zentralitätsgrad 2	15%	25%	53%

Tabelle 9-18: Die Verteilung der Frequenzen über die verschiedenen Grade der Neutralität

	Neutralität: Zentralitätsgrad 1 1479	Neutralität: Zentralitätsgrad 2 100	Neutralität: Zentralitätsgrad 3 715
Frequenz: Zentralitätsgrad 1	68%	71%	62%
Frequenz: Zentralitätsgrad 2	32%	29%	38%

Betrachtet man auch diese Tabellen wieder von links nach rechts, wird deutlich, dass der Anteil derjenigen Untersuchungseinheiten, die zu den höher frequenten Einheiten gehören (Zentralitätsgrad 1 der Frequenz) mit jedem der verschiedenen Zentralitätsgrade der jeweils vorab gewählten Eigenschaft (Repräsentiertheit, Plausibilität, Stabilität, Neutralität) abnimmt, dass also etwa der Anteil höher frequenter Einheiten bei der Gruppe derjenigen Untersuchungseinheiten, die einen Stabilitätsgrad von 3 aufweisen, deutlich geringer ist (47%) als bei der Gruppe der Untersuchungseinheiten, die einen Stabilitätsgrad von 2 bzw. 1 aufweisen (75% bzw. 85%). Die allgemeine Annahme, Frequenz spiegele grundsätzlich eine Art Zugehörigkeit zum Zentrum eines Wortschatzes wider, scheint hier also tendenziell bestätigt werden zu können.

Was bedeuten aber nun alle diese Untersuchungen für die Frage nach der Wechselwirkung der Selektionskriterien und ihrem Wirkungsgrad für eine Selektion?

Hinsichtlich der Wechselwirkung der Selektionskriterien können wir festhalten, dass die drei Eigenschaften der Neutralität, der Stabilität und der Plausibilität unabhängige Selektionseigenschaften zu sein scheinen, die in keiner deutlichen Wechselwirkung mit den anderen hier untersuchten Selektionseigenschaften stehen. Umgekehrt scheint es eine enge Verbindung zwischen der Repräsentiertheit und der Worthaftigkeit zu geben, da die Selektionsergebnisse, die mit jeweils der einen wie der anderen Eigenschaft erzielt werden, gleichgültig, welche der beiden zuerst betrachtet wird (s.o.), fast identisch sind. Über die Frequenz lassen sich hinsichtlich der Wechselwirkung nur sehr ungefähre Aussagen zu machen; wie die obige Analyse der Verteilung der Frequenz über die verschiedenen Zentralitätsgrade unterschiedlicher Eigenschaften gezeigt hat, scheint es jedoch eine schwache Wechselwirkung zwischen den jeweiligen Eigenschaften der Repräsentiertheit, der Plausibilität, der Stabilität und der Neutralität auf der einen und der Frequenz auf der anderen Seite zu geben, d.h. dass eine hohe Frequenz tendenziell mit einem vergleichsweise hohen Zentralitätsgrad der jeweiligen Eigenschaft einhergeht.

Bei der Frage nach dem Wirkungsgrad der einzelnen Selektionskriterien steht die Eigenschaft der Plausibilität mit ihrem stark selektiven Charakter an erster Stelle. Ähnlich selektiv scheinen die Eigenschaften der Neutralität und der Stabilität zu sein, weil sich ihre Zentralitätsgrade relativ gleichmäßig über die Gesamtmenge der Einheiten verteilen. Weniger selektiv hingegen wirkt die Eigenschaft der Frequenz allein aufgrund der Tatsache, dass sie die Untersuchungseinheiten nur in zwei Grade der Zentralität einteilen kann. Am geringsten ist der selektive Wirkungsgrad bei den Eigenschaften der Repräsentiertheit und der Worthaftigkeit, weil sie polarisierend wirken und bei ihnen eine sehr große Menge an

Untersuchungseinheiten dem Zentralitätsgrad 1 zugeordnet sind und nur wenige Einheiten den anderen beiden Zentralitätsgraden.

Zusammenfassend lässt sich dabei feststellen, dass die Eigenschaften, die am unabhängigsten von den anderen Eigenschaften sind und deren Selektionseffekt am stärksten ist, also die Eigenschaften der Plausibilität, der Neutralität und der Stabilität, gerade diejenigen Eigenschaften sind, die von der Einschätzung des Beurteilenden, im Falle der Neutralität der Lexikographen der Duden-Redaktion abhängen.³⁰⁰ Diese Beobachtung legt nahe, dass den Gegnern der rein frequentiell orientierten Ansätze bzw. den Vertretern einer transparenten Introspektion wie etwa Bonan-Garrigues 1993 recht zu geben ist.

9.3 Die Bildung einer Referenzselektion und ihre Analyse

9.3.1 Die konkreten Schritte der Erarbeitung einer Referenzselektion

Im folgenden soll es nun darum gehen, zu sehen, zu welchem Wortschatzumfang man gelangt, wenn man die hier betrachteten Eigenschaften als Selektionskriterien auf den untersuchten Wortbestand anwendet.

Dazu wollen wir eine Referenzselektion erstellen. Diese Referenzselektion soll als Grundlage für die Diskussion der Frage nach dem Umfang zentraler Wortschätze dienen und auch einige weitere Analysen erlauben. Dabei ist es wichtig hervorzuheben, dass es sich bei der Konstruktion einer solchen Referenzselektion um einen Versuch, ein Experiment handelt, dessen Ergebnisse vor allem Rückschlüsse zulassen sollen über die Eignung des angewandten Verfahrens, nicht aber einen konkreten zentralen Wortschatz repräsentieren. Ein solcher würde eine gründliche Diskussion der konkreten Anforderungen der Anwendung dieses Wortschatzes erfordern, wie sie durch die Definition des Begriffs *zentraler Wortschatz* (vgl. Kapitel 6) als einem wohldefinierten, anwendungsorientierten Konzept deutlich wird.

Wie bereits bei der Untersuchung möglicher Kookkurrenzen zwischen den einzelnen Eigenschaften deutlich geworden ist, spielt die Reihenfolge, in der man die einzelnen Eigenschaften zur Selektion heranzieht, eine ausschlaggebende Rolle. Weil der selektive Wirkungsgrad der verschiedenen hier betrachteten Eigenschaften unterschiedlich stark ist, wollen wir daher mit der Eigenschaft, die den geringsten selektiven Wirkungsgrad hat, nämlich der Repräsentiertheit beginnen; d.h. wir bilden eine erste Selektion derjenigen Einheiten der Untersuchungsmenge, die hinsichtlich der Repräsentiertheit einen Zentralitätsgrad von 1 aufweisen. Damit haben wir die Gesamtmenge an Untersuchungseinheiten von ursprünglich 2294 Einheiten auf 2060 Einheiten reduziert.

300. Was sich dem Vorwort des DUW entnehmen lässt, deutet darauf hin, dass die Sondermarkierungen, die den Einträgen zugewiesen sind, aufgrund der Beurteilung der Lexikographen zustande gekommen ist und nicht aufgrund einer detaillierten Korpusanalyse (die bei dem Umfang an zu beurteilenden Einträgen sehr aufwendig wäre).

Aus dieser Teilmenge wählen wir nun diejenigen Einheiten aus, die hinsichtlich ihrer Worthaftigkeit einen Zentralitätsgrad von 1 aufweisen, und erhalten damit eine weiterhin reduzierte Selektionsmenge von 1902 Einheiten.

In einem weiteren Selektionsschritt wählen wir aus dieser Teilmenge wiederum diejenigen Einheiten aus, die eine Frequenz aufweisen, die nicht zur Gruppe der niedrig-frequenten Einheiten gehören. Daraus ergibt sich eine Selektionsmenge von 1380 Einheiten.

Im weiteren begrenzen wir die Selektionsmenge auf diejenigen Einheiten, die hinsichtlich der Neutralität einen Zentralitätsgrad von 1 aufweisen, und erhalten damit eine weiter verkleinerte Selektionsmenge von 897 Einheiten.

Im nächsten Schritt begrenzen wir diese Selektionsmenge anhand der Eigenschaft der Stabilität und wählen nur diejenigen Einheiten der Untersuchungsmenge, die hinsichtlich der Stabilität einen Zentralitätswert von 1 aufweisen. Damit erhalten wir eine Selektionsmenge von 395 Einheiten.

Zum Schluss wenden wir die Eigenschaft mit der stärksten selektiven Wirkung an, nämlich die Plausibilität, und wählen aus der Selektionsmenge diejenigen Einheiten aus, die hinsichtlich der Plausibilität einen Zentralitätsgrad von 1 aufweisen. Damit erhalten wir eine endgültige Selektionsmenge von 71 Einheiten. Diese Selektionsmenge soll im weiteren als Referenzselektion bezeichnet werden. Sie ist auf der gedruckten Ausgabe beiliegenden CD-ROM in der Datei *Referenzselektion.xls* wiedergegeben.

9.3.2 Eine Annäherung an die Frage des Umfangs eines zentralen Wortschatzes des gesamten DUW-Bestandes

Die durch das oben beschriebene Verfahren der Erstellung einer Referenzselektion selektierte Anzahl von 71 Einheiten macht ca. 3 % des Bestandes von 2294 ursprünglich beschriebenen Untersuchungseinheiten aus.

Um bestimmen zu können, wie die entsprechende Anzahl von Untersuchungseinheiten für eine Selektion aus dem Gesamtbestand der im DUW beschriebenen Einheiten aussehen könnte, wollen wir in einer groben Hochrechnung berechnen, wieviele lexikalische Einheiten (nicht Stichworteinträge) der DUW insgesamt enthält. Dies wollen wir anhand der Anzahl von Seiten tun, die die jeweiligen Buchstaben im DUW einnehmen (vgl. Tabelle 9-19):

Tabelle 9-19: Anzahl Seiten pro Buchstabe im DUW

Buchstabe	Anzahl Seiten	
A	150	8.62%
B	96	5.51%
C	13	0.75%
D	71	4.08%
E	91	5.23%
F	77	4.42%
G	91	5.23%
H	102	5.86%
I	34	1.95%
J	12	0.69%
K	119	6.84%
L	56	3.22%
M	74	4.25%
N	40	2.30%
O	20	1.15%
P	87	5.00%
Q	6	0.34%
R	72	4.14%
S	225	12.92%
T	66	3.79%
U	51	2.93%
V	74	4.25%
W	59	3.39%
X	1	0.06%
Y	1	0.06%
Z	53	3.04%
	1741	100.00%

Geht man davon aus, dass die Anzahl von lexikalischen Einheiten pro Seite im Durchschnitt für alle Seiten im DUW gleich ist und dem Durchschnitt des Buchstaben O entspricht, der 2294 Einheiten auf insgesamt 20 Seiten beschreibt, dann erhält man eine ungefähre Anzahl von knapp 200 000 lexikalischen Einheiten, die im DUW insgesamt enthalten sind. Geht man weiterhin davon aus, dass der prozentuale Anteil derjenigen Einheiten, die durch das in Abschnitt 9.3.1 beschriebene Verfahren in die endgültige Selektion gelangen, für alle Buchstaben gleich ist und dem prozentualen Anteil der für den Buchstaben O vorgenommenen Selektion entspricht, dann kommt man für die Referenzselektion auf eine Anzahl von ca. 6 000 lexikalischen Einheiten. Diese Zahl liegt etwas über dem Umfang, der in der Grundwortschatzliteratur als obere Grenze für einen Grundwortschatz angegeben wird (vgl. Abschnitt 4.6).

Was lässt sich aus dieser vorläufigen Umfangsberechnung schließen? Zum einen wird deutlich, dass die Gesamtheit der hier untersuchten Eigenschaften durchaus zu einer Selektion führen kann, deren Umfang ungefähr dem in der Literatur bisher geschätzten entspricht. Zum anderen wird aber auch deutlich, dass zu dieser Umfangsreduktion die Eigenschaft der Plausibilität viel beiträgt. Eine Selektion, bei der die Plausibilität nicht berücksichtigt wird, würde zu einem Umfang von 395 (s. o.), d.h. rund 17 % geführt. Auf den Gesamtbestand des DUW umgerechnet wäre damit die Referenzselektion auf ca. 34 000 lexikalische Einheiten gekommen und hätte damit den geschätzten Umfang des Standardwortschatzes eines

durchschnittlichen Sprechers von ca. 8 000 bis 10 000 Wörtern (vgl. Glück ((Hg.) 2000, S. 799f.)) bei weitem überstiegen, selbst wenn man die Unterschiede in der Umfangsangabe in Betracht zieht, die durch die Verwendung von Wort gegenüber lexikalischer Einheit entstehen.

9.3.3 Das Verhältnis zwischen der Extension eines Lexems und seiner kumulativen Zentralität

Bei der Erfassung der Einträge des DUW und der Bildung von Untersuchungseinheiten war es ein Anliegen, möglichst detailliert der semantischen Aufteilung, die der DUW vornimmt, zu folgen und Untereinträge eines Stichwortes als eigene lexikalische Einheiten zu erfassen. Wenn wir weiter oben verschiedene Größenordnungen des Umfangs berechnet und besprochen haben, lag diesen Berechnungen immer die Grundeinheit im Sinne einer lexikalischen Einheit aus einer Form und einer Bedeutung zugrunde. Wie bereits Krohn (1992, S. 97) deutlich macht, kann diese feine Aufteilung zu einem Aufblähen des zentralen Wortschatzes führen. Wir wollen deshalb einmal sehen, wieviele der in der Referenzselektion erfassten lexikalischen Einheiten Teil ein und desselben DUW-Eintrages sind, um so erkennen zu können, ob sich in dieser Referenzselektion besonders viele Einzelbedeutungen polysemer Lexeme befinden, deren Abgrenzung voneinander so nicht nötig wäre und die dadurch eventuell den Umfang der Referenzselektion nur künstlich vergrößert.

Bei der Untersuchung der Referenzselektion auf lexikalische Einheiten polysemer DUW-Stichwörter müssen wir zwei Gruppen unterscheiden – Einheiten, die von polysemen DUW-Stichwörtern abstammen, aber als ihr alleiniger Vertreter in die Referenzselektion aufgenommen wurden (wie etwa *Onkel* (DUW 880.1.-)) und die man daher innerhalb der Selektion als ‘Monoseme’ betrachten kann, und Einheiten, die von polysemen DUW-Stichwörtern abstammen und gemeinsam mit anderen Einheiten, die andere Teilbedeutungen des DUW-Stichwortes repräsentieren (wie etwa die verschiedenen lexikalischen Einheiten der Form *offen* (DUW- ID 432.n.n)), in die Selektion aufgenommen wurden. Es sind diese letzteren Fälle, die zu einer künstlichen Vergrößerung der Selektion beitragen, wenn die semantische Aufgliederung des jeweiligen DUW-Stichwortes zu fein ausgefallen ist. Um diese lexikalischen Einheiten polysemer Stichwörter soll es im Folgenden gehen.

In der Referenzselektion finden sich 49 Einheiten, die sich mit einer oder mehreren anderen Einheiten der Selektion ein DUW-Stichwort teilen, also ein polysemes Lexem repräsentieren. Diesen wiederum liegen insgesamt 14 individuelle DUW-Stichwörter zugrunde. Nimmt man also an, die Angaben zum Umfang des Wortschatzes des Deutschen, zum produktiven Wortschatz eines einzelnen Sprechers oder zu Grundwortschätzen beruhen auf dem Lexem als Grundeinheit, das ja durchaus eine polyseme Bedeutungsstruktur aufweisen kann, und rechnet die Anzahl der Untersuchungseinheiten, die als polyseme Bedeutungsvarianten ein und dasselbe DUW-Stichwort vertreten, aus der Referenzselektion heraus und ersetzt sie stattdessen durch ihre polysemen Platzhalter, d.h. die polysemen DUW-Stich-

wörter, dann ergibt sich daraus eine Reduktion des Umfangs der Referenzselektion von 71 Einheiten auf 36, also ca. 1,5 % der betrachteten Gesamtmenge. Umgerechnet auf den Gesamtbestand des DUW würde sich ein derart erstellter zentraler Wortschatz demnach auf 3 000 Lexeme belaufen, also auf deutlich weniger als die ursprüngliche Hochrechnung. Legt man diesen Umfang dem Vergleich mit den Schätzungen in der bisherigen Literatur, die vom Wort bzw. Lexem als Grundeinheit ausgehen, zugrunde, ergibt sich ein Bild, das mit diesen Schätzungen recht gut übereinstimmt.

An dieser Stelle wollen wir auch der Vermutung nachgehen, polyseme Lexeme seien oft zentrale Elemente des Wortschatzes, gerade weil sie polysem verwendet werden können.³⁰¹ Dazu wollen wir uns ansehen, wieviele Einheiten in der Referenzselektion von polysemen DUW-Stichwörtern abstammen (gleichgültig ob sie zusammen mit ihren polysemen Bedeutungsvarianten in die Referenzselektion aufgenommen wurden oder ob sie als alleiniger Vertreter eines ursprünglich polysemen DUW-Stichwortes aufgenommen wurden) und wieviele Einheiten ursprüngliche Monoseme sind. Es zeigt sich, dass nur 11 der 71 lexikalischen Einheiten der Referenzselektion ursprüngliche Monoseme sind, d.h. bereits im DUW als monoseme Einträge verzeichnet sind. Weit mehr als die Hälfte aller Einheiten der Referenzselektion hingegen, d.h. 60 lexikalische Einheiten, sind aus polysemen DUW-Einträgen abgeleitet. Damit scheint sich die Annahme von der Korrelation zwischen polysemer Bedeutungsstruktur und Zentralität, die Carter 1998 seinen Tests auf *coreness* teilweise zugrunde legt (vgl. Abschnitt 5.5.1), zu bestätigen, selbst wenn man berücksichtigen muss, dass es sich hier um eine sehr kleine Datengrundlage handelt.

Eine Frage, die hier aus Platzgründen nicht diskutiert werden kann, ist diejenige, wie zweckmäßig die semantische Aufgliederung ist, die im DUW bei polysemen Lexemen vorgenommen wird. Sie scheint vor allem im Licht der praktischen Möglichkeiten einerseits sehr nützlich, weil etwa ohne ihre detaillierte Darstellung eine genaue Erhebung der lexikographischen Stabilität (vgl. Abschnitt 8.6) nicht möglich gewesen wäre. Andererseits sind ihre genauen Aufteilungen, etwa die Unterscheidung zwischen *Offizier* als Titel (DUW-ID 491.1.a) und *Offizier* als Bezeichnung für den Träger des Titels (DUW-ID 491.1.b) oft in anderen herangezogenen Quellen, etwa bei der Bestimmung der Frequenz oder bei der Beschreibung der Plausibilität, nicht nachzuvollziehen.

9.3.4 Die Zusammensetzung der Referenzselektion hinsichtlich der Wortarten

Wir wollen die Referenzselektion nun hinsichtlich der Frage überprüfen, wie es sich mit dem Verhältnis von grammatischer gegenüber lexikalisierte Information in der Referenzselektion verhält. Wir wollen dabei insbesondere der Frage nachgehen, ob es tatsächlich so ist, dass ein Großteil der Gruppe zentraler Wörter eines Wortschatzes von Funktionswör-

301. Vgl. Abschnitt 4.5.2.1 und den in Abschnitt 5.5.1 beschriebenen Extensionstest im Zusammenhang mit der *coreness* eines Lexems bei Carter 1998.

tern repräsentiert wird. Eine Antwort darauf soll uns eine Übersicht über die quantitative Zusammensetzung der Referenzselektion hinsichtlich der darin vertretenen Wortarten und ihr Vergleich mit der entsprechenden Zusammensetzung im Ausgangswortschatz der 2294 Untersuchungseinheiten geben (vgl. Tabelle 9-20):

Tabelle 9-20: Die Zusammensetzung der Referenzselektion hinsichtlich der darin vertretenen Wortarten

	Anzahl LE in der Referenzselektion	Anzahl LE im Ausgangsbestand
Substantive	28%	72%
Adjektive	25%	14%
Verben	11%	5%
Adverbien	19%	3%
Konjunktionen	12%	1%
Präpositionen	4%	1%
Interjektionen	1%	1%
(ohne Zuordnung)	0%	3%

Das Bild, das sich bietet, widerspricht dieser Annahme und bestätigt sie zugleich auch: Es widerspricht ihr dahingehend, als die Referenzselektion durchaus nicht von Funktionswörtern dominiert wird, sondern von den Substantiven, die 28 % der Referenzselektion ausmachen. Insgesamt macht die Menge der lexikalischen Wörter (also der Substantive, Adjektive, Verben und Adverbien) 83 % der gesamten Referenzselektion aus, während die Funktionswörter nur 17 % dieser Menge stellen. Zugleich bestätigt das Bild die Annahme von der Dominanz der Funktionswörter insofern, als ihr relativer Anteil in der Referenzselektion von 17 % gegenüber ihrem Anteil an der Gesamtmenge, also im Ausgangswortschatz mit 2294 Einheiten, wo sie etwas mehr als 3 % ausmachen, erheblich größer ist und sie also in der Referenzselektion erheblich stärker ins Gewicht fallen.

9.3.5 Ein Blick auf die semantische Zusammensetzung der Referenzselektion

Abschließend wollen wir die quantitative Analyse der Referenzselektion um eine kurze semantischen Analyse ergänzen. Sieht man sich die in der Referenzselektion enthaltenen Substantive an, fällt auf, dass unter ihnen vergleichsweise viele (d.h. 25 % aller Substantive) Bezeichnungen für Körperteile sind.³⁰² Berücksichtigt man darüber hinaus, dass viele der anderen Substantive ebenfalls eine vergleichsweise konkrete Bedeutung haben,³⁰³ zeigt sich, dass die Substantive der Referenzselektion in der Mehrzahl den Konkreta zuzurechnen sind.

Ein Blick auf die Verben wiederum zeigt, dass hier ausschließlich transitive bzw. reflexive Verben, nicht aber intransitive Verben vertreten sind, während der Ausgangswortschatz 31 intransitive Verben enthält. Ob sich daraus ableiten lässt, dass Verben mit mehr als nur einer Valenzstelle zentraler sind als Verben mit nur einer Valenzstelle, lässt sich daraus je-

302. *Oberarm, Oberkörper, Oberlippe, Oberschenkel und Ohr.*

303. Vgl. etwa *Obst, Öffnung, Oktober, Öl, Onkel, Organ und Ortschaft.*

doch nicht so unmittelbar ableiten, denn dazu ist zum einen die Datengrundlage zu klein, zum anderen kann eine solche Verknüpfung von Valenz und Zentralität nur isoliert betrachtet gelten, denn es ist denkbar, dass intransitive Verben, die andere, für die Zentralität relevante Eigenschaften aufweisen, vor dem Hintergrund einer größeren Datengrundlage durchaus auch in einen zentralen Wortschatz mitaufgenommen werden könnten.³⁰⁴

Bei den Adjektiven schließlich wird deutlich, dass sie oft morphologische Varianten entsprechender Substantive oder Verben in der Referenzselektion darstellen (*offen – öffnen, ordentlich – Ordnung – ordnen*), so dass zu vermuten ist, dass bei der Beschreibung der introspektiven Eigenschaften der Repräsentiertheit und der Plausibilität Interferenzen, d.h. Übertragungen in der Beurteilung der jeweiligen Untersuchungseinheiten stattgefunden haben, die zu ihrer gleichen Behandlung geführt haben.

Darüber hinaus trifft natürlich auch auf die Referenzselektion das Problem der Lückenhaftigkeit zu, das in der Grundwortschatzliteratur so oft als Kritik angeführt wurde. Denn natürlich kann man sich fragen, weshalb die Liste zwar die lexikalische Einheit *Öl* im Sinne von ‘Erdöl’ enthält (DUW-ID 700.2.a), nicht aber *Öl* im Sinne von ‘Speiseöl’ (DUW-ID 700.2.e), oder weshalb sie *östlich* (DUW-ID 1560.1.-) enthält, nicht aber *Osten* (DUW-ID 1499.1.-). Ein Vergleich der jeweiligen Untersuchungseinheiten hinsichtlich der sechs hier verwendeten Selektionseigenschaften zeigt, dass im Falle von *Öl* im Sinne von ‘Speiseöl’ (DUW-ID 700.2.e) der Plausibilitätsgrad von 2 und im Falle von *Osten* (DUW-ID 1499.1.-) ein Stabilitätsgrad von 2 die Nichtaufnahme in die Selektion bewirkt hat. Beim Umgang mit so vielen Daten und der Quantifizierung qualitativer Kriterien, wie wir sie hier vorgenommen habe, sind solche Grenzfälle, die der falschen Seite zugeordnet werden, unumgänglich, weshalb Schnörch (2002, S. 361f.) nur zuzustimmen ist, dass eine objektive Vollständigkeit einer Selektion nicht erreichbar ist, aber durch die Transparenz des Vorgehens und der Aufbereitung des Materials kompensiert werden kann.

9.4 Methodische Beobachtungen bei der praktischen Untersuchung

Nachdem wir nun ein mögliches Ergebnis des in dieser Arbeit vorgestellten Verfahrens besprochen haben, soll es im Folgenden um das Verfahren selbst gehen. Denn ein Teil des Ziels dieser Arbeit ist ja gerade die Entwicklung eines Verfahrens, das eine transparente, im Detail justierbare Selektion von lexikalischen Einheiten für einen zentralen Wortschatz erlaubt.

304. Anders Buchbinder (1973, S. 306), der eine direkte Verknüpfung zwischen mehrstelliger Valenz und Zentralität annimmt.

9.4.1 Die Durchführbarkeit der Trennung von Beschreibung und Bewertung

Ein zentraler Punkt des hier entwickelten Verfahrens ist die Trennung von Beschreibung und Bewertung der Eigenschaften, die zur Selektion von Einheiten des Wortschatzes herangezogen werden sollen. Eine solche Trennung soll es erlauben, die individuellen Schritte der Beschreibung nachvollziehbar zu machen und vor allem, die individuellen Entscheidungen, die bei jedem Schritt für jedes zu beurteilende Wort anfallen, im Nachhinein justieren zu können. Dies war durch das transparente Vorgehen durchaus möglich, wie das Beispiel der Nachjustierung der Beurteilung von Untersuchungseinheiten, die den Wortarten Interjektion, Konjunktion und Präposition angehören, zeigt.³⁰⁵

Zwei Bereiche haben sich im Laufe der Entwicklung des Verfahrens allerdings als schwierig erwiesen: Zum einen war die konsistente Beschreibung der Untersuchungseinheiten hinsichtlich so 'weicher' Eigenschaften wie der Repräsentiertheit, der Plausibilität, aber auch der Stabilität über lange Zeiträume hinweg nicht immer einfach. Daher waren oft mehrere Iterationen nötig, um Konsistenz zu gewährleisten. Dabei war es nützlich, soweit wie möglich die jeweilige Eigenschaft in konkret fassbare Merkmale umzuwandeln, wie das etwa bei der Stabilität und ihrer Aufteilung in verschiedene Identitäten geschah (vgl. Abschnitt 8.6). Der zweite Problembereich besteht darin, dass eine wirklich messerscharfe Trennung zwischen Beschreibung und Gewichtung nicht möglich ist, weil man, sobald man versucht, qualitative Beschreibungen miteinander vergleichbar zu machen, gezwungen ist, diese zu 'kalibrieren', also zu vereinheitlichen. Diese Vereinheitlichung geschieht durch die Vergabe von Werten (im vorliegenden Fall von Graden der Zentralität hinsichtlich einer bestimmten Eigenschaft), wodurch bereits ein Moment der Gewichtung eintritt.

Die Schwierigkeiten der konsistenten Beschreibung und deutlichen Trennung von Beschreibung und Gewichtung sind allerdings nicht spezifisch für das hier vorgestellte Verfahren. Sie stellen vielmehr ein allgemeines Problem der Beschreibung und quantitativen Analyse von Sprachdaten dar, das besonders bei der Arbeit mit offenen Systemen ohne klare innere Grenzen, wie es der Wortschatz darstellt, zum Tragen kommt. Abschließend kann man daher wohl sagen, dass die Vorteile der Transparenz die Nachteile der im Einzelfall nicht konsistenten Zuordnung überwiegen.

9.4.2 Die lexikalische Einheit als Grundeinheit

Ein weiterer Kernpunkt des hier vorgestellten Verfahrens ist die Orientierung an der lexikalischen Einheit als Grundeinheit. Mit ihr knüpfen wir an die Forderung von Schnörch 2002 an, die Einheit von einer Bedeutung und einer Form (vgl. Abschnitt 4.2) zur Grundeinheit zu machen, weil nur sie eine konkrete Beschreibung, Beurteilung, Zählung und Selektion zulässt. In der Praxis bereitete diese strenge Forderung einige Schwierigkeiten, weil

305. Vgl. Fußnote 296.

die Rückführung einer so spezifischen Bedeutung, wie sie eine lexikalische Einheit hat, auf bestimmte Eigenschaften, die aus anderen Quellen als dem DUW gewonnen wurden, etwa der Frequenz oder den Bedeutungsangaben in den für die Beschreibung der Stabilität herangezogenen Wörterbüchern, oft nicht direkt möglich war, sondern abgeleitet werden musste. Im Falle der Frequenz lag der Grund darin, dass die herangezogenen Frequenzangaben sich auf das Lexem bzw. die Wortform als Grundeinheit bezogen, also auf einen viel größeren Bedeutungsumfang als die jeweils zu beschreibende lexikalische Einheit des DUW-Bestandes. Doch selbst wenn die herangezogenen Daten auf derselben Grundeinheit beruhen, wie das bei den für die Beschreibung der Stabilität herangezogenen Wörterbüchern der Fall war, war eine direkte Zuordnung der Bedeutung der lexikalischen Einheit zu einer der im Wörterbuch angegebenen Lesarten nicht immer möglich, so dass der Umweg der semantischen Ableitung nötig wurde (vgl. Abschnitt 8.6), ein Schritt, der wiederum ein Moment der Subjektivität in das Verfahren einführt.

Doch auch hier müssen die Nachteile des Verfahrens den Vorteilen gegenübergestellt werden: Der große Vorteil der lexikalischen Einheit als Grundeinheit ist ihre gegenüber dem Lexem, das ja durchaus polysem sein kann, klare semantische Identität, die es sehr viel einfacher macht, die sich aus dem Verfahren ergebende Selektion von lexikalischen Einheiten konkret zu erfassen und zu verwenden. Von all den bisher als möglich diskutierten Grundeinheiten, dem Wort ganz allgemein, der Wortform, dem Begriff, dem Lexem (vgl. Abschnitt 4.2) stellt die lexikalische Einheit diejenige Einheit dar, die noch am deutlichsten abzugrenzen ist, sowohl auf der Form- als auch auf der Bedeutungsseite. Damit stellt sie die 'beste aller möglichen Grundeinheiten' dar und sollte allen zentralen Wortschätzen, bei denen Semantik eine Rolle spielt, zugrundegelegt werden.

9.4.3 Die Beschreibung und Bewertung der introspektiven Eigenschaft der Plausibilität

Wie wir in Abschnitt 9.2 gesehen haben, hat die Eigenschaft der Plausibilität einen hohen Wirkungsgrad, d.h. sie hat eine stark selektierende Wirkung und steht in keiner Wechselbeziehung zu einer der anderen untersuchten Eigenschaften. Zugleich stellt sie diejenige Eigenschaft dar, deren Erhebung am wenigsten abgesichert ist, weil sie auf dem Urteil nur eines Sprechers beruht. Wir wollen deshalb festhalten, dass die Eigenschaft der Plausibilität grundsätzlich ein nützliches Selektionskriterium darstellt, dass sie aber nur in Verbindung mit dem Urteil mehrerer Sprecher erhoben werden sollte. Ähnliches gilt für die Eigenschaft der Repräsentiertheit.

9.4.4 Begrenzungen durch die Vorauswahl des Buchstaben O

Ein letzter Aspekt der Beurteilung des hier vorgestellten Verfahrens gilt nur für die konkret vorliegende Versuchsanordnung, nicht aber für das Verfahren an sich: Datengrundlage dieses Versuchs ist ein willkürlich begrenzter Ausschnitt eines Wörterbuchs. Diese willkürli-

che Grenzziehung brachte es mit sich, dass bestimmte Probleme bei der praktischen Analyse gar nicht zutage traten. Dazu gehört beispielsweise das Problem der Wahl zwischen Synonymen bzw. lokalen Konkurrenten (vgl. Abschnitt 4.4.4). Dadurch, dass bei der vorliegenden Untersuchung nur Wörter betrachtet wurden, die mit dem Buchstaben O beginnen, tauchten bestimmte synonymische Varianten zu den in der Untersuchungsmenge enthaltenen lexikalischen Einheiten wie etwa das Wort *Sahne* als nicht-österreichische Variante zu *Obers* (DUW-ID 163) in der Untersuchungsmenge gar nicht auf. Auch kann der durch die Wahl der Stichwörter unter dem Buchstaben O entstandene Wortschatzausschnitt insofern nicht als repräsentativ für den Gesamtwortschatz des DUW oder gar des Deutschen gehalten werden, als es beispielsweise keine oder kaum Präfixverben gibt, die mit dem Buchstaben O beginnen. Daher war es wichtig, die qualitative Auswertung der Referenzselektion (vgl. Abschnitt 9.3) so weit wie möglich nur im Verhältnis zum Ausgangswortschatz, also zu den aus den Stichwörtern zu O abgeleiteten lexikalischen Einheiten zu betrachten und nicht zum Gesamtbestand des DUW. Allein die Frage nach dem Umfang der Referenzselektion und seiner Hochrechnung auf den Gesamtbestand des DUW musste hier eine Ausnahme bilden. Doch spielen hier Fragen der detaillierten qualitativen Zusammensetzung der Referenzselektion und des Ausgangswortschatzes sowie seine Repräsentativität in Bezug auf den Gesamtbestand des DUW keine Rolle, weshalb eine solche Hochrechnung möglich war.

Trotz dieser Begrenzungen des hier vorgestellten Verfahrens und der Beschränkungen der Interpretation, die die konkrete Versuchsanordnung mit sich bringt, zeigen jedoch die Ergebnisse der Auswertung, dass das Verfahren an sich aussagekräftig genug ist, um den Aufwand, den es mit sich bringt, zu rechtfertigen.

Kapitel 10: Zusammenfassung und Ausblick

10.1 Erkenntnisse und Kernaussagen

Wir haben nun also in einer Bestandsaufnahme gesehen, dass das Konzept zentraler Wortschätze nicht nur in der Sprachdidaktik Anwendung findet, sondern auch in Anwendungen der maschinellen Sprachverarbeitung. Aufbauend auf dieser Bestandsaufnahme haben wir eine Typologie zentraler Wortschätze erstellt und diskutiert, welche Probleme mit der Arbeit an solchen Wortschätzen verbunden sind. Dabei wurde deutlich, dass neben den für ein jedes lexikographische Projekt typischen Problemen der Auswahl und Aufbereitung des Wortschatzmaterials vor allem Fragen der terminologischen Abgrenzung und Festlegung des Konzeptes zentraler Wortschätze und praktische Fragen des adäquaten Verfahrens der Erstellung eines solchen Wortschatzes eine gründliche, anwendungsunabhängige Untersuchung erfordern.

Das Problem der terminologischen Abgrenzung haben wir bearbeitet, indem wir eine Definition des Konzeptes zentraler Wortschätze vorgelegt und ein entsprechendes Arbeitsmodell des Wortschatzes entworfen haben (vgl. Kapitel 6).

Den praktischen Fragen nach einem adäquaten Verfahren der Erstellung zentraler Wortschätze sind wir in der praktischen Analyse der Einträge des Buchstaben O des DUW nachgegangen (vgl. Kapitel 7 bis 9). Ziel war es dabei, einen Vorschlag für ein transparentes Selektionsverfahren zu entwickeln und der Frage nachzugehen, welchen Wirkungsgrad die einzelnen Selektionskriterien haben, ob sie in einer Wechselbeziehung zueinander stehen und welchen Einfluss sie auf den Umfang eines zentralen Wortschatzes haben.

Dabei haben wir uns auf die folgenden sechs Eigenschaften lexikalischer Einheiten konzentriert:

- ihre Repräsentiertheit,
- ihre Worthaftigkeit,
- ihre Plausibilität,
- ihre Frequenz,
- ihre Stabilität und
- ihre Neutralität.

Zunächst haben wir dann jede Einheit der aus dem DUW-Bestand gewonnenen Untersuchungsmenge hinsichtlich jeder dieser sechs Eigenschaften beschrieben. Erst in einem nächsten Schritt haben wir die jeweiligen Eigenschaften bewertet und in ein quantitatives System von Zentralitätsgraden überführt und die sich daraus ergebenden Schichtungsmöglichkeiten der Untersuchungsmenge analysiert.

Dabei hat sich gezeigt, dass vor dem Hintergrund der hier verwendeten Daten die Eigenschaften der Stabilität, der Neutralität und der Plausibilität den höchsten Wirkungsgrad hinsichtlich der Selektion haben. Umgekehrt sind die Eigenschaften der Repräsentiertheit und

der Worthaftigkeit kaum für eine Selektion zu einem zentralen Wortschatz geeignet, wenn die Datengrundlage aus dem Wortbestand eines Wörterbuchs besteht, weil ihr Selektionseffekt gering ist.

Was die Wechselwirkung der Selektionskriterien untereinander betrifft, so zeigen die Untersuchungsergebnisse, dass die Eigenschaft der Plausibilität Selektionsergebnisse liefert, die unabhängig von denen anderer Eigenschaften sind; ähnlich unabhängig scheinen auch die Eigenschaften der Stabilität und der Neutralität zu wirken.

Zusätzlich zur Untersuchung der jeweiligen Selektionseigenschaften haben wir eine Referenzselektion erstellt und an ihr Fragen des Umfangs und des Verhältnisses von Extension und Zentralität eines Lexems diskutiert. Dabei hat sich gezeigt, dass eine Selektion, die auf den hier verwendeten Selektionseigenschaften aufbaut, zu einem Umfang eines zentralen Wortschatzes führen würde, der mit ca. 6 000 lexikalischen Einheiten etwas größer ist als der, der in der Literatur etwa für Grundwortschätze angegeben wird. Während die Zahlen in der Literatur jedoch mit dem Wort als Grundeinheit operieren, beruht die hier vorgenommene Umfangsberechnung auf der (im Verhältnis zum Wort) kleineren lexikalischen Einheit, was den quantitativen Unterschied zum Teil erklären kann.

Von einem methodischen Standpunkt aus gesehen hat es sich dabei als wichtig und nützlich herausgestellt, zwischen der Beschreibung, Bewertung und Gewichtung von Eigenschaften eine klare Trennung zu vollziehen. Erst diese Trennung lässt ein kontrolliertes Experimentieren mit dem Einfluss der verschiedenen Eigenschaften auf das Ergebnis zu.

Des Weiteren war es wichtig, bei der Beschreibung der zu schichtenden Einheiten zunächst alle Einheiten zu berücksichtigen, nicht nur die klassischen lexikalischen Einheiten. Dadurch lässt sich eine vorgezogene oder versteckte Selektion vermeiden, und der Grad der Worthaftigkeit selbst kann als selektierende Eigenschaft in den Selektionsprozess miteinbezogen werden. Außerdem unterstützt ein solches Vorgehen die Transparenz des Verfahrens, weil die Beschreibung stärker von der Bewertung und Gewichtung getrennt wird. Gegen ein solches Vorgehen spricht allerdings, dass die Berücksichtigung wenig worthafter Einheiten zunächst kontra-intuitiv wirken kann. Zudem wird es dadurch notwendig, beim Vergleich von Angaben zum Umfang von zentralen Wortschätzen zu überprüfen, nach welchen Verfahren diese jeweils erstellt wurden. Dennoch zählt, zumindest im vorliegenden Fall, bei dem der Bestand eines Wörterbuches Ausgangsbasis war, die Transparenz des Verfahrens mehr als die mangelnde Vergleichbarkeit.

Als eine Schwäche des hier vorgestellten Verfahrens könnte man anmerken, dass die Bedeutungsunterschiede, die im DUW vorgenommen werden, ohne weitere Beurteilung übernommen wurden. Dadurch entstanden, wie sich bei der quantitativen Auswertung (vgl. Abschnitt 9.3.2) gezeigt hat, möglicherweise unnötig viele lexikalische Einheiten. Tatsächlich waren einige der Bedeutungsunterscheidungen insofern überflüssig, als dass sie in anderen Quellen der Beschreibung der Untersuchungseinheiten, etwa bei der Frequenz oder

bei der Stabilität, nicht reflektiert wurden. Andererseits ließ es der Rahmen dieser Arbeit nicht zu, Fragen der Bedeutungsabgrenzung im Detail zu problematisieren. Dies wäre eine wünschenswerte Ergänzung der weiteren Forschung zum Konzept zentraler Wortschätze.

Die Diskussionen zu lexikologischen Aspekten der Wortschatzschichtung wiederum haben gezeigt, dass es wichtig ist, zwischen dem Konzept zentraler Wortschätze einerseits und dem lexikologischen Wortschatzzentrum einer Sprache andererseits zu unterscheiden (vgl. die Diskussion in Abschnitt 5.5). Das Konzept zentraler Wortschätze bildet die methodische Grundlage für konkrete anwendungsorientierte Wortschatzausschnitte; das Konzept vom lexikologischen Wortschatzzentrum ist ein theoretisches Konstrukt, eine Abstraktion, die aus der Beobachtung des Systems gewonnen wird und die allenfalls als linguistische Annotierung in Wörterbüchern seinen Niederschlag finden kann. Beide Konzepte haben gewisse Berührungspunkte, man darf aber nicht automatisch davon ausgehen, dass sie deckungsgleich sind.

10.2 Weiterführendes

Mit der hier vorgenommenen Bestandsaufnahme, Weiterführung und praktischen Analyse ist, so hoffe ich, deutlich geworden, dass das Konzept zentraler Wortschätze ein lexikographisches Konzept ist, das in vielen Anwendungsbereichen zum Einsatz kommt. Die Erstellung eines solchen Wortschatzes ist ein komplexer lexikographischer Vorgang, der jedoch, wenn man darauf achtet, ihn transparent und revidierbar zu gestalten, zu begründeten und damit guten Ergebnissen führt. Dennoch ist damit erst ein Schritt in Richtung einer Neuorientierung getan. Verschiedene Fragen müssen noch beantwortet werden, sowohl in lexikographisch-praktischer als auch in lexikologisch-theoretischer Hinsicht.

Zum Beispiel verdient die Frage nach dem Verhältnis von lexikologischem Zentrum und anwendungsspezifischem zentralem Wortschatz (vgl. Abschnitt 5.5) noch eine gründlichere Untersuchung: Wie groß ist der Grad an Gemeinsamkeiten zwischen lexikologischem Zentrum und dem jeweiligen konkreten zentralen Wortschatz? Sind bestimmte Typen zentraler Wortschätze dem lexikologischen Zentrum näher als andere? Wie lassen sich die Tests auf *coreness*, wie sie Carter 1998 vorstellt, lexikographisch, d.h. für die konkrete Arbeit mit zentralen Wortschätzen nutzbar machen? Und in welcher Wechselwirkung stehen diese Tests zueinander, können sie einander beispielsweise ersetzen?

Von einem methodischen Standpunkt aus wiederum wäre es interessant, das hier vorgestellte Verfahren der Beschreibung und Bewertung der sechs Eigenschaften der Repräsentiertheit, der Worthaftigkeit, der Plausibilität, der Frequenz, der Stabilität und der Neutralität auf eine aus einem Textkorpus gewonnene Datengrundlage anzuwenden. Wie würden dabei etwa die Wechselwirkung und der Wirkungsgrad der Eigenschaften der Repräsentiertheit und der Worthaftigkeit aussehen? Und wie würde vor dem Hintergrund einer solchen Datengrundlage ein Vergleich zweier Selektionen aussehen, die nur mit den beiden introspektiven Eigenschaften der Repräsentiertheit und Plausibilität bzw. nur mit den Eigen-

schaften der Frequenz, Worthaftigkeit, Stabilität und Neutralität arbeiten? Würden sie zu sehr unterschiedlichen Ergebnissen führen?

Aber auch wenn man von der hier verwendeten Datengrundlage, also dem Bestand eines Wörterbuches, ausgeht, ist es nützlich, einigen Fragen noch genauer nachzugehen. So kann etwa das Verfahren zur Bestimmung der Stabilität, das sich in der vorliegenden Untersuchung auf den nicht ganz unproblematischen Fall des Vergleichs zweier Wörterbücher stützt, durch eine entsprechende Korpusanalyse ersetzt und die entsprechenden Ergebnisse einem Vergleich mit einer durch einen Wörterbuchvergleich erhobenen Stabilität unterzogen werden.

Schließlich wäre es wünschenswert, das hier vorgestellte Verfahren auf Daten aus einem Textkorpus anzuwenden und um eine gründliche und quantitativ breite Sprecherbefragung zu ergänzen, indem man etwa anhand der Eigenschaften der Frequenz eine Vorauswahl an lexikalischen Einheiten trifft und diese dann von einer breiten Sprechergruppe hinsichtlich der Eigenschaften der Repräsentiertheit und der Plausibilität beschreiben lässt. So ließe sich noch genauer die Rolle dieser beiden Eigenschaften bei der Selektion absichern.

Die Liste der mit dem Konzept zentraler Wortschätze verbundenen offenen Fragen ist also nach wie vor lang. Dass dieses Konzept dennoch eine wichtige Rolle in der angewandten Linguistik spielt, viele wesentliche Aspekte linguistischer Forschung in interessanter Weise verknüpft und zu sehr konkreten Ergebnissen führt, hoffe ich, mit dieser Arbeit gezeigt zu haben.

Anhang: Tabellen und Verzeichnisse

A: Tabellenauszüge aus der zugrundeliegenden Datenbank

Tabelle A-1: Beispielergebnis Arbeitsschritt 1 (Erfassung der DUW-Stichwörter)

DUW-ID	DUW-Stichwort
...	...
29.-.-	Obdachlose
...	...
44.-.-	obenauf
...	...
86.-.-	Oberdeutsch, Oberdeutsche
...	...
283.-.-	Obstanbau, Obstbau
...	...
334.-.-	och
...	...
759.-.-	Olim
...	...
860.-.-	Omnien
...	...
1059.-.-	opportun
...	...
1117.-.-	Oral History, Oralhistory
...	...
1375.-.-	2Ort
...	...
1396.-.-	Orthographie, Orthografie
...	...
1478.-.-	Ösophagus; Oesophagus
...	...
1660.-.-	Oxidation, Oxydation
...	...
1672.-.-	o. Z., O. Z.
...	...

Die vollständige Tabelle findet sich in der der gedruckten Ausgabe beigelegten Datei *DUW_Bestand.xls*.

Tabelle A-2: Beispielergebnis Arbeitsschritt 2 (Beschreibung der Repräsentiertheit der DUW-Stichwörter)

DUW-ID	DUW-Stichwort	Beschreibung der Repräsentiertheit
2.-.-	ö, Ö	Symbol oder Einzelbuchstabe ohne Möglichkeit einer konkreten Bedeutungszuordnung
...
7.-.-	Ω	Symbol oder Einzelbuchstabe ohne Möglichkeit einer konkreten Bedeutungszuordnung
...
13.-.-	ÖAMTC	Abkürzung ohne Möglichkeit der Angabe der Vollform, aber mit Bedeutungsangabe
...
16.-.-	OAU	Abkürzung ohne Möglichkeit, eine Vollform oder Gesamtbedeutung zuordnen zu können
...
17.-.-	OAU-Staaten	Kompositum, bei dem nur das Grundglied repräsentiert ist, nicht aber die Gesamtbedeutung
...
25.-.-	ÖBB	voll repräsentiert
...
29.-.-	Obdachlose	voll repräsentiert
...
208.-.-	Obi	Wort weder in Teilen noch als Ganzes repräsentiert
...
212.-.-	Obiter Dictum	Herkunftssprache zwar erkennbar, nicht aber die Gesamtbedeutung
...
249.-.-	oblong	Wort weder in Teilen noch als Ganzes repräsentiert
...
310.-.-	Obstruktionspolitik	Kompositum, bei dem nur das Grundglied repräsentiert ist, nicht aber die Gesamtbedeutung
...
311.-.-	obstruktiv	Wortbildungskonstruktion zwar erkennbar, nicht aber die Gesamtbedeutung
...
481.-.-	1Office	voll repräsentiert
...
880.-.-	1Onkel	voll repräsentiert
...
955.-.-	Open Shop	Herkunftssprache zwar erkennbar, nicht aber die Gesamtbedeutung
...
1037.-.-	Ophthalmologie	Wortbildungskonstruktion zwar erkennbar, nicht aber die Gesamtbedeutung
...
1173.-.-	Ordal	Wort weder in Teilen noch als Ganzes repräsentiert
...
1396.-.-	Orthographie, Orthografie	voll repräsentiert
...
1417.-.-	örtlich	voll repräsentiert
...
1461.-.-	OSB, O.S.B	Abkürzung ohne Möglichkeit, eine Vollform oder Gesamtbedeutung zuordnen zu können
...
1569.-.-	Ostrazismus	Wortbildungskonstruktion zwar erkennbar, nicht aber die Gesamtbedeutung

Die vollständige Tabelle findet sich in der der gedruckten Ausgabe beigelegten Datei *Beschreibung_Repräsentiertheit.xls*.

Tabelle A-3: Beispielergebnis Arbeitsschritt 3 (Detailerfassung)

DUW-ID	Form der LE	Bedeutung der LE	Grammatische Informationen	Sondermarkierung 1	Sondermarkierung 2	Sondermarkierung 3	Sonstige Besonderheiten
...
29.-.-	Obdachlose	{jemand (eine Frau), die obdachlos ist}	S f				
29.-.-	Obdachlose	{jemand (ein Mann), der obdachlos ist}	S m				
...
44.1.-	obenauf	obendrauf	Adv	landsch.			
44.2.a	obenauf	gesund, guter Laune	Adv				
44.2.b	obenauf	sich seiner Stärke bewusst	Adv				
...
86.-.-	Oberdeutsch	die oberdeutsche Sprache; mit bestimmtem Artikel: das Oberdeutsche	S n	Sprachw.			Mit bestimmtem Artikel: Oberdeutsche
86.-.-	das Oberdeutsche	die oberdeutsche Sprache	S n	Sprachw.			
...
283.-.-	Obstbau	Anbau von Obst	S m				
283.-.-	Obstanbau	Anbau von Obst	S m				
...
334.-.-	och	ach	Part	ugs.			
334.-.-	och	ach	Interj	ugs.			
...
759.-.-	zu Olims Zeiten	vor sehr langer Zeit	Adv	bildungsspr.	scherzh.		Phraseologismus
...
860.-.-	Omnien	Pl. v. Omnium					
...
1059.-.-	opportun	in der gegebenen Situation angebracht, von Vorteil	Adj	bildungsspr.			
...
1117.-.-	Oral History	Geschichte, Geschichtswissenschaft, die sich mit der Befragung lebender Zeugen befasst	S f				
1117.-.-	Oralhistory	Geschichte, Geschichtswissenschaft, die sich mit der Befragung lebender Zeugen befasst	S f				
...
1375.-.-	2Ort	Ahle, Pfriem	S m	veraltet			
1375.-.-	2Ort	Ahle, Pfriem	S n				

Tabelle A-4: Beispielergebnis Arbeitsschritt 3 (Vorfilterung)

DUW-ID	Form der LE	Bedeutung der LE	Gramm. Angaben	Weiter bearbeiten?	Kommentar für oder gegen eine weitere Bearbeitung
29.-.-	Obdachlose	{jemand (eine Frau), die obdachlos ist}	S f	ja	
29.-.-	Obdachlose	{jemand (ein Mann), der obdachlos ist}	S m	ja	
...
44.1.-	obenauf	obendrauf	Adv	ja	
44.2.a	obenauf	gesund, guter Laune	Adv	ja	
44.2.b	obenauf	sich seiner Stärke bewusst	Adv	ja	
...
86.-.-	Oberdeutsch	die oberdeutsche Sprache; mit bestimmtem Artikel: das Oberdeutsche	S n	ja	
86.-.-	das Oberdeutsche	die oberdeutsche Sprache		nein	regelhafte Kongruenz, je nach Verwendung des bestimmten Artikels
...
283.-.-	Obstbau	Anbau von Obst	S m	ja	idiosynkratische Alternativen, bei denen die dominierende Form erst noch bestimmt werden muss
283.-.-	Obstanbau	Anbau von Obst	S m	ja	idiosynkratische Alternativen, bei denen die dominierende Form erst noch bestimmt werden muss
...
334.-.-	och	ach	Part	ja	Zugehörigkeit zu unterschiedlichen Wortarten
334.-.-	och	ach	Interj	ja	Zugehörigkeit zu unterschiedlichen Wortarten
759.-.-	zu Olims Zeiten	vor sehr langer Zeit	Adv	ja	
860.-.-	Omnien	Pl. v. Omnium		nein	keine eigenständige Ws-Einheit, da nur Teil eines Paradigmas

Die der gedruckten Ausgabe beiliegende Datei *Nicht_weiter_bearbeitete_LE.xls* enthält alle 63 von der weiteren Bearbeitung ausgeschlossenen Einheiten.

Tabelle A-5: Beispiele zur Plausibilitätsbeschreibung

DUW-ID	Form der LE	Bedeutung der LE	Grammatische Informationen	Beschreibung der Plausibilität
...
29.-.-	Obdachlose	{jemand (ein Mann), der obdachlos ist}	S m	relativ plausibel
29.-.-	Obdachlose	{jemand (eine Frau), die obdachlos ist}	S f	relativ plausibel
...
44.1.-	obenauf	obendrauf	Adv	relativ plausibel
44.2.a	obenauf	gesund, guter Laune	Adv	relativ plausibel
44.2.b	obenauf	sich seiner Stärke bewusst	Adv	relativ plausibel
...
86.-.-	Oberdeutsch	die oberdeutsche Sprache; mit bestimmtem Artikel: das Oberdeutsche	S n	wenig plausibel
...
283.-.-	Obstbau	Anbau von Obst	S m	wenig plausibel
283.-.-	Obstanbau	Anbau von Obst	S m	relativ plausibel
...
334.-.-	och	ach	Interj	sehr plausibel
334.-.-	och	ach	Part	sehr plausibel
...
759.-.-	zu Olims Zeiten	vor sehr langer Zeit	Adv	wenig plausibel
...
1059.-.-	opportun	in der gegebenen Situation angebracht, von Vorteil	Adj	relativ plausibel
...
1117.-.-	Oralhistory	Geschichte, Geschichtswissenschaft, die sich mit der Befragung lebender Zeugen befasst	S f	wenig plausibel
...
1375.-.-	2Ort	Ahle, Pfriem	S n	wenig plausibel
1375.-.-	2Ort	Ahle, Pfriem	S m	wenig plausibel
...
1396.-.-	Orthografie	Rechtschreibung	S f	relativ plausibel
...
1478.-.-	Oesophagus	Speiseröhre, auch als Ösophagus geschrieben	S m	wenig plausibel
...
1660.1.-	Oxydation	Reaktion, Verbindung eines chemischen Elements od. einer chemischen Verbindung mit Sauerstoff	S f	wenig plausibel
1660.2.-	Oxydation	Vorgang, bei dem ein chemisches Element od. eine chemische Verbindung Elektronen abgibt, die von einer anderen Substanz aufgenommen werden	S f	wenig plausibel
...
1672.-.-	o. Z.	ohne Zensur	Adv	wenig plausibel
1672.-.-	o. Z.	ohne Zeichnung	Adv	wenig plausibel
...

Die der gedruckten Ausgabe beiliegende Datei *Beschreibung_Plausibilität.xls* gibt alle untersuchten Untersuchungseinheiten zusammen mit der Beschreibung ihrer Plausibilität an.

Tabelle A-6: Beispiele zur Frequenzbeschreibung

DUW-ID	Form der LE	Grammatische Informationen	Bedeutung der LE	Frequenzklasse	Kommentar zur Frequenzangabe
29.-.-	Obdachlose	S m	{jemand (ein Mann), der obdachlos ist}	12	Unterscheidet nicht zwischen männlicher und weiblicher Form
29.-.-	Obdachlose	S f	{jemand (eine Frau), die obdachlos ist}	12	Unterscheidet nicht zwischen männlicher und weiblicher Form
...
44.1.-	obenauf	Adv	obendrauf	14	Unterscheidet nicht zwischen den verschiedenen Bedeutungen
44.2.a	obenauf	Adv	gesund, guter Laune	14	Unterscheidet nicht zwischen den verschiedenen Bedeutungen
44.2.b	obenauf	Adv	sich seiner Stärke bewusst	14	Unterscheidet nicht zwischen den verschiedenen Bedeutungen
...
86.-.-	Oberdeutsch	S n	die oberdeutsche Sprache; mit bestimmtem Artikel: das Oberdeutsche	25	Ergibt keinen Treffer
...
283.-.-	Obstbau	S m	Anbau von Obst	17	
283.-.-	Obstanbau	S m	Anbau von Obst	18	
...
334.-.-	och	Interj	ach	16	Unterscheidet nicht zwischen Interj. und Partikel
334.-.-	och	Part	ach	16	Unterscheidet nicht zwischen Interj. und Partikel
...
759.-.-	zu Olims Zeiten	Adv	vor sehr langer Zeit	21	
...
1059.-.-	opportun	Adj	in der gegebenen Situation angebracht, von Vorteil	15	
...
1117.-.-	Oralhistory	S f	Geschichte, Geschichtswissenschaft, die sich mit der Befragung lebender Zeugen befasst	23	
...
1375.-.-	2Ort	S n	Ahle, Pfriem	8	Unterscheidet nicht zwischen verschiedenen Bedeutungen oder Genera
1375.-.-	2Ort	S m	Ahle, Pfriem	8	Unterscheidet nicht zwischen verschiedenen Bedeutungen oder Genera
...
1396.-.-	Orthografie	S f	Rechtschreibung	17	
1478.-.-	Oesophagus	S m	Speiseröhre, auch als Ösophagus geschrieben	23	
...

Die der gedruckten Ausgabe beiliegende Datei *Beschreibung_Frequenz.xls* enthält die Frequenzklassenangaben und Kommentare zu allen untersuchten Einheiten.

Tabelle A-7: Beispiele zur Stabilitätsbeschreibung

DUW-ID	Form der LE	Grammatische Informationen	Bedeutung der LE	LE als Stichwort in WH vorhanden?	Grammatische Form in WH identisch?	Graphische Form in WH identisch?	Bedeutung ableitbar aus WH?	LE als Stichwort in MUS vorhanden?	Graphische Form in MUS identisch?	Grammatische Form in MUS identisch?	Bedeutung ableitbar aus MUS?
...
29.-.-	Obdachlose	S m	{jemand (ein Mann), der obdachlos ist}	ja	nein	nein	nicht ableitbar	ja	ja	ja	ableitbar
29.-.-	Obdachlose	S f	{jemand (eine Frau), die obdachlos ist}	ja	nein	nein	nicht ableitbar	ja	ja	nein	nicht ableitbar
...
44.1.-	obenauf	Adv	obendrauf	ja	ja	ja	ableitbar	ja	ja	ja	ableitbar
44.2.a	obenauf	Adv	gesund, guter Laune	ja	ja	ja	ableitbar	ja	ja	ja	ableitbar
44.2.b	obenauf	Adv	sich seiner Stärke bewusst	ja	ja	ja	ableitbar	ja	ja	ja	ableitbar
...
86.-.-	Oberdeutsch	S n	die oberdeutsche Sprache; mit bestimmtem Artikel: das Oberdeutsche	ja	ja	ja	direkt	ja	ja	ja	direkt
...
283.-.-	Obstbau	S m	Anbau von Obst	ja	ja	ja	direkt	ja	ja	ja	direkt
283.-.-	Obstanbau	S m	Anbau von Obst	nein	nein	nein	nicht als WB-Eintrag vorhanden	nein	nein	nein	nicht als WB-Eintrag vorhanden
...

Die vollständige Liste aller Untersuchungseinheiten beschrieben nach ihrer Stabilität findet sich in der der gedruckten Ausgabe beigelegten Datei *Beschreibung_Stabilität.xls*.

Tabelle A-8: Beispiele zur Neutralitätsbeschreibung

DUW-ID	Form der LE	Gramm. Informationen	Bedeutung der LE	Sondermarkierung 1	Sondermarkierung 2	Sondermarkierung 3
...
29.-.-	Obdachlose	S m	{jemand (ein Mann), der obdachlos ist}			
29.-.-	Obdachlose	S f	{jemand (eine Frau), die obdachlos ist}			
...
44.1.-	obenauf	Adv	obendrauf	landsch.		
44.2.a	obenauf	Adv	gesund, guter Laune			
44.2.b	obenauf	Adv	sich seiner Stärke bewusst			
...
86.-.-	Oberdeutsch	S n	die oberdeutsche Sprache; mit bestimmtem Artikel: das Oberdeutsche	Sprachw.		
...
283.-.-	Obstbau	S m	Anbau von Obst			
283.-.-	Obstanbau	S m	Anbau von Obst			
...
334.-.-	och	Interj	ach	ugs.		
334.-.-	och	Part	ach	ugs.		
...
759.-.-	zu Olims Zeiten	Adv	vor sehr langer Zeit	bildungsspr.	scherzh.	
...
1059.-.-	opportun	Adj	in der gegebenen Situation angebracht, von Vorteil	bildungsspr.		
...
1117.-.-	Oralhistory	S f	Geschichte, Geschichtswissenschaft, die sich mit der Befragung lebender Zeugen befasst			
...
1375.-.-	2Ort	S n	Ahle, Pfriem			
1375.-.-	2Ort	S m	Ahle, Pfriem	veraltet		
...
1396.-.-	Orthografie	S f	Rechtschreibung			
...
1478.-.-	Oesophagus	S m	Speiseröhre, auch als Ösophagus geschrieben	Anat.		
...
1660.1.-	Oxydation	S f	Reaktion, Verbindung eines chemischen Elements od. einer chemischen Verbindung mit Sauerstoff			
1660.2.-	Oxydation	S f	Vorgang, bei dem ein chemisches Element od. eine chemische Verbindung Elektronen abgibt, die von einer anderen Substanz aufgenommen werden	Chemie	Physik	
...
1672.-.-	o. Z.	Adv	ohne Zensur			
1672.-.-	o. Z.	Adv	ohne Zeichnung			
...

Die der gedruckten Ausgabe beiliegende Datei *Beschreibung_Neutralität.xls* enthält alle Untersuchungseinheiten, die eine Sondermarkierung besitzen.

Tabelle A-9: Übersicht über die im DUW-Material verwendeten Sondermarkierungen

DUW-Sondermarkierung	Anzahl der damit markierten LE	DUW-Sondermarkierung	Anzahl der damit markierten LE
abwertend	23	Kochk.	3
Amtsspr.	6	Kunst	6
Anat.	10	landsch.	21
Archit.	1	Landw.	4
Astrol.	1	marx.	3
Astron.	2	Math.	9
Ballspiele	4	Med.	92
Bankw.	2	Met.	5
Bauw.	2	Milit.	15
Bergmannsspr.	3	Musik	17
bildungsspr.	77	Optik	3
Biol.	25	Pharm.	3
Bot.	5	Philos.	7
BRD	2	Physik	18
Börsenw.	4	Politik	4
Chemie	25	Psych.	6
DDR	8	Päd.	1
derb	1	Raumf.	3
dichter.	2	Rechtsspr.	15
EDV	14	regional	2
Eisenb.	1	Rel.	3
Elektronik	2	Rundf.	1
Elektrot.	3	röm. Myth.	1
emotional	4	salopp	8
ev. Kirche	4	scherzh.	18
Fachspr.	19	Schimpfwort	1
fam.	5	schweiz.	20
Ferns.	6	Seemannsspr.	8
Fernspr.	2	Seew	2
Film	5	selten	11
Flugw.	3	Sexualk.	3
Forstw.	1	Soziol.	2
Fot.	3	Sport	18
früher	19	Sprachw.	35
Fußball	3	spött.	3
Gartenbau	1	südd.	3
geh.	17	Technik	6
Geogr.	6	Theater	5
Geol.	13	Theol.	4
Geom.	3	ugs.	76
griech. Myth.	1	veraltend	17
hist.	1	veraltet	23
iron.	1	verhüll.	3
Jargon	13	Versicherungsw.	1
Judo	2	verstärkend	3
Jugendspr.	2	Winzerspr.	3
Jägerspr.	1	Wirtsch.	17
kath. Kirche	27	übertreibend	1
Kaufmannsspr.	10	Zool.	7
Kfz-T.	6	österr.	8
Kinderspr.	3		

Tabelle A-10: Einteilung der Sondermarkierungen in Gruppen und ihre Bewertung hinsichtlich der Neutralität

fachsprachlich		emotional konnotiert	regional	sozial	lexikographisch
Anat.	Kfz-T.	abwertend	BRD	Amtsspr.	früher
Archit.	Kochk.	bildungsspr.	DDR	fam.	selten
Astrol.	Kunst	derb	landsch.	Jargon	veraltend
Astron.	Landw.	dichter.	regional	Jugendspr.	veraltet
Ballspiele	marx.	emotional	schweiz.	Kinderspr.	
Bankw.	Math.	geh.	südd.	Seemannsspr.	
Bauw.	Med.	iron.	österr.	ugs.	
Bergmannsspr.	Met.	salopp			
Biol.	Milit.	scherzh.			
Bot.	Musik	Schimpfwort			
Börsenw.	Optik	spött.			
Chemie	Pharm.	verhüll.			
EDV	Philos.	verstärkend			
Eisenb.	Physik	übertreibend			
Elektronik	Politik				
Elektrot.	Psych.				
ev. Kirche	Päd.				
Fachspr.	Raumf.				
Ferns.	Rechtsspr.				
Fernspr.	Rel.				
Film	Rundf.				
Flugw.	röm. Myth.				
Forstw.	Seew				
Fot.	Sexualk.				
Fußball	Soziol.				
Gartenbau	Sport				
Geogr.	Sprachw.				
Geol.	Technik				
Geom.	Theater				
griech. Myth.	Theol.				
hist.	Versicherungsw.				
Judo	Winzerspr.				
Jägerspr.	Wirtsch.				
kath. Kirche	Zool.				
Kaufmannsspr.					

B: Literaturverzeichnis

[Albrecht 1992]

Albrecht, Jörn: *Wortschatz versus Terminologie: Einzelsprachliche Charakteristika in der Fachterminologie*. In: Albrecht, Jörn; Baum, Richard (Hg.): *Fachsprache und Terminologie in Geschichte und Gegenwart*. Tübingen 1992, S. 59–78. (= Forum für Fachsprachen-Forschung 14).

[Andreev 1980]

Andreev, Nikolaj D.: *Verteilungswörterbücher und Grundwortschatz*. In: Grothjahn, R. (Hg.): *Glottometrika 2*. Bochum 1980, S. 82–88. (= Quantitative Linguistics 2).

[Asch 1960]

Asch, Solomon; Nerlove, Harriet: *The Development of Double Function Terms in Children*. In: Kaplan, Bernard; Wapner, Seymour (Hg.): *Perspectives in Psychological Theory*. New York 1960, S. 47–60.

[Augst 1983]

Augst, Gerhard: *Rechtschreibgrundwortschatz – Ja oder nein? Überlegungen aus der Sicht der Linguistik*. In: *Deutsche Sprache* 11 (1983), S. 341–356.

[Augst 1985]

Augst, Gerhard: *'Kinderwort' – Der aktive Kinderwortschatz (kurz vor der Einschulung) nach Sachgebieten geordnet und mit einem alphabetischen Register*. 2. Aufl. Frankfurt a. M. 1985. (= Theorie und Vermittlung der Sprache 1).

[Augst 1987]

Augst, Gerhard: *Ist die degressive Struktur des Wortgebrauchs ein Argument für den Rechtschreibgrundwortschatz (RGW)?* In: Wagner, Klaus R. (Hg.): *Wortschatz-Erwerb*. Bern u.a. 1987, S. 115–127. (= Arbeiten zur Sprachanalyse 6).

[Augst 1988]

Augst, Gerhard: *Das Fremdwort – Ein Scheinphänomen?* In: *Sprachreport* 2 (1988), S. 4–5.

[Augst 1997]

Augst, Gerhard: *Wort – Wortfamilie – Wortfamilienwörterbuch*. In: Berens, Franz Josef; Wimmer, R. (Hg.): *Wortbildung und Phraseologie*. Tübingen 1997, S. 89–113.

[Bakonyi 1934]

Bakonyi, Hugo: *Die gebräuchlichsten Wörter der deutschen Sprache für den Fremdsprachenunterricht stufenmäßig zusammengestellt*. München 1934.

[Baldegger (Hg.) 1980]

Baldegger, Markus u. a. (Hg.): *Kontaktschwelle Deutsch als Fremdsprache*. Berlin 1980.

[Baldinger 1960]

Baldinger, Kurt: *Alphabetisches oder begrifflich geordnetes Wörterbuch?* In: *Zeitschrift für romanische Philologie* 76 (1960), 521–536.

[Barz 1998]

Barz, Irmhild: *Neologie und Wortbildung. Zum Neuheitseffekt von Wortneubildungen*. In: Teubert, Wolfgang (Hg.): *Neologie und Korpus*. Tübingen 1998, S. 11–30. (= Studien zur deutschen Sprache 11).

[Bauer 2005]

Bauer, Ulrich: *[Rezension zu:] Schnörch, Ulrich: Der zentrale Wortschatz des Deutschen. Strategien zu seiner Ermittlung, Analyse und lexikografischen Aufarbeitung*. Tübingen: Narr, 2002. In: *Informationen Deutsch als Fremdsprache* 32 (2005), S. 247–251.

[Beauchemin / Martel / Théoret 1992]

Beauchemin, Normand; Martel, Pierre; Théoret, Michel: *Dictionnaire de fréquence des mots du français parlé au Québec. Fréquence, dispersion, usage, écart, réduit*. New York u. a. 1992.

[Belica 1998]

Belica, Cyril: *Statistische Analyse von Zeitstrukturen in Korpora*. In: Teubert, Wolfgang (Hg.): *Neologie und Korpus*. Tübingen 1998, S. 31–42. (= Studien zur deutschen Sprache 11).

[Bergenholtz / Pedersen 1994]

Bergenholtz, Henning; Pedersen, Jette: *Zusammensetzung von Textkorpora für die Fachlexikographie*. In: Schaefer, Burkhard; Bergenholtz, Henning (Hg.): *Fachlexikographie. Fachwissen und seine Repräsentation in Wörterbüchern*. Tübingen 1994, S. 161–176. (= Forum für Fachsprachen-Forschung 23).

[Bergenholtz 1989]

Bergenholtz, Henning: *Probleme der Selektion im allgemeinen einsprachigen Wörterbuch*. In: Hausmann, Franz Josef u. a. (Hg.): *Wörterbücher. Ein internationales Handbuch zur Lexikographie*. Berlin u. a. 1989, S. 772–779. (= Handbücher zur Sprach- und Kommunikationswissenschaft (= HSK) 5.1).

[Bergenholtz 1994]

Bergenholtz, Henning: *Fachsprache und Gemeinsprache: Lemmaselektion im Fachwörterbuch*. In: Schaefer, Burkhard; Bergenholtz, Henning (Hg.): *Fachlexikographie. Fachwissen und seine Repräsentation in Wörterbüchern*. Tübingen 1994, S. 285–304. (= Forum für Fachsprachen-Forschung 23).

[Bergsland / Vogt 1962]

Bergsland, Knut; Vogt, Hans: *On the Validity of Glottochronology*. *Current Anthropology* 3(1962), S. 115–153.

[Besch 1983]

Besch, Werner: *Dialekt, Schreibdialekt, Schriftsprache, Standardsprache. Exemplarische Skizze ihrer historischen Ausprägung im Deutschen*. In: Besch, Werner u. a. (Hg.): *Dialektologie. Ein Handbuch zur deutschen und allgemeinen Dialektforschung*. Berlin u. a. 1983, S. 961–990. (= Handbücher zur Sprach- und Kommunikationswissenschaft (= HSK) 1.2).

[Bonan-Garrigues 1993]

Bonan-Garrigues, Mylène: *Méthode de paramétrage des dictionnaires et grammaires électroniques. Application à des systèmes interactifs en langue naturelle*. Thèse de doctorat en sciences du langage. Université Paris 7. Paris 1993.

[Braun 1990]

Braun, Peter: *Internationalismen – gleiche Wortschätze in europäischen Sprachen*. In: Braun, Peter; Schaefer, Burkhard; Volmert, Johannes (Hg.): *Internationalismen. Studien zur interlingualen Lexikologie und Lexikographie*. Tübingen 1990, S. 13–33. (= Reihe Germanistische Linguistik 102).

[Bruce / Guthrie / Wilks 1993]

Bruce, Rebecca; Guthrie, Louise; Wilks, Yorick: *Automatic Lexical Extraction – Theories and Applications*. In: Beckmann, Frank; Heyer, Gerhard (Hg.): *Theorie und Praxis des Lexikons*. Berlin u. a. 1993, S. 250–273. (= Grundlagen der Kommunikation und Kognition [o. Nr.]).

[Buchbinder 1973]

Buchbinder, Wolf A.: *Zur Objektivität bei der Auswahl der Lexik für die Oberschule*. In: *Deutsch als Fremdsprache* 10 (1973), S. 301–309.

[Bußmann 1990]

Bußmann, Hadumod: *Lexikon der Sprachwissenschaft*. 2. Aufl. Stuttgart 1990.

[Bußmann 1996]

Bußmann, Hadumod: *Routledge Dictionary of Language and Linguistics*. London u. a. 1996.

[Carlson u. a. 1985]

Carlson, Rolf u. a.: *Phonetic and orthographic properties of the basic vocabulary of five European languages*. In: *Speech Transmission Laboratory. Department of Speech Communication & Music Acoustics. Royal Institute of Technology, Sweden. Quarterly Progress and Status Report 1 (1985)*, S. 63–94.

[Carter 1998]

Carter, Ronald: *Vocabulary. Applied linguistics perspectives*. 2. Aufl. London 1998.

[Clark 1993]

Clark, Eve V.: *The lexicon in acquisition*. Cambridge 1993.

[Crystal 1997]

Crystal, David: *A Dictionary of Linguistics and Phonetics*. 4. Aufl. Oxford 1997.

[Crystal 2001]

Crystal, David: *Language and the Internet*. Cambridge 2001.

[Dell 1989]

Dell, Gary S.: *The Retrieval of Phonological Forms in Production: Tests of Predictions from a Connectionist Model*. In: Marslen-Wilson, William (Hg.): *Lexical Representation and Process*. Cambridge, Mass., 1989, S. 136–165.

[de Smet 1981]

de Smet, Gilbert A. R.: *[Rezension zu:] PETER KÜHN, Der Grundwortschatz. Bestimmung und Systematisierung*. Tübingen: Max Niemeyer Verlag 1979. IX, 192 S. (Reihe Germanistische Linguistik. Bd. 17). In: *Zeitschrift für Dialektologie und Linguistik* 48 (1981), S. 363–365.

[Dixon 1971]

Dixon, Robert: *A method of semantic description*. In: Steinberg, Danny; Jakobovits, Leon (Hg.): *Semantics. An interdisciplinary reader in philosophy, linguistics and psychology*. Cambridge 1971, S. 436–471.

[DUW]

Wissenschaftlicher Rat der Dudenredaktion (Hg.): *Duden. Deutsches Universalwörterbuch*. CD-ROM. 5. Aufl. Mannheim 2003.

[DUW (gedruckte Ausgabe)]

Wissenschaftlicher Rat der Dudenredaktion (Hg.): *Duden. Deutsches Universalwörterbuch*. 5. Aufl. Mannheim 2003.

[Dyen 1975]

Dyen, Isidore: *Linguistic Subgrouping and Lexicostatistics*. The Hague 1975. (= Janua Linguarum. Series Minor 175).

[Dyen / Kruskal / Black 1992]

Dyen, Isidore; Kruskal, Joseph; Black, Paul: *An Indoeuropean Classification. A Lexicostatistical Experiment*. Philadelphia 1992. (= Transactions of the American Philosophical Society Held at Philadelphia for Promoting Useful Knowledge 82.5).

[Embleton 1986]

Embleton, Sheila: *Statistics in Historical Linguistics*. Bochum 1986. (= Quantitative Linguistics 30).

[Embleton 2000]

Embleton, Sheila: *Lexicostatistics / Glottochronology: from Swadesh to Sankoff to Starostin to future horizons*. In: Renfrew, Colin; McMahon, April; Trask, Larry (Hg.): *Time depth in historical linguistics*. Cambridge 2000, S. 143–165. (= Papers in the Prehistory of Languages 1).

[Endres-Niggemeyer 1975]

Endres-Niggemeyer, Brigitte: *Das Lexikon französischer Massensliteratur in den mittleren und unteren Frequenzklassen. Eine vergleichende statistische Darstellung einer Schriftsprachverwendung größter Zugänglichkeit mit Anwendung auf dne Grundwortschatz des Francais Fondamental 2e degré*. Darmstadt 1975.

[Erk 1963]

Erk, Heinrich: *Ein Rangwörterbuch und viele Fragen*. In: *Deutschunterricht für Ausländer* 13 (1963), S. 86–93.

[Erk 1972]

Erk, Heinrich: *Zur Lexik wissenschaftlicher Fachtexte: Verben – Frequenz und Verwendungsweise*. München 1972. (= Schriften der Arbeitsstelle für wissenschaftliche Didaktik des Goethe-Instituts 4).

[Erk 1975]

Erk, Heinrich: *Zur Lexik wissenschaftlicher Fachtexte. Substantive – Frequenz und Verwendungsweise*. München 1975. (= Schriften der Arbeitsstelle für wissenschaftliche Didaktik des Goethe-Instituts 5).

[Erk 1982]

Erk, Heinrich: *Zur Lexik wissenschaftlicher Fachtexte. Adjektive, Adverbien und andere Wortarten. Frequenz und Verwendungsweise*. München 1982. (= Schriften der Arbeitsstelle für wissenschaftliche Didaktik des Goethe-Instituts 6).

[Fabre / Habert / Labbé 1997]

Fabre, Cécile; Habert, Benoît; Labbé, Dominique: *La polysémie dans la langue générale et les discours spécialisés*. In: *Sémiotiques* 13 (1997), S. 15–30.

[Fink / Strehl 2001]

Fink, Gerhard; Strehl, Linda: *Langenscheidt Grundwortschatz Latein*. München 2001.

[Fox 1992]

Fox, Christopher: *Lexical Analysis and Stoplists*. In: Frakes, William B.; Baeza-Yates, Ricardo (Hg.): *Information Retrieval. Data Structures & Algorithms*. Upper Saddle River, New Jersey, 1992, S. 102–130.

[Garibiani 2000]

Garibiani, Natalia: *2000 x Minuten-Training Russisch Grundwortschatz. Die neuen Schnell-Lern-Übungen für zwischendurch*. München 2000.

[Geens 1978]

Geens, Dirk: *On measurement of lexical differences by means of frequency*. In: Altmann, Gabriel (Hg.): *Glottometrika I*. Bochum 1978, S. 46–72. (= *Quantitative Linguistics*. 1).

[Glück (Hg.) 2000]

Glück, Helmut (Hg.): *Metzler-Lexikon Sprache*. Stuttgart 2000.

[Gougenheim 1958]

Gougenheim, Georges: *Dictionnaire fondamental de la langue française*. Paris 1958.

[Gougenheim / Rivenc / Michéa u. a. 1956]

Gougenheim, Georges; Rivenc, Paul; Michéa, René u. a.: *L'élaboration du Français fondamental (1er degré). Etude sur l'établissement d'un vocabulaire et d'une grammaire de base*. Paris 1956.

[Gross 1994]

Gross, Maurice: *Constructing Lexicon-Grammars*. In: Atkins, Beryl; Zampolli, Antonio (Hg.): *Computational Approaches to the Lexicon*. Oxford 1994, S. 213–263.

[Haderlein 1997]

Haderlein, Veronika: *Die Methode der Lexikostatistik als ein Mittel des Sprachenvergleichs. Eine Neubetrachtung am Beispiel der Sprachengruppe Deutsch-Niederländisch-Nynorsk-Schwedisch*. Abschlußarbeit zur Erlangung des Magistergrades im Fach Germanistische Linguistik an der Ludwig-Maximilians-Universität München. München 1997.

[Haderlein 2005]

Haderlein, Veronika: *Semantik bei der Arbeit mit zentralen Wortschätzen: Anforderungen und Möglichkeiten*. In: Schnorbusch, Daniel; Langer, Stefan (Hg.): *Semantik im Lexikon*. Tübingen 2005, S. 9–32.

[Häublein / Jenkins 2004]

Häublein, Gernot; Jenkins, Recs: *Thematischer Grund- und Aufbauwortschatz Englisch. Neue Ausgabe*. Stuttgart 2004.

[Handke 1995]

Handke, Jürgen: *The Structure of the Lexicon. Human versus Machine*. Berlin u. a. 1995. (= *Natural Language Processing* 5).

[Harman 1992]

Harman, Donna: *Ranking Algorithms*. In: Frakes, William B.; Baeza-Yates, Ricardo (Hg.): *Information Retrieval. Data Structures & Algorithms*. Upper Saddle River, New Jersey, 1992, S. 363–392.

[Hasan 1974]

Hasan, Finuta: *Der Grundwortschatz*. In: Graur, Alexandru (Hg.): *Einführung in die Sprachwissenschaft*. Berlin 1974, S. 178–191.

[Haß-Zumkehr 2001]

Haß-Zumkehr, Ulrike: *Deutsche Wörterbücher. Brennpunkt von Sprach- und Kulturgeschichte*. Berlin 2001.

[Hasselgren / Montn mery / Svensson u. a. 2003]

Hasselgren, Jon; Montn mery, Erik; Svensson, Markus u. a.: *HMS: A Predictive Text Entry Method Using Bigrams*. In: Association for Computational Linguistics (Hg.): *Proceedings of the Workshop on Language*

Modeling for Text Entry Methods. 10th Conference of the European Chapter of the Association of Computational Linguistics, April 12-17, 2003, Budapest, Hungary. o. O 2003, S. 43–49.

[online abrufbar unter: <http://www.cs.lth.se/Education/LTH/EDA170/Reports2003/HMS.pdf>]

[Helfrich 1993]

Helfrich, Uta: *Neologismen auf dem Prüfstand. Ein Modell zur Ermittlung der Akzeptanz französischer Neologismen*. Wilhelmsfeld 1993.

[Heller 1970]

Heller, Klaus: *Der Wortschatz unter dem Aspekt des Fachwortes – Versuch einer Systematik*. In: Wissenschaftliche Zeitschrift der Karl-Marx-Universität Leipzig 19 (1970), S. 531–544.

[Heller 1981]

Heller, Klaus: *Wortschatz unter dem Aspekt des Fachwortes*. In: Hahn, Walther von (Hg.): *Fachsprachen*. Darmstadt 1981, S. 218–238.

[Hendrickx 1987]

Hendrickx, Rudi: *The Construction of a Basic Vocabulary. A Socio-Linguistic or Statistical Problem?* Leuvense Bijdragen 76 (1987), S. 1–16.

[Henne 1972]

Henne, Herbert: *Semantik und Lexikographie. Untersuchungen zur lexikalischen Kodifikation der deutschen Sprache*. Berlin 1972.

[Herberg 1988]

Herberg, Dieter: *Neologismen – lexikologisch und lexikographisch betrachtet*. In: *Sprachpflege* 37 (1988), S. 109–112.

[Herberg 1992]

Herberg, Dieter: *Makrostrukturelle Beziehungen im Wortschatz und in Wörterbucheinträgen. Möglichkeiten und Grenzen des allgemeinen einsprachigen Wörterbuchs*. In: Brausse, Ursula; Viehweger, Dieter (Hg.): *Lexikonomie und Wörterbuch. Wege der Verbindung von lexikologischer Forschung und lexikographischer Praxis*. Tübingen 1992., S. 89–163. (= *Lexicographica*. Series Maior 43).

[Herbst 1986]

Herbst, Thomas: *Defining with a controlled defining vocabulary in foreign learner's dictionaries*. *Lexicographica* 2 (1986), S. 101–119.

[Heyer 1993]

Heyer, Gerhard: *On the Role of the Dictionary and Dictionary-Based Approaches in Language Products Technology*. In: Beckmann, Frank; Heyer, Gerhard (Hg.): *Theorie und Praxis des Lexikons*. Berlin u. a., S. 207–217. (= *Grundlagen der Kommunikation und Kognition* [o. Nr.]).

[Hoey 1991]

Hoey, Michael: *Lexis in Text*. Oxford 1991.

[Hoffmann 1972]

Hoffmann, Lothar: *Zur Definition des Wortes als Grundeinheit (statistisches Element) in lexikalischen Häufigkeitsuntersuchungen*. In: Wissenschaftliche Zeitschrift der Karl-Marx-Universität Leipzig 21 (1972), S. 39–47.

[Hoffmann 1984]

Hoffmann, Lothar: *Fachwortschatz, Grundwortschatz, Minimum*. In: *Deutsch als Fremdsprache* 21 (1984), S. 224–228.

[Hoffmann 1985]

Hoffmann, Lothar: *Kommunikationsmittel Fachsprache. Eine Einführung*. 2. Aufl. Tübingen 1985. (= *Forum für Fachsprachen-Forschung* 1).

[Hoffmann (Hg.) 1970]

Hoffmann, Lothar (Hg.): *Fachwortschatz Physik. Häufigkeitwörterbuch russisch, englisch, französisch*. Leipzig 1970.

[Hoffmann (Hg.) 1970a]

Hoffmann, Lothar (Hg.): *Medizinisches Häufigkeitswörterbuch russisch, englisch, französisch*. Leipzig 1970.

[Hoffmann (Hg.) 1973]

Hoffmann, Lothar (Hg.): *Fachwortschatz Chemie. Häufigkeitswörterbuch russisch, englisch, französisch*. Leipzig 1973.

[Hoffmann (Hg.) 1976]

Hoffmann, Lothar (Hg.): *Fachwortschatz Mathematik. Häufigkeitswörterbuch russisch, englisch, französisch*. Leipzig 1976.

[Hoffmann (Hg.) 1978]

Hoffmann, Lothar (Hg.): *Fachwortschatz Tierproduktion und Veterinärmedizin. Häufigkeitswörterbuch russisch, englisch, französisch*. Leipzig 1978.

[Häublein / Jenkins 2004]

Häublein, Gernot; Jenkins, Recs: *Thematischer Grund- und Aufbauwortschatz Englisch*. Stuttgart 2004.

[Hård af Segerstad 2003]

Hård af Segerstad, Ylva: *Towards a Better System for Text Prediction in Mobile Phones*. In: Petterson, J. S. (Hg.): *HumanIT 2003*. Karlstad, S. 125–137. (= Karlstad University Studies 2003).

[Online zugänglich unter: http://www.humanit.org/pdf/HumanIT_2003_Ch8_Hard_af_Segerstad.pdf]

[Jackendoff 1995]

Jackendoff, Ray: *The boundaries of the lexicon*. In: Everaert, Martin u. a. (Hg.): *Idioms: structural and psychological perspectives*. Hillsdale, New Jersey u. a., S. 133–165.

[Jackson / Moulinier 2002]

Jackson, Peter; Moulinier, Isabelle: *Natural Language Processing for Online Applications. Text Retrieval, Extraction and Categorization*. Amsterdam 2002.

[Jahr 1996]

Jahr, Silke: *Das Verstehen von Fachtexten. Rezeption. Kognition. Applikation*. Tübingen 1996. (= Forum für Fachsprachen-Forschung 34).

[Kaeding 1898]

Kaeding, Friedrich W.: *Häufigkeitswörterbuch der deutschen Sprache*. Steglitz [1898].

[Kalverkämper 1990]

Kalverkämper, Hartwig: *Gemeinsprache und Fachsprachen – Plädoyer für eine integrierende Sichtweise*. In: Stickel, Gerhard (Hg.): *Deutsche Gegenwartssprache. Tendenzen und Perspektiven*. Berlin u. a. 1990, S. 88–133. (= Jahrbuch des Instituts für deutsche Sprache, Mannheim, 1989).

[Katz 1967]

Katz, Jerrold: *Recent Issues in Semantic Theory*. *Foundations of Language* 3 (1967), S. 124–194.

[Kaufmann 1968]

Kaufmann, Gerhard: *Die Erarbeitung eines Grundwortschatzes Deutsch für das Fach Deutsch als Fremdsprache*. In: *Deutsch als Fremdsprache* 18 (1968), S. 7–20.

[Kennedy 1998]

Kennedy, Graeme: *An Introduction to Corpus Linguistics*. London u. a. 1998.

[Kittredge 2003]

Kittredge, Richard I.: *Sublanguages and controlled languages*. In: Mitkov, Ruslan (Hg.): *The Oxford Handbook of Computational Linguistics*. Oxford 2003, S. 430–447.

[Kloster 1997]

Kloster, Stephanie Xenia: *Konzeption eines onomasiologischen Wörterbuchs für erwachsene Lese-/SchreiberInnen*. Hildesheim 1997.

[Kosaras 1980]

Kosaras, István: *Grundwortschatz der deutschen Sprache. Einsprachiges Wörterbuch*. Budapest u. a. 1980.

[Krohn 1992]

Krohn, Dieter: *Grundwortschätze und Auswahlkriterien*. Göteborg 1992. (= Acta Universitatis Gothoburgensis. Göteborger Germanistische Forschungen 34).

[Kucera 1986]

Kucera, Antonín: *Terminologie der Technik und der Naturwissenschaften in einem einsprachigen Bedeutungswörterbuch – Ja oder Nein*. In: Weiss, Walter; Wiegand, Herbert Ernst; Reis, Marga (Hg.): *Textlinguistik contra Stilistik? – Wortschatz und Wörterbuch – Grammatische oder pragmatische Organisation von Rede? Akten des VII. Internationalen Germanisten-Kongresses*. Göttingen 1985. Tübingen 1986, S. 144-147. (= Kontroversen, alte und neue 3).

[Kühn 1979]

Kühn, Peter: *Der Grundwortschatz. Bestimmung und Systematisierung*. Tübingen 1979. (= Reihe Germanistische Linguistik 17).

[Kühn 1979a]

Kühn, Peter: *Kritik der bisherigen Grundwortschatzlexikographie*. In: *Zielsprache Deutsch* 10 (1979), S. 34–42.

[Kühn 1980]

Kühn, Peter: *Sprachverstehen und Grundwortschatzkodifikation*. In: Kühlwein, Wolfgang; Raasch, Albert (Hg.): *Sprache und Verstehen. Kongreßberichte der 10. Jahrestagung der Gesellschaft für Angewandte Linguistik GAL e.V.* Mainz 1979. Tübingen 1980, S. 55–61.

[Kühn 1990]

Kühn, Peter: *Das Grundwortschatzwörterbuch*. In: Hausmann, Franz Josef u. a. (Hg.): *Wörterbücher. Ein internationales Handbuch zur Lexikographie*. Berlin u. a. 1990, S. 1353–1364. (= Handbücher zur Sprach- und Kommunikationswissenschaft (= HSK) 5.2).

[Köpke 1964]

Köpke, Wulf: *Viele Fragen an die Sprachstatistik*. In: *Deutschunterricht für Ausländer* 14 (1964), S. 112–124.

[Labov 1976]

Labov, William: *Sprache im sozialen Kontext. Beschreibung und Erklärung struktureller und sozialer Bedeutung von Sprachvariation*. Band 1. Kronberg 1976. (= Monographien Linguistik und Kommunikationswissenschaft 33).

[Lakoff / Johnson 2003]

Lakoff, George; Johnson, Mark: *Metaphors we live by*. Chicago 2003.

[Lehmann 1991]

Lehmann, Hubert: *Towards a Core Vocabulary for a Natural Language System*. In: Association for Computational Linguistics (Hg.): *Proceedings of the fifth conference of the European chapter of the Association for Computational Linguistics*, Berlin, Germany, April 09-11, 1991. o. O. 1991, S. 303–305.

[Lehrer 1974]

Lehrer, Adrienne: *Homonymy and Polysemy: Measuring Similarity of Meaning*. In: *Language Sciences* 3 (1974), S. 33–39.

[Lerchner 1968]

Lerchner, Gotthard: *Zur Angemessenheit eines statistischen Wahrscheinlichkeitskalküls in der Sprachgeschichtsforschung*. In: Ruzicka, Rudolf (Hg.): *Probleme der strukturellen Grammatik und Semantik*. Leipzig 1968, S. 253–270.

[LODCE 1984]

Procter, Paul u. a.: *Longman Dictionary of Contemporary English*. [Korrigierter Nachdruck der Ausgabe von 1978.] Harlow 1984.

[Lutzeier 1985]

Lutzeier, Peter Rolf: *Linguistische Semantik*. Stuttgart 1985.

[Lutzeier (Hg.) 1993]

Lutzeier, Peter Rolf (Hg.): *Studien zur Wortfeldtheorie*. Tübingen 1993.

[Lyons 1980]

Lyons, John: *Semantik*. München 1980.

[Martin 1990]

Martin, Willy: *The Frequency Dictionary*. In: Hausmann, Franz Josef u. a. (Hg.): Wörterbücher. Ein internationales Handbuch zur Lexikographie. Berlin u. a. 1990, S. 1314–1322. (= Handbücher zur Sprach- und Kommunikationswissenschaft (= HSK) 5.2).

[McEnery / Wilson 2001]

McEnery, Tony; Wilson, Andrew: *Corpus Linguistics. An Introduction*. 2. Aufl. Edinburgh 2001.

[Mehl 1996]

Mehl, Stephan: *Probleme der Extraktion semantischer Relationen aus maschinenlesbaren Wörterbüchern*. In: Weber, Nico (Hg.): *Semantik, Lexikographie und Computeranwendungen*. Tübingen 1996, S. 261–273. (= Sprache und Information 33).

[Meyer 1993]

Meyer, Paul Georg: *Nicht-fachliches Vokabular im Fachtext: ein vernachlässigter Aspekt der Fachsprachenforschung*. In: Börner, Wolfgang; Vogel, Klaus (Hg.): *Wortschatz und Fremdsprachenerwerb*. Bochum 1993, S. 177–199. (= Fremdsprachen in Lehre und Forschung 14).

[Michéa 1953]

Michéa, René: *Mots fréquents et mots disponibles*. In: *Les Langues Modernes* 47 (1953), S. 338–344.

[Michéa 1964]

Michéa, René: *Les vocabulaires fondamentaux. Recherches et techniques nouvelles au service de l'enseignement des langues vivantes*. In: Actes [...] par la section régionale de l'Association. Association des Professeurs de langues vivantes de France, Strasbourg, 22-23 mars 1963. Paris 1964, S. 21–36.

[Morgan 1928]

Morgan, Bayard Quincy: *German Frequency Word Book. Based on Kaeding's „Häufigkeitswörterbuch der deutschen Sprache“*. New York 1928.

[Muller 1977]

Muller, Charles: *Principes et méthodes de statistique lexicale*. Paris 1977.

[Muret-Sanders 1899]

Sanders, Daniel; Schmidt, Imm.; Stoffel, Cornelis: *Muret-Sanders. Encyclopädisches englisch-deutsches und deutsch-englisches Wörterbuch. Ein Parallelwerk zu Sachs-Villatte's französisch-deutschem und deutsch-französischem Wörterbuche. Mit Angabe der Aussprache nach dem phonetischen System der Methode Tous-saint-Langenscheidt. Zweiter Teil: deutsch-englisch. Zweite Hälfte: K-Z. Berlin [1899]*.

[Muret-Sanders 1899a]

Sanders, Daniel; Schmidt, Imm.; Stoffel, Cornelis: *Muret-Sanders. Encyclopädisches englisch-deutsches und deutsch-englisches Wörterbuch. Ein Parallelwerk zu Sachs-Villatte's französisch-deutschem und deutsch-französischem Wörterbuche. Mit Angabe der Aussprache nach dem phonetischen System der Methode Tous-saint-Langenscheidt. Zweiter Teil: deutsch-englisch. Erste Hälfte: A-J. Berlin [1899]*.

[Nuopponen 2002]

Nuopponen, Anita: *Vocabularies for specific purposes: An overview*. In: Cruse, D. Alan u. a. (Hg.): *Lexikologie. Ein internationales Handbuch zur Natur und Struktur von Wörtern und Wortschätzen*. Berlin u. a., S. 856–866. (= Handbücher zur Sprach- und Kommunikationswissenschaft (= HSK) 21.1).

[Ogden 1930]

Ogden, Charles K.: *Basic English: A General Introduction*. London 1930.

[Ortmann 1979]

Ortmann, Wolf Dieter: *Hochfrequente deutsche Wortformen 3*. München 1979.

[Osgood / Suci / Tannenbaum 1957]

Osgood, Charles Egerton; Suci, George; Tannenbaum, Percy: *The Measurement of Meaning*. Urbana, IL, 1957.

[Petöfi / Bredemeier (Hg.) 1977]

Petöfi, János S.; Bredemeier, Jürgen (Hg.): *Das Lexikon in der Grammatik. Die Grammatik im Lexikon*. Hamburg 1977. (= Papiere zur Textlinguistik 2).

[Peyawari 1999]

Peyawari, Ahmad S.: *The Core Vocabulary of International English: A Corpus Approach*. Bergen 1999.

[Pfeffer 1964]

Pfeffer, Jay Alan: *Grunddeutsch. Basic (Spoken) German Word List. Grundstufe*. Englewood Cliffs, New Jersey 1964.

[Pfeffer 1970]

Pfeffer, Jay Alan: *Grunddeutsch. Basic (Spoken) German Dictionary. For Everyday Usage*. Englewood Cliffs, NJ 1970.

[Pfeffer 1975]

Pfeffer, Jay Alan: *Grunddeutsch. Basic (Spoken) German Idiom List. Grundstufe*. Englewood Cliffs, New Jersey 1975. (= Publications of the Institute for Basic German, University of Pittsburgh 3).

[Pfeffer 1984]

Pfeffer, Jay Alan: *Grunddeutsch. Texte zur gesprochenen Gegenwartssprache. Textkorpora I. Einführungs- und Registerband*. Tübingen 1984.

[Plickat 1983]

Plickat, Hans-Heinrich: *Deutscher Grundwortschatz*. Weinheim u. a. 1983.

[PONS 1999]

Terell, Peter; Schnorr, Veronika; Morris, Wendy u. a.: *PONS Collins Großwörterbuch für Experten und Universität. Deutsch – Englisch. Englisch – Deutsch*. Stuttgart u. a. 1999.

[Quasthoff 1997]

Quasthoff, Uwe: *Der deutsche Wortschatz – interaktive Sammlung*. Beilage zur CD Projekt Deutscher Wortschatz. Institut für Informatik, Universität Leipzig. [Leipzig] 1997.

[Quasthoff 1998]

Quasthoff, Uwe: *Deutscher Wortschatz im Internet*. In: LDV-Forum 15 (1998), S. 4–23.

[Quasthoff / Richter 2005]

Quasthoff, Uwe; Richter, Matthias: *Projekt Deutscher Wortschatz*. In: *Babylonia* Nr. 3 (2005), [o. Seitenzahl].

[Online verfügbar unter <http://www.babylonia-ti.ch/BABY305/quaride.htm>]

[Quasthoff / Wolff 1999]

Quasthoff, Uwe; Wolff, Christian: *Korpuslinguistik und große einsprachige Wörterbücher*. In: *Linguistik Online* 3 (1999), [o. Seitenzahl].

[Online verfügbar unter http://www.linguistik-online.de/2_99/quasthoff.html.]

[Quirk 1982]

Quirk, Randolph: *Style and Communication in the English Language*. London 1982.

[Reichmann 1990]

Reichmann, Oskar: *Das onomasiologische Wörterbuch*. In: Hausmann, Franz Josef u. a. (Hg.): *Wörterbücher. Ein internationales Handbuch zur Lexikographie*. Berlin u. a. 1989, S. 1057–1067. (= Handbücher zur Sprach- und Kommunikationswissenschaft (= HSK) 5.2).

[Renfrew / McMahon / Trask (Hg.) 2000]

Renfrew, Colin; McMahon, April; Trask, Larry (Hg.): *Time depth in historical linguistics*. Cambridge 2000. (= Papers in the Prehistory of Languages 1).

[Rettig 1989]

Rettig, Wolfgang: *Die Wortbildungszusammenhänge im allgemeinen einsprachigen Wörterbuch*. In: Hausmann, Franz Josef u. a. (Hg.): *Wörterbücher. Ein internationales Handbuch zur Lexikographie*. Berlin u. a. 1989, S. 642–649. (= Handbücher zur Sprach- und Kommunikationswissenschaft (= HSK) 5.1).

[Richards 1974]

Richards, Jack C.: *Word lists: Problems and prospects*. RELC Journal 5 (1974), S. 69–84.

[Richter 2004]

Richter, Matthias: *Korpusbasierte Lemmaselektion*. [Magisterarbeit eingereicht am Institut für Kommunikations- und Medienwissenschaft der Universität Leipzig]. Leipzig 2004.

[Online verfügbar unter <http://wortschatz.uni-leipzig.de/~mrichter/MR-Magisterarbeit.pdf>]

[Riesel / Schendels 1975]

Riesel, Elise.; Schendels, Eugenie: *Deutsche Stilistik*. Moskau 1975.

[Rosengren 1968]

Rosengren, Inger: *Ein Frequenzwörterbuch der modernen Zeitungssprache – wie und wozu?* In: Beiträge zur Linguistik und Informationsverarbeitung 14 (1968), S. 7–21.

[Rosengren 1969]

Rosengren, Inger: *Wort und Wortform*. In: Studia linguistica 23 (1969), S. 103–113.

[Rosengren 1972]

Rosengren, Inger: *Ein Frequenzwörterbuch der deutschen Zeitungssprache 1*. Lund 1972.

[Rosengren 1976]

Rosengren, Inger: *Der Grundwortschatz als theoretisches und praktisches Problem*. In: Moser, Hugo u.a. (Hg.): Probleme der Lexikologie und Lexikographie. Jahrbuch 1975 des Instituts für deutsche Sprache. Düsseldorf, S. 313–333 (= Sprache der Gegenwart. Schriften des Instituts für Deutsche Sprache 39).

[Rosengren 1977]

Rosengren, Inger: *Ein Frequenzwörterbuch der deutschen Zeitungssprache 2*. Lund 1977.

[Rosengren 1986]

Rosengren, Inger: *Welt und Sprache im Spiegel des Frequenzwörterbuchs*. In: Stedje, Astrid (Hg.): Die Bänder Grimm. Erbe und Rezeption. Stockholmer Symposium 1984. Stockholm, S. 155–170. (= Acta Universitatis Stockholmiensis. Stockholmer Germanistische Forschungen 32).

[Ruoff 1981]

Ruoff, Arno: *Häufigkeitswörterbuch gesprochener Sprache. Gesondert nach Wortarten alphabetisch, rückläufig alphabetisch und nach Häufigkeit geordnet*. Tübingen 1981. (= Idiomatologica 8).

[Ruoff 1983]

Ruoff, Arno: *Ein Häufigkeitswörterbuch gesprochener Sprache als Grundlage weiterführender Untersuchungen*. In: Haas, Walter; Näf, Anton (Hg.): Wortschatzprobleme im Alemannischen. 7. Arbeitstagung alemannischer Dialektologen. Freiburg i.Ü., 1.-3. Oktober 1981. Freiburg (Schweiz), S. 11–31. (= Germanistica Friburgensia 7).

[Raasch 1972]

Raasch, Albert: *Neue Wege zu einem Grundwortschatz*. In: Praxis des neusprachlichen Unterrichts 19 (1972), S. 235–244.

[Raasch 1995]

Raasch, Albert: [Rezension zu:] Dieter Krohn. *Grundwortschatze und Auswahlkriterien*. Göteborg 1992. In: Zeitschrift für romanische Philologie 111 (1995), S. 424–425.

[Schindler 2002]

Schindler, Wolfgang: *Lexik, Lexikon, Wortschatz: Probleme der Abgrenzung*. In: Cruse, D. Alan u. a. (Hg.): Lexikologie. Ein internationales Handbuch zur Natur und Struktur von Wörtern und Wortschatzen. Berlin u. a., S. 34–44. (= Handbücher zur Sprach- und Kommunikationswissenschaft (= HSK) 21.1).

[Schmidt 1955]

Schmidt, Wilhelm: *Zum Grundwortschatz und Wortbestand der deutschen Sprache*. In: Der Deutschunterricht 8 (1955), S. 530–540.

[Schneider 1988]

Schneider, Edgar W.: *Variabilität, Polysemie und Unschärfe der Wortbedeutung. Band 1: Theoretische und methodische Grundlagen*. Tübingen 1988. (= Linguistische Arbeiten 196).

[Schnörch 2002]

Schnörch, Ulrich: *Der zentrale Wortschatz des Deutschen. Strategien zu seiner Ermittlung, Analyse und lexikografischen Aufarbeitung*. Tübingen 2002. (= Studien zur deutschen Sprache 26).

[Schulz 2000]

Schulz, Matthias: *Zur Eignung von älteren Wörterbüchern und Wörterbuchvergleichen für eine Analyse von Wortschatz und dessen Wandel*. In: Sprachwissenschaft 25 (2000), S. 63–75.

[Schulz 2004]

Schulz, Matthias: [Rezension zu] ULRICH SCHNÖRCH, *Der zentrale Wortschatz des Deutschen. Strategien zu seiner Ermittlung, Analyse und lexikographischen Aufarbeitung, Studien zur Deutschen Sprache 26, Tübingen: Gunter Narr Verlag 2002, 422 Seiten*. In: Beiträge zur Namenforschung 39 (2004), S. 332–334.

[Schulz 1991]

Schulz, Petra: *Untersuchungen zum Grundwortschatz*. o. O. 1991. (= IWBS Report 164. [Herausgegeben von-IBM Deutschland GmbH. Wissenschaftliches Zentrum. Institut für Wissensbasierte Systeme Stuttgart].

[Schwarze / Wunderlich 1985]

Schwarze, Christoph; Wunderlich, Dieter: *Einleitung*. In: Schwarze, Christoph; Wunderlich, Dieter (Hg.): *Handbuch der Lexikologie*. Königstein 1985, S. 7–23.

[Sebastiani 2006]

Sebastiani, Fabrizio: *Classification of text, automatic*. In: Brown, Keith (Hg.): *The Encyclopedia of Language and Linguistics*. 2. Aufl. Band 14. Amsterdam 2006, S. 457–462.

[Online verfügbar unter <http://nmis.isti.cnr.it/sebastiani/Publications/ELL06.pdf>]

[Seebold 1993]

Seebold, Elmar: *Der deutsche Wortschatz und seine Geschichte*. Manuskript zur Vorlesung an der Ludwig-Maximilians-Universität München im Sommersemester 1993. München 1993.

[Seidenberg 1989]

Seidenberg, Mark S.: *Visual Word Recognition and Pronunciation: A Computational Model and Its Implications*. In: Marslen-Wilson, William (Hg.): *Lexical Representation and Process*. Cambridge, Mass. 1989, S. 25–74.

[Sinclair 1998]

Sinclair, John: *Korpusmorphologie. Ein Klassifikationsrahmen*. In: Teubert, Wolfgang (Hg.): *Neologie und Korpus*. Tübingen 1998, S. 111–128. (= Studien zur deutschen Sprache 11).

[Sowinski 2002]

Sowinski, Bernhard: *Dimensionen der Bedeutung IV: Stilistische Aspekte*. In: Cruse, D. Alan u. a. (Hg.): *Lexikologie. Ein internationales Handbuch zur Natur und Struktur von Wörtern und Wortschätzen*. Berlin u. a., S. 363–370. (= Handbücher zur Sprach- und Kommunikationswissenschaft (= HSK) 21.1).

[Stubbs 1986]

Stubbs, Michael: *Educational linguistics*. Oxford 1986.

[Swadesh 1950]

Swadesh, Morris: *Salish Internal Relationship*. In: *International Journal of American Linguistics* 16 (1950), S. 157–167.

[Swadesh 1955]

Swadesh, Morris: *Towards Greater Accuracy in Lexicostatistic Dating*. In: *International Journal of American Linguistics* 21 (1955), S. 121–137.

[Sweet 1899]

Sweet, Henry: *The Practical Study of Language*. London 1899.

[Teubert 1998]

Teubert, Wolfgang: *Korpus und Neologie*. In: Teubert, Wolfgang (Hg.): *Neologie und Korpus*. Tübingen 1998, S. 129–170. (= Studien zur deutschen Sprache 11).

[Tischler 1973]

Tischler, Johann: *Glottochronologie und Lexikostatistik*. Innsbruck 1973. (= Innsbrucker Beiträge zur Sprachwissenschaft 11).

[Trost 2003]

Trost, Harald: *Morphology*. In: Mitkov, Ruslan (Hg.): *The Oxford Handbook of Computational Linguistics*. Oxford 2003, S. 25–47.

[Tzoukermann / Klavans / Strzalkowski 2003]

Tzoukermann, Evelyne; Klavans, Judith; Strzalkowski, Tomek: *Information Retrieval*. In: MITKOV, Ruslan (Hg.): *The Oxford Handbook of Computational Linguistics*. Oxford 2003, S. 529–544.

[Vachek 1966]

Vachek, Josef: *The Linguistic School of Prague. An Introduction to its Theory and Practice*. Bloomington, Ind. 1966.

[van Ek 1977]

van Ek, Jan: *The Threshold Level. Ein Projekt des Europarates*. In: Hüllen, Werner; Raasch, Albert; Zapp, Franz Josef (Hg.): *Sprachminima und Abschlußprofile*. Berlin u. a. 1977, S. 91–103.

[Varile / Zampolli / Cole (Hg.) 1997]

Varile, Giovanni Batista; Zampolli, Antonio; Cole, Ronald (Hg.): *Survey of the state of the art in human language technology*. Cambridge 1997. (= *Linguistica computazionale* 12-13).

[Vernay 1981]

Vernay, Henry: *Überlegungen zu und Onomasiologie und 'Sprachsystem'*. In: Pöckl, Wolfgang (Hg.): *Europäische Mehrsprachigkeit. Festschrift zum 70. Geburtstag von Mario Wandruszka*. Tübingen 1981, S. 293–299.

[Volmert 1990]

Volmert, Johannes: *Interlexeme im Bereich des Buchstaben 'F'. Beobachtungen bei einem Vergleich von Wörterbüchern aus sechs europäischen Sprachen*. In: Braun, Peter; Schaefer, Burkhard; Volmert, Johannes (Hg.): *Internationalismen. Studien zur interlingualen Lexikologie und Lexikographie*. Tübingen 1990, S. 95–122.

[Wang 1994]

Wang, William S.-Y.: *Glottochronology, Lexicostatistics, and other numerical methods*. In: Asher, R. E.; Simpson, J. M. Y. (Hg.): *The encyclopedia of language and linguistics*. Oxford u.a. 1994, S. 1445–1450.

[Wängler 1963]

Wängler, Hans-Heinrich: *Rangwörterbuch hochdeutscher Umgangssprache*. Marburg 1963.

[Weinreich 1966]

Weinreich, Uriel: *Explorations in Semantic Theory*. In: Sebeok, T. A. (Hg.): *Current Trends in Linguistics*. The Hague 1966, S. 395–477.

[Wellmann 1987]

Wellmann, Hans: *Eine Brücke vom semasiologischen zum onomasiologischen Wörterbuch?* In: Kienpointner, Manfred; Schmeja, Hans (Hg.): *Sprache, Sprachen, Sprechen. Festschrift für Hermann M. Ölberg zum 65. Geburtstag am 14. Oktober 1987*. Innsbruck 1987, S. 195–218. (= *Innsbrucker Beiträge zur Kulturwissenschaft. Germanistische Reihe* 34.).

[Werner 1989]

Werner, Reinhold: *Probleme der Anordnung der Definitionen im allgemeinen einsprachigen Wörterbuch*. In: Hausmann, Franz Josef u. a. (Hg.): *Wörterbücher. Ein internationales Handbuch zur Lexikographie*. Berlin u. a. 1989, S. 917–930. (= *Handbücher zur Sprach- und Kommunikationswissenschaft* 5.1).

[Wichter 1994]

Wichter, Sigurd: *Experten- und Laienwortschätze. Umriß einer Lexikologie der Vertikalität*. Tübingen 1994. (= *Reihe Reihe Germanistische Linguistik* 144).

[Wiegand 1994]

Wiegand, Herbert Ernst: *Zur Unterscheidung von semantischen und enzyklopädischen Daten in Fachwörterbüchern*. In: Schaefer, Burkhard; Bergenholtz, Henning (Hg.): *Fachlexikographie. Fachwissen und seine Repräsentation in Wörterbüchern*. Tübingen 1994, S. 103–132. (= *Forum für Fachsprachen-Forschung* 23).

[Wieland 2000]

Wieland, Astrid: *Grundwortschatz Deutsch-Englisch für Pflegeberufe*. Stuttgart 2000.

[Wildhagen 1972]

Wildhagen, Karl: *English-German, German-English Dictionary. An encyclopaedic and strictly scientific representation of the vocabulary of the modern and present-day languages, with special regard to syntax, style, and idiomatic usage. Volume II: German-English*. 2. Aufl. Wiesbaden u. a. 1972.

[Winter 1977]

Winter, Eugene O.: *A clause-relational approach to English texts: A study of some predictive lexical items in written discourse*. In: *Instructional Science* 6 (1977), S. 1–92.

[Wolski 1995]

Wolski, Werner: *[Rezension zu:] Dieter Krohn: Grundwortschätze und Auswahlkriterien*. Göteborg 1992. In: *Zeitschrift für Dialektologie und Linguistik* 62 (1995), S. 339–342.

[Wängler 1963]

Wängler, Hans-Heinrich: *Rangwörterbuch hochdeutscher Umgangssprache*. Marburg 1963.

[ZERTIFIKAT DEUTSCH 1977]

Deutscher Volkshochschul-Verband (Hg.): *Das Zertifikat Deutsch als Fremdsprache*. 2. Aufl. Bonn-Bad Godesberg 1977.

[Zillig 2000]

Zillig, Werner: *Anmerkungen zum Begriff 'Grundwortschatz'*. In: Beckmann, Susanne; König, Peter-Paul; Wolf, Georg (Hg.): *Sprachspiel und Bedeutung. Festschrift für Franz Hundsnurscher zum 65. Geburtstag*. Tübingen 2000, S. 93–103.

[Zimmer 1990]

Zimmer, Dieter: *Wie viele Wörter hat der Mensch?* In: *Der Sprachdienst* 34 (1990), S. 80–84.

Curriculum Vitæ

Name: Veronika Haderlein

Geburtsdatum: 21. Mai 1972

Geburtsort: München

Schulbildung

1978 - 1985 Rudolf-Steiner-Schule München

1985 - 1992 Gymnasium am Anger München

Abschluss: Allgemeine Hochschulreife

Akademische Bildung

1992 - 1998 Studium der Germanistischen Linguistik, Neueren deutschen Literatur und Betriebswirtschaft

Abschluss: Magister Artium

Publikationen

2005 Semantik bei der Arbeit mit zentralen Wortschätzen - Anforderungen und Möglichkeiten. In: Langer, Stefan; Schnorbusch, Daniel (Hrsg.): Semantik im Lexikon. Tübingen 2005, S. 9-32. (= Tübinger Beiträge zur Linguistik 479).

