

Amerikanismen in der gegenwärtigen britischen Lexikographie
des Englischen

Erfassung, Markierung und Datenbasis

(Band 1 von 2)

Inauguraldissertation
zur
Erlangung der Doktorwürde

der
Philosophischen Fakultät
der
Rheinischen Friedrich-Wilhelms-Universität
zu Bonn

vorgelegt von
Veronika Schilling
aus
Bochum

Bonn 2005

Gedruckt mit Genehmigung der Philosophischen Fakultät der
Rheinischen Friedrich-Wilhelms-Universität Bonn

1. Berichterstatter: Professor Dr. F. W. Gester
2. Berichterstatter: Professor Dr. K. Reichl

Tag der mündlichen Prüfung: 20. Juli 2005

Diese Dissertation ist auf dem Hochschulschriftenserver der ULB Bonn http://hss.ulb.uni-bonn.de/diss_online elektronisch publiziert.

Vorwort

Die vorliegende Dissertation wurde von Professor Dr. F. W. Gester angeregt und in ihrem gesamten Verlauf in allen wissenschaftlichen Fragen stets hilfreich von ihm unterstützt. Für seine gute Betreuung der Arbeit bedanke ich mich herzlich. Mein Dank gilt auch Herrn Professor Dr. Karl Reichl, der das Korreferat übernehmen wird.

Für unseren interessanten fachbezogenen Informationsaustausch, nützliche Hinweise zur Forschungsliteratur und aufschlußreiche Stellungnahmen zu einzelnen Kapiteln der Arbeit danke ich Dr. Michaela Mahlberg (Liverpool). Gedankt sei auch Professor Dr. Joybrato Mukherjee (Gießen) für Auskünfte zu Wortfrequenzen in Korpora. Barbara Schuch danke ich für viele hilfreiche Informationen zur statistischen Bewertung der Ergebnisse meiner Fragebogenerhebung. Carsten Muhlack sei für die große Unterstützung bei der Abfrage der Amerikanismen im BNC gedankt.

Dank gebührt auch den Verlagsvertretern der von mir untersuchten Wörterbücher für ihre aufschlußreichen Auskünfte zu den analysierten Wörterbuchaufgaben und ihren Belegquellen, darunter Wolfgang Worsch (Langenscheidt KG), Steve Crowdy (Longman), Edith Strommen (Penguin Books Deutschland GmbH), Angus Stevenson (Oxford University Press), Patrick Gillard (Cambridge University Press) und Mary O'Neill (Chambers Harrap Publishers Ltd). Ebenfalls danken möchte ich Professor Wolfgang Teubert (Birmingham) für seine Informationen zu den von Harper Collins verwendeten Korpora.

Außerdem möchte ich mich herzlich bei den Kontaktpersonen in Großbritannien und den USA bedanken, die die von mir erstellten Fragebögen an geeignete Informanten vor Ort weitergaben (in alphabetischer Reihenfolge): Christoph und Petra Besch, Helen Bleck, Professor Dr. Martin Durrell, Dotty Frerker, Dr. Gabriele Hille-Coates und Dr. John Coates, Barbara Howard, Professor Dr. Virgil Howard sowie Brian Howard, Andrea, Helmut und Marie-Louise Müller, Frances Klein, Terry Proft, Professor Dr. Irmengard Rauch, Linda Roberts, Eberhard Thyen (vermittelt durch Ute Kalenberg), Sheila und Ken Turner, Familie Wagner, C. J. Wells, Dr. Jonathan West und Birgit Winkler. Ebenfalls zu Dank verpflichtet bin ich den Gewährspersonen selbst, ohne deren geduldige Mitarbeit die Fragebogenerhebung nicht möglich gewesen wäre.

Für ihre jahrelange Unterstützung möchte ich mich vor allem bei meinen Eltern und meinem Mann bedanken. Ihnen sei diese Arbeit gewidmet.

Inhaltsverzeichnis

Abkürzungsverzeichnis		4
1. Einleitung		7
2. Bestimmung des Begriffes ‚Amerikanismus‘		12
3. Die Datenbasen der Wörterbuchverlage (mit besonderer Berücksichtigung von Korpora)		15
3.1 Bedeutung von Korpora für die Lexikographie.....		15
3.1.1 Einleitung.....		15
3.1.2 Definition des Begriffes ‚Korpus‘		18
3.1.3 Vorteile des Korpusgebrauchs für die Lexikographie		19
3.1.4 Vorteile des Korpusgebrauchs für die Lexikographie im Vergleich zu anderen Methoden der Datengewinnung.....		25
3.1.5 Bei der Verwendung von Korpora auftretende Probleme		32
3.1.6 Arten von Korpora.....		36
3.2 Analyse der den Verlagshäusern für die untersuchten Wörterbücher zugrundeliegenden Belegquellen.....		51
3.2.1 Wörterbuchproduktionen des HarperCollins Verlages.....		51
3.2.2 Wörterbuchproduktionen des Addison Wesley Longman Verlages		65
3.2.3 Wörterbuchproduktionen von Oxford University Press		94
3.2.4 Wörterbuchproduktionen von Cambridge University Press.....		98
3.2.5 Wörterbuchproduktionen des Cassell Verlags.....		106
3.2.6 Wörterbuchproduktionen von Chambers Harrap.....		107
3.2.6 Wörterbuchproduktionen des Penguin Verlags.....		109
3.3 Abschließende Beurteilung der den Verlagshäusern für die untersuchten Wörterbücher zugrundeliegenden Belegquellen.....		111
4. Die ausgewählten Wörterbücher		118
4.1 Bedeutung der in den Wörterbüchern außerhalb des Lexikon-Teils gegebenen Informationen zur Berücksichtigung des AmE		118
4.2 Gattungsspezifische Unterschiede zwischen monolingualen Wörterbüchern für fortgeschrittene nicht-englischsprachige Lerner und Wörterbüchern für Muttersprachler der <i>Concise</i> -Sparte		120
4.3 Analyse der für die Berücksichtigung des AmE wichtigen Informationen außerhalb des Lexikon-Teils.....		124

4.3.1	Allgemeine Informationen.....	124
4.3.2	Beteiligung amerikanischer Lexikographen.....	129
4.3.3	Lemmata.....	130
4.3.4	Definitionen und Beispiele.....	138
4.3.5	Darstellung der Unterschiede zwischen dem Amerikanischen und Britischen Englisch in den einleitenden Kapiteln der Wörterbücher.....	140
5.	Begründung der Auswahl der untersuchten Amerikanismen.....	159
6.	Die Fragebogenuntersuchung.....	165
6.1	Einleitung.....	165
6.2	Informationen zur durchgeführten Befragung.....	166
6.2.1	Die Fragebogenkonzeption.....	166
6.2.2	Die Durchführung der schriftlichen Befragung.....	168
6.2.3	Anmerkungen zur Fragebogenerhebung.....	171
7.	Analyse der ausgewählten Amerikanismen in den zehn Wörterbüchern unter Berücksichtigung weiterer Quellen.....	174
7.1	Einleitung.....	174
7.2	Lexikalische Amerikanismen.....	175
7.2.1	block.....	175
7.2.2	junior college.....	182
7.2.3	unlisted.....	186
7.2.4	package.....	188
7.2.5	mailman.....	193
7.2.6	fall.....	196
7.2.7	Britisher.....	200
7.2.8	robin.....	206
7.2.9	chancellor.....	210
7.2.10	table (Verb).....	213
7.2.11	staff.....	217
7.2.12	baggage.....	220
7.2.13	taffy.....	228
7.2.14	granola.....	233
7.2.15	right off.....	236
7.2.16	dry goods.....	240
7.2.17	clipping.....	245

7.2.18	can.....	249
7.2.19	graft.....	257
7.2.20	billion.....	261
7.2.21	chipper.....	266
7.2.22	dumb.....	270
7.2.23	cop out.....	275
7.2.24	patsy.....	280
7.2.25	nerd.....	283
7.2.26	sashay.....	290
7.2.27	chutzpah.....	295
7.2.28	undershirt.....	302
7.2.29	outhouse.....	307
7.2.30	hockey.....	310
7.3	Grammatikalische Amerikanismen.....	315
7.3.1	Unterschiede im Gebrauch des Konjunktivs.....	315
7.3.2	Unterschiede im Gebrauch von Präpositionen.....	320
7.3.3	Unterschiede im Gebrauch von <i>have</i>	331
7.3.4	Unterschiede im Gebrauch des <i>simple past</i> im Zusammenhang mit bestimmten Adverbien.....	336
7.3.5	Unterschiede im Gebrauch von Artikeln.....	342
8.	Analyseergebnisse: Überblick und Bewertung.....	348
	Literaturverzeichnis.....	357
Anhang A: Die Wörterbucheinträge der untersuchten Amerikanismen in Kopie		
Anhang B: Muster der Fragebögen		
Anhang C: Bewertung von Amerikanismen durch die britischen Informanten		
 Bewertung von Britizismen durch die amerikanischen Informanten		
 Zusammenstellung der Fragebogenergebnisse		

Abkürzungsverzeichnis

1. Wörterbücher, Grammatiken und andere Nachschlagewerke

(ausführliche Angaben zu den Werken befinden sich im Literaturverzeichnis)

- AHD₃** = *The American Heritage Dictionary of the English Language*, 3rd ed.
- AHD₄** = *The American Heritage Dictionary of the English Language*, 4th ed.
- BBCED** = *BBC English Dictionary*
- CACD** = *The Cassell Concise Dictionary*
- CACOD** = *The Cassell Compact Dictionary*
- CADS** = *The Cassell Dictionary of Slang*
- CCD₃** = *Collins Concise Dictionary*, 3rd ed.
- CCD₄** = *Collins Concise Dictionary*, 4th ed.
- CCD₅** = *Collins Concise Dictionary*, 5th ed.
- CD₃** = *Collier's Dictionary A-J/ K-Z*, 3rd ed.
- CED₃** = *Collins English Dictionary*, 3rd ed.
- CED₄** = *Collins English Dictionary*, 4th ed.
- CGEL** = *A Comprehensive Grammar of the English Language*
- CGSAE** = *The Columbia Guide to Standard American English*
- CHD** = *The Chambers Dictionary*
- CH21₁** = *Chambers 21st Century Dictionary*, 1st ed.
- CH21₂** = *Chambers 21st Century Dictionary*, rev. ed.
- CIDE₁** = *Cambridge International Dictionary of English*, 1st ed.
- COBUILD₁** = *Collins COBUILD English Language Dictionary*, 1st ed.
- COBUILD₂** = *Collins COBUILD English Dictionary*, 2nd ed.
- COBUILD₃** = *Collins COBUILD English Dictionary for Advanced Learners*, 3rd ed.
- COD₉** = *Concise Oxford Dictionary*, 9th ed.
- COD₁₀** = *Concise Oxford Dictionary*, 10th ed. revised

COD₁₁	= <i>Concise Oxford Dictionary</i> , 11 th ed. revised
LDOCE₂	= <i>Longman Dictionary of Contemporary English</i> , 2 nd ed.
LDOCE₃	= <i>Longman Dictionary of Contemporary English</i> , 3 rd ed.
LDOCE₄	= <i>Longman Dictionary of Contemporary English</i> , 4 th ed.
LDELC₁	= <i>Longman Dictionary of English Language and Culture</i> , 1 st ed.
LDELC₂	= <i>Longman Dictionary of English Language and Culture</i> , 2 nd ed.
MEDAL	= <i>Macmillan English Dictionary for Advanced Learners</i>
MWCD₁₀	= <i>Merriam-Webster's Collegiate Dictionary</i> , 10 th ed.
NODE₁	= <i>New Oxford Dictionary of English</i> , 1 st ed.
NPED₁	= <i>The New Penguin English Dictionary</i> , 1 st ed.
OALD₄	= <i>Oxford Advanced Learner's Dictionary of Current English</i> , 4 th ed.
OALD₅	= <i>Oxford Advanced Learner's Dictionary of Current English</i> , 5 th ed.
OALD₆	= <i>Oxford Advanced Learner's Dictionary of Current English</i> , 6 th ed.
OCED	= <i>The Oxford Compact English Dictionary</i>
OCEL	= <i>The Oxford Companion to the English Language</i>
ODMS	= <i>The Oxford Dictionary of Modern Slang</i>
OED₂	= <i>The Oxford English Dictionary</i> , 2 nd ed.
OED₃	= <i>The Oxford English Dictionary</i> , 3 rd ed.
RHD₂	= <i>The Random House Dictionary of the English Language</i> , 2 nd ed.
RWRT	= <i>The Right Word at the Right Time</i>
TBD	= <i>Third Barnhart Dictionary of New English</i> , 3 rd ed.
WDEU	= <i>Webster's Dictionary of English Usage</i>
WED	= <i>Encarta World English Dictionary</i>
WNWD₃	= <i>Webster's New World Dictionary of American English</i> , 3 rd college ed.
WTNID₄	= <i>Webster's Third New International Dictionary of the English Language</i> , 4 th ed.

2. Korpora/ Datenbanken

ANC	= <i>American National Corpus</i>
BNC	= <i>British National Corpus</i>
BoE	= <i>Bank of English</i>
Brown Corpus	= <i>Brown University Standard Corpus of Present-Day Edited American English</i>
CLSC	= <i>Cambridge Language Survey Corpus</i>
COBUILD	= <i>Collins Birmingham University International Language Database</i>
FLOB Corpus	= <i>Freiburg-LOB Corpus of British English</i>
Frown Corpus	= <i>Freiburg-Brown Corpus of American English</i>
ICE	= <i>International Corpus of English</i>
ICLE	= <i>International Corpus of Learner English</i>
LCN	= <i>Longman Corpus Network</i>
LLC	= <i>Longman Learner's Corpus</i>
LLELC	= <i>Longman/ Lancaster English Language Corpus</i>
LOB Corpus	= <i>Lancaster-Oslo/Bergen Corpus of British English</i>
WBE	= <i>Wolverhampton Business English Corpus</i>

1. Einleitung

Wörterbücher sind mehr denn je von großer Bedeutung in einer Zeit, in der die internationale Staatengemeinschaft immer mehr zusammenwächst und die Notwendigkeit der Kommunikation zwischen Menschen verschiedenster Nationalitäten - durch den technologischen Fortschritt bedingt - mehr und mehr zunimmt. Wörterbücher sind eines der wichtigsten Hilfsmittel, eine fremde Sprache verstehen bzw. selber sprechen und schreiben zu können.

Unter den Sprachen kommt dem Englischen eine besondere Rolle zu. Das Englische ist die Sprache, die in der Welt am meisten verbreitet ist, denn sie wird auch von vielen hundert Millionen Menschen gesprochen, die eine andere Muttersprache haben.¹ Die englische Sprache besteht heute aus vielen Varietäten, wobei aber die beiden ältesten, größten und weitverbreitetsten, das Britische und Amerikanische Englisch², unbestreitbar als Bezugsnormen für andere Varietäten dienen und daher oft gegenübergestellt wurden und werden.³ Während das BrE, die Sprache des Britischen Empire, bis etwa gegen Ende des 19. Jahrhunderts einen stärkeren Einfluß auf das AmE ausübte als umgekehrt, scheint sich das Verhältnis seit der ersten Hälfte des 20. Jahrhunderts gewendet zu haben. Die Vereinigten Staaten sind zu einer Welt- und Wirtschaftsmacht geworden, die die internationale Politik und den internationalen Handel in entscheidendem Maße beeinflusst. Auf der Kinoleinwand und im Fernsehen dominieren weltweit schon seit langem amerikanische Filme, und bestimmte Musikrichtungen, die in den Vereinigten Staaten entstanden, wie z.B. der Jazz, haben in vielen Ländern Anklang gefunden. Die amerikanische Computerindustrie ist führend auf dem Weltmarkt und nicht zuletzt in Fragen der Kleidung und Nahrungsmittel wird nicht nur Großbritannien schon seit vielen Jahrzehnten von den Vereinigten Staaten beeinflusst. Dies sind die Hauptgründe, warum das AmE weltweit Verbreitung gefunden hat.⁴

Wenn man sich überdies vor Augen hält, daß die Sprecher des AmE allen anderen Muttersprachlern des Englischen zahlenmäßig ungefähr im Verhältnis zwei zu eins und den

¹ Den Angaben der Forschungsliteratur zufolge ist die Zahl der Sprecher des Englischen als Zweit- oder Fremdsprache um ein Vielfaches höher als die der Muttersprachler des Englischen: Der ersten Gruppe werden 1 - 2 Milliarden Sprecher, der zweiten ca. 350 Mio. Sprecher zugeordnet. Vgl. Crystal (1985), Strevens (1992:28), McArthur (1997:11) und Bobda (1998:15).

² Das Britische Englisch wird im folgenden mit BrE, das Amerikanische Englisch mit AmE abgekürzt. Unter BrE wird die Form der englischen Sprache verstanden, wie sie in Großbritannien verwendet wird. Unter AmE wird die Form der englischen Sprache verstanden, wie sie in den USA verwendet wird. Zur Problematik der Begriffe *American English* und *British English* vgl. das *Oxford Companion to the English Language* (im folgenden mit OCEL abgekürzt), S. 37, 41 und 156 und Algeo (1989b:151ff.).

³ Vgl. z.B. das OCEL, S. 41, Algeo (2001:xvii), Strevens (1992:32) und Ilson (1985b:7). Im OCEL wird darauf hingewiesen, daß der Wunsch nach kontrastiven Darstellungen des BrE und AmE die zugrundeliegende Überzeugung widerspiegeln, daß die Gemeinsamkeiten zwischen beiden Varietäten größer seien als die Unterschiede (S. 42).

⁴ Vgl. das OCEL, S. 38, Bailey (2001:493f.) und Bobda (1998:14f.).

Sprechern des BrE sogar fast im Verhältnis vier zu eins überlegen sind⁵, wird offensichtlich, daß das AmE in Wörterbüchern des Englischen als Gemeinsprache in angemessenem Umfang repräsentiert sein sollte.

Die Verfasserin dieser Arbeit war im besonderen daran interessiert, zu untersuchen, ob die britische Lexikographie des Englischen aus der jüngsten Zeit der heutigen Bedeutung des AmE durch eine umfangreiche Aufnahme ausschließlich oder vorwiegend im AmE auftretender Lexeme, Bedeutungen, Wendungen, grammatikalischer Formen u.ä. Rechnung getragen hat. Es erschien überdies lohnenswert, neben der Erfassung auch die Markierung der Amerikanismen⁶ in den Wörterbüchern zu überprüfen. In der Forschungsliteratur der letzten Jahre wird immer wieder kritisiert, daß die Markierungssysteme und Markierungen von Amerikanismen in britischen Wörterbüchern des Englischen stark voneinander abweichen.⁷ So führte der finnische Linguist Juhani Norri 1996 beispielsweise einen Vergleich zwischen sieben britischen und drei amerikanischen Wörterbüchern des Englischen bezüglich ihres Gebrauchs verschiedener diatopischer Markierungen, darunter auch *American/ U.S.*, durch und stellte fest, daß die Markierungen der untersuchten Lexeme in den zehn Wörterbüchern teilweise deutlich divergieren. Zudem bemängelte er, daß die jeweilige Markierungspraxis in den Einleitungen der Wörterbücher nur unzureichend dargelegt werde. Da Norri eine Reihe von britischen Wörterbüchern analysierte, die auch in der vorliegenden Arbeit berücksichtigt werden, es sich bei seinen Ausgaben aber um ältere Auflagen der hier untersuchten handelt, bot sich ein Vergleich zwischen Norris Ergebnissen und den in dieser Arbeit gewonnenen an. Anders als bei Norri konnten die in der vorliegenden Arbeit analysierten Markierungen zudem auch anhand von eigenen empirischen Untersuchungen überprüft werden (s.u.). Auf dieses Mittel griff auch Matsuda (1990) zurück, der mit Hilfe von Befragungen britischer Informanten herausfand, daß die Markierungen von Amerikanismen in verschiedenen britischen Wörterbüchern des Englischen und in der *Lexicographic Description of English* von Benson/ Benson/ Ilson (1986) nicht immer zutreffend sind. Da die Fragebogenerhebungen jedoch in den achtziger Jahren durchgeführt wurden und leider auch nur eine bestimmte soziale Gruppe, nämlich Universitätsangehörige, sowie eine relativ geringe Zahl von Informanten involvierte⁸, bot es sich an, seine Ergebnisse in dieser Arbeit zumindest teilweise zu überprüfen. Matsudas (1990:111) Einschätzung, daß Markierungsunterschiede in den Wörterbüchern u.a. dadurch entstehen können, daß dem Tempo des Sprachwandels von heute, vor allem auch dem Einfluß des AmE auf das BrE (der durch elektronische Kommunikationsformen wie dem Internet zusätzlichen

⁵ Vgl. das OCEL, S. 37.

⁶ Der Terminus "Amerikanismus" wird in Kap. 2 der Arbeit näher erläutert.

⁷ Vgl. z.B. Norri (1996:3 ff., 11ff.), Mittmann (2000:580ff.) und Pätzold (1997: 196).

Antrieb erhält) nicht genügend Aufmerksamkeit geschenkt wird⁹, ist zumindest ohne Einschränkung beizupflichten.

Da sich im Rahmen dieser Arbeit nicht alle kürzlich in Großbritannien veröffentlichten Wörterbücher des Englischen untersuchen ließen, mußte notwendigerweise eine Auswahl getroffen werden. In einem Land, in dem Englisch in den meisten Fällen die erste zu lernende Fremdsprache ist, bot sich eine Analyse der neueren Auflagen der einbändigen, monolingualen Lernerwörterbücher¹⁰ für fortgeschrittene Lerner in Hinblick auf ihren Einbezug des AmE an. Von diesen Lernerwörterbüchern wird beispielsweise in den höheren Klassen bzw. Jahrgangsstufen der weiterführenden Schulen sowie im Anglistikstudium an den Universitäten reger Gebrauch gemacht. Den fünf ausgewählten, in Deutschland sehr gängigen Lernerwörterbüchern wurden fünf ebenfalls sehr gebräuchliche Wörterbücher für Muttersprachler der *Concise-Sparte* gegenübergestellt, die über einen ähnlichen Umfang verfügen wie die Lernerwörterbücher. Hierdurch sollte überprüft werden, ob aufgrund der divergierenden Adressatenkreise und Zielsetzungen der zwei Wörterbuchtypen die Erfassung und Markierung von Amerikanismen in den zwei Gattungen unterschiedlich sein würde.

Die folgenden fünf Wörterbücher für fortgeschrittene Lerner wurden ausgewählt:

- *Oxford Advanced Learner's Dictionary*, 6. Aufl. (OALD₆),
- *Longman Dictionary of Contemporary English*, 3. Aufl. (LDOCE₃),
- *Longman Dictionary of English Language and Culture*, 2. Aufl. (LDEL₂),
- *Collins COBUILD English Dictionary for Advanced Learners*, 3. Aufl. (COBUILD₃),
- *Cambridge International Dictionary of English* (CIDE₁),

sowie fünf Wörterbücher für Muttersprachler der *Concise-Sparte*

- *Cassell Concise Dictionary* (CACD)
- *Concise Oxford Dictionary*, 10. Aufl. (COD₁₀)
- *Collins Concise Dictionary*, 5. Aufl. (CCD₅)
- *Chambers 21st Century Dictionary*, 2. Aufl. (CH21₂) und
- *New Penguin English Dictionary* (NPED₁).

⁸ Vgl. Kap. 5.

⁹ Vgl. auch Norri (1996:4).

¹⁰ Vgl. Jehles (1990:5) Bestimmung des Begriffs "Lernerwörterbuch".

Acht der zehn Wörterbücher basieren auf computerisierten Textkorpora. Korpora von mindestens 100 Mio. Wörtern sind für die überwältigende Mehrheit der britischen Lexikographen zu unverzichtbaren ‚Werkzeugen‘ geworden, die ihrer Ansicht nach sowohl in qualitativer wie quantitativer Hinsicht alle bisher zur Verfügung stehenden Informationsquellen zur Erstellung von Wörterbüchern übertreffen. In der Tat weist die Verwendung von Korpora sehr viele Vorzüge gegenüber herkömmlichen Methoden zur Datengewinnung auf, die sich auch - so die Annahme der Verfasserin dieser Arbeit - mit großer Wahrscheinlichkeit bei der Erfassung und Markierung von Amerikanismen in einem Wörterbuch positiv bemerkbar machen. Daher wurde der Frage nachgegangen, ob sich der mangelnde Korpusgebrauch bei zweien der berücksichtigten Wörterbuchverlage in Bezug auf das Thema dieser Arbeit nachteilig auswirkt. Der Nutzen eines Korpus ist allerdings an verschiedene Kriterien wie Größe, Aktualität und Zusammensetzung gebunden, denn diese Faktoren sind entscheidend dafür, welche Informationen dem Korpus entnommen werden können. Glücklicherweise hatte die Verfasserin dieser Arbeit die Möglichkeit, mehr oder weniger detaillierte Informationen zu den Korpora zu erhalten, die den o.g. Wörterbuchprodukten von *Oxford University Press*, *Longman*, *Harper Collins*, *Cambridge University Press* und *Chambers Harrap* zugrunde lagen. Somit war eine Beurteilung und ein Vergleich dieser Datenbasen bezüglich des Umfangs, der Variationsbreite und Aktualität von AmE Sprachdaten möglich. Aufgrund der dargelegten Bedeutung der Datenbasen, insbesondere der Korpora, die den Wörterbuchverlagen zur Verfügung standen, wird dieser Aspekt in der vorliegenden Arbeit ausführlich behandelt. Die eingehende Auseinandersetzung mit der Korpusmethodik spiegelt zudem die Einschätzung der Verfasserin dieser Arbeit wider, daß weitere Verbesserungen hinsichtlich der Erfassung und Markierung verschiedener Varietäten des Englischen in Wörterbüchern nur durch Verbesserungen im zugrundeliegenden Korpusmaterial erreicht werden können.¹¹

Bei der Untersuchung der zehn Wörterbücher hinsichtlich ihrer Aufnahme und Darstellung von Amerikanismen war es im Rahmen dieser Arbeit leider nicht möglich, die Gesamtzahl der jeweils in diesen Wörterbüchern erfaßten Amerikanismen zu ermitteln sowie die Markierungen aller in den Wörterbüchern auftretenden Amerikanismen zu überprüfen. Daher mußte eine bestimmte Anzahl von Amerikanismen ausgewählt werden, die in jedem Wörterbuch hinsichtlich ihrer Aufnahme und Markierung analysiert wurde. Bei diesen handelt es sich nur um solche der Lexik und Grammatik.¹²

Bei der Analyse der betreffenden Wörterbucheinträge wurde leider deutlich, daß die in den zehn untersuchten Wörterbüchern vorhandenen Auskünfte zu den ausgewählten Amerikanismen

¹¹ Vgl. Kap. 3.

¹² Vgl. Kap. 5.

stellenweise erheblich divergieren. Um die Qualität dieser Wörterbuchinformationen überprüfen zu können, zog die Verfasserin der vorliegenden Arbeit zum einen Angaben aus der Forschungsliteratur zu den Unterschieden zwischen dem BrE und AmE sowie Informationen aus anderen Wörterbüchern und weiteren Quellen heran, und führte zum anderen eine Fragebogenerhebung mit Muttersprachlern des BrE und AmE sowie Recherchen am *British National Corpus* (BNC) durch.

Im folgenden Kapitel wird zunächst eine kurze Einführung in die Definitionsgeschichte des Wortes "Amerikanismus" gegeben und schließlich die Bedeutung des Terminus dargelegt, die dieser Arbeit zugrunde liegt. Kapitel 3 beschäftigt sich mit der Verwendung von Korpora in der Lexikographie, auch im Vergleich zu anderen Methoden der Datengewinnung, sowie mit den Datenbasen der behandelten Wörterbuchverlage. Kapitel 4 geht zum einen auf gattungsspezifische Unterschiede zwischen monolingualen Wörterbüchern für fortgeschrittene Lerner und Wörterbüchern für Muttersprachler der *Concise*-Sparte ein und wertet zum anderen solche von den Wörterbuchherausgebern angeführten Informationen außerhalb der eigentlichen Einträge aus, die für den Untersuchungsgegenstand der Arbeit wichtige Aufschlüsse enthalten. In Kapitel 5 wird dargestellt, aufgrund welcher Kriterien die im Rahmen dieser Arbeit überprüften Amerikanismen ausgewählt wurden. In Kapitel 6 werden Informationen zu der Fragebogenerhebung gegeben, die von der Verfasserin durchgeführt wurde, um bei divergierenden Informationen zu den untersuchten Amerikanismen in den Wörterbüchern Auskünfte von *native speakers* zum Gebrauch dieser Wörter, Wortbedeutungen, Wendungen und grammatikalischen Konstruktionen zu erhalten. Kapitel 7 geht auf die Erfassung und Markierung der 35 im Rahmen dieser Arbeit behandelten Amerikanismen in den zehn Wörterbüchern ein. Zusätzlich werden in jedem Unterkapitel zu einem Amerikanismus Angaben aus anderen Quellen (Forschungsliteratur, Nachschlagewerke, Grammatiken etc.), die Resultate der Fragebogenuntersuchung sowie Ergebnisse von Recherchen am *British National Corpus* hinzugezogen. Kapitel 8 liefert eine Zusammenfassung und Bewertung der gewonnenen Erkenntnisse zur Erfassung und Markierung der ausgewählten Amerikanismen in den zehn Wörterbüchern mit Bezug auf die den Wörterbuchverlagen zur Verfügung stehenden Belegquellen.

2. Bestimmung des Begriffes ‚Amerikanismus‘

Der Terminus *Americanism* wurde 1781 von dem Schotten John Witherspoon geprägt.¹³ Witherspoon, der 1768 Rektor des College of New Jersey geworden war (das später zur Princeton University wurde), verwendete diesen Terminus zum ersten Mal in einem in der Zeitschrift *Pennsylvania Journal and The Weekly Advertiser* veröffentlichten Aufsatz.¹⁴ Er verstand unter *Americanism* - in Analogie zu dem Begriff *Scotticism* - "ways of speaking peculiar to this country" bzw. „an use of phrases or terms, or a construction of sentences, even among persons of rank and education, different from the use of the same terms or phrases, or the construction of similar sentences in Great-Britain“¹⁵. Durch die apologetische Formulierung „even among persons of rank and education“ wird deutlich, daß Witherspoon den Begriff *Americanism* noch als pejorativ ansah. Auch heute wird er beispielweise in Großbritannien noch manchmal in dieser Weise verwendet.¹⁶

Mathews (1963:13) zufolge war Witherspoon der erste, der ausführlich über die Unterschiede zwischen dem AmE und BrE schrieb. Mit ihm begann eine Zeit der intensiven Beschäftigung mit "Amerikanismen", die mit Namen wie John Pickering, John Russell Bartlett, Richard Grant White, Maximilian von Schele De Vere verbunden ist und auf die hier nicht näher eingegangen werden soll.¹⁷ Festzustellen ist jedoch, daß der Terminus Amerikanismus im Laufe seiner Verwendungsgeschichte zahlreiche Definitionen erfahren hat. So bevorzugt nicht zuletzt auch Mathews (1963:31) eine andere Bestimmung des Terminus *Americanism* als die von Witherspoon gewählte. Er kritisiert, daß gemäß Witherspoons Definition solche Ausdrücke oder Konstruktionen, die in den USA geprägt wurden, aber mittlerweile vollkommen in den gemeinenglischen¹⁸ Sprachschatz eingegangen sind, nicht mehr als Amerikanismen bezeichnet werden. Obwohl er sich im folgenden der Definition des OED anschließt ("A word or phrase peculiar to, or extending from, the United States"), kommt es ihm vor allem darauf an, in den USA entstandene Wörter wie z.B. *lengthy* und *breadstuff* als Amerikanismen bezeichnen zu können, auch wenn diese schon lange nicht mehr ausschließlich oder vorwiegend im AmE verwendet werden. Dazu heißt es bei Wächtler (1980:146): "Mathews beschränkt sich damit auf

¹³ Vgl. die 2. Auflage des *Oxford English Dictionary* (im folgenden mit OED₂ abgekürzt) sowie Mathews (1963:31), Görlach (1990a:257), Wächtler (1980:146) und das OCEL, S. 47.

¹⁴ Vgl. Mathews (1963:14).

¹⁵ Beide Textstellen sind zitiert nach Mathews (1963:16f.).

¹⁶ Vgl. die Antworten zur letzten Frage des von der Verfasserin dieser Arbeit konzipierten Fragebogens für die britischen Probanden in Anhang C.

¹⁷ Vgl. Mencken (1977:97-103), Crowell (1967), Maurer (1972), Görlach (1990a:257) und Görlach (1990b:1478-1482).

¹⁸ Mit dem Begriff "gemeinenglisch" werden solche Erscheinungsformen der englischen Sprache bezeichnet, die entweder allen Varietäten des Englischen oder zumindest dem BrE und AmE eigen sind.

einen rein historischen Amerikanismus-Begriff". Diese Definition von *Americanism* liegt auch dem von Mathews herausgegebenen *Dictionary of Americanisms on Historical Principles* zugrunde, das nur solche Wörter und Bedeutungen aufführt, die ihren Ursprung in den Vereinigten Staaten haben.¹⁹

Für heutige kommerzielle Gebrauchswörterbücher, die, wie ihr Name schon erkennen läßt, weder historisch ausgerichtet sind noch sprachwissenschaftliche Zwecke verfolgen, ist eine Definition des Terminus Amerikanismus "*par origine*" allerdings wenig sinnvoll, dagegen aber eine "*par usage*"²⁰ notwendig. Man kann daher Wächtler vollkommen zustimmen, der feststellt:

Für Zwecke der Sprachanwendung sind Amerikanismen zunächst solche Wörter, die im AmE häufiger sind als im BrE oder in irgendeiner anderen nationalen Varietät des Englischen bzw. in der Varietät, die das DCE [1. Auflage, meine Hinzufügung] 'world English' nennt. (1980:147)²¹

Umso mehr erstaunt es, daß das CIDE₁ die folgende Erklärung unter dem Eintrag *Americanism* anführt:

An Americanism is a word or expression which originated in the United States but is used by people in other countries, esp. those where English is spoken: '*Have a nice day!*' is an Americanism which many British people find extremely irritating. Far from polluting the English language, I would say Americanisms such as '*show off*' and '*beat up*' have enriched it.²²

Die Definition des Terminus Amerikanismus „*par origine*“ läßt sich also auch noch in einem Lernerwörterbuch aus dem Jahre 1995 finden. Auch im COD₁₀ wird der Aspekt des amerikanischen Ursprungs eines als Amerikanismus geltenden Wortes bzw. einer solchen Wendung ausdrücklich erwähnt: „A word or phrase peculiar to or originating in the United

¹⁹ Vgl. dessen Vorwort, S. v in Bd. 1, und das OCEL, S. 47.

²⁰ Vgl. Wächtler (1980:146f.). Algeo (1992c:287) unterscheidet zwischen einem *diachronic Americanism* („expression that originated in America, whatever its current use may be“) und einem *synchronic Americanism* („expression with characteristic form or use in America, whatever its origin may have been“).

²¹ Wächtler bezieht sich in seinem Aufsatz nur auf lexikalische Amerikanismen, wobei er aber nicht ausschließt, daß nicht auch Besonderheiten des AmE z.B. im Bereich der Aussprache, Orthographie und Grammatik als Amerikanismen bezeichnet werden könnten.

²² Aus den in dem Eintrag zu dem Begriff *Americanism* angefügten Beispielen des CIDE lassen sich immer noch die gemischten Gefühle der Briten gegenüber Amerikanismen heraushören. Daß diese Gefühle heute anscheinend immer noch bestehen, wird in der folgenden Äußerung Algeos (2001:xxvii) sichtbar: „Indeed, it is not unusual for educated and otherwise linguistically sophisticated English men and women to assume that any new linguistic feature they encounter is American in origin, and therefore objectionable on two grounds“. Vgl. auch z.B. die Ausführungen von Algeo (1989: 220), Algeo (1990), Bailey (2001:459), Ison (1985a:181), Ison (1990a:38-40), Gramley und Pätzold (1992:358), Horlacher (1990), und Schur (1987:1-4) zu der Einstellung der Briten gegenüber dem AmE und der Amerikaner gegenüber dem BrE, sowie die Ergebnisse der letzten Frage meiner zwei Fragebögen in Anhang C.

States.“²³ In den anderen acht in dieser Arbeit untersuchten Wörterbüchern liegt das Augenmerk dagegen auf einer Definition „par usage“. Allerdings könnten die Amerikanismus-Definitionen, vor allem im OALD₆, LDOCE₃ und COBUILD₃, noch deutlich verbessert werden, wenn man sie mit Wächtlers differenzierter Begriffsbestimmung vergleicht:

Ein lexikalischer Amerikanismus ist eine Wortform bzw. Wortbedeutung, die - vom heutigen gemeinenglischen (britischen und amerikanischen) Lexikonbestand her gesehen - nur oder vorwiegend oder als normbedingt bevorzugte (häufige) Zweitmöglichkeit im (nord-)amerikanischen Umlaufraum des Englischen gebraucht wird.(1980:156)

Der Terminus Amerikanismus wird im folgenden Teil der Arbeit in diesem Sinne verwendet, wobei die Verfasserin, ebenso wie Algeo (1989b:156), auch sprachliche Phänomene des AmE in der Grammatik, Aussprache, Orthographie und Pragmatik, die ausschließlich oder überwiegend im AmE vorkommen, unter den Begriff Amerikanismus faßt.²⁴ Diese werden z.B. in den Definitionen im CIDE₁, COD₁₀ und COBUILD₃ genauso wenig einbezogen wie solche Bedeutungen, die ausschließlich oder überwiegend im AmE geläufig sind. Wie unzulänglich die Bestimmungen von Nationalismen in Wörterbüchern oft sind, weist Algeo (1989b:144ff.) anhand seiner Analyse von neunzehn Wörterbüchern nach (darunter sind jedoch keine der in dieser Arbeit untersuchten). Er fordert zurecht: „...most dictionaries need a more consistent definition schema for nationalism terms than we now have. The lexicographer also needs a clearer concept of what constitutes a nationalism“ (1989b:157).

Die unzureichende Erklärung des Lemmas *Americanism* in manchen der in dieser Arbeit untersuchten Wörterbüchern hat allerdings keine ersichtlich negativen Auswirkungen auf die Erfassung und Markierung von Amerikanismen in der hier verwandten Bedeutung des Terminus. So wurden in das CIDE₁ auch solche Lexeme oder grammatikalischen Formen aufgenommen, die (fast) nur im AmE verwendet werden, dort aber nicht entstanden sind (z.B. *fall* n. für 'autumn' und *gotten*), und im COD₁₀ und COBUILD₃ sind z.B. auch solche Bedeutungen enthalten, die häufiger, aber nicht ausschließlich im AmE benutzt werden (vgl. beispielweise die Einträge zu *table* v. und *graft* n.).

²³ Im COD₁₁ und in der zweiten Auflage des CIDE wird das Wort *Americanism* interessanterweise nicht mehr definiert.

²⁴ Vgl. auch das OCEL, S. 47, in dem allerdings nicht die orthographischen Besonderheiten des AmE mit dem Begriff *Americanism* bezeichnet werden.

3. Die Datenbasen der Wörterbuchverlage (mit besonderer Berücksichtigung von Korpora)

3.1 Bedeutung von Korpora für die Lexikographie

3.1.1 Einleitung

Over the past ten to fifteen years, the discipline of lexicography has changed almost beyond recognition. This change is due to the technological revolution which has computerized the lexicographers' working environment to a very high degree and which has permitted a veritable quantum leap in the amount and variety of resources that can be brought to bear on the lexicographical process. The most important of these resources are computerized corpora of real, mostly written, but now increasingly also spoken, running text.

Meijs (1996:99)

Heute werden kaum noch monolinguale Wörterbücher des Englischen in Großbritannien produziert, die nicht auf umfangreichen computerisierten Textkorpora basieren²⁵. Alle großen britischen Wörterbuchverlage wie *Oxford University Press*, *Addison Wesley Longman*, *HarperCollins*, *Chambers Harrap* und *Cambridge University Press* verfügen mittlerweile über Korpora im Umfang von mind. je 100 Mio. Wörtern.²⁶ Von Willem Meijs (1996:100) Mitte der 90er Jahre dazu befragt, inwieweit sich die lexikographische Praxis in den letzten zwanzig Jahren durch Korpora verändert habe, gaben alle leitenden Redakteure der o.g. Verlagshäuser übereinstimmend an, daß sie Lexikographie ohne computerisierte Korpusdaten heutzutage für praktisch undenkbar halten trotz unterschiedlicher Vorstellungen hinsichtlich des Aufbaus und der Nutzung dieser Korpora.²⁷ Auch Rundell/ Stock (1992a:14) betrachten Korpora als unverzichtbare ‚Werkzeuge‘ jedes Lexikographen²⁸ und sprechen von einer ‚Korpusrevolution‘

²⁵ Vgl. Kennedy (1998:91), Moon (1998:347), McEnery/ Wilson (1996:129), Rundell (1996a:200) und Rundell (1996b:1). Somit hat Sinclair (1985:83) recht behalten, der schon 1985 einen Boom in der Entwicklung von Korpora für die neunziger Jahre und danach vorhersagte. Vgl. auch Biber/ Conrad/ Reppen (1996:115), Svartvik (1996:11), Svartvik (1992:8), Aijmer/ Altenberg (1991:2) zum stark wachsenden Gebrauch von Korpora in verschiedenen Forschungsgebieten und Sprachen. Zur Definition des Begriffes ‚Textkorpus‘ oder kurz ‚Korpus‘ s. Kapitel 3.1.2.

²⁶ Vgl. dazu auch Meijs (1996:99).

²⁷ Vgl. dazu ebenfalls Sinclair (1993:5), Svartvik (1996:9f.) Rundell/ Stock (1992c:51), Teubert (1998:151,156) und Moon (1998:347). Kennedy (1998:91) geht noch einen Schritt weiter, indem er die Ansicht vertritt, daß jegliche Art lexikalischer Beschreibung auf einem Korpus basieren sollte: „It is now almost inconceivable that worthwhile and comprehensive lexical descriptions can be undertaken without a corpus.“ Und bei Rundell (1996a:200) heißt es sogar: „there can be little doubt that electronic text corpora will eventually be seen as an essential prerequisite for all forms of language description“. (Vgl. auch Rundell 1996a: 214 und Rundell 1996b:1,8.)

²⁸ Vgl. dazu ebenso Rundell/ Stock (1992b:21), Rundell/ Stock (1992c:51) und Rundell (1996b:1). Mit dem Begriff ‚Lexikograph‘ sollen im folgenden der Einfachheit halber sowohl männliche als auch weibliche Vertreter dieser Berufsgruppe bezeichnet werden.

innerhalb der Lexikographie in Großbritannien seit Anfang der 80er Jahre²⁹, denn durch Korpora seien dem Lexikographen Sprachdaten an die Hand gegeben worden, die in quantitativer wie qualitativer Hinsicht alle Informationsquellen in den Schatten stellten, die ihm früher für seine Arbeit zur Verfügung gestanden hätten, und so zu ganz neuen Einsichten in den Sprachgebrauch führten.³⁰ Inwieweit sich die Verbesserungen in der Güte der Wörterbuchinformationen auch hinsichtlich der Erfassung und Markierung³¹ von Amerikanismen bemerkbar machen, wird in den nächsten Kapiteln Untersuchungsgegenstand sein.

Die Verwendung von Korpora in der Lexikographie ist mittlerweile so gängig, daß Rundell (1996a:214) z.B. solchen neuen Lernerwörterbüchern, die nicht auf Korpora basieren, aufgrund der seiner Einschätzung nach mangelnden Glaubwürdigkeit ihrer Informationen schlechtere Absatzchancen auf dem Wörterbuchmarkt voraussagt. Verlage, die ihre Wörterbücher dagegen mit Hilfe von Korpora erstellen, werben ausdrücklich mit der Zuverlässigkeit der in den Wörterbüchern gegebenen Informationen. So heißt es beispielsweise auf dem Buchrücken der sechsten Auflage des OALD in der Lizenzausgabe des Cornelsen Verlags: „Die Oxford University Press bürgt mit dem British National Corpus und der Oxford Corpus Collection für die Authentizität und sprachliche Korrektheit der Stichwörter und Anwendungsbeispiele.“ Inwieweit bei der Anfertigung eines jeden Wörterbucheintrags die von den Verlagen zitierten Korpora tatsächlich herangezogen wurden³² oder wie die Lexikographen im Einzelfall die Korpusdaten verwandten, bleibt Außenstehenden leider verborgen.³³ Nach Ansicht Krishnamurthys (2002a:6) bestand beispielsweise bei solchen Verlagen, die vor dem Gebrauch von Korpora frühere Auflagen eines Wörterbuchs veröffentlicht hatten, ein gewisser Widerstand, den Wörterbuchtext auf der Basis der Korpusbelege grundlegend zu verändern, da viel Geld in die Erstellung des bestehenden Textes geflossen sei. Oft ist aber bedauerlicherweise auch der für ein Wörterbuchprojekt festgesetzte enge Zeitrahmen dafür mitverantwortlich, daß sich die Lexikographen nicht immer eingehend mit Korpusdaten auseinandersetzen können.

²⁹ Vgl. den Titel der Artikelreihe von Rundell/ Stock in *English Today* 1992 sowie Rundell/Stock (1992c:51), Rundell (1996a:198, 200, 212, 214) und Rundell (1996b:1). Siehe auch Agor (2000:3), Krishnamurthy (2002a:1) und Tognini Bonelli (2001:48).

³⁰ Vgl. Rundell/ Stock (1992a:14) und Tognini Bonelli (2001:48). Hierauf wird unter 3.1.3. noch näher eingegangen.

³¹ Unter den Begriff der ‚Markierung‘ wird - anders als bei Hausmann (1989:649ff.) - sowohl das einzelne „Markierungsetikett“ als auch die Gesamtheit der „Markierungsetikette“ gefaßt.

³² So kann die Verfasserin dieser Arbeit auf Grund ihrer Mitarbeit am *Harrap Compact German Dictionary*, einem bilingualen Wörterbuch (Englisch-Deutsch, Deutsch-Englisch), beispielweise berichten, daß bei diesem Wörterbuchprojekt das *British National Corpus* (im folgenden mit BNC abgekürzt) nicht herangezogen wurde, obwohl es dem Verlag *Chambers Harrap* zur Verfügung steht. Fairerweise muß man allerdings hinzufügen, daß auch nirgendwo in diesem Wörterbuch behauptet wird, man hätte zu seiner Abfassung von Korpora Gebrauch gemacht.

³³ Zum *corpus-driven* bzw. *corpus-based approach* vgl. Moon (1998:347), Ooi (1998:51f.) und vor allem Tognini-Bonelli (2001:65ff.).

Anders als im Fall von früheren Wörterbuchausgaben, deren Datengrundlagen sich der Beurteilung von außen entzogen, haben Rezensenten nun sogar teilweise die Möglichkeit, dieselben Korpora zu benutzen und zu analysieren, die die Grundlage oder einen Teil der Datenbasis für die neusten Auflagen bilden³⁴. Die Wörterbuchverlage geben zudem selbst Informationen über die Größe und die Zusammensetzung ihrer Korpora, so daß mit Blick auf das Thema dieser Arbeit folgende Aspekte untersucht werden können:

1. Inwieweit findet das AmE Eingang in die Korpora? Je größer der Anteil AmE Daten an der Gesamtdatenmenge des Korpus ist, umso umfangreicher ist wahrscheinlich auch die Aufnahme von Amerikanismen im Wörterbuch, deren Bedeutung sich zudem anhand der größeren Vorkommenshäufigkeit mancher Amerikanismen im Korpus in einigen Fällen besser ermitteln lassen wird. Gleichzeitig läßt sich der tatsächliche Gebrauch bestimmter Amerikanismen im AmE bei einer umfangreicheren AmE Datenmenge besser einschätzen. Das kann zu einer genaueren Markierung besonders der Amerikanismen führen, die im AmE irgendeiner Beschränkung unterliegen, z.B. hinsichtlich des Registers bzw. der Situation, in der sie verwendet werden.

2. Wie ‚repräsentativ‘³⁵ sind die Korpora für die nationalen Varietäten einer Sprache, die sie widerspiegeln sollen, in diesem Falle für das AmE? Die Variationsbreite der benutzten AmE Quellen für ein Korpus ist bezüglich dieser Frage von Wichtigkeit, ob also fiktionale und non-fiktionale Texte, geschriebenes und gesprochenes Englisch einbezogen wurden, wieviele Sachgebiete die Quellen umfassen etc. Davon hängt die Zahl der in einem Wörterbuch erfaßten Amerikanismen einer Textsorte, eines Sachgebietes oder verschiedener Registererebenen ab.

3. Wie aktuell sind die verwandten Quellen? Sprache ist in einem ständigen Wandel begriffen. Innerhalb von weniger als einem Jahrzehnt kann z.B. eine Wortform oder eine Wortbedeutung, die vorwiegend im AmE verwendet wurde, in den gemeinenglischen Wortschatz übergegangen sein.

Der derzeitige ‚Korpusboom‘ wäre selbstverständlich ohne die großen Fortschritte der letzten Jahrzehnte im Bereich der Computertechnologie nicht denkbar, die zu einer vorher nicht für möglich gehaltenen Datenverarbeitungs- und speicherkapazität von Computern führte. Anfang der achtziger Jahre konnte man die Anzahl einschlägiger elektronischer Korpora, die Sprachwissenschaftler im Laufe der vorangegangenen zwei Jahrzehnte erstellt hatten, noch an einer Hand abzählen. Diese frühen computerisierten Korpora konnten zunächst auch nur auf

³⁴ So z.B. im Fall des BNC, das seit 1994 für Wörterbuchproduktionen der Verlage *Oxford University Press*, *Addison Wesley Longman* und *Chambers Harrap* verwendet wird, oder der *Bank of English*, mit deren Hilfe der *HarperCollins* Verlag seine Wörterbücher erstellt.

³⁵ Der Begriff ‚repräsentativ‘ wird hier und im folgenden nicht im statistischen Sinne des Wortes gebraucht.

Großrechnern verwendet werden, die gemessen an heutigen Standards sehr langsam arbeiteten.³⁶ Erst die Entwicklung leistungsstarker Mikrocomputer seit Mitte der siebziger Jahre und das Aufkommen von CD-ROMs sowie Scannern in den achtziger Jahren machte Forschungen, die auf Korpora basierten, einem viel größerem Publikum möglich.³⁷ Seit den neunziger Jahren ist die Vielzahl der weltweit zu findenden abgeschlossenen oder noch andauernden Korpusprojekte sowie die Bandbreite an elektronischen Archiven und anderen Datenbanken kaum noch überschaubar. Dies liegt nicht zuletzt daran, daß zum einen die Zahl maschinenlesbarer bzw. online verfügbarer, kostengünstiger (und teilweise sogar kostenloser) Texte in den letzten Jahren enorm zugenommen hat und zum anderen die zur Korpuserstellung notwendige Hardware und Software weiterhin immer preiswerter, bedienerfreundlicher und leistungsstärker wird.

Im folgenden soll zunächst erläutert werden, was man unter einem ‚Korpus‘ versteht. In weiteren Unterkapiteln wird es u.a. um die Vorteile und Probleme des Korpusgebrauchs in der Lexikographie, um den Aufbau von Korpora und um zukünftige Entwicklungen im Korpusbereich gehen, die für die Verbesserung der Qualität der Wörterbücher von großem Interesse sind. Alle diese Aspekte werden hier mit Blick auf den Untersuchungsgegenstand dieser Arbeit diskutiert.

3.1.2 Definition des Begriffs ‚Korpus‘

Aufgrund der Fülle der Forschungsliteratur zum Thema ‚Korpus‘ gibt es auch eine Vielzahl von Definitionen dieses Begriffs³⁸, die teilweise deutlich voneinander abweichen. Im OED₂ heißt es zu der Bedeutung des Wortes in dem hier untersuchten Sinne: ‚A body of written or spoken material upon which a linguistic analysis is based‘. Diese Definition beinhaltet nur einen Teil der Aspekte, die für eine Erläuterung des Terminus, gerade auch im Hinblick auf die in dem Kapitel 3.2. behandelten Korpora der Wörterbuchverlage, wichtig erscheinen. Eine sehr umfassende und somit auch präzise Begriffsbestimmung liefert Tognini-Bonelli (2001:55) in Auseinandersetzung mit Definitionen anderer Sprachwissenschaftler: „a corpus is taken to be a computerised collection of authentic texts, amenable to automatic or semi-automatic processing or analysis. The texts are selected according to explicit criteria in order to capture the regularities of a language, a language variety or a sub-language.“ In dieser Erläuterung des Terminus kommt u.a.

³⁶ Kennedy (1998:7) gibt dazu folgendes Beispiel: In den 70er Jahren habe es eine Stunde oder länger dauern können, bis der Großrechner alle Belege für ein Wort wie z.B. *when* in einem Korpus, das eine Million Wörter umfasste, ausgegeben hätte. Heutzutage kann ein PC diese Leistung (bei weitaus größeren Korpora) je nach Frequenz des abgefragten Wortes oder der Wendung innerhalb von Minuten oder sogar Sekunden erbringen.

³⁷ Vgl. Kennedy (1998:4) und Rundell (1996b:6).

³⁸ Vgl. zu dieser Problematik z.B. Bergenholtz/ Mugdan (1989:141) und Bergenholtz (1996:257ff.). Definitionen des Terminus ‚Korpus‘ finden sich u.a. bei Clear (1993:163), Aston/ Burnard (1997:4), Kennedy (1998:1), Bergenholtz/ Mugdan (1989: 141), Sinclair (1991: 171), Sinclair (1998:113f.), Francis (1992:17), Bergenholtz

zum Ausdruck, daß ein Korpus so weit als möglich 'repräsentativ' für eine bestimmte Sprache oder Varietät sein soll, damit sich aus der Analyse des Korpus Rückschlüsse auf diese ziehen lassen. Auf der Annahme, daß ein solches Vorgehen legitim sei, beruhen nicht zuletzt die neusten Wörterbuchproduktionen. Entscheidend sind aber dann die Kriterien zur Erstellung eines Korpus, die gewährleisten sollen, daß das Korpus beispielweise für eine bestimmte Sprache oder nationale Varietät exemplarischen Charakter hat. Auf diese Kriterien, die auch in Tognini-Bonellis Begriffsbestimmung Erwähnung finden, wird in den folgenden Kapiteln noch näher einzugehen sein.

Die oben angeführte Definition von Tognini-Bonelli läßt sich noch dahingehend ergänzen, daß neben oder anstelle von vollständigen Texten auch Teiltexthe bei der Zusammenstellung mancher Korpora verwendet werden. Hinzuzufügen ist auch, daß das für ein Korpus gesammelte Sprachmaterial aus geschriebenen und/ oder gesprochenen Quellen stammen kann.³⁹

3.1.3 Vorteile des Korpusgebrauchs für die Lexikographie

Schon seit Samuel Johnsons *Dictionary of the English Language* (1755) wurde von empirischen Daten für die Erstellung von größeren Wörterbüchern Gebrauch gemacht. Man sammelte in mühevoller Kleinarbeit Zitate aus authentischen, meist literarisch bedeutenden Quellen, die in erster Linie dazu benutzt wurden, die Bedeutungen der Stichwörter zu illustrieren.⁴⁰ Das umfangreichste in dieser Weise durchgeführte Wörterbuchprojekt führte zur Erstellung der ersten Auflage des OED von 1858 - 1928. Selbst heute im Zeitalter computerisierter Korpusdaten werden Wörterbücher immer noch mit Hilfe dieser Methode der Datengewinnung produziert.⁴¹ So ist auf dem Schutzumschlag des im Jahre 2000 neu erschienenen *The New Penguin English Dictionary* (NPED₁) von „examples drawn from the works of major writers, from Shakespeare to Jane Austen to P G Wodehouse to Salman Rushdie“ die Rede. In diesem Wörterbuch wird an keiner Stelle auf die Verwendung eines Korpus als Grundlage für die Anfertigung der Einträge im NPED₁ verwiesen. Dasselbe gilt auch für *The Cassell Concise Dictionary* (CACD). Damit sind diese beiden Wörterbücher die einzigen unter den zehn in dieser Arbeit untersuchten, die nicht auf Korpora basieren. Da die Verwendung von Korpora für die

(1996:259), Bergenholtz/ Mugdan (1990:1619), Schaefer (1990:1642), Tognini-Bonelli (2001:2ff., 52ff.) und im OED (1992:265f.).

³⁹ In der deutschen Forschungsliteratur wird manchmal der Terminus ‚Textkorpus‘ dem Begriff ‚Korpus‘ mit der Begründung vorgezogen, daß er verdeutliche, daß es sich bei einem Korpus um eine Sammlung von Texten und/ oder Teiltexthen handele und nicht um eine Auflistung isolierter Sätze oder einzelner Wörter. Da im weiteren bezüglich der Bedeutung von ‚Korpus‘ die Definition von Tognini-Bonelli zugrunde gelegt wird, erscheint es der Verfasserin nicht notwendig, in den folgenden Kapiteln ausdrücklich von ‚Textkorpora‘ zu sprechen. Dasselbe gilt für Termini wie ‚Referenzkorpus‘ oder ‚Auswahlkorpus‘, die dasselbe bezeichnen, was im weiteren unter dem Begriff ‚Korpus‘ verstanden wird.

⁴⁰ Vgl. Meijs (1996:100), McEnery/Wilson (1996:90), Rundell/ Stock (1992a:10), Rundell/Stock (1992b:21), Sinclair (1985:83), Francis (1992:18ff.), Biber/ Conrad/ Reppen (1998: 21f.) und Sinclair (1991:41).

Erstellung von Wörterbüchern jedoch von großem Nutzen ist, und der Gebrauch anderer Methoden zur Datengewinnung demgegenüber einige Nachteile aufweist, wie im folgenden noch darzulegen sein wird, stellt sich die Frage, ob die Informationen zu den im NPED₁ und CACD verzeichneten Stichwörtern qualitativ vergleichbar sind mit denen der anderen acht im Rahmen dieser Arbeit analysierten Wörterbüchern. Dieser Frage soll in Kap. 7 nachgegangen werden.

Für eine gewinnbringende Auswertung eines Korpus ist der Einsatz von Softwareprogrammen, die die Erstellung einer sog. Konkordanz für eine beliebige Wortform oder Phrase ermöglichen, unverzichtbar. Unter einer Konkordanz im korpuslinguistischen Sinne versteht man die Auflistung aller (oder einer Auswahl von) Vorkommen eines bestimmten sprachlichen Phänomens zusammen mit dem jeweils vorher spezifizierten Quantum an Kontext. Das am häufigsten benutzte Konkordanzformat - im Englischen *KWIC (Key Word In Context) concordance* genannt - gestaltet sich folgendermaßen: dem Beleg für eine Wortform wird je eine Zeile der Konkordanz zugeordnet, in deren Mitte sich die Wortform selbst befindet.⁴² Rechts und links von ihr steht der dazugehörige Kontext, der je nach Wunsch auch länger als nur einzeilig gewählt werden kann. Oft gibt es zudem die Option, Wörter rechts oder links neben der Wortform oder Phrase alphabetisch sortiert anzeigen zu lassen.⁴³ Der Lexikograph kann also innerhalb von Sekunden (oder allenfalls Minuten bei hochfrequenten Wörtern) alle Belege für ein Wort oder eine Wendung aus vielen Millionen laufender Wörter im Korpus aufrufen und zwischen verschiedenen Darstellungsweisen dieser Belege wählen.

Mit Hilfe der Konkordanz kann sich der Lexikograph die Daten eines Korpus auf verschiedene Weise zunutze machen⁴⁴:

1. Die Auflistung der Vorkommen eines Wortes kann dem Lexikographen helfen, sein Vorwissen hinsichtlich des semantischen Gehalts dieses Wortes zu überprüfen. Dabei kommen manchmal neue Bedeutungen ins Blickfeld, die dem intuitiven Sprachurteil des Lexikographen

⁴¹ Vgl. dazu McEnery/Wilson (1996:90).

⁴² Zur Definition des Begriffes ‚Konkordanz‘ bzw. *KWIC concordance* in der Korpuslinguistik vgl. z.B. Rundell (1996a:201) Biber/ Conrad/ Reppen (1998:26), Bergenholtz / Mugdan (1990:1620), Sinclair (1985:85), Sinclair (1991:43,170,173), Kilgarriff (1997:11), Aston/ Burnard (1997:7) und Meijs (1996:106).

⁴³ Dies vereinfacht das Erkennen von Kollokationen. Vgl. dazu McEnery/Wilson (1996:91).

⁴⁴ Der Gebrauch von Korpora ist natürlich auch für linguistische Forschungsbereiche von großem Nutzen, so z.B. für komparative Studien der regionalen Varietäten einer Sprache, für vergleichende Analysen verschiedener Sprachen und für Untersuchungen von Sprachwandel und Unterschieden zwischen verschiedenen Textsorten oder Sprachbenutzergruppen. Außerdem sei erwähnt, daß Korpora besonders im *Natural Language Processing (NLP)* eine große Rolle spielen, aber zunehmend auch in den Übersetzungswissenschaften, für die Erstellung von Grammatiken und im *English Language Teaching (ELT)* - Sektor an Bedeutung gewinnen. Vgl. z.B. Kennedy (1998:1,9), Rundell (1996a: 212ff.) und vor allem Aston/ Burnard (1997:12ff.). Auf weitere Anwendungsgebiete von Korpusdaten gehen Atkins/ Clear/ Ostler (1992:13ff.), Leech (1993:10), Lenders/ Willée (1998:45ff.) McEnery/Wilson (1996:Kap. 4 und 5), Tognini-Bonelli (2001:1), Rundell (1996b:1, 7f.), Sinclair (1993:5), Clear (1993:183ff.) und Lenz (2000:15f.) ein.

sonst verborgen geblieben wären.⁴⁵ Insbesondere übertragene Bedeutungen sind ohne die Hinzuziehung von Korpusbeispielen schwer zu ermitteln. Dasselbe gilt für die Differenzierung zwischen vermeintlichen Synonymen: der Kontext und die Genres, in denen diese gebraucht werden, geben Aufschluß über subtile Unterschiede in Bedeutung und Gebrauchweise. Außerdem gibt die semantische Analyse der Korpusbelege dem Lexikographen mehr Sicherheit bei der Entscheidung, wie er die verschiedenen Bedeutungsaspekte eines Wortes in einem Wörterbucheintrag darstellen will: ob er, abhängig von der Zielgruppe des Wörterbuchs, z.B. weniger und abstraktere Definitionen anbieten möchte, die eine Bandbreite von Bedeutungsnuancen abdecken und weniger Platz im Wörterbuch einnehmen, oder ob er konkretere und damit mehr Definitionen zur Verfügung stellen will, die dem Benutzer Möglichkeiten sowohl zur Dekodierung als auch zur Enkodierung eines Wortes geben.

2. Die Satzkonstruktionen, in denen ein Wort für gewöhnlich vorkommt, dienen als Grundlage dafür, das grammatische Verhalten dieses Wortes zu beschreiben.⁴⁶ Bei umstrittenem Sprachgebrauch kann durch Korpusbelege aufgezeigt werden, wie Muttersprachler ein Wort tatsächlich benutzen. So läßt sich laut Rundell/Stock (1992b:24) anhand von Analysen des *Longman/Lancaster English Language Corpus* u.a. nachweisen, daß ein Drittel der Belege der angeblich im AmE bevorzugten Präposition *than* nach dem Adjektiv *different* aus britischen Texten stammt und amerikanische Autoren für gewöhnlich nach diesem Adjektiv *from* verwenden. Erkenntnisse letzterer Art führen zu einer exakteren varietätenspezifischen Markierung des grammatischen Verhaltens eines Wortes. Ist das anhand der Korpusbeispiele ersichtliche grammatische Verhalten eines Wortes zu differenzieren, um durch die syntaktische Kodierung in einem Wörterbuch wiedergegeben werden zu können, besteht immerhin noch die Möglichkeit, es in den für das Wort angeführten Beispielen zu verdeutlichen. Außerdem kann man durch die Analyse eines Korpus feststellen, inwieweit scheinbar ähnliche grammatische Strukturen in unterschiedlichen Kontexten verwendet werden und unterschiedliche Funktionen haben (vgl. z.B. den Gebrauch von „Have you got...?“ und „Do you have...?“, s. Kap. 7.3.3).

⁴⁵ Vgl. dazu auch Rundell/Stock (1992b:22f., 28). Die Bedeutungen eines Wortes oder einer Wendung sind keine fest fixierten Einheiten, deshalb ist es für Lexikographen auch oft schwer, Bedeutungsunterscheidungen zu treffen. Vgl. Sinclair (1991:7) und Kilgarriff (1997). Bei Meijs (1996:109) heißt es dazu treffend: „Meaning distinctions are not god-given entities“.

⁴⁶ Im Englischen bezeichnet man seit John Rupert Firth das gemeinsame Auftreten eines Wortes mit bestimmten grammatischen Wortklassen als *colligation* (vgl. Aston/ Burnard 1997:14). Rundell/Stock (1992b:24) und Rundell (1996a:203) geben als Beispiel das Verb *decide* an, dessen Komplementierung (vgl. *decide to, decide that, decide whether* etc.) und die jeweilige Häufigkeit dieser Satzergänzungen mittels Korpusanalysen mit sehr viel größerer Sicherheit bestimmt werden könnten. Für das Thema dieser Arbeit ist jedoch vor allem das in verschiedenen Varietäten divergierende syntaktische Verhalten von Lexemen interessant, das anhand von Korpusuntersuchungen nachgewiesen werden kann, vgl. Rundells (1996a:203) Ausführungen zu *protest*, das im AmE transitiv gebraucht werden kann (s. auch Kap. 7.3.2 dieser Arbeit).

3. Der Kontext kann auch aufschlußreich dafür sein, mit welcher Intention ein Wort häufig gebraucht wird (z.B. um Ironie oder Mißbilligung auszudrücken). Zudem sind in den meisten Korpora die Texte, aus denen die Korpora bestehen, mit sog. ‚situational tags‘ ausgestattet, die Auskünfte über die spezifischen Besonderheiten eines Textes (regionale Zuordnung, Erscheinungsdatum, Sachgebiet, Textsorte und Geschlecht des Autors) geben, die für die Markierungspraxis wichtig sein können.

4. Das Korpus kann als Quelle für Beispiele dienen, die in Wörterbucheinträgen den Bedeutungserklärungen zu einem Lexem oder einer Wendung hinzugefügt werden. Der Vorteil von Korpusbelegen gegenüber selbst ausgedachten Beispielen ohne Korpusbezug besteht darin, daß der Lexikograph anhand der authentischen Sprachdaten die typischen Gebrauchsweisen eines sprachlichen Phänomens relativ schnell und problemlos erkennen und Beispiele aus dem Korpus auswählen kann, die diese am besten veranschaulichen. Nach Meijs (1996:102) sind die Ansichten der Lexikographen hinsichtlich der Verwendung von Anwendungsbeispielen, die dem Korpus direkt entnommen wurden, allerdings geteilt. Redakteure der COBUILD Publikationen legten großen Wert darauf, authentische Sätze aus dem Korpus als Anwendungsbeispiele in ihren Wörterbucheinträgen zu gebrauchen, die - wenn nötig - nur geringfügig verändert (z.B. gekürzt) würden⁴⁷. Dagegen dienten den meisten anderen Wörterbuchmachern Sätze aus dem Korpus oft eher als Anregung für die selber erstellten Beispiele in den Wörterbüchern. Auch im letzteren Fall sind Korpusbeispiele den Angaben von Rundell/Stock (1992b:28) zufolge sehr nützlich, denn sie erlaubten es, Beispiele dahingehend zu optimieren, daß sie die typischen Verwendungsweisen eines Wortes aufzeigten.

5. Informationen zur Vorkommenshäufigkeit eines Wortes im Korpus - differenziert nach unterschiedlichen Bedeutungen, Wortarten-Status oder sogar Flexionsform - können für Lexikographen sehr nützlich sein.⁴⁸ So kann beispielsweise die Frequenz vermeintlicher Synonyme verglichen werden.⁴⁹ Ebenso hilfreich sind Auskünfte zur Frequenz bestimmter morphologischer Formen oder grammatischer Konstruktionen (auch z.B. in Abgrenzung zu anderen, ähnlichen Formen oder Konstruktionen). Untersuchungen zur Auftretenshäufigkeit eines sprachlichen Phänomens in verschiedenen Textsorten, Altersgruppen u.ä. können ebenfalls sehr aufschlußreich sein. Ein Vergleich der Frequenz von Wörtern, Bedeutungen, Wendungen, grammatischen Konstruktionen etc. in verschiedenen Varietäten des Englischen ist für die

⁴⁷ Vgl. auch Sinclair (1991:5), der sich entschieden gegen die Verwendung von erfundenen Beispielen ausspricht. Als einen Grund für diese Position gibt er an, daß es sehr schwer sei, sich den Kontext auszudenken, in den ein Stichwort in einem Anwendungsbeispiel eingebettet ist. Solche Beispiele kämen in der Sprachrealität höchstwahrscheinlich nicht vor. Vgl. zur Problematik des Gebrauchs selbstgebildeter Beispiele auch Kap. 3.1.4 dieser Arbeit.

⁴⁸ Vgl. zur Verwendung von Frequenzinformationen in der Lexikographie vor allem Scholfield (o.J.), Summers (1996) und Schaefer (1990:1640f.).

Vergabe von diatopischen Markierungen in einem Wörterbuch von großer Wichtigkeit, nicht zuletzt auch für die Kennzeichnung von Amerikanismen. Anhand der Informationen zur Frequenz ist es zudem z.B. einfacher, festzustellen, welche Wörter, Bedeutungen, Wendungen oder grammatische Konstruktionen aufgrund ihrer Seltenheit im Korpus nicht in jedem Fall in ein mittelgroßes oder kleines Wörterbuch aufgenommen werden müssen. Überdies wird den Lexikographen mit Hilfe von Frequenzinformationen die Entscheidung erleichtert, in welcher Reihenfolge die Bedeutungen eines Wortes oder einer Wendung im Wörterbuch angeführt werden sollten, und ob man eine flektierte Form auf Grund ihrer Vorkommenshäufigkeit auch als separates Stichwort angeben sollte (beispielsweise *amusing* als Adjektiv und als -ing -Form von *amuse*). Werden Anwendungsbeispiele dem Korpus direkt entnommen, kann die hohe Frequenz bestimmter, mit dem jeweiligen sprachlichen Phänomen verbundener Gebrauchsmuster die Auswahl und Anordnung typischer Belege vereinfachen.

Aus Korpusanalysen entspringende Häufigkeitsangaben zu Lexemen werden zunehmend in Lernerwörterbücher integriert (vgl. z.B. das LDOCE₃ und das COBUILD₃) und sind für den Sprachenlerner sehr hilfreich, weil er durch diese vor Augen geführt bekommt, welche Wörter besonders häufig vorkommen und daher vorrangig gelernt werden sollten.⁵⁰ Die Redakteure von *Longman* bedienten sich des BrE und AmE Korpusmaterials des Verlags außerdem auch dazu, dem Benutzer des LDOCE₃ anschauliche graphische Darstellungen zur Häufigkeit und somit Gebräuchlichkeit bestimmter Wörter bzw. Bedeutungen im BrE und AmE an die Hand zu geben.

6. Durch ein Korpus kann man Informationen über Neologismen in einer Sprache oder regionalen Varietät erhalten.⁵¹ Hinsichtlich der im Korpus vertretenen regionalen Varietäten läßt sich z.B. untersuchen, inwieweit Wörter, Wendungen, Bedeutungen, grammatische Konstruktionen u.ä., die in ihrer Verwendung auf eine bestimmte Varietät beschränkt waren, auch in andere Varietäten Eingang finden. Erkenntnisse dieser Art sind auch für die Markierung von Amerikanismen in einem Wörterbuch entscheidend. Rundell (1996a:205) führt als Beispiel an, daß AmE Baseball-Ausdrücke im BrE auf immer größere Akzeptanz trafen. Zur Beobachtung solcher Entwicklungen eignen sich am besten diachrone *vergleichbare Korpora*⁵², d.h. in diesem Fall mehrere möglichst ähnlich strukturierte und gleich große

⁴⁹ Vgl. Rundell (1998:335).

⁵⁰ Bei solchen Häufigkeitsangaben wird von seiten der Lexikographen natürlich auch darauf geachtet, daß das jeweilige Wort in vielen verschiedenen Quellen nachgewiesen ist, also nicht den idiosynkratischen Sprachgebrauch einzelner Autoren widerspiegelt.

⁵¹ Voraussetzung dafür ist aber, daß das Korpus die aktuellen Entwicklungen in der Sprache widerspiegelt, wie es beispielsweise ein *monitor corpus* tut. Vgl. Kap. 3.1.6.2.

⁵² Unter einem *vergleichbaren Korpus* versteht Sinclair (1998:126) eines, „das ähnliche Texte in mehr als einer Sprache oder Varietät enthält“. In diesem Zusammenhang wird gelegentlich in der Forschungsliteratur auch der Terminus *Parallelkorpus* benutzt, unter dem aber in der Regel folgendes gefaßt wird: „eine Sammlung von Texten, von denen neben der originalsprachlichen Fassung eine oder mehrere Übersetzung(en) vorliegen“ (Sinclair 1998:126).

varietätenspezifische Korpora, die zeitgleich erstellt werden, und diachrone Vergleichsmöglichkeiten innerhalb einer Varietät und zwischen Varietäten bieten, wenn nach einer gewissen Zeitspanne ein zweiter Satz von Korpora desselben Aufbaus und derselben Größe erstellt wird. (vgl. auch Kap. 3.1.6.1.).

7. Mithilfe eines Korpus läßt sich feststellen, in welchen mehr oder weniger festen Wortverbindungen Wörter häufig auftreten⁵³. Dieser Aspekt wird als wesentlich für das Verstehen der Bedeutung und Funktion von Wörtern erachtet⁵⁴. Kollokationen geben wichtige Hinweise in dem Prozeß der Unterscheidung bestimmter Bedeutungen eines Wortes. Somit kann die Analyse der Kollokationen eines Lexems auch zur Präzisierung der Bedeutungserklärungen zu diesem Wort führen sowie zu einer genaueren Darstellung der Verwendungsweise des Lexems. Die nachgewiesene Häufigkeit sog. ‚recurrent word combinations‘ (darunter versteht man nicht nur Kollokationen, sondern auch idiomatische Wendungen, *phrasal verbs* und Sprachformeln) in den verschiedensten Textsorten eines Korpus, vor allem aber in der gesprochenen Sprache, legen eine eingehendere Behandlung dieser Wortverbindungen vor allem in Lernerwörterbüchern nahe.

Die oben angeführten Anwendungsbereiche von Korpora kommen selbstverständlich auch der Erfassung und Markierung von Amerikanismen in Wörterbüchern zugute. Besteht ein Korpus, wie z.B. das BNC, allerdings (fast) nur aus BrE Sprachmaterial, ergibt sich das Problem, daß dort nur Daten zu solchen Amerikanismen zur Verfügung stehen, die auch im BrE verwendet werden. Da diese Amerikanismen im BrE aber per definitionem weniger häufig benutzt werden als im AmE, findet man für sie oft eine nicht ausreichende Zahl an Belegen vor, um treffsichere Aussagen über ihre Bedeutung und ihren Gebrauch machen zu können. Außerdem muß der Lexikograph schon a priori vor der Analyse verschiedener varietätenspezifischer Sprachdaten mittels Introspektion entscheiden, was als Amerikanismus anzusehen ist: Dies führt wieder in die Zeiten traditioneller Methoden der Vergabe von Markierungen zurück. Ein direkter Vergleich zwischen der Vorkommenshäufigkeit, Bedeutung und Verwendung eines Amerikanismus im AmE und BrE (sofern er in beiden Varietäten auftritt) ist also unerlässlich, denn nur dann ist es möglich, in einem Wörterbuch präzise Aussagen über diesen Amerikanismus zu treffen (so ist - wie die Untersuchungen der Verfasserin zeigen⁵⁵ - das Lexem ‚Britisher‘ heutzutage nicht nur im BrE, sondern auch im AmE selten und müßte daher wahrscheinlich mit *old-fashioned* markiert werden; fraglich ist auch, ob eine Kennzeichnung als

⁵³ Vgl. Svartvik (1996:10).

⁵⁴ Vgl. dazu McEnery/Wilson (1996:92) und Sinclair (1991:6f.). Bei letzterem heißt es: „[...] each meaning can be associated with a distinctive formal patterning. So regular is this that in due course we may see formal patterns being used overtly as criteria for analysing meaning [...]“. Siehe auch Sinclair (1993:12) und Tognini-Bonelli (2001:4f.).

Amerikanismus immer noch gerechtfertigt ist⁵⁶). Empfehlenswert ist also, daß auch AmE Daten in großer Zahl und nach vorher festgelegten Prinzipien in ein Korpus aufgenommen werden bzw. unterschiedliche Korpora für verschiedene Varietäten des Englischen zur Verfügung stehen, die aber in Größe und Aufbau vergleichbar sein sollten. Der letztere Fall bietet am ehesten die Möglichkeit eines fundierten Vergleichs der Bedeutung, Verwendung und Auftretenshäufigkeit von Wörtern und Wendungen in den verschiedenen Varietäten, wobei sich die erstellten Korpora im Zuge der wachsenden Bedeutung anderer Varietäten als dem BrE oder AmE nicht auf diese beiden beschränken sollten.⁵⁷

Leider kann man weder den Informationen der hier besprochenen Wörterbuchverlage noch der Forschungsliteratur entnehmen, wie hoch der Frequenzunterschied der Verwendung eines sprachlichen Phänomens zwischen dem AmE und anderen Varietäten des Englischen sein muß, damit dieses von den Lexikographen beispielsweise als *especially AmE* oder *AmE* markiert werden kann.

3.1.4 Vorteile des Korpusgebrauchs für die Lexikographie im Vergleich zu anderen Methoden der Datengewinnung

Die Vorteile von Korpora gegenüber herkömmlichen Belegensammlungen sind offensichtlich⁵⁸:

1. Statt einer willkürlich getroffenen Auswahl von einzelnen Belegen, die für ein Wörterbuch zusammengetragen werden, handelt es sich bei einem Korpus (im Idealfall) um eine umfangreiche und für eine oder mehrere Varietäten des Englischen weitestgehend ‚repräsentative‘ Sammlung von Texten. Für die Erstellung von Wörterbüchern wie dem *Dictionary of the English Language* (1755) beschränkte sich die Auswahl der Quellen, denen die Belege entnommen wurden, auf sog. ‚authorities‘ (z.B. bekannte literarische Werke sowie Abhandlungen aus den Fachbereichen der Philosophie, Geschichtswissenschaften, Naturwissenschaften und Theologie) und zwar zumeist auf solche, die nach subjektiver Beurteilung für besonders wertvoll gehalten wurden.⁵⁹ Somit hatte die normative Funktion des Wörterbuchs einen weitaus größeren Stellenwert als eine größtmögliche ‚Repräsentativität‘ der

⁵⁵ Vgl. Kap. 7.2.7.

⁵⁶ Vgl. dazu auch die Amerikanismus-Definition in Kap. 2.

⁵⁷ Vgl. zum Einsatz von Korpora für Varietätenstudien Kap. 3.1.6.1.

⁵⁸ Vgl. dazu McEnery/Wilson (1996:90f.), McEnery/ Wilson (o.J.:18), Sinclair (1985:85), Sinclair (1991:4, 39f., 43, 100, 169), Rundell (1996a:198f.,201), Rundell/ Stock (1992a:10ff.), Rundell/ Stock (1992c:51), Summers (1993:185f.), Biber/ Conrad/ Reppen (1998:22f.), Biber/ Conrad/ Reppen (1994:176), Aston/ Burnard (1997:6), Agor (2000:3), Clear (1993:164ff.), Clear et al. (1996:307f.), Fillmore (1992:39), Hanks (1990:34ff.), Ooi (1998:48, 51), Quirk (1992:459,467), Bergenholtz/ Mugdan (1990:1618f.), Johansson (1995:244), Krishnamurthy (2002a:1) und Schaefer (1990:1642).

⁵⁹ Vgl. Samuel Johnsons Vorwort zu dem von ihm herausgegebenen Wörterbuch. Schon der Untertitel zum *Dictionary of the English Language* macht den selektiven Charakter der Quellenauswahl augenscheinlich: „A Dictionary of the English Language in which the words are deduced from their originals and illustrated in their different significations by examples from the best writers“.

im Wörterbuch enthaltenen Sprachdaten für die jeweilige Sprache. Die gesprochene Sprache fand in diesen Wörterbüchern so gut wie keinen Niederschlag, und Textsorten wie z.B. Zeitungsartikel wurden erst in jüngerer Zeit einbezogen. Außerdem war es bei sehr großen Wörterbüchern, wie beispielsweise dem OED, früher unverzichtbar, daß eine hohe Anzahl linguistisch in der Regel nicht geschulter freier Mitarbeiter daran beteiligt war, Belege für die in dem Wörterbuch zu verzeichnenden Stichwörter zu suchen. So blieb es weitgehend der Entscheidung dieser Leser überlassen, welche Daten sie für die Aufnahme in ein Wörterbuch für wichtig hielten. Ungewöhnliche und seltene sprachliche Phänomene fielen eher ins Auge, weswegen ihnen häufig größere Aufmerksamkeit geschenkt wurde als solchen, die gängig und bekannt waren. Letztere wurden oft vernachlässigt oder übersehen. Im Gegensatz zu Belegsammlungen ist eine Konkordanz zum einen exhaustiv, d.h. sie listet alle zur Verfügung stehenden Vorkommen eines sprachlichen Phänomens auf (wodurch bei ausgewogener Korpusstruktur im Kontrast zu Belegsammlungen aussagekräftige Frequenzen ermittelt werden können, die für fast alle Bereiche der Sprachbeschreibung von großer Wichtigkeit sind, s.o.). Weil keine Belege übersehen werden, ist die Wahrscheinlichkeit größer, daß z.B. alle wichtigen Bedeutungen eines Wortes repräsentiert sind. Die Anzahl der Belege für jedes Lexem (oder jede seiner Bedeutungen oder Verwendungen) wird nicht von vornherein begrenzt, wie das bei den meisten Belegsammlungen üblich ist. Zum anderen sind die mit Hilfe der Konkordanz gewonnenen Ergebnisse leicht reproduzierbar. Außerdem sind Belegsammlungen durch die Begrenzung des Kontextes auf den Satzrahmen gekennzeichnet, während man bei Korpusuntersuchungen auch Zugriff auf größere Textabschnitte oder ganze Texte erhalten kann, deren Berücksichtigung bezüglich des zu analysierenden sprachlichen Phänomens bei vielen lexikographischen Fragestellungen von Vorteil ist. Im Kontrast zu Belegsammlungen wird bei der Erstellung eines Korpus, das bei Wörterbuchproduktionen herangezogen werden soll, auf die Einbeziehung einer großen Bandbreite verschiedener Textsorten und Sachgebiete geachtet, damit das Korpus als möglichst aussagekräftig für alle im Korpus vertretenen nationalen Varietäten einer Sprache gelten kann. Dabei spielt auch die beeindruckende Größe der heutzutage für lexikographische Zwecke genutzten Korpora eine Rolle, die im Gegensatz zu herkömmlichen Belegsammlungen eine weitaus größere Fülle von Daten verfügbar macht. Durch die systematische Analyse des in der Regel sehr viel umfangreicheren und nach bestimmten Kriterien ausgewählten Datenmaterials eines Korpus können verlässlichere und präzisere Informationen in den Wörterbucheinträgen gegeben werden (s.o.). Der auf Computerkorpora basierende Ansatz, der speziell für Sprachwissenschaftler entwickelte Softwareprogramme verwendet, schafft die Voraussetzung für eine Vielzahl neuer Forschungsmöglichkeiten zum Sprachgebrauch. Ein schon bereits angeführtes Beispiel betrifft die Untersuchung von

Wortkombinationen. Die Möglichkeit, neben individuellen Wörtern auch Wortkombinationen abrufen zu können, und statistische Tests wie den *mutual information* Test zur Überprüfung der Stärke der Verbindung einzelner Wortkombinationselemente einsetzen zu können, führt zu einer systematischeren Behandlung von Wendungen und Kollokationen in den Wörterbüchern, als vorher denkbar war. Obwohl dem Lexikographen aufgrund seiner muttersprachlichen Sprachkompetenz die oben schon erwähnten sog. ‚recurrent word combinations‘ vertraut sind und er sie auch benutzt, werden sie erst durch die Korpusbelege augenfällig. So konstatiert Rundell (1996a:203): „Corpus evidence, then, enables lexicographers to ‚reach the parts that earlier dictionaries couldn’t reach““, wohingegen sein Urteil zu Belegsammlungen lautet: „[...] citation banks could only ever supply a patchy and somewhat lopsided body of evidence for lexicographers to base their description of the language on.“ (1996a:199).⁶⁰

2. Durch den Gebrauch von computerisierten Korpora und speziellen, auf lexikographische Zwecke zugeschnittenen Softwareprogrammen, wie z.B. den Konkordanzprogrammen, können Wörterbücher viel schneller produziert und überarbeitet werden, wodurch sie eher auf dem neuesten Stand der Sprachentwicklung bleiben.

Obwohl Korpora Belegsammlungen als Datenbasis grundsätzlich überlegen sind, eignen sich letztere aber nach Ansicht vieler Lexikographen als zusätzliche Datenquelle, z.B. um Neologismen aufzuspüren.⁶¹ Dies spielt vor allem dann eine Rolle, wenn den Produkten eines Wörterbuchverlags ein statisches Korpus zugrunde liegt, d.h. ein Korpus, dem nach Abschluß seiner Entwicklungsphase keine neue Daten zugeführt werden.⁶² Aber auch im Falle der Verwendung eines Korpus, das laufend vergrößert wird, ist es den Angaben Rundells (1992a:14) zufolge zum Zweck der Analyse von Neologismen nutzbringend, Belegsammlungen heranzuziehen. So könne man zusätzliche Daten gewinnen, die gerade im frühen Stadium des Gebrauchs eines Neologismus wertvoll seien (zu diesem Zeitpunkt werde die Verwendungsweise des neuen sprachlichen Phänomens für gewöhnlich erläutert, was dem Lexikographen bei der Bedeutungsfindung helfe). Außerdem könnten die in die sog. ‚reading programmes‘ involvierten Leser dazu beitragen herauszufinden, ob der jeweilige Neologismus dauerhaft in die Gemeinsprache oder eine bestimmte Varietät integriert wird. Auf die Bedürfnisse des Lexikographen ausgerichtete Belegsammlungen sind nach Rundell (1992a:13f.) aber auch generell dazu dienlich, Zusatzinformationen für solche Wörter zu liefern, die im Korpus verhältnismäßig selten vorkommen, wie z.B. im Fall von Fachbegriffen und Fremdwörtern.

⁶⁰ Vgl. Rundell/ Stock (1992c:51) und Biber/ Conrad/ Reppen (1998:23).

⁶¹ Vgl. Agor (2000:3), Atkins/ Levin (1995:85), Bergenholtz/ Mugdan (1990:1619), Summers (1993:185f.) und Rundell/ Stock (1992a:10ff.). Teubert (1998:160ff.) plädiert dagegen für ein Spezialkorpus für Neologismen.

Im Rahmen der grundlegenden Neubearbeitung des OED bedienen sich die Redakteure daher zur Erstellung der Einträge für bestimmte Wörter oder Wendungen, die nur mangelhaft belegt sind, der Hilfe der Öffentlichkeit. In regelmäßigen Abständen wird eine sog. ‚Appeals List‘ im Internet veröffentlicht, in der bestimmte Wörter und Wendungen verzeichnet sind, für die Belege aus einem bestimmten Zeitraum gesucht werden. Gerade für die Überarbeitung eines bedeutenden historischen Wörterbuchs wie des OED scheint ein solches Vorgehen unerlässlich zu sein. Das *New Penguin English Dictionary* (NPED₁), das kein historisches Wörterbuch ist, sondern zu der *Concise*-Sparte gehört, verwendet dagegen, wie schon bereits erwähnt, als einziges empirisches Datenmaterial nur eine Belegsammlung. Sie wird überdies wohl ausschließlich dazu genutzt, Beispiele für die Verwendung vieler in dem Wörterbuch verzeichneter Stichwörter zu liefern. Es sind jedoch im NPED₁ auch zahlreiche Beispiele anzutreffen, denen keine Quellenangabe angefügt ist: diese sind wahrscheinlich erdacht (obwohl das an keiner Stelle im Wörterbuch explizit erwähnt wird). Die Zusammensetzung der Belegsammlung ist außerdem insofern sehr einseitig, als möglichst viele literarische Werke einbezogen wurden. Bedenkt man diese Umstände und bezieht auch die oben ausgeführten allgemeinen Überlegungen zu den Nachteilen von Belegsammlungen gegenüber Korpora mit ein, so ist es eher unwahrscheinlich, daß die Qualität der im NPED₁ enthaltenen Informationen der solcher Wörterbücher, die auf Korpora basieren, entspricht. Dazu kommt, daß die Herausgeber des NPED₁ abgesehen von der Belegsammlung, die wohl mehr eine periphere Rolle bei der Erstellung des Wörterbuchs gespielt hat, auch andere herkömmliche Methoden der Datengewinnung benutzt haben werden, die ebenfalls einige Mängel aufweisen. Auf diese soll im folgenden eingegangen werden.

Weitere, bisher noch nicht erwähnte Methoden in der Lexikographie betreffen die Auswertung von anderen Wörterbüchern oder von Grammatiken, die Introspektion des Lexikographen und die Informantenbefragung. Informantenbefragungen, besonders solche größeren Umfangs, sind in der Lexikographie aufgrund ihres zeitlichen und finanziellen Aufwands selten. Sie müssen außerdem sorgfältig vorbereitet und durchgeführt werden, damit ihre Ergebnisse verwertbar sind. Diese Form der Datenerhebung wird daher voraussichtlich nie eine bedeutende Rolle in der Lexikographie einnehmen.⁶³ Die Auswertung bereits existierender Wörterbücher oder Grammatiken ist dagegen immer noch gängige Praxis.⁶⁴ Ob andere Nachschlagewerke allerdings heute immer noch als die beliebteste Quelle lexikographischer

⁶² Zum Begriff ‚statisches Korpus‘ s. auch Kap. 3.1.6.2.

⁶³ Vgl. Sinclair (1991:39). Siehe zu den Vorteilen und Nachteilen der Informantenbefragung Kap. 6.

⁶⁴ Leider werden Übernahmen aus anderen Quellen nur in den seltensten Fällen gekennzeichnet. Die folgenden Darlegungen zu dieser Methode stützen sich vor allem auf Sinclair (1991:37ff.) und Bergenholtz/ Mugdan (1990:1612f.).

Daten gelten können, wie Sinclair (1991:37) andeutet, muß aufgrund der Wertschätzung, die Korpora in der gegenwärtigen Lexikographie genießen, bezweifelt werden. Im Gegensatz zur Verwendung von Korpusdaten wirft die Nutzung schon vorhandener Wörterbuchausgaben oder Grammatiken nicht nur urheberrechtliche oder zumindest ethische Probleme auf, sondern ist auch mit nicht unerheblichen Nachteilen behaftet. So heißt es bei Bergenholtz/ Mugdan (1990:1612):

Wörterbücher, Grammatiken und andere Quellen liefern sekundäre Aussagen über häufig nicht genau spezifizierte Daten. [...] Es ist sogar mit groben Fehlern in der Datenauswertung zu rechnen [...], wie auch die Primärdaten bei der Weiterverarbeitung unabsichtlich verfälscht werden können.

Außerdem läßt sich anhand von bereits bestehenden Wörterbüchern nicht ersehen, welche der aufgeführten Wörter, Wendungen und Bedeutungen schon nicht mehr in Gebrauch sind. Da Wörterbücher veralten, besteht zudem bei ihrem Gebrauch als Datenfundus das Problem, daß dem Lexikographen Neologismen entgehen. Diesbezüglich können dagegen z.B. Monitorkorpora⁶⁵ sowie andere Korpora, die kontinuierlich neues Sprachmaterial aufnehmen, und auch ständig aktualisierte Belegsammlungen verlässliche Auskünfte geben.

Darüber hinaus sind in Wörterbüchern ohne empirische Datenbasis auch oftmals Wortformen und/ oder Bedeutungen verzeichnet, die zwar theoretisch existieren könnten, deren Vorkommen aber selbst in sehr umfangreichen, ausgewogenen Korpora oder in großen Belegsammlungen nicht nachzuweisen ist. Die Übernahme solcher lexikographischer Konstrukte in ein neues Wörterbuch oder in eine neue Wörterbuchauflage ist nicht wünschenswert, schon allein aus dem Grunde, weil wertvoller Platz für die Aufnahme oder Beschreibung anderer Stichwörter verloren geht. Dasselbe gilt auch für den schon oben angesprochenen Einbezug von solchen aus einem anderen Wörterbuch entnommenen Wörtern, Wendungen und Bedeutungen, die eigentlich schon veraltet sind. Die Herausgeber eines Wörterbuchs der in dieser Arbeit untersuchten *Concise*-Sparte beispielsweise sind im Gegensatz zu den Editoren historischer Wörterbücher weniger daran interessiert, Archaismen und seltene oder ungewöhnliche sprachliche Phänomene zu verzeichnen, als aktuellen Sprachgebrauch darzustellen, um nicht zuletzt auch bei Textproduktionen Hilfestellung zu leisten.⁶⁶ Grundsätzlich sollte keine Form oder Bedeutung aus einer anderen Quelle ungeprüft übernommen werden.

Weitere Nachteile des Rückgriffs auf bereits vorhandene Wörterbücher betreffen zum einen das Problem, daß Fehler in der Organisation eines Eintrags schwer zu entdecken sind,

⁶⁵ Vgl. die Ausführungen in Kap. 3.1.6.2 zu den Monitorkorpora.

⁶⁶ Eine Ausnahme bildet das NPED, dessen Herausgeber explizit darauf hinweisen, daß das Wörterbuch auch ungeläufigeren, literarischen oder veralteten Wortschatz berücksichtigt hat. Auch Cassell preist das CACD u.a. damit an, daß Archaismen in das Wörterbuch aufgenommen worden seien.

wenn man diesen Eintrag nicht von Grund auf selbst erstellt hat. Zum anderen können neue Entwicklungen in der linguistischen Beschreibung, z.B. Informationen zur Pragmatik, nicht ohne weiteres in herkömmliche lexikographische Formate eingebunden werden.

Aber auch die Methode der Introspektion⁶⁷ des Lexikographen ist abzulehnen, wenn sie die hauptsächliche oder gar einzige Basis für eine verlässliche Beschreibung der in einem Wörterbuch verzeichneten sprachlichen Phänomene darstellen soll. Ein wichtiger Kritikpunkt wird von Bergenholtz/ Mugdan (1990:1614) folgendermaßen formuliert: „Es ist eine allgemein bekannte und unumstrittene Tatsache, daß das Sprachverhalten eines kompetenten Sprechers im Widerspruch zu seinen introspektiven Urteilen über dieses Verhalten stehen kann.“ Bei dem Versuch, sich die eigene Sprachkompetenz bewußt zu machen, wird vieles leicht übersehen. Was für die Intuitionen über den eigenen Sprachgebrauch gilt, trifft um so mehr zu, wenn der Lexikograph mit Hilfe der Introspektion Aussagen über den Sprachgebrauch aller Angehöriger einer Sprachgemeinschaft machen will. Eine derartige Sprachbeschreibung ist ausschließlich von subjektiven Einschätzungen geprägt und kann häufig im Kontrast zur Realität stehen. So neigen Muttersprachler laut Krishnamurthy (1996a:119, 1996b:141) beispielsweise dazu, sich zu überlegen, was sprachlich möglich ist, während im Gegensatz dazu durch ein Korpus aufgezeigt werden kann, welche sprachlichen Phänomene tatsächlich häufig oder typisch sind. Wie in diesem und dem vorherigen Kapitel schon mehrfach angeführt wurde, bestätigen Korpusbenutzer zudem immer wieder, daß die Analyse von Korpusmaterial nicht nur das in Erinnerung ruft, was man eigentlich über die sprachlichen Elemente der Muttersprache schon weiß, sondern immer auch neue und oft überraschende Erkenntnisse mit sich bringt. Nicht selten werde man dazu gezwungen, die eigenen subjektiven Auffassungen zu überdenken⁶⁸. Den Aussagen von Bergenholtz/ Mugdan (1990:1614) zufolge sind vor allem die mit Hilfe der Introspektion gewonnenen Angaben zur regionalen Verteilung sowie zeitlichen und stilistischen Einordnung sprachlicher Phänomene unzuverlässig, da die Treffsicherheit solcher Informationen durch die jeweiligen sprachlichen Erfahrungen des einzelnen eingeschränkt ist. Auf das Thema dieser Arbeit bezogen heißt das beispielsweise, daß es einem britischen Lexikographen je nach Vertrautheit mit dem AmE schwer fallen dürfte, nur aus der Introspektion heraus zu entscheiden, ob und in welcher Form ein sprachliches Element als Amerikanismus markiert werden sowie

⁶⁷ Vgl. die kritische Auseinandersetzung mit dieser Methode in Bergenholtz/ Mugdan (1990:1613ff.), Bergenholtz/ Mugdan (1989:142), Clear (1993:165), Sinclair (1985:82), Sinclair (1991:39), Rundell/ Stock (1992b:28f.), Rundell/ Stock (1992a:14), Rundell (1996a:199, 201), Johansson (1991:306), Biber/ Conrad/ Reppen (1998:25,41), Svartvik (1992:9), Fillmore (1992), McEnery/ Wilson (o.J.:3), Meijs (1996:101f., 109), Moon (1998: 348), Ooi (1998:47ff.), Biber/ Conrad/ Reppen (1994:169f., 174), Tognini-Bonelli (2001:86,91), Krishnamurthy (2002a:1) und Summers (1993:183f.). Die ausschließliche Verwendung der Introspektion für die Erstellung von Wörterbüchern nennt man auch oft „armchair lexicography“, vgl. Ooi (1998:48).

⁶⁸ Vgl. Rundell/ Stock (1992b:28f.), Krishnamurthy (1996a:126), Meijs (1996:102), Tognini-Bonelli (2001:86), Aston/ Burnard (1997:6) und Biber/ Conrad/ Reppen (1998:25).

andere Kennzeichnungen (stilistische, zeitliche) erhalten sollte. Zudem ist es sehr fraglich, ob ein britischer Lexikograph, der über keine genauere Kenntnis des AmE verfügt, sich nur mit Hilfe der Introspektion alle oder auch nur die frequentesten in dieser Varietät üblichen, vom BrE abweichenden sprachlichen Varianten bewußt machen kann. Oft wird zudem die begrenzte Verbreitung bestimmter sprachlicher Elemente der eigenen regionalen Varietät erst in der Konfrontation mit anderen Varietäten bewußt. Das mangelnde Wissen um die sprachlichen Gepflogenheiten in anderen Varietäten des Englischen außer dem BrE könnte dazu führen, daß solche sprachlichen Phänomene, die auf das BrE beschränkt sind, nicht dementsprechend markiert werden. Generell ist festzuhalten, daß die Introspektion als Quelle für Informationen zur Frequenz eines sprachlichen Elements nicht verläßlich ist, während man mittels eines ‚nicht-opportunistischen Korpus‘⁶⁹ verhältnismäßig genaue Frequenzangaben erhalten kann.

Auf die Problematik selbstgebildeter Beispielsätze wurde ansatzweise schon in Kapitel 3.1.3 eingegangen. Bergenholtz/ Mugdan (1990:1615) wenden sich mit folgender Argumentation gegen das Erfinden von lexikographischen Beispielen:

Während ein kompetenter Sprecher über die Fähigkeit verfügt, in seiner Sprache angemessen zu kommunizieren, gibt es keinen Grund zu der Annahme, daß er sein Sprachverhalten außerhalb einer realen Kommunikationssituation wirklichkeitsgetreu reproduzieren kann. Die Gleichstellung konstruierter Äußerungen mit realen Texten ist daher äußerst fragwürdig, und in der Tat fallen selbstgebildete Beispielsätze oft recht sonderbar aus [...].

Zu unterscheiden ist allerdings zwischen solchen selbst erstellten Beispielen, die ausschließlich mittels Introspektion entstanden sind, und anderen, die auf die Analyse von Korpusbelegen zurückgehen. Letztere können durch die Berücksichtigung von empirischem Sprachmaterial typische Verwendungsweisen eines Wortes abbilden und sind somit nicht vollkommen abzulehnen.

Vor dem Hintergrund der dargelegten Mängel ist die Methode der Introspektion zur Gewinnung lexikographischer Daten nicht zu befürworten. Sie weist gegenüber der Nutzung von Korpora zahlreiche Nachteile auf. Abgesehen von den bereits genannten Kritikpunkten genügt die Introspektion im Gegensatz zur „Korpusmethode“ auch nicht dem Postulat wissenschaftlicher Forschung, daß Ergebnisse verifizier- bzw. falsifizierbar sein müssen. Wenn es allerdings um die Analyse von Korpusdaten durch den Lexikographen geht, ist die Introspektion unverzichtbar. Unter der Masse der Korpusbelege muß er solche herausfiltern, die atypisch bzw. ungrammatisch sind, denn muttersprachliche Sprecher und Schreiber gehen mit der Sprache manchmal in höchst kreativer Weise um oder machen auch Fehler. Außerdem kann

⁶⁹ Vgl. die Definition in Kap. 3.1.6.2.

der Lexikograph intuitiv beurteilen, ob bestimmte sprachliche Elemente, die er als kompetenter Sprecher kennt, im Korpus nicht vertreten sind. Und letztlich ist natürlich das Analysevermögen und die Professionalität des Lexikographen dafür entscheidend, wie die vorliegenden Korpusdaten zum Zwecke der Erstellung von Wörterbucheinträgen ausgewertet werden (wobei Fehlinterpretationen nicht ausgeschlossen sind). Insofern wird die Introspektion bei dem Prozeß der Wörterbucherstellung wohl nie ganz überflüssig sein.

3.1.5 Bei der Verwendung von Korpora auftretende Probleme

Ungeachtet des großen Nutzens, den die Analyse geeigneter Korpora für die Lexikographie hat, sollen auch die Schwierigkeiten, die im Umgang mit Korpora vorkommen können, angesprochen werden. Durch die eigenen Untersuchungen am BNC kann die Verfasserin der vorliegenden Arbeit viele der in der Forschungsliteratur geschilderten Probleme bei der Verwendung von Korpora bestätigen. Eines betrifft zum Beispiel den Umstand, daß es selbst in umfangreichen Korpora für manche sprachliche Elemente zu wenig (oder gar keine) Belege gibt.⁷⁰ In diesem Fall reicht die Datenbasis nicht dafür aus, verlässliche Informationen zum Gebrauch eines sprachlichen Phänomens zu geben. Den Angaben Krishnamurthys (1996b:143) zufolge betrachtete COBUILD zu dem Zeitpunkt, als die BoE erst 167 Mio. laufende Wörter umfasste, eine Auftretenshäufigkeit von weniger als 15 Belegen für ein Wort, eine Wendung oder Bedeutung (d.h. von weniger als einem Vorkommen in 10 Mio. Wörtern) als unzureichend für eine weitere Analyse. Einige der von der Verfasserin im BNC abgefragten Amerikanismen kamen dort ebenfalls weniger als einmal pro 10 Mio. Wörter vor, manche insgesamt sogar nur fünf Mal oder weniger in dem 100-Mio.-Korpus (z.B. *chipper* und *junior college*). Die geringe Auftretenshäufigkeit von Amerikanismen im BNC liegt selbstverständlich in der Zusammensetzung des Korpus begründet, denn das BNC enthält (in dem Maße wie es für die Verantwortlichen des Projekts kontrollierbar war) nur Daten von Muttersprachlern des BrE. Damit wird deutlich, daß der Aufbau eines Korpus dafür entscheidend ist, welche Informationen dem Korpus entnommen werden können.⁷¹ Aber auch die Gesamtgröße eines Korpus spielt natürlich eine große Rolle dabei, wieviel Belege sich für einzelne sprachliche Phänomene finden lassen. Eine Reihe von Linguisten, wie z.B. Fillmore (1992:35), sind zwar davon überzeugt, daß kein Korpus, gleichgültig welchen Umfangs, Daten zu allen Bereichen des Wortschatzes oder der Grammatik des Englischen enthalten kann, aber der allgemeine Trend geht zu immer

⁷⁰ Vgl. Moon (1998:347) und Bergenholtz/ Mugdan (1990:1621).

⁷¹ Die niedrige Frequenz einer Reihe von untersuchten Amerikanismen im BNC macht zweierlei Dinge augenscheinlich. Zum einen, daß diese Amerikanismen im BrE wohl nicht sehr geläufig sind, woraus sich Rückschlüsse auf die in den Wörterbüchern jeweils notwendigen diatopischen Markierungen ziehen lassen (vgl. auch Kap. 7), und zum anderen, daß Lexikographen für die verlässliche Darstellung des AmE in einem Wörterbuch auf authentische Sprachdaten AmE Muttersprachler nicht verzichten können.

größeren Korpora, um mehr Beispiele für den Gebrauch seltenerer sprachlicher Elemente zur Verfügung stellen zu können. Dieses Vorgehen bringt allerdings andere Probleme mit sich, denn für hochfrequente Wörter kann die Zahl der verfügbaren Belege unüberschaubar groß werden.⁷² So erhielt die Verfasserin bei der Abfrage mancher Wortformen im BNC unermeßlich viele Treffer, z.B. für *staff* 22174 oder für *fall* 11028. Die hohe Anzahl an Belegen für diese beiden Wortformen bezieht sich nur auf die jeweilige Verwendung des Wortes als Substantiv im Singular. Es ist nämlich bei Abfragen im BNC durchaus möglich, die Suche auf eine bestimmte Wortart sowie Flexionsform einzuschränken, aber eine semantische Binnendifferenzierung wird zumindest von dem an das BNC gebundene Rechercheprogramm SARA nicht geleistet. Das bedeutet, daß man nicht ausschließlich nur Gebrauchsbeispiele für die gewünschte Bedeutung eines Wortes abrufen kann. Auf der anderen Seite wäre eine aufgrund der Angaben in bereits existierenden Wörterbüchern a priori gesetzte semantische Vorsortierung der Korpusbelege eher kontraproduktiv, denn es sollte gerade die primäre Aufgabe des Lexikographen, der mit Korpusdaten arbeitet, sein, die Bedeutungen eines Wortes erst aufgrund der gründlichen Durchsicht der Belege zu ermitteln. Es ist jedoch eine kaum zu bewältigende Aufgabe, anhand der Analyse von beispielsweise 22174 Konkordanzzeilen für das Substantiv *staff* im Singular die Anzahl der verschiedenen Bedeutungen festzustellen, Definitionen zu entwerfen und die unterschiedlichen Gebrauchsweisen nach der Reihenfolge ihrer Frequenz anzuordnen. Meijs (1996:107) führt - auf Forschungen von COBUILD Bezug nehmend - an, daß ein Korpusnutzer nur ungefähr bis zu 1000 Beispiele für den Gebrauch eines sprachlichen Phänomens verarbeiten kann. Die Masse der Belege für manche Wortformen im BNC war auch der Grund, warum die Verfasserin der vorliegenden Arbeit auf die Untersuchung einiger Amerikanismen in diesem Korpus verzichten mußte.

Es gibt jedoch bereits verschiedene Lösungsversuche zu dem Problem des Umgangs mit großen Treffermengen für bestimmte Wortformen, die allerdings nicht alle gleichermaßen zufriedenstellend sind. Ein sehr pragmatischer Ansatz besteht darin, die Zahl der zur Verfügung stehenden Belege zu reduzieren, indem man entweder z.B. nur die ersten 200 oder 500 Konkordanzzeilen zur Kenntnis nimmt oder den Computer nur jedes soundsovielte Beispiel auswählen läßt, das dann zur weiteren Analyse herangezogen wird. Weiterhin kann durch statistische Verfahren eine Zufallsauswahl erstellt und auf diese Weise die Gesamtzahl der Treffer eingeschränkt werden. Alle drei Methoden widersprechen jedoch dem Sinn eines großen Korpus, eine möglichst große Bandbreite verschiedener Gebrauchsbeispiele zum Zweck der

⁷² Vgl. zu dieser Problematik und den Lösungsmöglichkeiten Biber/ Conrad/ Reppen (1998:28,35), Biber/ Conrad/ Reppen (1994:176ff.), Moon (1998:347), Rundell/ Stock (1992c:50f.), Sinclair (1985:85), Sinclair (1991:43) McEnery/ Wilson (1996:92,174) und vor allem Meijs (1996:107ff.).

Untersuchung sprachlicher Phänomene zu liefern. Außerdem können nach der Reduzierung der Datenmenge beispielsweise keine verlässlichen Aussagen mehr über die relative Häufigkeit bestimmter Wortbedeutungen im Vergleich mit den Bedeutungen anderer Wörter getroffen werden, da nicht in allen Fällen der Zahl der Belege sinnvollerweise verringert werden kann oder muß. Ein weiterer Ansatz basiert auf der Annahme, daß die Kollokate eines Lexems Aufschluß über dessen Bedeutung(en) geben können. Dazu heißt es bei Biber/ Conrad/ Reppen (1998:35f.):

There is a strong tendency for each collocate of a word to be associated with a single sense or meaning (although more than one collocate can be associated with that same sense). Thus, identifying the most common collocates of a word provides an efficient and effective means to begin analyzing senses.

Konkordanzprogramme bieten üblicherweise die Möglichkeit, die links oder rechts von der jeweiligen Wortform befindlichen Wörter in alphabetischer Reihenfolge anzeigen zu lassen. Auf diese Weise werden häufig auftretende Kollokationsmuster augenfällig, die nach entsprechender Gruppierung für die Identifizierung der verschiedenen Bedeutungen eines Wortes sehr hilfreich sein können. Meijs (1996:108) gibt allerdings zu Recht zu bedenken, daß Kollokate nicht direkt nebeneinander stehen müssen und die den Zwischenraum ausfüllenden sprachlichen Elemente sehr unterschiedlich sein können. Als die Verfasserin dieser Arbeit bei den am BNC durchgeführten Untersuchungen den zuletzt beschriebenen Ansatz zur Differenzierung verschiedener Wortbedeutungen verwendete, stellte sie fest, daß sich zwar sehr viele, aber bei weitem nicht alle Belege auf diese Weise sortieren lassen.⁷³

Ein anderes generelles Problem im Umgang mit Korpora betrifft den Umstand, daß die Bedeutung eines Wortes sich nicht unbedingt immer klar aus dem Kontext eines Anwendungsbeispiels erschließen lassen muß. Dies gilt nicht nur für Fachtermini⁷⁴, sondern auch für Wörter des Gemeinwortschatzes, wie z.B. bei der Untersuchung von *nerd* im BNC deutlich wurde. Insofern ist eine größere Auswahl an Belegen sehr wichtig, damit der Lexikograph eine Reihe von verschiedenen Anwendungsbeispielen zur Bedeutungsfindung von Wörtern heranziehen kann. Leider stößt man bei Korpusrecherchen auch immer wieder auf bizarre oder fehlerhafte Verwendungsweisen eines Wortes⁷⁵, die wenig hilfreich sind und unberücksichtigt gelassen werden müssen.

⁷³ Clear (1994) hat jedoch unter Anwendung statistischer Tests eine Methode entworfen, die tatsächlich eine weitgehende Automatisierung der Identifizierung unterschiedlicher Wortbedeutungen auf der Basis von Kollokationen erreichen könnte.

⁷⁴ Vgl. Rundell/ Stock (1992b: 21f.). Klein (2001) gibt zudem die Auffassung von Vertretern der Fachsprachenforschung wieder, daß Fachwörter in Presstexten manchmal falsch verwendet werden, so daß die jeweiligen Belege im Korpus irreführend seien.

⁷⁵ Vgl. Sinclair (1991:5) und Bergh/ Seppänen/ Trotta (1998:43).

Hat man anhand der Durchsicht der Korpusbelege verschiedene Bedeutungen eines Wortes ausfindig gemacht, stößt man zudem manchmal auf Anwendungsbeispiele, die sich nicht eindeutig nur einer dieser Bedeutungen zuordnen lassen. So wurden beispielsweise unter den Treffern für das Adjektiv *dumb* im BNC einige aufgelistet, bei denen sich die Verfasserin dieser Arbeit nicht sicher war, ob das Wort in dem AmE Sinne oder in einer anderen Bedeutung verwendet worden war.⁷⁶

Andere Schwierigkeiten im Umgang mit Korpora betreffen das sog. *part-of-speech* (POS) *tagging*, d.h. die Markierung aller in einem Korpus vorkommenden Wörter nach ihrer Wortart und Flexionsform. Obwohl die Verknüpfung der Suche nach einem Wort mit einem bestimmten *grammatical tag* sehr nützlich sein kann, um z.B. zu vermeiden, daß Belege für eine orthographisch gleiche Wortform mit einer anderen Wortklassenzugehörigkeit (vgl. z.B. *fall* als Verb im Gegensatz zu seiner Verwendung als Substantiv) ebenfalls abgerufen werden, bringt dieses Vorgehen auch Nachteile mit sich. So werden Anwendungsbeispiele, in denen die gesuchte Wortform mit einem sog. *portmanteau code*⁷⁷ versehen oder sogar falsch getaggt wurde, bei dieser Recherche nicht erfaßt. Zu der Genauigkeit von Tagging-Systemen gibt es in der Forschungsliteratur unterschiedliche Angaben. Krishnamurthy (1996b:144) spricht im Zusammenhang mit der *Bank of English* nur von einer Zuverlässigkeitsquote des automatischen Taggers von 80-90 %, während Meijs (1996:104) davon ausgeht, daß bei den meisten Tagging-Programmen 95-97 % der Markierungen richtig sind und Rundell/ Stock (1992c:50) den besten Systemen eine 98-99 %ige Genauigkeit zugesteht.⁷⁸ Ein weiteres Problem bei der Benutzung von *grammatical tags* im Zusammenhang mit der Suche nach bestimmten Wörtern kann sich dadurch ergeben, daß der Lexikograph oder Sprachwissenschaftler bei der Gesamtzahl unterschiedlich getaggtter Wortformen eines Lexems vielleicht vergißt, einzelne Flexionsformen abzufragen.⁷⁹

Während einige der geschilderten Schwierigkeiten im Umgang mit Korpora eher als korpusinhärent anzusehen sind, kann die Vergabe falscher *grammatical tags* sowie die mehrfache Auflistung identischer Belege (in der ersten Fassung des BNC wurde bei manchen Recherchen bis zu fünfmal derselbe Treffer angezeigt) oder das Auftauchen automatisch generierter Infomeldungen unter den abgerufenen Konkordanzzeilen bei einer Überarbeitung der Korpusdaten und des Tagging im nachhinein verhältnismäßig einfach korrigiert werden.

⁷⁶ Vgl. Kap. 7 und zu der angesprochenen Problematik allgemein Kilgarriff (1991).

⁷⁷ *Portmanteau codes* werden benutzt, wenn das Tagging-Programm zwischen zwei möglichen Tags unentschieden ist.

⁷⁸ Meijs (1996:105) schließt sich Clears (1993:182) Einschätzung an, daß das Problem der Treffsicherheit der Tagger nicht zuletzt darin wurzelt, daß sich bestimmte sprachliche Phänomene nicht einfach einer bestimmten vorher festgelegten Kategorie zuordnen lassen.

⁷⁹ Rundell/ Stock (1992c:50) geben als Beispiel an, daß es bei der lexikographischen Arbeit an dem Eintrag für *eat* leicht passieren könne, daß man alle Belege für *eating*, *eats* und *eaten* im Korpus abrufen, aber *ate* aus Versehen übergeht.

Dasselbe gilt für Rechtschreibfehler oder Druckfehler, die dazu beitragen können, daß sich Anwendungsbeispiele für eine bestimmte Wortform durch die Falschschreibung der gesuchten Wortform der Abfrage entziehen.⁸⁰ Aufgrund eigener Recherchen scheint es der Verfasserin der vorliegenden Arbeit, daß in der ersten Ausgabe des BNC die Fehlertrefferquote insgesamt relativ hoch ist. Durch diesen Umstand wird leider auch der *z-score* verfälscht, der automatisch vom SARA - Programm mitgeliefert wird, um die relative Häufigkeit einer Wortform im Korpus anzugeben. Den Informationen Burnards (2001:10f.) zufolge sind aber viele der in der ersten Ausgabe des BNC auftretenden Fehler in der zweiten, überarbeiteten Fassung (*BNC-World Edition*) ausgemerzt worden.

3.1.6 Arten von Korpora

Im folgenden soll nur auf solche Typen von Korpora eingegangen werden, die für das Thema dieser Arbeit von Interesse sind.

3.1.6.1 Korpora für komparative Studien von regionalen bzw. nationalen Varietäten des Englischen

Wie in Kapitel 3.1.3 bereits schon angedeutet, bieten Korpora verschiedener nationaler Varietäten, die in Größe und Aufbau vergleichbar sind, die verlässlichste Basis für die Behandlung dieser Varietäten in den Wörterbüchern. Die Parallelität der Korpora hinsichtlich der ausgewählten Genres bzw. Sachgebiete und der Menge der ihnen zugeordneten Daten sowie des Gesamtumfangs des Sprachmaterials gewährleistet, daß z.B. Unterschiede in der Frequenz bestimmter sprachlicher Phänomene zwischen den jeweiligen nationalen Korpora höchstwahrscheinlich nicht auf Zufälligkeiten im Stichprobenverfahren zurückgehen. Allerdings haben, soweit es die Verfasserin anhand der zur Verfügung stehenden Informationen beurteilen kann, die in dieser Arbeit betrachteten Wörterbuchverlage bisher kaum von solchen parallel aufgebauten Korpora Gebrauch macht. Mit der Fertigstellung des ANC (s.u.), das nach dem Modell des BNC konzipiert wurde, wird sich dies aber zumindest für die Verlage *Longman* und *Oxford University Press* ändern.

Parallel aufgebaute Korpora verschiedener nationaler Varietäten entstanden bisher in erster Linie zum Zweck linguistischer Vergleichsstudien. Im Jahr 1964 wurde das *Brown University Standard Corpus of Present-Day Edited American English* (kurz *Brown Corpus*), das erste

⁸⁰ Vgl. zur Beurteilung von falschen *grammatical tags* und fehlerhaften Korpusdaten Moon (1998:347) sowie zur Problematik der Recherchen nach Wörtern mit alternativen Schreibweisen Moon (1998:347) und Knowles (1990:1650).

maschinenlesbare Korpus für sprachwissenschaftliche Forschungen, fertiggestellt.⁸¹ Das Ziel der beiden verantwortlichen Linguisten, Nelson Francis und Henry Kučera, war es gewesen, ein synchrones Korpus von ca. einer Million Wörtern zu schaffen⁸², das repräsentativ für das in geschriebenen und gedruckten Quellen verwandte Englisch in den USA im Jahr 1961 sei. Um diesem Anspruch gerecht zu werden, wurden 15 Textkategorien festgelegt, die sowohl non-fiktionale als auch fiktionale Prosa umfassen und eine große Bandbreite stilistischer und thematischer Aspekte des zeitgenössischen geschriebenen AmE enthalten sollen. Den 15 Textkategorien wurden insgesamt 500 Textproben zugeordnet, die aus jeweils ungefähr 2000 laufenden Wörtern bestehen. Auch bei der Auswahl der Textproben war man um Repräsentativität bemüht.⁸³ Die Benutzung des Begriffs *Standard* in der offiziellen Bezeichnung des Korpus sollte die Hoffnung reflektieren, daß das *Brown Corpus* für vergleichende sprachwissenschaftliche Untersuchungen genutzt würde, indem es sorgfältig ausgewähltes und aufbereitetes Datenmaterial in einem standardisiertem Format verfügbar mache und somit als Muster für die Erstellung anderer varietätenspezifischer Korpora dienen könne.⁸⁴ Tatsächlich entstand in den Jahren 1970 bis 1978 an den Universitäten Lancaster und Oslo in Zusammenarbeit mit dem *Norwegian Computing Centre for the Humanities* in Bergen ein mit dem *Brown Corpus* vergleichbares Korpus, das sog. *Lancaster-Oslo/Bergen Corpus of British English* (kurz *LOB Corpus*).⁸⁵ Es ist genauso aufgebaut wie das *Brown Corpus*, enthält also auch 500 Textbeispiele, die alle 1961 veröffentlicht wurden und aus jeweils ca. 2000 Wörtern bestehen. Die Gesamtgröße des Korpus liegt daher ebenfalls bei ungefähr einer Million Wörtern. Die benutzten Textkategorien sind dieselben wie im Falle des *Brown Corpus* und die Prinzipien des geschichteten Stichprobenverfahrens zur Gewinnung der Textproben sind ebenfalls identisch, außer daß natürlich andere bibliographische Quellen benutzt wurden. Unterschiede zwischen den beiden Korpora, die aber nach Kennedy (1998:28) die Vergleichbarkeit nicht wesentlich einschränken, betreffen zudem z.B. unterschiedliche Gewichtungen bei der Auswahl der Zeitungen sowie kleine Abweichungen hinsichtlich der Zahl der Texte in einigen Kategorien. So finden sich beispielsweise in der Kategorie N (*adventure and western fiction*) des *LOB*

⁸¹ Vgl. zu diesen Informationen und detaillierteren Auskünften zum *Brown Corpus* Francis/ Kučera (1979), Lenz (2000:6), Sinclair (1998:112f.), Teubert (1998:139f.), und Rundell (1996b:2). Ausführungen zum *Brown Corpus* und *LOB Corpus* enthalten z.B. Oostdijk (1988:14ff.), Kennedy (1998:23ff.), Siemund (1995:352f.), Svartvik (1996:6), Taylor/ Leech/ Fligelstone (1991:326f., 341f.) und Viereck (1984:6).

⁸² Es sind exakt 1.014.312 Wörter (vgl. Francis/ Kučera (1979:2), die Seitenangaben beziehen sich hier und im folgenden auf die maschinenlesbare Ausgabe auf der ICAME-CD-Rom).

⁸³ Vgl. dazu vor allem Francis/ Kučera (1979:2) und Kennedy (1998:26).

⁸⁴ Vgl. Francis/ Kučera (1979:2).

⁸⁵ Die Erstellung des *LOB Corpus* stand unter der Federführung von Stig Johansson und Geoffrey Leech. Vgl. zu diesem Korpus auch die Informationen in Johansson/ Leech/ Goodluck (1978) und Johansson/ Hofland (1989).

Corpus weniger Western- und mehr Abenteuergeschichten als im *Brown Corpus*.⁸⁶ An dieser Stelle ist auf ein generelles Problem bei der Benutzung derselben Textkategorien für vergleichbare Korpora, die verschiedene Varietäten des Englischen repräsentieren sollen, hinzuweisen. Während der kulturelle Kontext, in den das AmE und das BrE eingebettet ist, noch verhältnismäßig viele Gemeinsamkeiten aufweist, können andere Varietäten wie z.B. das Indische Englisch durch ganz andere kulturelle Normen geprägt sein, die in den gemeinsamen Textkategorien keinen Niederschlag finden.⁸⁷ Die Schwierigkeit bei der Erstellung von vergleichbaren Korpora besteht darin, daß man einerseits eine größtmögliche Ähnlichkeit zum Zwecke der Vergleichbarkeit schaffen will, andererseits aber die für jede Varietät spezifischen charakteristischen Strukturen abbilden möchte. So heißt es bei Johansson/ Hofland (1982:3): „The true representativeness of the LOB Corpus arises from the deliberate attempt to include relevant categories and subcategories of texts rather than from blind statistical choice.“ Kleinere Abweichungen vom *Brown Corpus*, die aber die Vergleichbarkeit nicht wirklich gefährdeten, wurden daher in Kauf genommen.

Auf diesen beiden Korpora basieren Hunderte sprachwissenschaftlicher Untersuchungen, vor allem auch sehr viele zu den Unterschieden und Gemeinsamkeiten zwischen dem AmE und BrE, da die beiden Korpora zum ersten Mal einen systematischen Vergleich zwischen den geschriebenen Formen der beiden Varietäten ermöglichten. Auch heute werden das *Brown* und das *LOB Corpus* immer noch für komparative linguistische Studien herangezogen, und nicht nur diese beiden, sondern auch andere sog. „first-generation corpora“, die nach dem Muster des *Brown Corpus* für das Indische, Neuseeländische und Australische Englisch entstanden.⁸⁸ Die sorgfältige Planung der Struktur des *Brown Corpus* hatte einen Standard in der korpuslinguistischen Arbeit gesetzt, der bis in die 90er Jahre hineinwirkte. So wurden 1991 an der Universität Freiburg unter der Leitung von Christian Mair das *Freiburg-LOB Corpus of British English* (kurz *FLOB Corpus*) und das *Freiburg-Brown Corpus of American English* (kurz *Frown Corpus*) erstellt.⁸⁹ Diese beiden Korpora entsprechen ihren Vorgängermodellen *LOB* und *Brown* in Struktur und Größe, repräsentieren aber die Sprache von 1991 bzw. 1992, indem sie bei der Auswahl der Textbeispiele ausschließlich auf gedruckte Quellen dieser Jahre

⁸⁶ Zu weiteren Informationen zu den Unterschieden zwischen den beiden Korpora vgl. Johansson/ Leech/ Goodluck (1978:3ff.: die Seitenangaben beziehen sich hier auf die maschinenlesbare Ausgabe auf der ICAME-CD-Rom).

⁸⁷ Vgl. Kennedy (1998:30).

⁸⁸ Vgl. Kennedy (1998:29f.). Einen kurzen Überblick über sprachwissenschaftliche Studien, die u.a. auf der Basis von *Brown*, *LOB* und dem *Kolhapur Corpus* durchgeführt wurden, gibt Meyer (1994:56). Daß sich anhand von vergleichbaren Korpora auch kulturelle Unterschiede zwischen zwei Sprachgemeinschaften ermitteln lassen, stellen Leech/ Fallon (1992) mit Hilfe des *Brown* und *LOB* unter Beweis.

⁸⁹ Vgl. zum *FLOB* bzw. *Frown Corpus* Hundt/ Sand/ Siemund (1998), Hundt/ Sand/ Skandera (1999), Kennedy (1998:30), Mair (1997), Sand/ Siemund (1992) und Siemund (1995:354).

zurückgreifen. Mithilfe von *FLOB* und *Frown* sollte Sprachwissenschaftlern u.a. die Möglichkeit gegeben werden, die Interdependenz von synchroner regionaler (d.h. zwischen dem BrE und AmE) und stilistischer Variation einerseits und diachronen Entwicklungen andererseits zu untersuchen. Diese beiden neuen Korpora eröffnen also eine Reihe von weiteren interessanten Forschungsmöglichkeiten zum Vergleich des geschriebenen AmE und BrE (vor allem im Bereich des grammatischen Wandels), weisen aber genauso wie die sog. „first-generation corpora“ (z.B. *LOB*, *Brown*) eine Reihe beträchtlicher Nachteile auf, die sie beispielsweise als Datenbasis für ein monolinguales Handwörterbuch des Englischen ungeeignet scheinen lassen. Während man das Datenmaterial von *LOB* und *Brown* schon als veraltet bezeichnen könnte, trifft dies auf *FLOB* und *Frown* zwar vielleicht noch nicht zu, doch lassen sich andere Mängel für alle Korpora, die nach dem Muster des *Brown Corpus* entstanden sind, feststellen: Zum einen ist eine Million Wörter viel zu wenig, um den Wortschatz einer Varietät angemessen wiederzugeben. Den Angaben von Johansson/ Hofland (1982:39) zufolge enthält das *LOB Corpus* nur 50.000 verschiedene ‚graphische Wörter‘⁹⁰ (und noch viel weniger verschiedene Lexeme), eine bescheidene Anzahl im Vergleich mit dem Gesamtwortschatz. Teubert (1998:140) gibt zudem folgende Informationen zu den Wortfrequenzen in einem Korpus dieser Größe:

[...] in einem Korpus von einer Million Wörtern sind es nicht mehr als gut tausend Wörter, die mindestens zehnmal, und weit weniger als 10.000 Wörter, die wenigstens fünfmal belegt sind. Alles aber, was darunter liegt, liegt im Bereich des Zufälligen, den man nur begrenzen kann, wenn man neben dem Parameter Häufigkeit auch noch einen Parameter Distribution oder Verteilung der Belege über eine Mindestzahl von Texten ansetzt.

Sinclair (1991:24) resümiert daher, daß Korpora wie das *Brown* oder *LOB Corpus* sich nur zur Untersuchung frequenter Wörter eignen, die in einer Reihe von Textsorten vorkommen⁹¹. Für die Analyse weniger häufig auftretender sprachlicher Phänomene benötigt man dagegen sehr viel größere Korpora, um sich ein verlässliches Bild über den Gebrauch machen zu können. Da die in dieser Arbeit behandelten Amerikanismen zu der eben genannten Gattung gehören, stellte sich die Recherche nach ihnen im *LOB* und *Brown Corpus* auch als wenig fruchtbar heraus. Korpora dieser Größe konnten der Verfasserin als Informationsquellen zur Überprüfung der in den Wörterbüchern gegebenen Auskünften zu den hier analysierten Amerikanismen nicht nützlich sein. So kommen beispielsweise die Substantive *nerd* und *chutzpah*, die Verben *cop out* und *sashay* sowie das Adjektiv *unlisted* weder im *LOB* noch im *Brown Corpus* vor, obwohl die letzteren vier Lexeme, den Angaben des OED nach zu schließen, 1961 zumindest im AmE schon

⁹⁰ Unter einem ‚graphischen Wort‘ („graphic word“) verstehen Johansson/ Hofland (1982:7) „a sequence of alphanumeric characters surrounded by spaces“.

⁹¹ Vgl. dazu auch Aston/ Burnard (1997:15 f.) und Holmes (1994:30).

viele Jahre im Gebrauch gewesen sein mußten. Wörter wie *block* (*LOB*: 45 Treffer, *Brown*: 66 Treffer) oder *fall* (*LOB*: 122 Treffer, *Brown*: 147 Treffer) sind in den Korpora zwar oft zu finden, aber hier tritt ein anderes Problem auf: Die Software kann nur nach Wortformen sortieren und nicht zwischen verschiedenen Homographen, Wortklassen oder Bedeutungen unterscheiden. Das bedeutet, daß man die aufgelisteten Belege sehr genau prüfen muß, um feststellen zu können, wie viele von ihnen sich tatsächlich auf den Amerikanismus beziehen, der im Zentrum des Interesses steht.

Trotz der Einbeziehung einer Bandbreite von Textsorten und Registern in die Korpora, die nach dem Muster des *Brown Corpus* entstanden, finden in diesen zudem sehr wichtige Genres keinen Niederschlag, wie z.B. die gesprochene Sprache oder unveröffentlichtes geschriebenes Material und bestimmte Textgattungen wie z.B. Werbetexte. Diese Schwäche der Korpora schränkt die Aussagekraft der aus ihnen gewonnenen Ergebnisse weiter ein.⁹²

Obwohl die „first-generation corpora“ sowie *FLOB* und *Frown* also einige Mängel aufweisen, haben einige ihrer Strukturmerkmale sogar den Aufbau des *International Corpus of English* (kurz *ICE*) beeinflusst. Dies ist das bisher ehrgeizigste Projekt zum Zwecke komparativer Studien zwischen verschiedenen Varietäten des Englischen auf der ganzen Welt.⁹³ Es begann 1990 und sieht die Erstellung von 20 vergleichbaren Korpora vor. Das verwendete Sprachmaterial stammt aus den Jahren 1990-96⁹⁴ und bezieht solche Autoren bzw. Sprecher ein, die älter als 18 Jahre sind und ihre Schulbildung in englischer Sprache in dem Land, in dem das jeweilige Korpus entsteht, durchlaufen haben. Zu den an dem Projekt beteiligten Ländern gehören daher sowohl solche, in denen das Englische die dominierende Sprache bzw. die Muttersprache ist, wie in Großbritannien, USA, Kanada, Australien und Neuseeland, als auch solche, in denen das Englische eine Amtssprache ist und eine Zweitsprache für einen Großteil der Bevölkerung darstellt, wie beispielsweise in Indien, Nigeria oder Singapur. Um die Vergleichbarkeit zu gewährleisten, sind alle 20 Korpora auf gleiche Art strukturiert. Gewisse Parallelitäten zum Modell des *Brown Corpus* sind dabei nicht zu übersehen⁹⁵: Jedes Korpus besteht aus 500 Textproben, die jeweils ca. 2000 laufende Wörter umfassen. Insgesamt enthält

⁹² Siehe auch Sinclair (1998:112f.). Die für die Erstellung der Korpora verantwortlichen Linguisten waren sich der oben genannten Mängel durchaus bewußt, vgl. z.B. Johansson/ Hofland (1982:39f.). Die an dem Aufbau des *LOB* und *Brown Corpus* beteiligten Sprachwissenschaftler weisen außerdem daraufhin, daß man sich zwar bemüht habe, nur Texte der jeweils im Zentrum des Interesses stehenden nationalen Varietät aufzunehmen, es aber nicht auszuschließen sei, daß auch Sprachmaterial anderer Varietäten in die Korpora gelangt seien, vgl. z.B. Johansson/ Hofland (1982:2,33) und Francis/ Kučera (1979:2).

⁹³ Die folgenden Informationen zu *ICE* gehen auf Kennedy (1998:54ff.), Kennedy (1996), Greenbaum (1990), Greenbaum (1991), Greenbaum (1996), Leitner (1992), Nelson (1996), Oostdijk (2000:127), Tognini-Bonelli (2001:7f.) und die Webseite <http://www.hku.hk/english/research/ice.htm> zurück.

⁹⁴ Die Angaben zum Entstehungszeitraum der benutzten Quellen variieren in der Forschungsliteratur, doch die an dieser Stelle angeführte Information spiegelt wohl den aktuellsten Stand wider, da sie auf die Auskünfte der ständig überarbeiteten Webseite zum *ICE* Projekt zurückgreift (<http://www.hku.hk/english/research/ice.htm>).

also jedes Korpus etwa eine Million Wörter. Obwohl man selbstverständlich die technischen und finanziellen Schwierigkeiten nicht vergessen darf, die sich im Rahmen der Erstellung der Korpora in den jeweiligen Ländern ergeben mögen, ist diese Größe der Korpora genauso unbefriedigend wie im Falle der nach dem Muster des *Brown Corpus* entstandenen Korpora. Alles, was diesbezüglich bereits ausgeführt wurde (s.o.), gilt hier ebenfalls. Obwohl der Aufbau von *ICE* eine große korpuslinguistische Leistung darstellt, indem er zum einen die Grundlagen für die ersten systematischen sprachwissenschaftlichen Untersuchungen vieler regionaler Varietäten des Englischen überhaupt legt und zum anderen erstmalig Vergleiche vor allem in den Bereichen der Morphologie und Syntax zwischen einer Vielzahl von Varietäten des Englischen ermöglicht, eignen sich die 20 Korpora nicht für solche lexikalischen Untersuchungen, die nicht hochfrequente Wörter betreffen. Insofern kam leider auch *ICE* für Korpusanalysen der Verfasserin zum Zwecke der Überprüfung der in dieser Arbeit behandelten Amerikanismen nicht in Betracht. Außerdem ist bisher nur das britische, ostafrikanische, neuseeländische und singapurische Korpus bereits fertiggestellt - der Abschluß des US-amerikanischen Korpus beispielsweise steht hingegen noch aus (Stand: Mai 2002). Die Größe der *ICE* Korpora ist auch in erster Linie der Grund, warum sich diese nicht für lexikographische Zwecke eignen⁹⁶. Was dagegen die Zusammensetzung der Korpora angeht, haben sich gegenüber dem *Brown* Modell einige wichtige Verbesserungen durchgesetzt: eine davon ist, daß 60 % des verwandten Sprachmaterials aus gesprochenen Quellen stammen. Damit wird der Wichtigkeit der gesprochenen Sprache Tribut gezollt. Dieser Teil der Korpora zeichnet sich genauso wie der auf geschriebene Texte zurückgehende Teil durch den Einbezug einer Bandbreite verschiedener Genres aus. Unter den geschriebenen Texten finden sich im Gegensatz zu den nach dem Vorbild des *Brown Corpus* entwickelten Korpora außerdem nicht nur verschiedene Kategorien gedruckter, sondern auch ungedruckter Texte.⁹⁷

Besonders vielversprechend sowohl für sprachwissenschaftliche Vergleiche zwischen dem BrE und AmE als auch vor allem für die Lexikographie ist die derzeitige Entwicklung eines

⁹⁵ Vgl. auch Leitner (1992:33, 37ff.).

⁹⁶ Der Lexikographie eine Datenbasis zur Verfügung zu stellen war allerdings auch nicht Ziel der für den Aufbau der *ICE* - Korpora verantwortlichen Linguisten. Manche der nationalen Korpusteams werden zwar wahrscheinlich dem ‚Kernkorpus‘ von einer Million Wörtern weiteres Material zum Zwecke der weiteren Erforschung der eigenen regionalen Varietät hinzufügen, doch da hierfür keine Richtlinien festgesetzt wurden, ist eine Vergleichbarkeit dieser zusätzlichen Daten zwischen den Korpora voraussichtlich nicht mehr gegeben. Greenbaum (1996:6f.) erwähnt allerdings z.B. die Möglichkeit, daß neben dem ‚Kernkorpus‘ ein *monitor corpus* aufgebaut werden könne, das vor allem für Lexikographen bei der Suche nach Neologismen in der jeweiligen regionalen Varietät hilfreich wäre. Vgl. auch Kennedy (1998:54), Nelson (1996:29), Leitner (1992) und Greenbaum (1991:7).

⁹⁷ Vgl. zu weiteren Unterschieden zwischen dem *Brown Corpus* bzw. seinen Nachfolgekorpora und *ICE* z.B. Kennedy (1998:55f.). Durch *FLOB* und *Frown* mag Greenbaum (1996:11) allerdings zu dem Vorschlag angeregt worden sein, in regelmäßigen Abständen erneut Korpora für jede Varietät nach dem Muster der ersten *ICE* Korpora zu erstellen, um diachrone Studien zu ermöglichen.

AmE Pendant zum BNC⁹⁸. Den Angaben auf der Webseite der Projektträger zufolge⁹⁹ (Stand: April 2004) soll dieses *American National Corpus* (kurz: ANC) ein ‚Kernkorpus‘ von wenigstens 100 Mio. Wörtern enthalten, das dieselben Genres wie das BNC einbezieht und daher in Größe und Struktur dem BNC entsprechen wird.¹⁰⁰ Durch den zur Verfügung stehenden Datenumfang werden somit im Kontrast zu den oben beschriebenen Korpora auch Vergleiche zwischen der Vorkommenshäufigkeit und dem Gebrauch weniger frequenter Lexeme im BrE und AmE ermöglicht, die für lexikographische Zwecke eine große Rolle spielen. Der einzige bisher ersichtliche strukturelle Unterschied zwischen dem ANC und BNC, der ihre Vergleichbarkeit vielleicht mindern könnte, betrifft den Entstehungszeitraum der benutzten Quellen. Denn das ‚Kernkorpus‘ des ANC soll sich den Angaben der verantwortlichen Linguisten zufolge in erster Linie auf Textproben stützen, die nach 1990 entstanden sind. Erst bei späteren Phasen der Datengewinnung (s.u.) wird man eventuell auf Quellen zurückgreifen, die bis in die sechziger Jahre zurückgehen oder sogar noch älter sein können. Der größte Teil der Textbeispiele des BNC stammt dagegen aus den Jahren 1975-1993 - eine Reihe fiktiver Texte wurde sogar schon zwischen 1960 und 1974 veröffentlicht. Da die Informationen bezüglich der Zeitstruktur des ANC jedoch noch vage sind, sollte man zunächst das Erscheinen des Korpus abwarten, um sich ein Urteil zu dieser Fragestellung bilden zu können. Warum die für das ANC verantwortliche Projektleitung hinsichtlich dieses Aspektes von dem Muster des BNC abweicht, wird nicht begründet. Denkbar ist, daß an dieser Stelle die Verwendung des ANC als Datenbasis zur Erforschung der eigenen nationalen Varietät, wie sie sich heutzutage darstellt, Vorrang hatte vor Überlegungen der Vergleichbarkeit zwischen dem BNC und ANC.

Im Gegensatz zum Modell BNC wird, wie schon angedeutet, zudem eine Erweiterung der Datenbasis des ANC unabhängig vom ‚Kernkorpus‘ angestrebt. Zusätzlicher Bestandteil des ANC soll eine Datensammlung von möglicherweise mehreren hundert Mio. Wörtern werden, die dem Zweck dienen soll, eine möglichst breite und große Auswahl von Texten zu bieten. Es ist geplant, daß diese Daten dem ‚Kernkorpus‘ nach seiner Erstellung kontinuierlich hinzugefügt werden, so daß das ANC etwa alle fünf Jahre durch 10 % weiteres Sprachmaterial ergänzt würde. Welcher Art dieses Sprachmaterial im Einzelnen sein wird, erfährt man bisher aber noch nicht. Es wird nur erwähnt, daß man den Aufbau eines Subkorpus erwägt, das auf Quellen wie e-

⁹⁸ Vgl. zum BNC Kap. 3.2.2.2.

⁹⁹ Vgl. <http://americannationalcorpus.org/>. An der Erstellung des *American National Corpus* sind international renommierte Korpuslinguisten beteiligt wie z.B. Randi Reppen (Projektleitung) und Charles Fillmore. Zu den Mitgliedern des Beratergremiums gehört zudem u.a. Patrick Hanks, einer der Herausgeber des *New Oxford Dictionary of English*, Sue Atkins, die schon im Beraterstab für das BNC Mitglied war und Steve Crowdy, einer der Projektmanager des BNC, der maßgeblich für den Aufbau des *BNC Spoken Corpus* verantwortlich zeichnete.

¹⁰⁰ Wie der Name des Korpus jedoch schon deutlich macht, werden für den Aufbau des ANC nur Texte solcher Personen berücksichtigt, die in den USA geboren oder erzogen wurden und derzeit in den USA leben. Vgl. Ide/Reppen/ Suderman (2002:1).

mails, interne Berichte, und möglicherweise Texte von Popmusik und Rap sowie Drehbücher von Seifenopern zurückgreift.¹⁰¹ Das ‚Kernkorpus‘ soll jedoch genauso wie das BNC statisch bleiben, um die Komparabilität der beiden Korpora sicherzustellen.¹⁰² Für lexikographische Zwecke wäre es sicherlich sehr hilfreich, neben einem statischen, vergleichbarem Korpus für das BrE und AmE für jede Varietät kontinuierlich „aktuelle“ Sprachdaten nach bestimmten Kriterien weiterzusammeln, um somit über Änderungen im Sprachgebrauch informiert bleiben zu können. Dieses Vorgehen ist wohl auch von den Projektträgern des ANC geplant, wie anhand des Zeitplans auf der schon erwähnten Webseite deutlich wird.

Die Existenz des ANC wird für britische Lexikographen ein großer Gewinn sein und auf der anderen Seite des Atlantiks dazu führen, daß auch in den USA mehr Lexikographen Erfahrungen mit der Verwendung von Korpora für die Erstellung von Wörterbüchern sammeln und den Wert von Korpusdaten kennenlernen. Dem im *Guardian Weekly* veröffentlichten Artikel von Agor (2000) zufolge sehen nämlich viele Vertreter amerikanischer Verlagshäuser bisher keine Notwendigkeit, Korpora für ihre Wörterbuchproduktionen heranzuziehen, da sie sich keinen entscheidenden Vorteil davon versprechen. So ging die Initiative für die Erstellung des ANC auch nicht wie im Falle des BNC u.a. von Wörterbuchverlagen aus, sondern von Computerlinguisten, die in erster Linie an einer umfangreichen und ausgewogenen Datenbasis für Forschungen im Bereich der maschinellen Sprachverarbeitung interessiert waren¹⁰³. Unter den Konsortiumsmitgliedern für das ANC sind daher eine Reihe wissenschaftlicher Institutionen und Softwarefirmen zu finden, während man dagegen nur auf einen amerikanischen Verlag stößt (bzw. einen Verlag, der seinen Hauptsitz in den USA hat: *HarperCollins*). Es sind allerdings vier britische Verlage vertreten (*Pearson Education*¹⁰⁴, *Cambridge University Press*, *Oxford University Press* und *Macmillan Publishers*), was erneut unterstreicht, welchen großen Stellenwert Korpora für die britische Lexikographie haben. Die Mitglieder des Konsortiums sind verpflichtet, den Aufbau des ANC in der Anfangsphase finanziell zu unterstützen (und, wenn möglich, elektronisches Datenmaterial beizusteuern), haben aber dafür von Beginn an bis fünf Jahre nach dem Erscheinen des ersten ANC Korpusmaterials Exklusivrechte auf das ANC bezüglich seiner kommerziellen Verwendung. Die Verlagshäuser unter den

¹⁰¹ Fillmore et al. (1998) schlagen in ihrem Artikel, in dem sie für die Erstellung eines mit dem BNC vergleichbaren Korpus für das AmE plädieren, zudem den Einbezug von Textbeispielen anderer wichtiger Sprachen Nordamerikas wie des Spanischen oder kanadischen Französisch vor. Für einen Teil dieser Daten sollten Parallelübersetzungen ins Englische und/ oder in andere in das Korpus aufgenommene Sprachen angefertigt werden. Ob diese Absicht noch besteht, ist unklar; auf der aktuellen Webseite zum ANC zumindest wird auf diesen Vorschlag nicht mehr eingegangen.

¹⁰² Dies wird eine ‚geologische‘ Datenorganisation genannt.

¹⁰³ Vgl. Fillmore et al. (1998).

¹⁰⁴ Pearson Education entstand 1998 nach der Fusion zwischen *Addison Wesley Longman*, *Financial Times Management* und *Simon & Schuster's* pädagogischem Verlagszweig. ELT-Produkte werden jedoch mit dem *Addison Wesley Longman* - Impressum verkauft.

Konsortiumsmitgliedern für das BNC haben ähnliche Vereinbarungen für die Publikation einer Reihe von Wörterbüchern genutzt¹⁰⁵, so daß man auch im Falle des ANC davon ausgehen kann, daß die Zugriffsmöglichkeit auf dieses Korpus neue Wörterbücherausgaben oder zumindest überarbeitete Auflagen hervorbringen wird. Die Qualität der in ihnen enthaltenen Informationen, vor allem hinsichtlich der Unterschiede zwischen dem AmE und BrE Sprachgebrauch, wird bei der Verwendung beider Korpora gegenüber Vorgängerprodukten mit großer Wahrscheinlichkeit deutlich verbessert sein.

Um auch anderen regionalen Varietäten des Englischen in der lexikographischen Beschreibung in ausreichender Weise gerecht werden zu können, würde sich die Erstellung weiterer varietätenspezifischer Korpora nach dem BNC Modell anbieten. Außer dem statischen ‚Kernkorpus‘ von etwa 100 Mio. Wörtern sollte, wie schon bereits weiter oben vorgeschlagen, für jedes vergleichbare Korpus im Idealfall zusätzlich ein ‚dynamisches Korpus‘ zur Verfügung stehen, in das nach vereinbarten Richtlinien kontinuierlich neue Sprachdaten einfließen. Nur so kann nach Abschluß des Kernkorpus neueren Sprachentwicklungen innerhalb einer Varietät Rechnung getragen werden, die zudem auch dann mit solchen in anderen Varietäten verglichen werden können, wenn die Konzeption der ‚dynamischen Korpora‘ ähnlich ist. In jedem Falle garantiert aber das ‚Kernkorpus‘ die Vergleichbarkeit zwischen den Korpora.

Ob und wann weitere varietätenspezifische vergleichbare Korpora nach dem Muster des BNC und ANC entstehen werden, ist ungewiß. Ihre Existenz würde aber, wie schon angedeutet, zu einer adäquateren Markierung und Erfassung dieser Varietäten in kommenden ‚Wörterbuchgenerationen‘ führen. Später könnte man dann erwägen, in einem zeitlichen Abstand von etwa 20 Jahren für jedes der vergleichbaren Korpora wiederum ein gleich strukturiertes Korpus zu schaffen, so daß man fundierte Aussagen über diachrone Prozesse innerhalb der jeweiligen Varietät machen könnte, die sich mit solchen in anderen Varietäten vergleichen ließen.

Ein erster Ausschnitt des ANC ‚Kernkorpus‘ im Umfang von mehr als 10 Mio. Wörtern steht Konsortiumsmitgliedern oder Forschungseinrichtungen seit Oktober 2003 zur Verfügung. Er enthält sowohl geschriebenes als auch gesprochenes AmE, seine Struktur ist jedoch den Angaben der Verantwortlichen zufolge noch nicht ausgewogen, denn die darin aufgenommenen Texte sind die, die als erste bei den Projektmitarbeitern eintrafen und somit zunächst von ihnen verarbeitet wurden. Der vollständige Aufbau des ANC ‚Kernkorpus‘ wird voraussichtlich erst im Herbst 2005 beendet sein. Die Verfasserin der vorliegenden Arbeit konnte somit das ANC leider nicht mehr in ihre Untersuchungen einbeziehen.

¹⁰⁵ Vgl. Kap. 3.2.2.2.

3.1.6.2 *sample corpus* und *monitor corpus*

Da die Wörterbuchverlage der hier untersuchten Wörterbücher auf unterschiedliche Typen von Korpora zurückgreifen, nämlich *sample* bzw. *monitor corpora*¹⁰⁶, soll an dieser Stelle auf die Beschaffenheit sowie die Vorteile und Nachteile dieser beiden Korpusarten eingegangen werden.

Das *sample corpus* zeichnet sich zum einen dadurch aus, daß seine Größe im vorhinein festgelegt wird. Es ist eine statische Textsammlung, die charakteristisch für die betreffende Sprache im Ganzen (oder einen Teilbereich dieser Sprache) in einem bestimmten Zeitraum sein soll. Damit dieses Ziel erfüllt werden kann, wird dem Aufbau des *sample corpus* sehr viel Aufmerksamkeit geschenkt. Es wird angestrebt, daß eine möglichst ‚repräsentative‘ Auswahl an Textsorten im Korpus vorhanden ist, denen jeweils eine bestimmte Anzahl von Textproben mit einer festgelegten Größe zugeordnet wird. Ein *sample corpus* enthält in der Regel keine ganzen Texte (es sei denn, diese entsprechen in ihrem Umfang der festgelegten Größe oder sind sogar noch kürzer). Das BNC, das von *Oxford University Press, Longman* und *Chambers Harrap* als Basis für die Erstellung von Wörterbüchern genutzt wird, ist ein *sample corpus*. Auch bei dem *Longman/Lancaster English Language Corpus* und dem *Longman Spoken American Corpus*, die für die lexikographische Arbeit des *Longman* Verlags benutzt werden, handelt es sich um diesen Korpusstyp.

Im Jahr 1981 stellte John Sinclair sein Konzept eines neuen Korpusmodells, des *monitor corpus*, vor und charakterisierte Korpora dieser Art als „gigantic, slowly changing stores of text“ (Sinclair 1982:4). Der Umfang des *monitor corpus* ist im Gegensatz zum *sample corpus* unbegrenzt. Große Textmengen werden kontinuierlich gesammelt und durch Softwareprogramme bearbeitet, die solche Daten herausfiltern, die für die Lexikographen von Interesse sind, während der Rest verworfen wird. Erfasst werden z.B. neue sprachliche Phänomene oder Gebrauchsweisen sowie Verschiebungen in der Benutzungshäufigkeit sprachlicher Elemente. Wichtig sind auch Texte, in denen sich Gebrauchsbeispiele für sprachliche Phänomene finden, die im Korpus bisher nicht zufriedenstellend belegt sind. Außerdem kann der Datenzufluß dazu dienen, ältere Texte des Korpus durch aktuellere zu ersetzen oder die Auswahl verschiedener Textsorten, Sachgebiete, Varietäten des Englischen etc. im Korpus durch die Hinzunahme geeigneter neuer Texte ausgewogener zu gestalten. Es ist nicht zuletzt auch möglich, durch die Anhäufung von Texten einer bestimmten Gattung ein spezialisiertes Subkorpus zu schaffen, das aus dem *monitor corpus* ausgegliedert werden kann. Eine dynamische Textsammlung dieser Art wächst also beständig, ihre Zusammensetzung wird

¹⁰⁶ Ausführungen zu diesen beiden Gattungen von Korpora finden sich bei Aston/ Burnard (1997:21f.), Clear (1988), Clear (1992:28), Greenbaum (1996:7), Johansson (1991:305), Kennedy (1998: 21f., 60ff., 91f.),

durch die Hinzunahme neuer Textbeispiele und das Entfernen mancher älterer laufend verändert. Im Gegensatz zum *sample corpus* enthält das *monitor corpus* vollständige Texte. Das Korpus *Bank of English*, das als Grundlage für die von *HarperCollins* herausgegebenen Wörterbücher dient, ist ein *monitor corpus*.

Beide Korpusarten haben sowohl Vorteile als auch Nachteile. In der Entwicklung eines *monitor corpus* liegt ein Hauptaugenmerk auf dem Datenvolumen. Auf die Notwendigkeit einer sehr umfangreichen Datenbasis zum Zwecke der Analyse sprachlicher Phänomene mit geringer Häufigkeit wurde schon in Kapitel 3.1.6.1 hingewiesen. Der Lexikograph braucht eine große Anzahl von Korpusbelegen für ein sprachliches Element, um es in einem Wörterbuch adäquat beschreiben zu können. Ooi (1998:56) macht in diesem Zusammenhang - auf Forschungen Sinclairs Bezug nehmend - darauf aufmerksam, daß beispielsweise die Identifizierung signifikanter phraseologischer Strukturmuster sehr umfangreiche Datenmengen erforderlich macht.¹⁰⁷ Ein *monitor corpus* kann diesen Ansprüchen im Gegensatz zum *sample corpus* auf Dauer gerecht werden, da seine Größe nicht begrenzt ist. So heißt es bei Aston/ Burnard (1997:22): „Such a corpus [ein *monitor corpus*] would allow the user to detect phenomena which would be inadequately represented in even a very large sample corpus“. Es braucht kaum erwähnt zu werden, daß dies auch der Erfassung und Beschreibung von Amerikanismen in Wörterbüchern zugute kommen würde. Der Umstand, daß in einem *monitor corpus* mehr Gebrauchsbeispiele für seltenere sprachliche Phänomene zu finden sein müßten als in einem *sample corpus*, wird Linguisten zudem in den Stand setzen, auch bei solchen statistischen Tests zum Verhalten von Wörtern verlässliche Ergebnisse zu erzielen, deren Verwendung eine Mindestzahl von beobachteten Häufigkeiten voraussetzt. Ein weiterer Vorteil des *monitor corpus* - der allerdings für die Lexikographie eine weniger bedeutende Rolle spielt - ist, daß die Aufnahme ganzer Texte auch die Analyse textübergreifender Merkmale wie der Erzählstruktur oder der „intratextuellen“ Variation ermöglicht. Für die Erstellung von *sample corpora* werden, wie bereits erwähnt, meist begrenzte Textproben gleicher Größe verwendet, die für solche Untersuchungen weniger geeignet sind. Allerdings bringt das Ziel des Einbezugs ganzer Texte auch Schwierigkeiten mit sich¹⁰⁸, denn abgesehen von dem Problem der Copyright-Beschränkungen muß zunächst erst einmal definiert werden, was unter einem vollständigen Text zu verstehen ist (Ist beispielsweise die Ausgabe einer Zeitung ein vollständiger Text oder jeder

McEnery/ Wilson (1996:22f., 91), Ooi (1998:55f.), Renouf (1993), Rundell/ Stock (1992c:48f.), Sinclair (1998:123ff.), Sinclair (1993:8), Sinclair (1991:23ff.), Teubert (1999:2) und Teubert (1998:150ff.).

¹⁰⁷ So seien zwar 1 Mio. Wörter ausreichend, um die Strukturmuster eines gebräuchlichen Verbs wie z.B. *fit* aufzuzeigen, 10 Mio. aber nötig, um die Verwendung dieses Wortes als Verb mit Präposition (*fit into*) in ausreichender Weise zu belegen und 100 Mio. unverzichtbar, um genügend Gebrauchsbeispiele für eine aus drei Wörtern bestehende Verbalphrase (z.B. *fit into place*) zu erlangen.

¹⁰⁸ Vgl. zu dem folgenden Aston/ Burnard (1997:22) und Burnard (2001:4).

Beitrag in der Zeitung?). Noch schwerwiegender ist jedoch die Frage nach dem strukturellen Gleichgewicht des Korpus, wenn ganze Texte unterschiedlichster Größe aufgenommen werden, denn ein solches Vorgehen kann dazu führen, daß die sprachlichen Charakteristika umfangreicher Texte in unverhältnismäßig hohem Maße im Korpus vertreten sind und Untersuchungsergebnisse verzerren. Dieser Punkt leitet zu dem generellen Problem der mangelnden Ausgewogenheit eines *monitor corpus* über. Dazu heißt es bei Kennedy (1998:61): „The data collection for a monitor corpus is often opportunistic and necessarily not ‚balanced‘. Quantity of text replaces planning of sampling as the main compilation criterion.“ Daß diese Aussagen Kennedys nicht ganz von der Hand zu weisen sind, machen Ausführungen der für die Entwicklung des *monitor corpus Bank of English* verantwortlichen Linguisten, Sinclair und Clear, deutlich. So konstatiert Clear (1988:385) beispielsweise, daß der Vorschlag für das Modell eines *monitor corpus* zum Teil durch die zunehmende Verfügbarkeit englischer Texte in maschinenlesbarer Form motiviert war. Clear (ebd.) und Sinclair (1991: 25)/ (1998:125) betonen beide, daß es nun eine Überfülle an Sprachmaterial gebe, das im Computersatz hergestellt wurde und das man deshalb problemlos in großen Einheiten durch die bereits erwähnten „Softwarefilter“ fließen lassen könne. Und bei Sinclair (1991: 25) heißt es explizit, daß das meiste des in das Korpus eingespeisten Materials aus maschinenlesbaren Quellen stamme. Für ihn (ebd.) hat die Datenmenge deutlich Vorrang vor der sorgfältigen Planung eines Korpus: „There is less need to build up sample corpora slowly and carefully, hoard them, and subject them to intensive examination, when there is no limit to be placed on text length.“¹⁰⁹ Teubert (1998:159) faßt den so beschriebenen Typ des *monitor corpus* deshalb auch unter die Kategorie des *opportunistischen Korpus* und definiert dieses so (1998:157): „Ein opportunistisches Korpus ist ein Korpus, das qualitative Ansprüche auf dem Umweg über Quantität erfüllt, indem es auf solche maschinenlesbar vorhandenen Texte zugreift, die mit einem Minimum an Aufwand bezogen auf die Textlänge zu integrieren sind.“ Daß ein auf diese Weise erstelltes Korpus in struktureller Hinsicht mit der Qualität eines *sample corpus* konkurrieren kann, dessen Sprachdaten systematisch und proportional bestimmten Textsorten entnommen werden, ist allerdings - wie bereits dargelegt - zu bezweifeln.¹¹⁰ Nicht wenige Forscher sind der Ansicht, daß ein begrenztes, aber sorgfältig konstruiertes Korpus eher eine verlässliche Basis für Generalisierungen darstellt als ein Korpus, das zwar viel größer ist, aber dessen Zusammensetzung nicht die Vielfalt der in einer Sprache vorhandenen Textsorten widerspiegelt.

¹⁰⁹ Die riesige Datenmenge eines ständig wachsenden *monitor corpus* bringt auch den Nachteil mit sich, daß die Editierung und Kodierung der Korpustexte so weit wie möglich automatisiert werden muß, um Kosten zu sparen. Dies kann jedoch zu Qualitätsminderungen führen, vgl. Clear (1988:386), Clear et al. (1996:306) und Teubert (1998:158).

Den Aussagen Rundell/ Stocks (1992:49) zufolge werden die aus der Analyse eines Korpus gewonnenen Ergebnisse um so unzuverlässiger, je mehr das Korpus von einer bestimmten Datenquelle dominiert wird. Wegen der ständigen Veränderungen in der Größe und Zusammensetzung eines *monitor corpus* könnte sich dieses nach Einschätzung einiger Forscher auch als für quantitative Untersuchungen nicht so geeignet erweisen, beispielsweise bei Vergleichen zwischen der relativen Häufigkeit bestimmter sprachlicher Phänomene in verschiedenen Varietäten einer Sprache.¹¹¹ Dieser Umstand hätte erhebliche Auswirkungen auf die Vertrauenswürdigkeit der auf der Analyse eines *monitor corpus* beruhenden diatopischen Markierungen in einem Wörterbuch, also nicht zuletzt auch auf die Markierung von Amerikanismen im COBUILD₃ und CCD₅. Nach Ansicht Teuberts (1998:143) wird das Problem der Ausgewogenheit eines sehr umfangreichen *monitor corpus* vom Korpusersteller zum Korpusbenutzer verlagert, der sich aus der Menge des angebotenen Datenmaterials die für seine Untersuchungen geeigneten Texte zum Aufbau eines Spezialkorpus auswählen muß.¹¹² Dies ist für Linguisten, die ein bestimmtes begrenztes Forschungsziel verfolgen, sicherlich ein gangbarer und vielleicht sogar reizvoller Weg, aber für den Lexikographen, der eine große Bandbreite verschiedenster sprachlicher Phänomene beschreiben soll, nicht praktikabel. Neben den schon genannten Vorzügen sind die sorgfältige Strukturierung sowie gleichbleibende Beschaffenheit und Größe des statischen *sample corpus* nicht nur insofern vorteilhaft, als daß anhand des Korpus gewonnene Untersuchungsergebnisse reproduzierbar sind und die Resultate verschiedener Analysemethoden gegenübergestellt werden können, sondern auch ähnlich aufgebaute Korpora zu komparativen Studien herangezogen werden können. Der letztere Punkt ist - wie schon in Kap. 3.1.6.1 ausgeführt - von besonderer Bedeutung für das Thema dieser Arbeit, da Korpora verschiedener nationaler Varietäten, die in Größe und Aufbau vergleichbar sind, nach Einschätzung der Verfasserin der vorliegenden Arbeit die verlässlichste Basis für die Behandlung dieser Varietäten in den Wörterbüchern bieten.

Es gibt jedoch auch unter den Befürwortern des *monitor corpus* zumindest einen Linguisten, der fordert, daß die Zusammensetzung eines Korpus dieses Modells langfristig möglichst unverändert bleiben sollte.¹¹³ Dies hängt mit der diachronen Perspektive eines *monitor corpus* zusammen¹¹⁴. Da ein *monitor corpus* kontinuierlich weiter wächst, kann der Sprachstand zu einem beliebigen Zeitpunkt mit früheren oder späteren Zuständen verglichen und Sprachwandel dokumentiert werden. Um die dafür notwendige Vergleichbarkeit zu

¹¹⁰ Jedoch können aufgrund der beschränkten Größe auch in ein *sample corpus* nicht alle erdenklichen Genres oder örtlich unterschiedlichen Formen des Englischen einbezogen werden, vgl. auch Kennedy (1998:61).

¹¹¹ Vgl. z.B. Kennedy (1998:61) und McEnery/ Wilson (1996:23,91).

¹¹² Vgl. auch Teubert (1998:157), Teubert (1999:2), Sinclair (1991:25) und Johansson (1991:305).

¹¹³ Vgl. Teubert (1998:159) und Teubert (1999:2).

gewährleisten, postuliert Teubert (1999:2), daß jedes der einem bestimmten Zeitabschnitt zugeordneten Subkorpora in gleicher Weise strukturiert sein muß. So kann beispielsweise verhindert werden, daß Unterschiede in der Gebrauchshäufigkeit eines sprachlichen Elements während einer gewissen Zeitspanne fälschlicherweise als Sprachwandel gedeutet werden, wenn es sich eigentlich nur um Zufälligkeiten im Stichprobenverfahren handelt. Die Umsetzung von Teuberts Konzept vorausgesetzt, ermöglicht es also ein *monitor corpus*, Veränderungen in der Sprache zu verfolgen.¹¹⁵ Ein großer Nutzen dieses Korpusstyps liegt daher für Lexikographen nicht zuletzt in der Extraktion von Neologismen.¹¹⁶ Während ein *monitor corpus* durch die kontinuierliche Aufnahme neuer Texte den aktuellen Stand einer Sprache widerspiegeln kann, wird ein *sample corpus* im Laufe der Zeit eine immer unzuverlässigere Basis zur Beschreibung neuerer Sprachentwicklungen. Daher charakterisiert Sinclair (1993:8) das *monitor corpus* auch als „moving mirror on the language“, wogegen das *sample corpus* oft als „snapshot (of the language)“¹¹⁷ bezeichnet wird (ein „Schnappschuß“, der leider auf die Dauer veraltet). Für die Praxis bedeutet das beispielsweise, daß die Lexikographen der o.g. Verlagshäuser, die zum Zwecke der Erstellung ihrer Wörterbücher auf das BNC (ein *sample corpus*) zurückgreifen, mittlerweile zusätzlich auch andere Datenquellen heranziehen müssen, die aktuelleres Sprachmaterial anbieten, denn die jüngsten Texte im BNC (und das sind zudem nicht viele) stammen aus dem Jahr 1994. Tun sie das nicht, könnten die Informationen in den von ihnen herausgegebenen Wörterbuchprodukten - nicht zuletzt auch solche zu Amerikanismen - nicht mehr auf dem neuesten Stand sein. Diese Sichtweise wird auch durch Clear (1996:310) Ausführungen gestützt. Seinen Aussagen zufolge haben Analysen des Korpus *Bank of English* ergeben, daß schon innerhalb von 10 Jahren viele bemerkenswerte sprachliche Veränderungen auftreten können. So habe man beispielweise nach Ablauf dieser Zeitspanne festgestellt, daß solche sprachlichen Phänomene, die vorher als Amerikanismen galten, vollkommen in das BrE integriert worden seien. Der auch auf anderen Sprachebenen verzeichnete Sprachwandel veranlaßt Clear (ebd.) deshalb zu folgendem Resümee: „The surprising extent and speed of this [gemeint ist der Sprachwandel] underlines the importance that any corpus used as evidence of

¹¹⁴ Siehe dazu auch Clear (1988:385), Sinclair (1998:125) und Sinclair (1991:25).

¹¹⁵ Auf diesen Umstand gründet sich auch der Name des *monitor corpus*: „A monitor corpus [...] is so-called because it provides the means to monitor changing patterns of usage over time.“ (Kennedy 1998:61) Vgl. zu dem Aspekt der Beobachtung von Sprachwandel anhand eines *monitor corpus* auch Kennedy (1998:92f.), Aston/Burnard (1997:22), Greenbaum (1996:7), McEnery/ Wilson (1996:22, 91), Ooi (1998:55), Rundell/ Stock (1992c:48), Sinclair (1993:8) und Sinclair (1998:125).

¹¹⁶ Teubert (1998:159) und Renouf (1993) erläutern im einzelnen, wie der Lexikograph oder Linguist Neologismen in *monitor corpora* erkennen kann.

¹¹⁷ Rundell/ Stock (1992c:49), Kennedy (1998:60f.), Ooi (1998:55) und McEnery/ Wilson (1996:22). Vgl. auch Burnard (2001:10). Bei Kennedy (1998:61) heißt es zudem treffend zum *sample corpus*: „It freezes the language at a point in time“. Der Umstand, daß das *sample corpus* zu einer bestimmten Zeit zusammengestellt wird, kann

current English must be constantly updated: a static corpus is automatically a historical corpus, and a snapshot of English in the past.“

Bezüglich der beständigen Aktualität seiner Daten weist das *monitor corpus* also deutliche Vorteile gegenüber dem *sample corpus* auf. Trotzdem bleibt abzuwarten, ob Sinclair (1993:8) recht hat, wenn er prognostiziert, daß für die meisten lexikographischen Aufgaben bezüglich der Behandlung der heutigen Sprache in Zukunft ein *monitor corpus* gefordert werden wird. Wie bereits dargelegt, besitzt auch das Korpusmodell des *sample corpus* eine Reihe von Vorzügen für die Lexikographie. Der wichtigste liegt in der größeren Ausgewogenheit des *sample corpus* hinsichtlich seiner Zusammensetzung begründet, die dazu führen mag, daß die auf der Basis eines solchen Korpus beruhenden Generalisierungen verlässlicher sein könnten als die anhand eines *monitor corpus* gewonnenen. Die Überprüfung der von der Verfasserin ausgewählten Amerikanismen in den Wörterbüchern wird dazu vielleicht mehr Aufschluß geben können. Kennedy (1998:61f.) faßt die Stärken und Schwächen der beiden Korpusmodelle folgendermaßen zusammen:

[...] mega-corpora of the monitor type do promise to provide rich insights (especially for lexicographers and diachronic linguists) into the processes of language change and the use of low frequency words. The place of static sample corpora nevertheless seems to be assured for research on high to medium frequency vocabulary, and for phonological, morphological, syntactic and much discourse-focused research.

Clear (1988:28) äußert schon 1988 den interessanten Gedanken, daß man die Vorzüge beider Korpusarten miteinander verbinden könnte. Dies könnte verwirklicht werden, indem man ein Korpus entwickelt, das wie im Falle des *monitor corpus* beständig durch neue Daten angereichert wird, dessen Aufbau aber nach strikten Strukturprinzipien erfolgt, die jeweils auch bei dem Einbezug neuer Daten beachtet werden. Damit sich dieses Modell auch für komparative Untersuchungen zwischen verschiedenen nationalen Varietäten des Englischen und damit ebenfalls für die Behandlung dieser Varietäten in den Wörterbüchern als besonders fruchtbringend erweist, müßte für jede Varietät des Englischen ein vergleichbares Korpus dieser Art geschaffen werden, wobei gewährleistet sein sollte, daß die Größenverhältnisse der einzelnen Teile eines Korpus trotz der kontinuierlichen Zunahme der Datenmenge identisch bleiben.

Wie dieses Kapitel und die vorhergehenden verdeutlichen, sind die Aspekte ‚Korpusgröße‘ und ‚Korpusaufbau‘ entscheidend für die Verlässlichkeit der aus der Analyse eines Korpus gewonnenen Informationen - nicht zuletzt auch für solche, die in den Wörterbüchern zu Amerikanismen gegeben werden.

außerdem den Nachteil haben, daß gewisse „Modewörter“ häufiger im Korpus vorkommen als es sonst der Fall gewesen wäre (vgl. Aston/ Burnard 1997:40).

3.2 Analyse der den Verlagshäusern für die untersuchten Wörterbücher zugrundeliegenden Belegquellen¹¹⁸

3.2.1 Wörterbuchproduktionen des HarperCollins Verlages

Das erste britische Wörterbuch, das ausschließlich auf der Grundlage eines Korpus entstand, war das *Collins COBUILD English Language Dictionary* (COBUILD₁).¹¹⁹ Es erschien 1987 und war die erste Veröffentlichung des COBUILD Projektes, das unter der Leitung von Professor John Sinclair stand. Das COBUILD Projekt wurde 1980 gemeinsam von *Collins*¹²⁰ und einem im Englischen Seminar der Universität Birmingham ansässigen Forschungsteam mit dem Ziel ins Leben gerufen, ein umfangreiches Korpus des zeitgenössischen Englischs zu erstellen und zu analysieren, und die Ergebnisse in einer Reihe von Nachschlagewerken für Lehrer und Lerner des Englischen zu publizieren.¹²¹ Das Akronym COBUILD steht für *Collins Birmingham University International Language Database*.¹²²

Die Erstellung dieses ersten großen lexikographischen Korpus - *Birmingham Corpus* bzw. *COBUILD Corpus* genannt - das zur Herausgabe von COBUILD₁ führte, hatte bahnbrechende Veränderungen in der lexikographischen Praxis der meisten britischen Wörterbuchverlage zur Folge. So heißt es bei Rundell (1996a:200):

The beginning of the 1980s [...] saw the creation of the first large lexicographic corpus, the Birmingham Collection of English Texts¹²³, a pioneering development that led to a radical transformation in the craft of lexicography. For the first time, dictionary text would be based primarily on the analysis of large amounts of naturally-occurring data, and within a decade most serious lexicographic work in English had become, in some form or another, corpus-based.

¹¹⁸ Da die für die Wörterbuchproduktion herangezogenen Datenbasen der Verlagshäuser einem ständigen Wandel unterworfen sind, hat sich die Verfasserin dieser Arbeit bemüht, die folgenden Ausführungen möglichst häufig durch Zeitangaben zu ergänzen. Trotz der beschriebenen Problematik wurde versucht, auch neueste Entwicklungen zum Thema zu berücksichtigen. Dazu wandte sich die Verfasserin oft persönlich per E-Mail an die Verlage, denn die Auskünfte im Internet, in den Verlagsbroschüren oder in den Wörterbüchern selbst waren vielfach nicht aktuell genug, und zudem auch häufig nicht sehr informativ.

¹¹⁹ Vgl. Meijs (1996:99,106).

¹²⁰ Der britische Verlag *Collins* und der US-amerikanische Verlag *Harper & Row* fusionierten erst 1989. Der neue Verlag wurde *HarperCollins* genannt.

¹²¹ Vgl. Renouf (1987:1), Krishnamurthy (2002a:3) und Kennedy (1998:46).

¹²² Gleichzeitig dient COBUILD als Abkürzung für das *Collins COBUILD English Language Dictionary*, *Collins COBUILD English Dictionary* und *Collins COBUILD English Dictionary for Advanced Learners*.

¹²³ Die *Birmingham Collection of English Text* wird in der Regel nicht als Korpus bezeichnet, da sie nicht nur aus dem *Birmingham Corpus* bestand, sondern auch aus einer Reihe anderer größerer oder kleinerer Textsammlungen.

Das heute vom *HarperCollins* Verlag und COBUILD zur Wörterbucherstellung benutzte Korpus, die *Bank of English* (BoE), hat sich aus dem *COBUILD Corpus* entwickelt. Deshalb soll kurz auf dieses Korpus näher eingegangen werden.¹²⁴

Das erste Teilkorpus des *COBUILD Corpus* war das sogenannte ‚Main‘ *COBUILD Corpus*. Die Struktur des ‚Main‘ *COBUILD Corpus*, dessen Erstellung Ende 1980 begann, sollte sowohl den Bedürfnissen von Englischlernern, Englischlehrern und anderen Nutzern Rechnung tragen als auch dem Forschungsinteresse von Sprachwissenschaftlern Genüge tun. Angestrebt wurde „an authoritative body of linguistic evidence which can support generalisations and against which hypotheses can be tested“ (Renouf 1987:2). Zu diesem Zweck unterlag die Zusammensetzung des Korpus bestimmten Richtlinien: Das Korpus sollte zu 75 % aus geschriebener und zu 25 % aus gesprochener Sprache bestehen und in erster Linie die Gemeinsprache (nicht so sehr Fachsprachen) widerspiegeln. Weiterhin wurde für wichtig erachtet, daß das Korpus möglichst aktuellen Sprachgebrauch (ab dem Jahr 1960, vorzugsweise aber aus jüngster Zeit) umfasse. Es sollten weder Dramen noch Lyrik aufgenommen werden, dagegen allerdings fiktionale Prosatexte. Außerdem war es das Ziel der Verantwortlichen bei COBUILD, nur die Standardsprache einzubeziehen (also keine Dialekte), und überdies ausschließlich den Sprachgebrauch von Muttersprachlern, die über 16 Jahre alt waren, zu berücksichtigen. Die Buchautoren, deren Texte man ausschnittsweise in das Korpus aufnehmen wollte, sollten zu 75 % männlich und zu 25 % weiblich sein. Hinsichtlich der Einbeziehung verschiedener nationaler Varietäten des Englischen in das Korpus entschied man, das Hauptaugenmerk auf das Britische Englisch zu legen (70 % des Sprachmaterials). 20 % der Sprachdaten sollten aus dem AmE stammen und 5 % aus anderen Varietäten.¹²⁵ Gründe dafür, warum die Gewichtung mancher Korpuskomponenten so ausfällt wie dargestellt, werden von Renouf (1987:3) leider nicht angegeben.

Für das Thema dieser Arbeit ist vor allem der Anteil des in das Korpus aufgenommenen AmE Sprachmaterials von Interesse. Er ist im Falle des ‚Main‘ *Corpus* verhältnismäßig

¹²⁴ Die folgenden Informationen stammen vor allem aus Renouf (1987) und (1984), die von 1980-83 Projektkoordinatorin in Birmingham war, gehen aber auch auf die Ausführungen in Sinclair/ Kirby (1990:21ff.), Clear (1990:93ff.), Clear (1994:33), Clear (1996:303ff.), Kennedy (1998:46ff.), Krishnamurthy (2002a:3), Krishnamurthy (1996b:141), Rundell (1996b:2) und auf der COBUILD Webseite (<http://titania.cobuild.collins.co.uk/about.html>, Stand: Mai 2003) zurück. Ausführlichere Auskünfte finden sich ebd.

¹²⁵ Vgl. Renouf (1987:3). Leider erfährt man an dieser Stelle nicht, um welche anderen Varietäten es sich handelt. Den Auskünften zu den Kodierungen, die in den Konkordanzen vorkommen, läßt sich allerdings entnehmen, daß z.B. auch Daten aus dem Australischen, Kanadischen und Südafrikanischen Englisch entnommen wurden. Da die Summe der genannten Prozentsätze zum Einbezug nationaler Varietäten nicht 100% ergibt, sondern nur 95%, fragt sich der Leser, wie die restlichen 5% zuzuordnen sind. Bei Krishnamurthy (2002b:3) findet sich allerdings die Angabe, daß 10% des Datenmaterials aus anderen Varietäten als dem AmE und BrE stammten.

niedrig.¹²⁶ Andere Korpora, wie z.B. das *Longman/ Lancaster English Language Corpus* (LLELC), verfügen über eine größere Menge AmE Sprachdaten. Das LLELC, mit dessen Erstellung der Longman Verlag 1988 begann¹²⁷, enthält nur zu 50 % BrE, aber zu 40 % AmE und zu 10 % andere Varietäten. Das ‚Main‘ *COBUILD Corpus* ist mit einer Größe von 7,3 Mio. Wörtern überdies um ungefähr Dreiviertel kleiner als das LLELC (ca. 30 Mio. Wörter), so daß die Menge der AmE Sprachdaten hier nur bei ca. 1,5 Mio. liegt, im LLELC aber bei ca. 12 Mio. Allerdings darf man hinsichtlich der Größe des ‚Main‘ *Corpus* nicht vergessen, daß es sich hier um das erste bedeutende sog. ‚Mega-Korpus‘¹²⁸ handelt, mit dem die ‚magische Größe‘ der bisher verfügbaren Standardkorpora (*Brown*, *LOB*) von einer Million Wörtern überwunden werden konnte und das daher neue Standards bezüglich des Datenumfangs von Korpora erst schuf. Was den Anteil gesprochener Sprache im ‚Main‘ *Corpus* angeht, so ist anzumerken, daß er nicht in ausreichendem Maße die Bedeutung dieses Sprachmodus im Alltag widerspiegelt.¹²⁹ Da den Projektmitarbeitern jedoch nicht genügend Zeit zur Verfügung stand, eigene Sprachaufnahmen zu machen, weil die Beschaffung von mündlichem Sprachmaterial immer mit einem großen zeitlichen und finanziellen Aufwand verbunden ist, mußten sie sich mit der Datenmenge zufrieden geben, die von der Universität Birmingham und anderen Institutionen, wie z.B. dem British Council, der BBC und der Universität Sussex, bereitgestellt wurde. Der Umfang mündl. Sprachmaterials hätte nur auf Kosten der Ausgewogenheit erhöht werden können, doch stand hier das strukturelle Gleichgewicht des Korpus im Vordergrund. Die gesprochene Sprache des AmE fand allerdings keinen Niederschlag im *Main Corpus*, was sich mit großer Wahrscheinlichkeit nachteilig auf den Einbezug von informeller Sprache oder Slang-Ausdrücken des AmE z.B. im *COBUILD*₁ ausgewirkt hat.

Um innerhalb der festgesetzten Richtlinien für den Aufbau des Korpus eine große Bandbreite sprachlicher Phänomene berücksichtigen zu können, wurden Listen angefertigt, auf denen das *COBUILD* Team die seiner Ansicht nach in der englischen Sprache zentralen Textsorten und Sachgebiete aufführte. Diese Listen lagen der Textauswahl für das Korpus zugrunde. Die Ausgewogenheit des Korpus bezüglich solcher Strukturelemente, für die a priori keine Prozentzahlen festgesetzt worden waren, sollte während des Entstehungsprozesses des Korpus sichergestellt werden.

¹²⁶ Renouf (1984:6) äußert sich zum Einbezug verschiedener nationaler Varietäten des Englischen in die *Birmingham Collection of English Text* folgendermaßen: „The language in the Collection was to be predominantly British, since we were better equipped to deal with the analysis of this variety, but the American English component had to be sufficiently large to allow us to perceive major differences between the two varieties.“

¹²⁷ Vgl. zum LLELC Kap. 3.2.2.1.1 dieser Arbeit.

¹²⁸ Vgl. Kennedy (1998:46).

¹²⁹ Diese Kritik trifft aber auch auf die meisten anderen der in dieser Arbeit behandelten Korpora zu.

Im Dezember 1981 hatte das Korpus die Zielgröße von 6 Mio. Wörtern erreicht. Während der lexikographischen Arbeit mit dem Korpus stellte sich jedoch heraus, daß die Zusammensetzung des Korpus noch verbessert werden könnte, indem man geschriebenes journalistisches sowie gesprochenes Material hinzufügte. Durch diesen Schritt wollte man eine noch größere Ausgewogenheit und Breite des Korpusmaterials erreichen. Das Korpus wurde also durch ca. 1,3 Mio. Wörter ergänzt, so daß es im August 1982 insgesamt 7,3 Mio. Wörter umfasste.

Nach Beendigung der Arbeit am ‚*Main*‘ *Corpus* führen die Verantwortlichen fort, Sprachdaten zu sammeln, weil sie zu Forschungszwecken weitere große Korpora erstellen wollten. Aus diesem Sprachmaterial entstand das sog. ‚*Reserve*‘ *Corpus*, das eine Größe von 13 Mio. Wörtern hatte. Es enthielt z.B. den restlichen Text der Bücher, von denen die ersten 70.000 Wörter ins ‚*Main*‘ *Corpus* eingegangen waren; außerdem solche Bücher, deren Themenbereiche im ‚*Main*‘ *Corpus* nicht ausreichend vertreten waren. Überdies wurden Sprachdaten aus elf Zeitschriften und einer Reihe von Broschüren und Prospekten einbezogen. Zeitungen oder gesprochenes Material blieben dagegen unberücksichtigt. Unter den für das ‚*Reserve*‘ *Corpus* berücksichtigten Büchern und Zeitschriften war auch AmE Sprachmaterial, wie man den Informationen im Anhang des Artikels von Renouf (1987:23ff.) entnehmen kann. Das ‚*Reserve*‘ *Corpus* wurde als dynamisches Korpus konzipiert, bei dem nicht die ausgewogene Zusammensetzung, sondern die Vielfalt seiner Strukturelemente im Vordergrund stand: „What was aimed for was variety, not balance, since it was conceived to be an on-going accumulation of data“ (ebd., S. 11). Das ‚*Reserve*‘ *Corpus* weist in seiner Anlage also schon Ähnlichkeiten mit der *Bank of English* auf, indem der Aspekt der Ausgewogenheit zugunsten der Sammlung einer ständig wachsenden Datenmenge zurücktritt. Die Entwicklung des ‚*Main*‘ *Corpus* zeichnete sich dagegen noch dadurch aus, daß man bestrebt war, durch eine vorher festgelegte Grundstruktur und Größe den Aufbau des Korpus zu steuern und das Gleichgewicht in der Zusammensetzung des Korpus während des Entstehungsprozesses zu überprüfen. Das ‚*Main*‘ *Corpus* kann daher dem Typ des *sample corpus* zugerechnet werden.

Das ‚*Reserve*‘ *Corpus* wurde jedoch auch für die Erstellung des COBUILD₁ herangezogen, da man feststellte, daß das ‚*Main*‘ *Corpus* zu wenige Belege für weniger frequente Wörter enthielt. Lag die Vorkommenshäufigkeit einer Wortform im ‚*Main*‘ *Corpus* bei 50 oder darunter (dies betraf ca. 90 % der Wortformen (types¹³⁰) in diesem Korpus), griff man zusätzlich auf die Daten des ‚*Reserve*‘ *Corpus* zurück.

¹³⁰ Unter *Types* versteht man nach Lenders/ Willée (1998²:138) „die verschiedenen Wortformen, aus denen ein Text besteht“. *Tokens* sind dagegen „die in einem Text vorkommenden laufenden Wortformen“.

Schon während der Entstehung des ‚Main‘ Corpus im Jahre 1981 vertrat John Sinclair den Angaben Renoufs (Renouf 1989:89) gemäß die Auffassung, daß ein begrenztes Korpus, gleichgültig welchen Umfangs, nur für die gängigsten Wörter eine ausreichende Menge an Daten bieten würde. Seine Zielvorstellung war es, ein *monitor corpus* zu entwickeln. Mit der Umsetzung dieses Plans begann das COBUILD Team 1991, als eine große Korpusentwicklungsinitiative mit dem Namen *The Bank of English* (BoE) ins Leben gerufen wurde, die auf der bereits vorhandenen Datenbasis aufbaute.¹³¹ Die zunächst angestrebte Größe des Korpus von 200 Mio. Wörtern wurde schon 1995 überschritten; den jüngsten verfügbaren Informationen zufolge¹³² enthält die BoE, die stetig weiter wächst, bereits 524 Mio. Wörter (Stand Juni 2004). Im Gegensatz zu seinen Vorläufern (‚Main‘ Corpus, ‚Reserve‘ Corpus) setzt sich die BoE aus einer Reihe von Subkorpora zusammen, deren Homogenität jeweils dadurch gewährleistet werden soll, daß in einem Subkorpus eine bestimmte nationale Varietät des Englischen und/ oder eine spezifische Textsorte vertreten ist (Krishnamurthy 2002a:1, Appendix). Die Aufgliederung der BoE in Einzelkomponenten hat zum einen den Vorteil, daß man kontrastive Analysen zwischen einzelnen oder Gruppen von Subkorpora durchführen kann, zum anderen erleichtert diese Art der Datenorganisation aber auch die Kontrolle der Korpusentwicklung¹³³. Nach Auskunft von Teubert (persönliche Mitteilung, Juni 2002) wird angestrebt, das Korpus etwa einmal im Jahr zu aktualisieren; tatsächlich haben sich in der Geschichte der BoE aber auch schon größere Zeitabstände zwischen einzelnen *updates* ergeben.¹³⁴ Neues Sprachmaterial für das Korpus wird jedoch kontinuierlich gesammelt, die monatliche Zuwachsrate lag z.B. Ende 1998 bei 3 Mio. Wörtern¹³⁵. Der Forschungsliteratur und der Webseite zur BoE kann man folgende Informationen über die Strukturprinzipien des Korpus entnehmen¹³⁶: Eine wichtige Leitlinie bei der Entwicklung der BoE heißt beispielsweise „the bigger the better“ (Clear 1996:305), womit allerdings nicht nur eine ständige Vergrößerung des Korpus, sondern auch die Steigerung der Anzahl verschiedener *text types* in der BoE gemeint ist. Den Auskünften auf der Webseite zur BoE zufolge wird die Zusammensetzung des Korpus kontinuierlich überprüft, so daß neue Quellen auch deswegen in die BoE aufgenommen werden sollen, um die Ausgewogenheit des Korpusmaterials zu erhalten (unklar ist dabei, wie der

¹³¹ Vgl. Krishnamurthy (1996b:141), Rundell (1996b:2), Clear (1996:304) und Kennedy (1998:47).

¹³² Vgl. Krishnamurthy (2002a:3) und die Webseite von *Collins Cobuild* (http://titania.cobuild.collins.co.uk/boe_info.html, Stand Juni 2004).

¹³³ Vgl. z.B. Sinclair (1998:124ff.).

¹³⁴ So z.B. zwischen den Aktualisierungen der Jahre 1996 und 2000. Den Informationen Krishnamurthys zufolge wurden der BoE in diesem Zeitraum wegen interner Umstrukturierungsmaßnahmen bei *HarperCollins* keine neuen Daten hinzugefügt (vgl. Krishnamurthy 2002b:7). Einen Überblick über die Zusammensetzung der BoE in den Jahren 1993 - 2000 liefert Krishnamurthy (o.J.) in einem ‚In-house working paper‘, s. Literaturverzeichnis.

¹³⁵ Diese Information geht auf einen pers. Brief vom *HarperCollins* Verlag Ende 1998 zurück.

¹³⁶ Vgl. Clear (1996:305), Ooi (1998:55f.) und die Webseite von *Collins Cobuild* (http://titania.cobuild.collins.co.uk/boe_info.html, Stand Mai 2003). Siehe auch die Ausführungen zum Typ des *monitor corpus*.

Begriff der ‚Ausgewogenheit‘ definiert wird). Es wird jedoch nicht angestrebt, den Anteil verschiedener *text types* im Korpus festzulegen:

No fine and delicate tuning of the relative proportions of varying text types is attempted, because we see no evidence that our samples can be significantly improved in that way, given that even a corpus of hundreds of millions of words is of a pitifully small size in comparison to the amount of English language being generated daily. We do, however, see firm evidence that the addition of large tranches of new corpus data from new sources enhances every aspect of our analysis.

Clear (1996:305)¹³⁷

Auf den Aspekt der Größenordnung der in der BoE vertretenen ‚Texttypen‘ und der Ausgewogenheit der Struktur des Korpus wird im folgenden noch einzugehen sein. Eine weitere Zielvorstellung bez. der BoE ist, daß sie das gegenwärtige allgemein gebräuchliche Englisch widerspiegelt. Um die Aktualität der in der BoE enthaltenen Sprachdaten sicherzustellen, werden dem Korpus beständig neuere Texte hinzugefügt und ältere entfernt. So enthält die BoE im wesentlichen z.Zt. nur Sprachmaterial, das aus den Jahren 1990 - 2004 stammt (Stand Juni 2004). Wie in Kapitel 3.1.6.2 bereits dargelegt wurde, eignen sich *monitor corpora* wie die BoE auf Grund dieses Strukturprinzips sehr gut zur Extraktion von Neologismen, aber die stetig wechselnde Zusammensetzung des Korpus bringt auch Probleme mit sich. Denn zum einen lassen sich durch diesen Umstand mittels Korpusanalysen gewonnene Ergebnisse nach dem nächsten *update* nicht mehr reproduzieren, also auch nicht mehr überprüfen.¹³⁸ Zum anderen ist die Datenkontrolle sehr viel schwieriger als bei einem statischen Korpus und es kann daher leichter zu einem Ungleichgewicht zwischen Korpusstrukturelementen kommen (vgl. Kap. 3.1.6.2). Daß dieser Kritikpunkt auch auf die BoE zutrifft, wird sogar von einer Mitarbeiterin des COBUILD Teams bestätigt, die nach einem Vergleich der Struktur von *sample corpora* wie dem ICE und BNC mit der der BoE äußert: „The Bank of English [...] is comparatively unbalanced and indiscriminate [...]“ (Moon 1998:348). Eine nähere Analyse der Informationen von Krishnamurthy (2002a:1f., Appendix) zur Zusammensetzung der BoE im Januar 2002 bringt ein ähnliches Ergebnis zutage, wie im folgenden dargelegt werden soll:

¹³⁷ Die kontinuierliche Vergrößerung des Korpus wird von Krishnamurthy (2002b:5) auch damit begründet, dass für die Entwicklung von Wörterbüchern des Englischen für Muttersprachler und für die Erstellung bilingualer Wörterbücher zunehmend auch auf das BoE zugegriffen werde, womit eine sehr viel größere Datenmenge erforderlich werde.

¹³⁸ Vgl. auch Sinclair (1998:47). Prof. Teubert, der den Lehrstuhl für Korpuslinguistik an der Universität Birmingham inne hat, bedauert in einer persönlichen Mitteilung, daß er selber über den derzeitigen Aufbau der BoE nicht informiert sei, weil sich die Zusammensetzung des Korpus beständig ändere. Seiner Meinung nach

Bank of English (Jan. 2002):

<i>Subcorpora</i>	<i>percentage of BoE</i>	<i>No of words</i>	<i>No of Texts</i>	<i>Dates</i>
American Books	7.23 %	32.4 Mio.	327	1990>
NPR American Radio	4.96 %	22.2 Mio.	726	1990-93
BBC World Service Radio	4.15 %	18.6 Mio.	143	1990-91
British Books	9.67 %	43.4 Mio.	578	1990>
British Ephemera ¹³⁹	1.03 %	4.6 Mio.	2359	1991-96
British Magazines	9.84 %	44.2 Mio.	1113	1992-00
British Spoken	4.48 %	20.1 Mio.	2670	1991-96
Economist	3.50 %	15.7 Mio.	229	1991-99
Independent	6.26 %	28.1 Mio.	240	1995-99
Times	11.57 %	51.9 Mio.	322	1999-01
Guardian	7.20 %	32.3 Mio.	332	1995-99
New Scientist	1.76 %	7.9 Mio.	188	1992-99
Australian Newspapers	7.79 %	34.9 Mio.	406	1995-99
American Ephemera	0.78 %	3.5 Mio.	2786	1995-96
American Newspapers	2.23 %	10 Mio.	281	1994-96
Sun and News of the World	9.98 %	44.8 Mio.	597	1997-01
American Academic Textbooks	1.41 %	6.3 Mio.	31	1990-96
American Spoken	0.45 %	2 Mio.	16	1994-97
Strathy Canadian Corpus	3.55 %	15.9 Mio.	234	1980>
Wolverhampton Business English Corpus	2.15 %	9.7 Mio.	23	1999-00
Total	ca. 100% ¹⁴⁰	448.5 Mio.	13.601 ¹⁴¹	1980>

Tab. 3.1 *Bank of English* 2002

(nach Krishnamurthy 2002a:1f., Appendix)

Rechnet man beispielsweise den Anteil des Sprachmaterials in der BoE zusammen, der aus Zeitungen stammt, kommt man auf ca. 45 %.¹⁴² Bezieht man auch die Menge der Sprachdaten aus dem *Economist*, *New Scientist* und anderen britischen Zeitschriften sowie den zwei

sollte das Problem der mangelnden Dokumentation der durchgeführten *updates* bei der Erstellung des *Global English Monitor Corpus* (s. u.) besser gelöst werden.

¹³⁹ Die Kategorie 'Ephemera' enthält, sowohl was das BrE als auch das AmE Subkorpus betrifft, „junk mail, advertising leaflets, tourist brochures, etc.“ (Krishnamurthy 2002a:1, Appendix).

¹⁴⁰ Bei der Addition aller Werte erhält man ganz genau nur 99,99 %. Dieser Umstand liegt wahrscheinlich an Rundungsfehlern bei den Einzelangaben.

¹⁴¹ Krishnamurthy (2002a:1f., Appendix) gibt hier 14.813 an, aber bei der Summierung aller Zahlen dieser Spalte sind es nur 13.601 Texte.

¹⁴² Davon gehen allein 35.01% auf britische Zeitungen zurück.

Radiosubkorpora ein, wird deutlich, daß das Korpus fast zu 70 % auf ‚Mediensprache‘ basiert. In der Einleitung zum COBUILD₃ (S. xiv) wird dieses Faktum bestätigt, denn dort heißt es: „Two-thirds of the corpus is made up of media language: newspapers, magazines, radio and TV“. Damit ist dieses Genre deutlich überrepräsentiert, was sich auch bei den aus Korpusanalysen erwachsenden Ergebnissen bemerkbar macht. So führt Summers (1996:4) beispielsweise an, daß im COBUILD₂ das Wort *newscaster* zu den 700 im Englischen am häufigsten verwendeten Wörtern gezählt wird und alle Bezeichnungen für Nationalitäten, so auch das Wort *Yemeni*, dieselbe Frequenzmarkierung erhalten wie das Lexem *American*. Summers (ebd.) macht den hohen Prozentsatz an Nachrichtenmaterial in der BoE für diese Entscheidungen des COBUILD Teams verantwortlich. Zu untersuchen bleibt, inwieweit die Markierung von Amerikanismen im COBUILD₃ und CCD₅ dadurch beeinflusst wird. Der Grund für die Aufnahme einer umfangreichen Menge von Sprachmaterial aus journalistischen Texten liegt wahrscheinlich darin, daß Zeitungen und Zeitschriften heutzutage am Computer erstellt werden und ihre Sprachdaten daher bereits in elektronischer Form vorliegen, wodurch deren Verarbeitung schneller und kostengünstiger durchgeführt werden kann als es z.B. bei der Kategorie ‚Ephemera‘ der Fall ist.¹⁴³ Durch das periodische Erscheinen der Vertreter dieses Genres ist zudem ein fortwährender Datenzufluß möglich. Vergleicht man die Zusammensetzung der BoE im Januar 2002 mit der des Korpus im Oktober 2000, so fällt auf, daß es ausschließlich Sprachmaterial aus Zeitungen ist, das durch neueres ersetzt oder ergänzt wurde. So ist die Kritik von Biber/ Conrad/ Reppen (1994:188), daß bei der Entwicklung der BoE der Einbezug leicht verfügbarer Materialien eine große Rolle spielt, wohl nicht ganz von der Hand zu weisen. Interessanterweise handelt es sich bei den beiden der BoE erst kürzlich hinzugefügten Subkorpora, dem *Strathy Canadian Corpus* und *Wolverhampton Business English Corpus* (WBE) ebenfalls um solche, deren Sprachdaten schon in elektronischer Form bereitstanden und deren Akquisition vermutlich nicht schwierig war. Die für den Aufbau des WBE verantwortlichen Computerlinguisten bedienten sich beispielsweise des World Wide Web, indem Texte von 23 verschiedenen Webseiten mit Fokus auf dem Bereich Wirtschaft zu einem Korpus zusammengestellt wurden. Den Auskünften zu diesem Korpus zufolge stammen die verwandten Texte allerdings nicht nur von Sprechern muttersprachlicher Varietäten des Englischen, wie dem BrE und AmE, sondern wurden teilweise von Autoren aus Ländern wie z.B. der Schweiz, Belgien und den Niederlanden verfaßt. Der Nutzen dieses zuletzt genannten Teils des WBE für die BoE erscheint fraglich. Zum Zwecke einer sinnvollen Auswertung des WBE sollte jedoch

¹⁴³ So heißt es bei Sinclair (1998:124): „Kurzlebige [= ‚Ephemera‘], basierend auf einer großen Vielfalt von Pamphleten und Broschüren, muß immer wieder neu eingegeben werden, und das ist mühsam und teuer.“ Das wird die Ursache dafür sein, daß der Anteil dieser Kategorie am gesamten Korpus gering ist.

zumindest die Herkunft jedes Textes gekennzeichnet sein, damit sich die Lexikographen bei ihrer Arbeit darüber bewußt sind, um welche Varietät des Englischen es sich im Einzelfalle handelt.

Der Prozentsatz gesprochener Daten in der BoE (ca. 14 %, Stand Januar 2002) ist gering im Vergleich zu der Masse mündlichen Materials, das täglich produziert wird.¹⁴⁴ Er liegt jedoch höher als im BNC (10 %) und im *Cambridge International Corpus* (5,4 %, Stand: Mai 2004). Die BoE weist mit der Anzahl von 62,9 Mio. laufenden Wörtern aus dem gesprochenen Englisch deutlich mehr Daten dieses Sprachmodus auf als das *Cambridge International Corpus* (39 Mio., Stand: Mai 2004) oder das BNC (10 Mio.). Betrachtet man allerdings die Quellen des mündlichen Materials in der BoE genauer, so stellt man fest, daß allein 40,8 Mio. Wörter auf die zwei Radiosubkorpora im Korpus, nämlich das NPR *American Radio* und das BBC *World Service Radio*, zurückgehen.¹⁴⁵ Bei diesen Daten handelt es sich aber oft um bereits im vorhinein geplante oder sogar schriftlich vorbereitete Sprachäußerungen, deren sprachliche Merkmale den Informationen Rundells (1996a:208f.) zufolge eher denen geschriebener Texte ähneln. Spontan gesprochenes Material findet sich zwar im ‚British Spoken‘ Subkorpus (20,1 Mio. Wörter), jedoch bleibt unklar, wie hoch der Anteil von Sprachdaten persönlicher Gespräche (in der Familie, unter Freunden u.ä.) darin ist.¹⁴⁶ In der Forschungsliteratur¹⁴⁷ wird hervorgehoben, daß sich die Strukturmerkmale dieser Gattung von denen anderer fundamental unterscheiden. Die Zusammensetzung des ‚American Spoken‘ Subkorpus ist in jedem Falle kritisch zu beurteilen, da dieses Subkorpus hauptsächlich aus Material besteht, das auf Informationen des Weißen Hauses an die Presse und auf universitäre Ausschußsitzungen zurückgeht.¹⁴⁸ Die Vielfalt der im britischen Pendant vorhandenen Genres ist hier nicht vertreten, ganz zu schweigen von ‚face-to-face conversations‘. Auf Grund der in der BoE verwandten Quellen für das gesprochene AmE stellt sich die Frage, ob Ausdrücke aus der informellen Sprache oder dem Slang dieser Varietät in ausreichendem Maße im Korpus repräsentiert sind. Das ‚American Spoken‘ Subkorpus (2 Mio. Wörter) ist zudem sehr klein, es hat nur ein Zehntel der Größe seines britischen Gegenstücks. Dies mag wiederum an der Schwierigkeit der Beschaffung des Materials liegen. Das amerikanische Radiosubkorpus (22,2 Mio. Wörter) ist dagegen sogar größer als das britische (18,6 Mio. Wörter).

¹⁴⁴ Vgl. auch Moon (1998:348).

¹⁴⁵ Diese zwei Subkorpora wurden für die Erstellung des *BBC English Dictionary* benötigt (vgl. Krishnamurthy 2002b:5).

¹⁴⁶ Vgl. die Informationen zum Aufbau der Subkorpora der BoE in Krishnamurthy (2001:11). Die Beschaffung dieser Kategorie mündlicher Sprachdaten ist zugegebenermaßen auch mit besonders viel Aufwand verbunden.

¹⁴⁷ Vgl. z.B. Moon (1998:349ff.).

¹⁴⁸ Vgl. die Auskünfte zum ‚American Spoken‘ Subkorpus in Krishnamurthy (2001:11).

Insgesamt liegt der Anteil AmE Sprachdaten an der BoE im Januar 2002 bei mindestens ca. 17 % (76,4 Mio. Wörter).¹⁴⁹ Neben den zwei bereits behandelten Subkorpora des gesprochenen AmE existieren noch folgende andere Subkorpora, die nur aus AmE Sprachmaterial bestehen: 'American Books' (7,23 %), 'American Ephemera' (0,78 %), 'American Newspapers' (2,23 %) und 'American Academic Textbooks' (1,41 %). Der Umfang der Daten, die aus amerikanischen Zeitungen entnommen wurden, ist im Vergleich zu der Menge des aus britischen Zeitungen gewonnenen Sprachmaterials (ca. 35 %) sehr gering. Zudem wurden im Gegensatz zum britischen Pendant in das amerikanische Subkorpus nur regionale und lokale Zeitungen einbezogen, auf die man über das World Wide Web Zugriff haben konnte. Die Gattung 'American Academic Textbooks' hat kein britisches Gegenstück, das Subkorpus 'British Magazines' aber auch kein amerikanisches. Bisher besteht also keine Parallelität hinsichtlich der Kategorien, durch die das BrE und AmE im Korpus vertreten werden sollen. Laut Krishnamurthy (2002b:7) besteht jedoch die Absicht, in der Zukunft ein 'British Academic Textbooks subcorpus' zu erstellen sowie Daten aus amerikanischen Zeitschriften zu gewinnen, um somit kontrastive Analysen zwischen den entsprechenden AmE und BrE Subkorpora zu ermöglichen.

Wie bereits erwähnt, sind mindestens ca. 17 % der in der BoE enthaltenen Sprachdaten dem AmE zuzurechnen. Das BrE ist demgegenüber mit mindestens ca. 69 % im Korpus vertreten.¹⁵⁰ Durch diese Gewichtung wird sehr offensichtlich, daß sich das Interesse des COBUILD Teams, wie schon bei der Erstellung der *Birmingham Collection of English Texts*, vor allem auf das BrE konzentriert. Zieht man jedoch die bedeutende Rolle, die das AmE heutzutage in der Welt einnimmt, in Betracht, erscheint der Anteil des AmE an der BoE etwas niedrig zu sein. Der Prozentsatz AmE Sprachmaterials im *Longman/ Lancaster English Language Corpus* (40 %) und im *Cambridge International Corpus* (ca. 32 %, Stand Mai 04) ist deutlich höher. Die Angaben im COBUILD₃ (S. xiv) zum Umfang AmE Sprachdaten in der BoE treffen leider nicht ganz zu. Dort heißt es: „Although most of the sources are British, approximately 25 % of our data comes from American English sources, and about 5 % from other native varieties of English - such as Australian and Singapore.“ Da das COBUILD₃ bereits im Jahre 2001 erschienen ist, war die Zusammensetzung der BoE zur Zeit der Erstellung des Wörterbuchs anders als den angeführten Informationen zufolge (anders also als in dem oben abgedruckten Überblick dargestellt). Aber auch im Oktober 2000, als noch am COBUILD₃ gearbeitet wurde, lag der

¹⁴⁹ Da der Verfasserin nicht bekannt ist, in welcher Größenordnung das AmE Eingang in das WBE gefunden hat, können die Daten dieses Korpus leider nicht berücksichtigt werden.

¹⁵⁰ Auch der Umfang des in das WBE aufgenommenen BrE Sprachmaterials ist leider nicht bekannt und kann deshalb nicht berücksichtigt werden.

Anteil des AmE an der BoE nur bei ca. 18 %.¹⁵¹ Es wird vom COBUILD Team wahrscheinlich angestrebt, daß 25 % aller Sprachdaten im Korpus aus dem AmE stammen sollen, aber de facto wurde dieser Zielprozentsatz seit 1993 nicht mehr erreicht.¹⁵² Betrachtet man die Entwicklung der BoE von 1993 bis 2002, so fällt auf, daß sich zwar die absolute Menge an AmE Sprachmaterial über die Jahre mehr als verdoppelt hat (von ca. 32 Mio. Wörtern im Jahre 1993 bis zu 76,4 Mio. Wörtern im Januar 2002), der prozentuale Anteil dieser Varietät am Gesamtkorpus jedoch von ca. 27 % im Jahre 1993 auf ca. mindestens 17 % im Januar 2002 zurückgegangen ist. Positiv muß allerdings vermerkt werden, daß in den Jahren 1996 bis 2000 die Zahl verschiedener Genres aus dem AmE erhöht wurde, so daß den Lexikographen bei der Erstellung des COBUILD₃ nicht nur zwei bzw. drei¹⁵³, sondern insgesamt wahrscheinlich alle sechs der Genres, die AmE Sprachdaten enthalten und seit Oktober 2000 Bestandteil der BoE sind, zur Verfügung gestanden haben¹⁵⁴. Die Zusammensetzung der BoE im Oktober 2000, die wahrscheinlich als Basis für die Erstellung des COBUILD₃ diente, unterscheidet sich vom Aufbau der BoE im Januar 2002 bezüglich des AmE nur insofern, als vermutlich, wie bereits erwähnt, auch AmE Sprachmaterial in das WBE eingegangen ist (Informationen zu Art und Umfang der AmE Sprachdaten im WBE waren leider nicht verfügbar). Unterschiede zwischen den Fassungen der BoE im Oktober 2000 und Januar 2002 scheinen insgesamt lediglich darin zu bestehen, daß für das jüngste *update* - wie bereits oben erwähnt - eine Reihe von älteren Texten aus den Zeitungen *Independent*, *Times* und *Sun and News of the World* durch neuere ersetzt wurden. Außerdem enthielt die BoE im Oktober 2000 noch ein weiteres Zeitungssubkorpus (Sprachmaterial aus der überregionalen Tageszeitung *Today*), dagegen aber noch nicht das *Strathy Canadian Corpus* und *Wolverhampton Business English Corpus* (WBE). Mit dem *Strathy Canadian Corpus* wurden nun auch Sprachdaten aus dem Kanadischen Englisch in die BoE einbezogen. Vorher stand außer dem BrE und AmE Sprachmaterial nur solches aus dem Australischen Englisch zur Verfügung. Schon in einem 1996 veröffentlichten Aufsatz gab Clear (1996:314) an, daß es das Ziel des COBUILD Teams sei, die Zahl der in der BoE enthaltenen Varietäten zu erhöhen. Der Anteil des Australischen Englisch an der BoE liegt im Januar 2002 bei mindestens 7,79 % und der des Kanadischen Englisch bei mindestens 3,55 %¹⁵⁵. Damit stammen mindestens ca. 11 % der Sprachdaten in der BoE aus anderen Varietäten als dem AmE und BrE. Der in der Einleitung des COBUILD₃ (S. xiv) in diesem Zusammenhang angegebene

¹⁵¹ Vgl. Krishnamurthy (2001:11).

¹⁵² Vgl. die Informationen in Krishnamurthys (o.J.) ‚In-house working paper‘.

¹⁵³ Von 1993 - 1995 bestanden nur die Subkorpora ‚American Books‘, ‚American Radio (NPR)‘ und ‚Wall Street Journal‘. Das Sprachmaterial aus dem *Wall Street Journal* wurde 1996 aus der BoE entfernt, weil es nach Ansicht des Cobuild Teams veraltete (vgl. Krishnamurthy (2002b:7)

¹⁵⁴ Der Anteil der vier neu hinzugefügten AmE Genres am Gesamtkorpus ist indessen verhältnismäßig gering.

Prozentsatz (s.o.) trifft also nicht mehr zu, der Umfang dieses Sprachmaterials ist jetzt deutlich höher als angeführt.¹⁵⁶ Eine angemessene Berücksichtigung anderer Varietäten des Englischen außer dem BrE und AmE ist zu begrüßen. Den Worten Clears (s.o.) nach zu urteilen, werden die Bemühungen des COBUILD Teams in dieser Richtung noch weitergehen.¹⁵⁷ Wünschenswert wäre dabei, daß die jeweilige Varietät durch eine Reihe verschiedener Genres repräsentiert würde. Im Falle des Australischen Subkorpus in der BoE handelt es sich nämlich derzeit ausschließlich um eine Sammlung von Daten aus regionalen Tageszeitungen. Das *Strathy Canadian Corpus* hingegen enthält zwar eine Vielfalt geschriebener Genres, aber Kategorien wie unveröffentlichte Texte oder verschiedene gesprochene Genres sind unterrepräsentiert. Erfreulicherweise wird an einer diesbezüglichen Erweiterung des *Strathy Canadian Corpus* bereits gearbeitet.

Dem COBUILD Team ist die Wichtigkeit der Korpuszusammensetzung für die aus einem Korpus zu gewinnenden Informationen durchaus bewußt, aber die Herausgeber des COBUILD₃ werden ungeachtet der Unausgewogenheiten im Aufbau der BoE nicht müde, mit der Autorität und Verlässlichkeit der Daten dieses Korpus zu werben: „So the information about English has been checked against the large amount of corpus data that gives COBUILD its reliability and authority“ (COBUILD₃, S. ix).¹⁵⁸ Sie vertreten die Ansicht, daß die BoE die ganze Bandbreite des heutigen Englisch widerspiegeln: „Of course, any corpus will reflect only those language varieties it contains, but the Bank of English reflects the whole range of today’s language [...]“ (ebd., S. xlii). Derjenige, der sich nicht näher mit der Korpuszusammensetzung der BoE beschäftigt hat, wird ihnen diese Aussage wahrscheinlich glauben.

Die BoE gehört zu den größten lexikographischen Korpora des Englischen weltweit.¹⁵⁹ Daher wäre es trotz der oben dargestellten strukturellen Mängel des Korpus von großem Interesse gewesen, eigene Untersuchungen zu den in dieser Arbeit behandelten Amerikanismen anhand der BoE durchzuführen und die Ergebnisse ggf. mit den Resultaten der Analyse des BNC zu vergleichen. Der Zugriff auf die gesamte BoE ist jedoch Personen vorbehalten, die dem COBUILD Team, der Universität Birmingham oder dem *HarperCollins* Verlag angehören.

¹⁵⁵ Leider liegen auch zum Umfang des Sprachmaterials im WBE, das aus anderen Varietäten außer dem BrE und AmE stammt, keine Informationen vor.

¹⁵⁶ Verwunderlich ist, daß in der Einleitung des COBUILD₃ (S. xiv) erwähnt wird (s.o.), daß Singapurisches Englisch in die BoE aufgenommen worden sei. Ein entsprechendes Subkorpus wird in den der Verfasserin vorliegenden Übersichten zur Zusammensetzung der BoE von 1993 - 2002 nicht aufgelistet.

¹⁵⁷ Krishnamurthy (2002b:5,7,15) macht ebenfalls deutlich, dass das COBUILD Team durch den Einbezug einer größeren Anzahl entsprechender Quellen die Darstellung des Englischen als Weltsprache in der BoE in der Zukunft stärken will.

¹⁵⁸ Vgl. auch S. xiii und xv der Wörterbucheinleitung zum COBUILD₃.

¹⁵⁹ Vgl. Rundell (1996b: 4) und die Webseite des *Centre for Corpus Linguistics* in Birmingham (<http://www.corpus.bham.ac.uk/mission.html>, Stand: Mai 2003). Wie bereits erwähnt, enthält die BoE

Außenstehenden wird lediglich ermöglicht, einen Ausschnitt des Korpus im Umfang von 56 Mio. Wörtern online zu verwenden. Durch diese massive Einschränkung, die verständlicherweise die kommerziellen Interessen des *HarperCollins* Verlag wahren soll, erschien eine Analyse der BoE nicht mehr sehr lohnenswert. Aber die Auswertung des *Collins Concise Dictionary* (5th ed.) und des COBUILD₃ wird zeigen, ob die folgende Einschätzung aus dem *Collins English Dictionary Catalogue* von 1999 bezüglich der BoE und der auf diesem Korpus basierenden Produkte zutrifft: “Using these authentic examples and evidence, *Collins Dictionaries* create the most up-to-date and accurate reference books available today”.

Die wichtigsten Daten zur *Bank of English* im Überblick:

<i>Bank of English</i>	
Typ	<i>monitor corpus</i>
Größe	448.5 Mio Wörter (Stand Jan. 2002)
Varietäten des Englischen ¹⁶⁰	BrE: ca. 69 % (= ca. 312 Mio. Wörter) AmE: ca. 17 % (= ca. 76 Mio. Wörter) Andere: ca. 11 % (= ca. 51 Mio. Wörter)
Alter der Quellen	1980 >
Art der Quellen	Entstehungsdatum der überwiegenden Mehrheit der Texte aber von 1990 - geschrieben: ca. 86 % (Bücher (fiktionale und non-fiktionale Texte, unter letzteren auch wissenschaftliche Bücher), Zeitungen, Zeitschriften, ‘Ephemera’, Material aus dem <i>World Wide Web</i> zum Thema Wirtschaft) gesprochen: ca. 14 % (Radiosendungen, Vorlesungen/ Vorträge (‘lectures’), Treffen/ Besprechungen (‘meetings’), ‘interviews’, private Telefongespräche, gesellschaftl. Ereignisse,

mittlerweile bereits 524 Mio. Wörter, vgl. die Internetseite <http://www.cobuild.collins.co.uk/Pages/latest.aspx> (Stand Juni 2004).

¹⁶⁰ Das dem WBE entnommene Sprachmaterial wird an dieser Stelle nicht berücksichtigt, da unklar ist, welche Varietäten in welchem Umfang in das Korpus einbezogen wurden.

	Informationen des Weißen Hauses an die Presse, universitäre Ausschußsitzungen u.a.)
Struktur	<p>Gliederung des Korpus in 20 Subkorpora (Stand Oktober 2002)</p> <p>Aktualisierung des Korpus einmal im Jahr angestrebt (Hinzufügung neuerer Texte, Entfernung älterer Texte), kontinuierliches Sammeln von Sprachmaterial für die BoE</p> <p>Ziel: Vergrößerung des Korpus und Steigerung der Anzahl der <i>text types</i>, aber keine Festlegung des Anteils verschiedener <i>text types</i> im Korpus</p>

Tab. 3.2 *Bank of English*

Im *Centre for Corpus Linguistics* (CCL) in Birmingham, das mittlerweile anscheinend durch das *Centre for Corpus Research* ersetzt wurde, arbeitet man seit Ende 2001 an einem neuen Projekt, der Erstellung der *World Bank of English* (auch *Global English Monitor Corpus* genannt).¹⁶¹ Die *World Bank of English* ist ebenfalls als *monitor corpus* geplant¹⁶² und wird aus Texten englischsprachiger Zeitungen aus der ganzen Welt bestehen. Den Angaben Teuberts (pers. Mitteilung vom 11.01.02) gemäß sollen über Jahrzehnte hinweg dieselben Zeitungen in das Korpus aufgenommen werden, um Unterschiede zwischen verschiedenen Varietäten des Englischen und vor allem Sprachwandel dokumentieren zu können. Auf der o.g. Internetseite befindet sich die Information, daß das Korpus bereits 20 Milliarden Wörter enthält.

Unter der Voraussetzung, daß Umfang und Art des aus den Zeitungen verschiedener Varietäten entnommenen Sprachmaterials vergleichbar sind, kann die *World Bank of English* mit Sicherheit einen Beitrag dazu leisten, sprachliche Gemeinsamkeiten und Unterschiede dieser Varietäten über einen großen Zeitraum hinweg zu beobachten. Allerdings darf nicht vergessen werden, daß sich die Ergebnisse der Analysen nur auf ein bestimmtes Genre beziehen, also nicht als exemplarisch für alle Genres der Varietäten gelten können.

¹⁶¹ Die folgenden Ausführungen zur *World Bank of English* gehen auf Informationen auf der Webseite des *Centre for Corpus Linguistics* (<http://www.english.bham.ac.uk/ccl/mission.htm>) (Stand Dezember 2004) und auf persönliche Auskünfte des Leiters des Zentrums, Prof. Teubert, und einer seiner Mitarbeiterinnen, zurück.

¹⁶² Da der diachronische Aspekt bei der Erstellung des Korpus eine große Rolle spielt (s.u.), verwundert es etwas, daß es sich bei der *World Bank of English* auch um ein *monitor corpus* handeln soll, bei dem älteres Material durch neueres ersetzt wird.

3.1.2 Wörterbuchproduktionen des Addison Wesley Longman Verlages

Den Lexikographen des Verlags *Addison Wesley Longman* (im folgenden der Einfachheit halber nur mit ‚*Longman*‘ bezeichnet¹⁶³) steht bei ihrer Arbeit das *Longman Corpus Network* (LCN) zur Verfügung. Es setzt sich aus fünf verschiedenen Korpora zusammen: dem *Longman/Lancaster English Language Corpus*, dem *Longman Learner's Corpus*, dem *Spoken British Corpus*, dem *Longman Written American Corpus* und dem *Longman Spoken American Corpus*.¹⁶⁴ Jedes dieser Korpora ist unterschiedlich aufgebaut und dient unterschiedlichen Zwecken; zusammen sollen diese Korpora jedoch verlässliche Auskünfte über die englische Sprache geben und vor allem als Basis von Wörterbüchern, Grammatiken und Lehrbüchern des Verlages dienen.

Zur Erstellung des LDOCE₃ wurde laut Herausgeber sowohl das *Longman Corpus Network* als auch das BNC herangezogen. Selbst das jüngste Korpus des *Longman Corpus Network*, das *Longman Spoken American Corpus*, konnte nach Angaben der Redaktion bereits bei der Arbeit an der dritten Auflage dieses Wörterbuchs verwandt werden¹⁶⁵. Da allerdings in diesem Zusammenhang nur von einer Gesamtzahl von über 140 Mio. Wörtern im *Longman Corpus Network* und BNC¹⁶⁶ die Rede ist, wird deutlich, daß einige Korpora im *Longman Corpus Network* noch nicht ihre jetzige Größe erreicht hatten. Bei diesen handelt es sich wahrscheinlich um die zwei ‚dynamischen‘ Korpora im *Longman Corpus Network*, das *Longman Written American Corpus* und das *Longman Learner's Corpus*. Die Erstellung des *Longman/Lancaster English Language Corpus* war nämlich schon 1991 und die des BNC 1994¹⁶⁷ abgeschlossen, so daß diese beiden Korpora mit insgesamt 130 Mio. Wörtern vollständig genutzt werden konnten, während die anderen Korpora im *Longman Corpus Network*, darunter auch die beiden ‚amerikanischen‘ Korpora, demzufolge zum Zeitpunkt der Entstehung des LDOCE₃ insgesamt nur über 10 Mio. Wörter umfasst haben konnten. Da das *Longman Learner's Corpus* nach Aussage des *Longman* Verlages zu dieser Zeit aus ca. 5 Mio. Wörtern bestand¹⁶⁸,

¹⁶³ Die Verlage *Addison Wesley* und *Longman* fusionierten 1995.

¹⁶⁴ Summers (1993) liefert eine ausführliche Beschreibung des *Longman/Lancaster English Language Corpus* (vor allem auf diesen Artikel stützen sich die folgenden Ausführungen zu diesem Korpus). Crowdy (1992) gibt ebenfalls Auskünfte zum *Longman/Lancaster English Language Corpus*, aber auch zum *Longman Learner's Corpus*. Informationen zu allen Korpora des *Longman Corpus Network*, die allerdings nicht sehr umfangreich waren, konnten im Internet auf der Webseite des *Addison Wesley Longman Verlages* unter <http://www.longman.com/dictionaries/corpus/1ccont.html> abgefragt werden (Stand: November 2004). Weitere Hinweise zum *Longman Corpus Network* oder nur zum *Longman/Lancaster English Language Corpus* fanden sich bei Summers (1996:1,3,4), Kennedy (1998: 48-50), Meijs (1996:103f.), Biber / Conrad/ Reppen (1994:188), Rundell (1996b:2), Taylor et al. (1991:343) und im *Longman* ELT Produktkatalog 2000.

¹⁶⁵ Vgl. Karen Sterns Aufsatz (S. 5), der im Internet zu finden ist unter <http://www.longman-elt.com/dictionaries/r3stern.shtml> (Stand: Juni 2000).

¹⁶⁶ Vgl. zum BNC Kap. 3.2.2.2 dieser Arbeit.

¹⁶⁷ Der *Longman* Verlag fügte laut Steve Crowdy (persönliche Mitteilung vom 25.04.01) dem BNC im Jahr 2001, allerdings nur für den Gebrauch im eigenen Hause, eine Datenmenge von 20 Mio. Wörtern aus den Jahren 2000-2001 hinzu (wie diese strukturiert ist, wird leider nicht erläutert).

¹⁶⁸ Vgl. die elektronische Nachricht des *Longman* Verlages vom 21.05.01.

können die AmE Korpora nur eine Größe von über 5 Mio. Wörtern gehabt haben. Die neue Auflage des LDOCE¹⁶⁹ wird daher von einer verbreiterten AmE Datenbasis profitiert haben, indem Entscheidungen zur Erfassung und Markierung von Amerikanismen mit größerer Sicherheit getroffen werden konnten. Die Redakteure der zweiten Auflage des LDEL_C (1998 erschienen) hatten auch schon die Möglichkeit, auf einen sehr erweiterten AmE Datenbestand zurückzugreifen: Die AmE Korpora des Verlages bestanden zum Zeitpunkt der Erstellung des LDEL_C₂ aus insgesamt 120 Mio. Wörtern¹⁷⁰. Während es sich bei dem *Longman Spoken American Corpus* um ein statisches Korpus handelt (das seine Größe aus diesem Grund nicht verändert haben dürfte) enthält das *Longman Written American Corpus* mittlerweile 150 Mio. laufende Wörter¹⁷¹ und ist somit sogar schon anderthalbmal so groß wie das BNC₁. Das *Longman Corpus Network* umfasst nach Angaben von Steve Crowdy (Stand: April 2001) insgesamt ca. 300 Mio. laufende Wörter.¹⁷² Bis Februar 2004 hat sich diese Größenangabe für das *Longman Corpus Network* nicht verändert.¹⁷³

3.2.1.1 Das Longman Corpus Network

3.2.1.1.1 Longman/ Lancaster English Language Corpus

Die Arbeit am *Longman/ Lancaster English Language Corpus* (im folgenden mit ‚LLELC‘ abgekürzt)¹⁷⁴ begann im Januar 1988¹⁷⁵. 1987 war die 2. Auflage des LDOCE erschienen, die den Angaben der Herausgeber zufolge auf dem *Longman Citation Corpus*¹⁷⁶ basierte. Dieses Korpus bestand laut Redaktion aus einer 25,5 Mio. Wörter umfassenden Zitatensammlung (für die 2. Auflage waren zu dem ursprünglich 25 Mio. Wörter großen Bestand eine halbe Million Wörter hinzugefügt worden, davon 15.000 Neologismen) und aus weiteren zwei Mio. Wörtern, die per Zufallsprinzip aus aktuellen britischen und amerikanischen Zeitungen ausgewählt worden waren.¹⁷⁷ In ihrem Artikel über das LLELC setzt sich Della Summers (1993: 183, 185, 186, 190)

¹⁶⁹ Das LDOCE₄ erschien erst 2003 und konnte deshalb leider nur noch in Ansätzen in dieser Arbeit berücksichtigt werden.

¹⁷⁰ Vgl. die elektronische Nachricht des *Longman* Verlags vom 21.05.01.

¹⁷¹ Diese Information stammt aus der persönlichen Mitteilung Steve Crowdys vom 25.04.01.

¹⁷² Seinen Informationen zur Größe der einzelnen Korpora zufolge müßte das *Longman Corpus Network* ohne die zum BNC neu hinzugefügten Daten ca. 295 Mio. Wörter umfassen, mit diesen dagegen schon 315 Mio. Wörter.

¹⁷³ Vgl. die Informationen auf der Internetseite <http://www.longman.com/ldoce/members/stephen.html>.

¹⁷⁴ „Lancaster“ wurde in den Namen des Korpus mit aufgenommen, weil Professor Geoffrey Leech von der Universität Lancaster (der u.a. auch für die Erstellung des *LOB Corpus* von 1970-1976 an der Universität Lancaster verantwortlich war) laut Della Summers entscheidende Anregungen bei der Entwicklung des LLELC gegeben hatte und das Longman Team die schon lang währende Zusammenarbeit mit ihm würdigen wollte, vgl. Summers (1993:182).

¹⁷⁵ Vgl. Summers (1993:181). Rundell (1996b:2) weist darauf hin, daß das Korpus seit 1990 einsatzfähig ist.

¹⁷⁶ Vgl. das Kapitel *General Introduction* des LDOCE₂, S. F8.

¹⁷⁷ Ebd. Summers (1993:186) gibt jedoch an, daß für grammatische Angaben in der 2. Auflage des LDOCE vor allem auf ein 5 Mio. Wörter großes Zeitungskorpus zurückgegriffen wurde. An anderer Stelle (1993:183,190)

kritisch mit den Nachteilen von Belegsammlungen oder Zeitungskorpora als alleinige Grundlagen zur Erstellung von Korpora auseinander. Das Bewußtsein darüber, daß auch erfahrene Redakteure dazu tendieren, auf der Suche nach sprachlichem Material (z.B. in Zeitungen und Zeitschriften) für eine Zitatensammlung gängige sprachliche Phänomene zugunsten von neuen oder ungewöhnlichen Wörtern bzw. Wortbedeutungen zu vernachlässigen und Zeitungskorpora durch ihre Festlegung auf ein Genre eine Ausgewogenheit hinsichtlich verschiedenster Aspekte (z.B. der Sachgebiete, Textgattungen, Themen, unterschiedlichen Schwierigkeitsgrade der Texte, Wiedergabe geschriebener und gesprochener Sprache) vermissen lassen, war vermutlich ein entscheidender Faktor für den Wunsch nach einem neuen, möglichst ausgewogenen und ‚repräsentativen‘ Korpus mit genau festgelegten Strukturprinzipien. Für wie wichtig das *Longman* Team die Zusammensetzung eines Korpus in Hinblick auf die Qualität der Informationen, die aus dem Korpus gewonnen werden können, hält, verdeutlicht Della Summers (1993:186) mit diesen Worten: „We believe it to be self-evident that the way the corpus is put together determines the nature and quality of the information that can be derived from a corpus regarding the meaning and behaviour of words and phrases“.

Zudem hatte der *Collins* Verlag in Zusammenarbeit mit einem Forschungsteam des Englischen Seminars der Universität Birmingham 1987 (in demselben Jahr als auch die 2. Auflage des LDOCE erschien), wie bereits erwähnt, ein neues Wörterbuch für fortgeschrittene Lerner des Englischen veröffentlicht, das sich auf ein ca. 20 Mio. Wörter großes Korpus stützte. Zumindest das sog. ‚Main‘ *COBUILD Corpus* war der Versuch, ein Korpus zu schaffen, das nicht aus subjektiv ausgewählten Zitaten, reinen Zeitungstexten o.ä. besteht, sondern gemäß bestimmter Prinzipien aus verschiedensten Quellen zusammengesetzt sein sollte, um die ‚Repräsentativität‘ des Korpus so gut als möglich zu gewährleisten. Kennedy (1998:47) konstatiert, daß das *COBUILD Corpus* in den 80er Jahren das größte und modernste verfügbare Korpus war, das Forscher aus aller Welt für Forschungsvorhaben nach Birmingham zog. Der Konkurrenzdruck mag die Entscheidung des *Longman* Teams für die Erstellung eines möglichst ausgewogenen und ‚repräsentativen‘ Korpus beschleunigt haben. Zitatebanken haben laut Summers (1993:185 f.) beim *Longman* Verlag aber immer noch ihren Platz, um Belegmaterial für Neologismen zu liefern, die ihren Eingang in neue Auflagen bestehender Wörterbücher und in das alle zwei Jahre erscheinende *Longman Register of New Words* finden sollen.¹⁷⁸

spricht Summers ebenfalls von 5 Mio. (und nicht 2 Mio.) Wörtern in dem sog. *Longman Newspaper Corpus*, das 2 Mio. Wörter aus Zeitungstexten und 3 Mio. Wörter aus Zeitschriftentexten enthalten habe.

¹⁷⁸ Dieses Neologismenwörterbuch existiert mittlerweile allerdings der elektronischen Nachricht des *Longman* Verlags vom 21.05.01 zufolge nicht mehr. Rundell (1992a:10) gibt die Auskunft, daß im Rahmen des *Longman Wordwatch* Programmes die allgemeine Öffentlichkeit darum gebeten wird, dem *Longman* Team über Neologismen und Sprachwandel i.allg. zu berichten. In der Lehrerhandreichung zu dem von Langenscheidt-Longman herausgegebenen Heft *DCE Dictionary Skills* (1999) heißt es (S. 8), wöchentlich würden 50

Folgende Zielvorstellungen wurden im Zusammenhang mit dem LLELC Projekt formuliert: Es sollte ein wohlausgewogenes Korpus des Englischen des 20. Jahrhunderts entstehen, das vor allem Britisches und Amerikanisches Englisch, aber auch andere wichtige Varietäten des Englischen in geschriebener und gesprochener Form einbezieht und zwischen 30 und 50 Mio. Wörter umfasst.¹⁷⁹ Das Korpus sollte in erster Linie als Basis für die Erstellung von Wörterbüchern, Grammatiken und Sprachlehrbüchern jeglicher Art dienen, außerdem aber auch gegen einen geringen Kostenbeitrag Sprachforschern aus akademischen Instituten zur Verfügung gestellt werden¹⁸⁰. Somit wollte *Longman* wie schon zuvor das COBUILD Team die „magische Größe“ der ersten Standardkorpora (*Brown, LOB*) von 1 Million Wörtern überwinden. Entschieden wendet sich Summers (1993:182) gegen eine Vorgehensweise bei der Sammlung für Korpusmaterial, die sich weniger nach vorher genau bedachten und im Vorfeld festgelegten Prinzipien richtet, sondern eher nach der Frage, welche Texte z.B. schon in maschinenlesbarer Form verfügbar sind oder hinsichtlich der urheberrechtlichen Bestimmungen keine Probleme bereiten.

Die Planung des LLELC sah von Anfang an auch den Einbezug gesprochener Sprache vor (ähnlich wie es beim *COBUILD Corpus* gewesen war, vgl. Kap. 3.2.1 dieser Arbeit), der vom *Longman* Team als sehr wichtig erachtet wurde, doch wurde die detaillierte Planung und Sammlung des gesprochenen Materials zunächst auf die Zeit nach dem Beginn der Erstellung des geschriebenen Teils verschoben. Summers (1993:184) führt als Gründe dafür an, daß das *Longman* Team sich zum einen dessen bewußt gewesen sei, daß die logistischen Probleme bei der Sammlung einer großen Menge gesprochenen Materials sehr viel größer sind als bei der Beschaffung geschriebenen Materials, zum anderen die Faktoren, die für den Aufbau eines gesprochenen Korpus eine Rolle spielen, für noch nicht zureichend überblickbar hielt. Es ist anzunehmen, daß die Mitarbeiter von *Longman* erst einmal bei der Erstellung des auf schriftlichem Material basierenden Teils Erfahrungen für dieses aus dem Nichts zu schaffende,

Neologismen der Wordwatch-Datenbank bei Longman zugeführt. Laut Crowdy (pers. Mitteilung vom 25.04.01) läuft dieses Programm zwar noch, soll aber bald in veränderter Form von neuem gestartet werden.

¹⁷⁹ Es blieb schließlich bei ca. 30 Mio. Wörtern. In der Lehrerhandreichung zu dem Heft *DCE Dictionary Skills* wird allerdings von „weit über 30 Mio. Wörtern“ im LLELC gesprochen (S. 8). Diese Angabe widerspricht Steve Crowdys Aussagen in der persönlichen Mitteilung vom 25.04.01. Dies ist ein gutes Beispiel dafür, wie die Informationen in verschiedenen Quellen divergieren.

¹⁸⁰ Vgl. Summers (1993:184). Der als zweites genannte Zweck kommt Forschern zugute, ist jedoch natürlich auch für den *Longman* Verlag von Interesse - nicht nur durch die Gebühr, die die Forscher für die Benutzung des Korpus bezahlen müssen. So ist es z.B. die Hoffnung des Longman Teams, daß linguistische Analysen anhand des Korpus die Ausgewogenheit der Struktur des LLELC entweder bestätigen oder in Frage stellen werden und Forscher somit Anregungen zur Verbesserung des Aufbaus des LLELC oder bezüglich der Strukturprinzipien kommender Korpora geben können (ebd. S. 191), was wiederum der Qualität der Wörterbücher nutzen würde. Nebenbei spielen vielleicht auch Werbungsgründe eine Rolle und der Wunsch, hinter dem Angebot des COBUILD Verlags für Forscher nicht zurückzustehen (s.o. und z.B. Renouf 1989:87). Allerdings hat *Longman* den Angaben Della Summers' zufolge bereits in den 80er Jahren akademischen Forschungsinstituten den Zugriff auf Wörterbücher und Thesauren des Verlages gestattet, vgl. Summers (1993:184) und Kennedy (1998:35).

vollkommen neue Korpus sammeln wollten, bevor sie sich an die schwierigere Konzeption eines aus mündlichem Material bestehenden Teils begaben. Erst Anfang 1990 wurde mit der detaillierten Planung der Gestaltung des *Spoken Corpus* begonnen. Als 1991 das BNC Projekt initiiert wurde, band man die Forschungsarbeit zu dem *Spoken Corpus* in das Projekt ein: Das *Longman* Team wurde mit der Erstellung des *BNC Spoken Corpus* beauftragt. So ist dieses Korpus, das ursprünglich Teil des LLELC werden sollte, nun mit dem *BNC Spoken Corpus* identisch, wird aber gleichzeitig auch als Teil des *Longman Corpus Network* angesehen (zum *Spoken Corpus* s. das Kapitel zum BNC).

Wie bereits angedeutet, erschien es dem *Longman* Team unverzichtbar, daß das gesamte Korpus so weit als möglich ‚repräsentativ‘ sein sollte, um verlässliche lexikalische Informationen für die Erstellung der Verlagsprodukte liefern zu können¹⁸¹. Die Frage „repräsentativ in Bezug auf was?“ wird von Summers (1993:186) folgendermaßen beantwortet: „Representative of the standard language in a very general sense, not restricted to a regional variety (e.g. British English or a local dialect) or a narrow range of text types (e.g. scientific texts, newspaper writing, language of a particular social class)“. Dem *Longman* Team ist bewußt, daß die genannte Grundgesamtheit („the standard language in a very general sense“) kaum präzise definiert werden kann. Dennoch hält es eine so vage Beschreibung der Grundgesamtheit für gerechtfertigt, da das Korpus hauptsächlich für die Erstellung von Lernerwörterbüchern genutzt würde und die Lerner sehr divergierende Ansprüche an die Wörterbücher stellten (ebd., S. 186). Mit ‚Repräsentativität‘ ist aber nicht gemeint, daß die volle Variationsbreite der Grundgesamtheit einbezogen werden soll, die vom *Longman* Team als ungeheuer groß eingeschätzt wird (ebd. S. 186). In der Praxis bedeutet der Begriff ‚repräsentativ‘ für die Redaktion zum einen, daß in das Korpus hineingenommen wird „what we judge to be the typical and central aspects of the language“ (ebd. S. 189f.) und zum anderen, daß eine ausreichende Anzahl von Korpusbelegen zu Wörtern oder Wendungen für Lexikographen oder andere Sprachforscher zur Verfügung gestellt werden soll, damit präzise Aussagen über die Verwendung dieser sprachlichen Phänomene getroffen werden können (ebd., S. 190). Die erste der beiden Aussagen erweckt den Eindruck, als ob das, was letztlich als ‚repräsentativ‘ betrachtet und somit in das Korpus aufgenommen wird, sich nach Einschätzungen der Redaktion richtet, aber bei der Beschreibung der Struktur des LLELC wird immer wieder das Bemühen der Redaktion deutlich, möglichst objektive Auswahlkriterien für das Korpusmaterial zu finden. Eine statistische Vorgehensweise bei der Erstellung des LLELC wird auch mit dem Hinweis

¹⁸¹ Vgl. auch Summers (1996:1). Die nach demographischen Prinzipien erstellten Teile der auf gesprochenem Material basierenden Korpora werden von ihr als besonders repräsentativ angesehen, da die demographischen

darauf abgelehnt, daß ein Korpus von 30 Mio. Wörtern schon im Vergleich mit der Gesamtheit der geschriebenen Textmenge statistisch gesehen klein ist. Stattdessen entschied sich das *Longman* Team gemäß Summers (1993:190ff.) nach kritischer Auseinandersetzung mit anderen Auswahlmethoden für ein Stichprobenverfahren, das als wesentliches Organisationsprinzip die Einbeziehung einer großen Bandbreite ‚objektiv‘ definierter Text- oder Dokumenttypen vorsieht¹⁸². Inwieweit die Verwendung des Begriffs ‚objektiv‘ in diesem Zusammenhang seine Berechtigung hat, ist fraglich, denn die Einteilung von Texten in Sachgruppen und insbesondere in Niveaustufen (Angabe des ‚level‘) ist bis zu einem gewissen Grade subjektiv. Auch der Auswahl und Definition der relevanten Sachgruppen liegen subjektive Entscheidungen zugrunde. Es ist aber begrüßenswert, daß die *Longman* Redaktion auf statistische Methoden zurückgreifende linguistische Analysen des LLELC, wie sie schon bereits von Biber an anderen Korpora durchgeführt wurden, ausdrücklich anregt, um die strukturelle Ausgewogenheit des Korpus überprüfen und ggf. Mängel korrigieren zu können. Das *Longman* Team ist ebenfalls daran interessiert, auf diese Weise eventuell nützliche Hinweise für bessere Strukturprinzipien zukünftiger Korpora zu gewinnen (ebd., S. 191).

Im folgenden wird nun genauer auf die Struktur des LLELC eingegangen. Solche Aspekte, die für das Thema der vorliegenden Arbeit für nicht so relevant gehalten werden, sollen nur gestreift werden. Es erscheint aber sinnvoll, einen kurzen Überblick über die Zusammensetzung des LLELC zu geben, denn, wie schon oben gesagt¹⁸³, hat die Struktur eines Korpus entscheidende Bedeutung für die Verlässlichkeit der Informationen, die aus ihm gewonnen werden können. Je ‚repräsentativer‘ das Korpus, desto ‚repräsentativer‘ sind vermutlich auch die Aussagen, die im Wörterbuch über die Bedeutung und den Gebrauch von Amerikanismen gemacht werden können.

Das erste Kriterium beim Aufbau des LLELC war die Unterscheidung zwischen fiktionalem und non-fiktionalem Sprachmaterial¹⁸⁴. Diese Kategorisierung hat, wie die Redaktion betont, den Vorteil, daß sie objektiven Kriterien unterliegt. Das LLELC besteht zu ca. 42 % aus fiktionalen und zu ca. 58 % aus non-fiktionalen Texten. Das *Longman* Team gibt selber an, daß der Prozentsatz veröffentlichter fiktionaler Texte im Vergleich zu non-fiktionalen Texten, auf das gesamte gedruckte Sprachmaterial bezogen, eigentlich sehr viel niedriger liegt, verteidigt aber diesen hohen Anteil damit, daß fiktionale Texte eine größere Wirkung hätten als

Verhältnisse in der Gesamtbevölkerung hinsichtlich der Kategorien Geschlecht, Alter, ethnische Herkunft und Bildung statistisch korrekt widerspiegelt würden (vgl. ebd.).

¹⁸² Vgl. ebd., S. 190: „[...] using the notion of a broad range of objectively defined document or text types as its main organizing principle“. Mit „text type“ oder „document type“ ist folgendes gemeint: „text from a particular subject area, together with a cluster of relatively objectively identifiable features such as time, region, medium, and level“ (ebd., S. 192).

¹⁸³ Vgl. z.B. Kap. 3.1.6.2 dieser Arbeit.

non-fiktionale Texte und eine höhere Zahl von Lesern aufweisen könnten (was sich anhand von Ausleihstatistiken sowie Bestsellerlisten aus den letzten 10 Jahren und aufgrund der Tatsache, daß auf Leselisten für Prüfungen im Bereich Englisch als Zweit- oder Fremdsprache mehr fiktionale Texte vertreten seien, beweisen lasse). Kennedy (1998:49) wendet dagegen zurecht ein, daß das journalistische Genre heutzutage eine noch größere (und unterschiedlichere Leserschaft) habe, und sein Einfluß daher als gravierender einzuschätzen sei. Leider gibt es keine gesicherten Erkenntnisse darüber, in welchen Genres die Auftretenshäufigkeit von Amerikanismen besonders hoch ist. Es läßt sich jedoch vermuten, daß in der ‚Mediensprache‘ die Zahl der Amerikanismen höher ist. Eine höhere Zahl von Beispielen zum Gebrauch verschiedener Amerikanismen kommt der Präzision der Definitionen dieser Amerikanismen in den Wörterbüchern zugute. Allerdings ist es natürlich nicht das Ziel der Redaktionen, bei den Überlegungen zur Größe der Textmenge aus bestimmten Genres auf diesen Aspekt zu achten, vielmehr kommt es vielen Verlagen darauf an, ein möglichst ‚repräsentatives‘ Korpus zu schaffen. Dies ist in Hinblick auf die Markierung der Amerikanismen in den Wörterbüchern von Bedeutung: wie oft werden bestimmte Wörter, Wendungen und grammatische Besonderheiten des AmE im BrE Teil des Korpus verwandt? Welche Markierung ist jeweils am zutreffendsten?

Das wichtigste Strukturprinzip des LLELC ist die Kategorisierung des schriftlichen Materials in 10 große Sachbereiche (sog. ‚Superfields‘), denen fixe Zielprozentsätze hinsichtlich ihrer Größe zugewiesen werden.

Superfield

1 Natural & pure sciences	6.0 %
2 Applied science	4.3 %
3 Social science	14.1 %
4 World affairs	10.4 %
5 Commerce & finance	4.4 %
6 Arts	7.9 %
7 Belief & thought	4.7 %
8 Leisure	5.7 %
9 Fiction	40.0 % (discretionary percentage)
10 Poetry, drama, humour	2.3 %

[Total: 99.8 % - Ed.]

(Summers 1993:193)

¹⁸⁴ Vgl. zum Aufbau des LLELC Summers (1993:192 ff.).

Die Zielprozentsätze für die non-fiktionalen ‚Superfields‘ sind laut Summers den Verhältnissen in der Buchproduktion der letzten 20 Jahre in Großbritannien nachempfunden. Auch hier kommt das Bemühen zum Ausdruck, bei strukturellen Entscheidungen den Korpusaufbau betreffend möglichst objektive Kriterien walten zu lassen. Der Aspekt der Objektivität spielte nach Summers u.a. auch eine Rolle bei der Entscheidung, als wichtigstes Strukturprinzip des Korpus die Klassifikation in Sachbereiche und nicht in Genres anzustreben¹⁸⁵, denn die Zuordnung von Texten zu den Sachbereichen könnte als weitestgehend objektiv angesehen werden und in Zweifelsfällen könne man zur Kontrolle auf Bibliotheksklassifikationssysteme zurückgreifen. Ein anderer wichtiger Grund für die Entscheidung, diesen Ansatz zu wählen, war nach Summers die Überzeugung, daß eine nach Sachbereichen gegliederte Klassifizierung eher lexikalisch homogene Kategorien hervorbringen würde als eine nach Genres strukturierte. Da das Korpus vor allem lexikographischen Zwecken diene, sei dieser Ansatz gewählt worden.

Vier Kriterien wurden verwendet, um Texte bestimmten sog. ‚Dokumenttypen‘ zuzuordnen: Region, Zeit, Medium und Level („Primary document features“). Die ersten 3 Kriterien sind objektiv bestimmbar, wie Summers (1993:193) betont. Für die verschiedenen Kategorisierungen innerhalb der Bereiche Region, Zeit, Medium und Level wurden Zielprozentsätze festgelegt, die für das ganze Korpus gelten sollten.

Region

British	50 %
American	40 %
Other	10 %

Time

Imaginative

1900-49	30 %
1950-69	30 %
1970-91	40 %

Informative

1900-49	10 %
1950-69	20 %
1970-91	70 %

Medium

Books	80 %
Periodicals	13.3 %
Ephemera	6.7 %

¹⁸⁵ Das wichtigste Kategorisierungsprinzip des Brown und des LOB Korpus z.B. ist dagegen die Einteilung in Genres, die aber auch nicht in allen Fällen konsequent durchgeführt zu sein scheint: auch hier gibt es Klassen, die eher als Sachbereiche gewertet werden könnten, z.B. ‚Religion‘ oder ‚Skills and Hobbies‘.

Level

<i>Imaginative</i>		<i>Informative</i>	
literary	45%	technical	30 %
middle	35 %	lay	50 %
popular	20 %	popular	20 %

(Vgl. Summers 1993:193ff., die Jahreszahl 1991 unter der Kategorie *Time* wurde von mir aufgrund einer persönlichen Mitteilung von Steve Crowdy vom 25.04.01 ergänzt.)

Für das Thema dieser Arbeit ist in erster Linie der im Vergleich zu anderen Korpora hohe Anteil des AmE im LLELC von Interesse (40 %). Zu dem Einbezug der verschiedenen Varietäten des Englischen in das Korpus schreibt Summers (1993:193):

Although there is some bias towards British English in these percentages, the American English percentage would be seen in the UK as very high. However, since there is a very strong commitment at Longman to the concept of English as a world language and to the full and accurate coverage in our dictionaries of American as well as British English, it was essential that American English was adequately covered in the corpus.

Diese begrüßenswerte Sichtweise findet sich bei vielen anderen Verlagen nicht. Auf die Schwierigkeiten, zu ermitteln, ob die Muttersprache eines Autors tatsächlich das AmE ist oder ob die Sprache eines amerikanischen Autors nicht aufgrund von Auslandsaufenthalten, familiären Bindungen o.ä. durch andere Varietäten des Englischen beeinflusst wurde, wird allerdings nicht eingegangen.

In den 10 % des Korpus, die auf andere Varietäten des Englischen außer dem BrE und dem AmE entfallen, sind laut Summers (1993: 193) u.a. Texte aus dem Australischen und Irischen Englisch sowie aus afrikanischen Varietäten des Englischen als Muttersprache enthalten. Die Liste der *Region Codes* (ebd., S. 198) legt nahe, daß außerdem Quellen aus dem Kanadischen, Neuseeländischen, Karibischen und Indischen Englisch Eingang in diese 10 % des Korpus gefunden haben. Genaue Prozentsätze zum jeweiligen Einbezug der verschiedenen Varietäten werden aber nicht genannt. Die Redaktion führt auch keine explizite Begründung dafür an, warum die Anteile der verschiedenen Varietäten genau auf 50% BrE, 40 % AmE und 10% andere Varietäten festgesetzt wurden.

Hinsichtlich des Kriteriums ‚Zeit‘ ist anzumerken, daß das *Longman* Team sich gegen die Erstellung eines synchronen Korpus entschied mit der Begründung, daß das LLELC als ‚allgemeines Korpus‘ (‚general corpus‘) geplant sei. Ein synchrones Korpus veralte schnell und sei daher als Grundlage für ein ‚allgemeines Korpus‘ ungeeignet, deshalb werde ein diachronischer Ansatz gewählt und sprachliches Material ab dem Jahr 1900 aufgenommen (vgl.

Summers, 1993:193 f.). Als weiterer Grund für diese Entscheidung wird angegeben, daß viele Texte, vor allem fiktionale, die am Anfang des 20. Jahrhunderts geschrieben wurden, noch immer häufig gelesen würden. Daher werden auch 60% der fiktionalen Texte Quellen entnommen, die aus den Jahren 1900 – 1969 stammen (vgl. die oben angeführte Tabelle). Dieser hohe Prozentsatz wird auf ähnliche Weise begründet wie der große Anteil fiktionaler Texte am Gesamtkorpus: er trage dem Einfluß und der ‚Langlebigkeit‘ fiktionaler Texte Rechnung. Dieses Argument aufgreifend könnte man allerdings mit gutem Grund auch viele bedeutende Werke der englischen Literatur, z.B. solche von Shakespeare, Dickens oder Austen, einbeziehen, die vor dem Jahr 1900 entstanden sind und als mindestens ebenso einflußreich betrachtet werden können. Auch in diesem Zusammenhang beruft sich Summers auf die in Leselisten für Englischlerner verzeichnete Literatur und gibt abermals die vornehmliche Ausrichtung der Verlagsprodukte auf den ELT Bereich zu erkennen: Auf allen Leselisten, außer denen für fortgeschrittenste Lerner, würden nur sehr selten Werke, die vor 1900 erschienen seien, genannt. Ob dies tatsächlich so ist, sei dahingestellt, aber es bleibt die Frage, diesen Argumentationsfaden weiterspinnend, warum denn dann nicht auch den fortgeschrittensten Lernern die Möglichkeit gegeben werden soll, Wörter nachzuschlagen, auf die sie bei der Lektüre von vor 1900 entstandenen Werken stoßen. Der Prozentsatz der non-fiktionalen Texte, die Quellen entnommen sind, die aus den Jahren 1900-1969 stammen, fällt dagegen deutlich niedriger aus (30%). Dies wird damit begründet, daß nonfiktionale Texte, abgesehen von wenigen Klassikern, schneller überholt sind und in vielen Fällen nach einer bestimmten Zeit nicht mehr benutzt werden. Insgesamt besteht das Korpus jedoch zu ca. 43% (dies entspricht 12.9 Mio. Wörtern von 30 Mio.) aus Texten, die älter sind als 1970. Dieses Faktum wirft die Frage auf, ob das LLELC das heutige Englisch in ausreichender Form repräsentiert, wenn es als Basis von Wörterbüchern der englischen Gegenwartssprache dienen soll. Möglicherweise sind viele der im Korpus vorkommenden Wörter, Wortbedeutungen, Wendungen oder grammatischen Besonderheiten, die aus den älteren Quellen stammen, schon veraltet oder auch einige neue sprachliche Phänomene aus Mangel an einem höheren Anteil modernerer Texte im Korpus gar nicht oder zumindest in nicht ausreichender Anzahl vertreten, um diese angemessen in einem Wörterbuch zu beschreiben. Dies könnte gravierende Folgen für die Verlässlichkeit der in den Wörterbüchern gegebenen Informationen und somit auch für die Erfassung und Markierung von Amerikanismen haben, wenn sich das *Longman* Team bei der Wörterbucherstellung nur auf das LLELC stützen würde, was aber nicht der Fall ist (s. u.). Gerade in den letzten beiden Jahrzehnten hat der Einfluß des AmE auf das BrE (vor allem durch die Medien bedingt) stark zugenommen, so daß viele ursprünglich aus dem AmE stammende Wörter, Wendungen und grammatische Phänomene heute schon als gemeinenglisch angesehen werden. Im Vergleich zu den Korpora ‚der ersten

Generation‘ (z.B. *Brown, LOB*), die nur 1 Mio. Wörter umfassten, stellen natürlich auch die 16,5 Mio. Wörter des LLELC, die aus Quellen der Jahre ab 1970 entnommen wurden, eine große Anzahl dar. Die Materialsammlung für das LLELC endete jedoch mit dem Jahr 1991, so daß den neuesten Entwicklungen in der englischen Sprache ab 1992 bis heute nicht mehr Rechnung getragen werden konnte. Wieviel Prozent der Texte zudem aus dem Zeitraum von Ende der 80er Jahre bis Anfang der 90er Jahre stammen, also wirklich jüngeren Datums sind, bleibt unklar. Die schon oben erwähnten Zitatensammlungen ermöglichen es vielleicht, dem Mangel an ganz aktuellem Sprachmaterial zumindest in begrenztem Maße abzuhelpfen, indem sie Belegmaterial für Neologismen liefern, die dann auch Eingang in neue Auflagen bestehender Wörterbücher finden. Zudem enthält auch das *Longman Written American Corpus* und der neu zum BNC hinzugefügte Teil ganz aktuelle Sprachdaten.

Wenn man sich die Angaben zu dem Kriterium ‚Medium‘ betrachtet, fällt auf, daß der Prozentsatz, der den Einbezug von Büchern betrifft, im Vergleich zu den Prozentsätzen für Zeitschriften/ Zeitungen und sog. ‚Ephemera‘ (z.B. unveröffentliche Texte oder solche, deren Autor ungenannt bleibt, Werbematerial, Finanzberichte, Informationsblätter der Regierung bis hin zu Busfahrkarten) relativ hoch ist. Gründe für diese Entscheidung werden nicht gegeben. Aus einer pragmatischen Perspektive betrachtet ist die Sammlung von Material für die Kategorie ‚Ephemera‘ natürlich häufig viel zeitaufwendiger und kostspieliger als die Beschaffung von Büchern, aber sowohl vom Standpunkt der Rezeption als auch der Produktion betrachtet scheint diese Gattung leicht unterrepräsentiert zu sein. Dasselbe trifft mindestens ebenso für den Einbezug von Zeitungen und Zeitschriften zu, wenn man sich auf den Aspekt der Rezeption beruft: das journalistische Genre ist heutzutage das am meisten gelesene und einflußreichste Genre unter allen geschriebenen Genres.¹⁸⁶ Auch für das Thema dieser Arbeit wäre ein höherer Anteil an Zeitungs- bzw. Zeitschriftentexten interessant, da eine höhere Vorkommenshäufigkeit von Amerikanismen in diesem Genre wahrscheinlich ist.

Abgesehen von den „Primary document features“ wurden auch noch weitere Kriterien verwendet, um Texte zu klassifizieren. Für den Einbezug bestimmter Kategorien innerhalb dieser sog. ‚Secondary document features‘ legte man allerdings keine Zielprozentsätze fest; Informationen zu diesen Textmerkmalen wurden lediglich zusammen mit jedem Text gespeichert. Für das Thema der vorliegenden Arbeit könnten u. U. besonders die mit den Textmerkmalen „Author“, „Origin“ oder „Biography“ verknüpften Informationen interessant sein, gesetzt den Fall, sie könnten weitere Hinweise dazu geben, welche Varietät des Englischen die Muttersprache des Autors ist und ob seine Sprache durch irgendwelche Einflüsse (z.B. lange

¹⁸⁶ Vgl. dazu auch Kennedy (1998: 49).

Auslandsaufenthalte in anderen englischsprachigen Ländern) modifiziert wurde. Hier könnten, wenn diese Informationen erhältlich waren, auch das Geburtsdatum des Autors oder Auskünfte zu seinem sozialen Status verzeichnet sein, die zusammen mit Informationen zum Geschlecht und zur ethnischen Herkunft (unter dem Textmerkmal „Gender“ und „Race“) interessante Untersuchungsergebnisse zu der Frage ermöglichen würde, welche Amerikanismen von welcher Alters-, Geschlechts-, ethnischen oder sozialen Gruppe besonders häufig verwendet würden.

Die Standardgröße der den Quellen entnommenen Textausschnitte lag bei bis zu 40.000 Wörtern. Das *Longman* Team wollte deshalb keine ganzen Texte in das Korpus einbeziehen, weil man eine größere Vielfalt von Textmaterial für wichtiger erachtete als die vollständige Aufnahme von Texten (nur wenn der Text weniger als 40.000 Wörter enthielt, wurde er im Ganzen erfaßt). Mithilfe von Zufallsverfahren erreichte man, daß eine ungefähr gleiche Anzahl von Textausschnitten dem Anfang, der Mitte und dem Ende der Quellen entnommen wurde.

Geoffrey Leech machte laut Summers (1993:199) Anfang 1989 den Vorschlag, die bis zu diesem Zeitpunkt bereits gesammelten Daten daraufhin zu überprüfen, ob die erforderlichen Mengen für die einzelnen Kategorien erfaßt würden und ob man bestimmte Datengruppen aufgrund linguistischer Analysen anderen Kategorien zuordnen müsse. Die *Longman* Redaktion begrüßte diese Anregung und hoffte, daß dieser Ansatz in Zukunft von Linguisten aufgegriffen werde, sah sich aber aus Kosten- und Zeitgründen außerstande, ihn selber umzusetzen.

Das LLELC kann als Vorläufer des *BNC Written Corpus* betrachtet werden. Die Strukturprinzipien des LLELC haben großen Einfluß auf den Aufbau des geschriebenen Teils des BNC gehabt¹⁸⁷, das nur kurze Zeit nach dem LLELC entstand.

Die wichtigsten Daten zum LLELC im Überblick:

<i>Longman/ Lancaster English Language Corpus</i>	
Typ	<i>sample corpus</i> , statisch, diachron ¹⁸⁸
Größe	ca. 30 Mio Wörter
Varietäten des Englischen	BrE: 50 % (= ca. 15 Mio. Wörter) AmE: 40 % (= ca. 12 Mio. Wörter) Andere: 10 % (= ca. 3. Mio. Wörter)
Alter der Quellen	1900 – 1970: ca. 43 % 1970 – 1991: ca. 57 %

¹⁸⁷ Vgl. Summers (1994:277) und Meijs (1996:103f.). Diese Information wurde auch in der persönlichen Mitteilung Steve Crowdys vom 25.04.01 bestätigt.

¹⁸⁸ Die Charakterisierung des LLELC als ‚diachron‘ scheint insofern gerechtfertigt, als das gesammelte Material aus einem Zeitraum von fast 100 Jahren stammt.

Art der Quellen	geschrieben (Bücher: Literatur und Sachtexte, Zeitungen, Zeitschriften, 'Ephemera'); auf gesprochenen Quellen basierender Teil des Korpus jetzt Teil des BNC
Struktur	fiktionale Texte: ca. 42% non-fiktionale Texte: ca. 58 % Gliederung der Texte in 10 große Sachbereiche, (Zielprozentsätze für die non-fiktionalen ‚Superfields‘ sind den Verhältnissen in der Buchproduktion der letzten 20 Jahren in Großbritannien nachempfunden) Klassifizierung der Texte nach den Kriterien Region, Zeit, Medium und Level

Tab. 3.3 Longman/Lancaster English Language Corpus

3.2.1.1.2 Longman Learner's Corpus

Das *Longman Learner's Corpus*¹⁸⁹ (im folgenden mit ‚LLC‘ abgekürzt) besteht aus Samples von geschriebenen Texten, die von Lernern des Englischen stammen. Bei diesen Texten handelt es sich um Aufsätze, Briefe, Berichte, Lösungen zu Prüfungsaufgaben und Tagebüchern. Das LLC umfasst z.Zt. (Stand November 2004) ca. 10 Mio. Wörter aus über 10.000 Texten. Über acht verschiedene Sprachkompetenzniveaus und 160 verschiedene Nationalitäten von Lernern sind im Korpus vertreten und jeweils gekennzeichnet. Das LLC soll vornehmlich der Erstellung von Wörterbüchern und Sprachlehrmaterialien dienen, steht aber auch für wissenschaftliche Zwecke zur Verfügung. Mithilfe des Korpus läßt sich z.B. feststellen, welche Aspekte der englischen Grammatik und Lexik besonders schwer erlernbar scheinen oder welche spezifischen sprachlichen Probleme eine bestimmte Gruppe von Lernern hat (z.B. französische fortgeschrittene Lerner des Englischen). Das LLC wurde z.B. verwendet, um die *Usage Notes* im LDOCE₃ und im *Longman Active Study Dictionary* zu schreiben. Mithilfe des LLC konnte die Redaktion vermutlich auch erkennen, wo die Lerner Schwierigkeiten hatten, zwischen BrE und AmE Gebrauch von bestimmten Wörtern und Wendungen zu unterscheiden, und durch

¹⁸⁹ Die folgenden Informationen zum *Longman Learner's Corpus* stammen aus Kennedy (1998:42), Crowdy (1992:128), Rundell (1996b:6), der Internet-Webpage <http://www.longman-elt.com/dictionaries/lclearn.shtml> und dem *Longman* Verlagskatalog 2000, S. 96. Das Longman Learner's Corpus war das erste Lernerkorpus überhaupt und wurde speziell für die Erstellung von Lernerwörterbüchern konzipiert, vgl. Rundell (1996b:6).

entsprechende *Usage Notes*, wie z.B. zu „professor“ oder „flat“, Hilfestellung leisten. Während das LLC zum Zeitpunkt der Erstellung des LDOCE₃ erst ca. 5 Mio. Wörter umfasste, waren es während der Arbeit an der zweiten Auflage des LDEL_C schon 8 Mio. Wörter.¹⁹⁰

Die wichtigsten Daten zum LLC im Überblick:

<i>Longman Learner's Corpus</i>	
Typ	Spezialkorpus (Lernerkorpus), dynamisch
Größe	ca. 10 Mio Wörter (Stand Nov. 2004)
Varietäten des Englischen	Lernerenglisch
Alter der Quellen	ab Ende der 80er Jahre ¹⁹¹
Art der Quellen	geschrieben (Aufsätze, Briefe, Berichte, Lösungen zu Prüfungsaufgaben und Tagebücher)
Struktur	Gliederung des Korpusmaterials nach 8 Kompetenzstufen und 160 Nationalitäten von Lernern

Tab. 3.4 *Longman Learner's Corpus*

3.2.1.1.3 *Spoken British Corpus*

Das *Spoken British Corpus* ist Teil des BNC und wird im Kapitel zu diesem Korpus behandelt (vgl. Kap. 3.2.2.2.).

3.2.1.1.4 *Longman Written American Corpus*

Das *Longman Written American Corpus*¹⁹² ist ein *dynamisches* Korpus, das 150 Mio. Wörter aus Zeitungen, Zeitschriften, Bestseller-Romanen, technischen und naturwissenschaftlichen Texten

¹⁹⁰ Vgl. die elektronische Nachricht des *Longman* Verlags vom 21.05.01.

¹⁹¹ Diese Information wurde Rundell (1996b:6) entnommen. In einer persönlichen Mitteilung vom 25.04.01 gibt Steve Crowdy das Jahr 1990 an.

¹⁹² Vgl. dazu die leider nicht sehr umfangreichen Informationen auf der Internet-Webpage <http://www.longman-elt.com/dictionaries/lcawritt.shtml> und im *Longman* Verlagskatalog 2000, S. 96. Die Größe des *Longman Written American Corpus* wird im Internet z.B. noch mit 100 Mio. Wörtern beziffert (Stand: Nov. 2004), in der persönlichen Mitteilung von Steve Crowdy vom 25.04.01 aber schon mit 150 Mio. Wörtern. Bemerkenswerterweise heißt es im *Longman* Verlagskatalog 2000, daß das *Longman Written American Corpus* bereits 200 Mio. Wörter enthalte. Bei derartigen Diskrepanzen zwischen den Angaben kommen Zweifel auf, inwieweit man sich auf die Korrektheit der Verlagsinformationen überhaupt verlassen kann. Die Auskünfte auf den Internetseiten der Verlage scheinen jedenfalls nicht regelmäßig auf den neuesten Stand gebracht zu werden. Leider wird auch nur auf manchen Web-Seiten das Datum der Veröffentlichung bzw. der letzten Überarbeitung

sowie Bildbänden enthält (Stand April 2001). Laut Redaktion wird die Zusammensetzung des Korpus ständig überarbeitet und ergänzt.

Den Angaben Steve Crowdys zufolge¹⁹³ besteht das *Longman Written American Corpus* eigentlich aus zwei Teilkorpora: einem Auswahlkorpus, das 50 Mio. Wörter umfasst und nach den Strukturprinzipien des *BNC written corpus* ausgerichtet ist, und einem Korpus von 100 Mio. Wörtern, das aus Daten von Nachrichtentexten besteht, die ab 1994 gesammelt wurden.¹⁹⁴ Die Aussage auf der Web-Seite des Verlags zum Aufbau des *Longman Written American Corpus* ist daher irreführend: „the design of the Longman Written American Corpus is based on the general design principles of the Longman Lancaster English Language Corpus and the written component of the British National Corpus“. Hier entsteht der falsche Eindruck, als ob das ganze Korpus strukturell ausgewogen sei. Den Auskünften Crowdys zufolge trifft dies aber nur auf den kleineren Teil des Korpus (50 Mio. Wörter) zu. Weiterhin verwirrt, daß nicht nur die Strukturprinzipien des *BNC written corpus*, sondern auch die des LLELC für den Aufbau des *Longman Written American Corpus* eine Rolle gespielt haben sollen, denn die Zusammensetzung des LLELC und die des geschriebenen Teils des BNC ist hinsichtlich einiger Aspekte durchaus unterschiedlich. Außerdem wird bei der Beschreibung der Quellen des *Longman Written American Corpus* auf der Webseite beispielsweise die Kategorie ‘Ephemera’ gar nicht erwähnt, die sowohl in bezug auf das LLELC als auch das BNC für einen unverzichtbaren Bestandteil gehalten wird.

Nach Aussage der Lexikographen ermöglicht es das *Longman Written American Corpus* im Zusammenspiel mit den anderen zur Verfügung stehenden Korpora, die Vorkommenshäufigkeiten und Verwendungsweisen von Wörtern, Wortbedeutungen und grammatischen Phänomenen im BrE und AmE zu vergleichen und zu kontrastieren. Das Korpus wird daher wahrscheinlich auch zur Erstellung von *Usage Notes* im LDOCE₃ zum Unterschied zwischen dem BrE und AmE herangezogen worden sein.

Die den beiden Teilkorpora ständig neu hinzugefügten Daten dürften garantieren, daß auch neueste Entwicklungen im AmE in das Korpus Eingang gefunden haben. Leider ist der auf Ausgewogenheit ausgerichtete Teil des *Longman Written American Corpus* nur halb so groß (50 Mio. Wörter) wie der in seiner Zusammensetzung einseitigere „news data“- Teil des Korpus (100 Mio. Wörter). Das könnte z.B. dazu führen, daß Belege für bestimmte selten vorkommende Lexeme, die nicht in Nachrichtentexten anzutreffen sind, vielleicht nicht in der wünschenswerten

der Verlagsinformationen angegeben. Der Nutzen des Internets zum Abrufen von Informationen wird durch diese Mängel geschmälert.

¹⁹³ Diese Informationen stammen aus der persönlichen Mitteilung vom 25.04.01.

¹⁹⁴ Beide Teilkorpora sind als ‚dynamisch‘ zu bezeichnen, was in der elektronischen Nachricht des *Longman* Verlages vom 21.05.01 bestätigt wurde.

Anzahl in dem Auswahlkorpus von 50 Mio. Wörtern vorliegen, um diese präzise definieren zu können.¹⁹⁵ Umgekehrt könnte z.B. das doppelt so große Korpus der Nachrichtentexte dazu verleiten, Lexeme, die in diesem Genre oft vorkommen, für das AmE insgesamt als häufig zu betrachten. Für zukünftige Wörterbuchproduktionen wäre eine Vergrößerung der Datenbasis des Auswahlkorpus wünschenswert, die sicherlich auch erfolgen wird, da es sich bei diesem Teilkorpus ebenfalls um ein dynamisches Korpus handelt.

Die wichtigsten Daten zum *Longman Written American Corpus* im Überblick:

<i>Longman Written American Corpus</i>	
Typ	1. Teilkorpus: Spezialkorpus (aus geschriebenem AmE bestehend), dynamisch 2. Teilkorpus: Spezialkorpus (aus geschriebenem AmE Nachrichtentexten bestehend), dynamisch
Größe	ca. 150 Mio. Wörter insgesamt (Stand April 2001) 1. Teilkorpus: ca. 50 Mio. Wörter 2. Teilkorpus: ca. 100 Mio. Wörter
Varietäten des Englischen	AmE
Alter der Quellen Art der Quellen	1. Teilkorpus: 1990 - 2. Teilkorpus: ab 1994 - 1. Teilkorpus: geschrieben (Zeitungen, Zeitschriften, Bestseller-Romane, technische und naturwissenschaftliche Texte sowie Bildbände) 2. Teilkorpus: geschrieben (genauere Informationen nicht verfügbar)
Struktur	1. Teilkorpus: nach den „allgemeinen“ Strukturprinzipien des BNC (und LLELC?) 2. Teilkorpus: genauere Informationen nicht verfügbar

Tab. 3.5 *Longman Written American Corpus*

¹⁹⁵ Auch das LLELC verfügt vergleichsweise (man denke an das BNC oder die BoE) nur über eine relativ kleine Datenbasis, so daß sich in bezug auf dieses Korpus ebenfalls die Frage stellt, ob z.B. Belege für seltenere Lexeme in genügender Zahl vorkommen, um exakte Definitionen erstellen zu können. Für das BrE bietet jedoch das BNC eine sehr viel größere Datenmenge, die in Zweifelsfällen herangezogen werden kann.

3.2.1.1.5 *Longman Spoken American Corpus*

Das *Longman Spoken American Corpus*¹⁹⁶ ist das jüngste Korpus des *Longman Corpus Network* und hat eine Größe von 5 Mio. Wörtern, die aus Alltagsgesprächen von insgesamt über 1000 Amerikanern aus 12 Regionen der kontinentalen USA stammen. Für jede Region wurde die gleiche Anzahl von Informanten ausgewählt und ein ausgewogenes Verhältnis zwischen den Gewährspersonen aus ländlichen und städtischen Gebieten innerhalb dieser Regionen angestrebt.¹⁹⁷ Insgesamt handelte es sich um ca. 100 Informanten¹⁹⁸. Jede Gewährsperson erhielt ein tragbares Tonbandgerät, mit dem sie über eine Zeitspanne von mindestens 4 Tagen Alltagsgespräche in 4-Stunden-Einheiten aufzeichnete. Zu jeder der Unterhaltungen wurden Informationen zur Gesprächssituation und zu den demographischen Daten der Gesprächsteilnehmer festgehalten. Um sicherzustellen, daß die Gespräche sich so natürlich wie möglich gestalteten, ließ man die Tonbänder vier Stunden durchgehend laufen, auch wenn dadurch neben den Unterhaltungen ebenfalls lange Pausen und unverständliches gesprochenes Material aufgezeichnet wurden (die man später löschte). Amerikanische Mitarbeiter wurden mit der Transkription der Aufnahmen betraut, bei der man alle persönlichen Angaben, die zufällig in den Gesprächen erwähnt wurden, wegließ. Allen Teilnehmern am Projekt garantierte man Anonymität.¹⁹⁹

Der *Longman* Verlag war darum bemüht, eine repräsentative Auswahl von Informanten zu treffen, die die Verhältnisse in der Gesamtbevölkerung hinsichtlich der Kategorien Geschlecht, Alter, ethnische Herkunft und Bildung widerspiegeln sollte. Dabei orientierte man sich an der jüngsten US-amerikanischen Volkszählungsstatistik.²⁰⁰

¹⁹⁶ Vgl. zu diesem Korpus die Informationen auf der Internet-Webseite <http://www.longman-elt.com/dictionaries/lcaspoke.shtml> und den Aufsatz von Karen Stern, der ebenfalls im Internet zu finden ist unter <http://www.longman-elt.com/dictionaries/r3stern.shtml> (Stand: Nov. 2004).

¹⁹⁷ Über die genaue Aufgliederung der kontinentalen USA in diese Regionen erhält man leider keine weiteren Auskünfte.

¹⁹⁸ Diese Information stammt aus einer persönlichen Mitteilung Steve Crowdys vom 25.04.01. Wenn aus jeder Region exakt die gleiche Anzahl von Gewährspersonen rekrutiert wurden, müßten es eigentlich entweder 96 Informanten (8 pro Region) oder 108 (9 pro Region) gewesen sein.

¹⁹⁹ Aus den Informationen zum *Longman Spoken American Corpus* geht leider nicht genau hervor, ob die Gesprächspartner der am Projekt beteiligten Informanten vor den Aufnahmen darüber Auskunft erhielten, daß ihre Unterhaltungen aufgezeichnet wurden. Das Wissen darum, daß die eigenen Gesprächsbeiträge aufgenommen werden, könnte Auswirkungen auf die Natürlichkeit der sprachlichen Äußerungen der Gesprächsteilnehmer gehabt haben (ungeachtet der Garantien, daß die Anonymität dieser Personen gewahrt bleiben würde). Man kann vermuten, daß in so einer „Beobachtungssituation“ dazu tendiert wird, auf sehr informelle Sprache oder Slang-Ausdrücke zu verzichten.

²⁰⁰ Die folgende Tabelle wurde dem Aufsatz Sterns auf der Web-Seite <http://www.longman-elt.com/dictionaries/r3stern.shtml> entnommen. Dadurch, daß die demographischen Daten jeder einzelnen an den Aufnahmen beteiligten Person abrufbar sind, werden interessante soziolinguistische Untersuchungen zur Vorkommenshäufigkeit bzw. Verwendungsweise bestimmter Wörter, Wendungen und grammatischer Phänomene in bestimmten Bevölkerungsgruppen ermöglicht.

Gender:	Male 50 %	Female 50 %			
Age²⁰¹:	18-24 20 %	25-34 20 %	35-44 20 %	45-60 20%	60 + 20%
Ethnicity:	White 75 %	Black 13 %	Hispanic 8%	Asian 4%	
Education:	Degree/ Higher Degree 33 %	College 33 %	High School 33%		

Die Sammlung des gesprochenen Materials für das *Longman Spoken American Corpus* wurde von Professor Jack Du Bois und seinem Team von der *University of California* in Santa Barbara durchgeführt.

Nach Angaben der Redaktion ist das Korpus eines der wichtigsten Hilfsmittel der amerikanischen Lexikographen im *Longman* Team in dem Bemühen, in den Lernerwörterbüchern das im Alltag verwendete AmE wiederzugeben. Das *Longman Spoken American Corpus* soll zusammen mit dem *Longman Written American Corpus* diese Varietät in all ihren Erscheinungsformen erfassen. Ob die dem *Longman* Verlag zur Verfügung stehenden Korpora des AmE diese Varietät tatsächlich umfassend repräsentieren, ist fraglich. Verglichen mit dem *Longman Written American Corpus* (150 Mio. Wörter) ist das *Longman Spoken American Corpus* (5 Mio. Wörter) nämlich sehr klein. Wenn man an die große Bedeutung der gesprochenen Sprache im Alltag denkt, wäre ein größeres Korpus des gesprochenen AmE sehr wünschenswert gewesen. Allerdings ist zu bedenken, daß die Erstellung eines gesprochenen Korpus mit diesen Strukturprinzipien sehr zeitaufwendig und teuer ist und nicht einmal das BNC Projekt, das zusätzlich zur Kostenbeteiligung der Verlage die finanzielle Unterstützung der britischen Regierung erhielt, mit einem größeren gesprochenen ‚demographischen‘ Korpus aufwarten konnte.

Nach Angaben des *Longman* Teams kann man mithilfe des *Longman Spoken American Corpus* sehr gut beispielsweise neueste Slang-Ausdrücke²⁰² und viele neue Verwendungsweisen bereits bekannter Wörter im AmE ermitteln. Mithilfe des Korpus lassen sich den Angaben des *Longman* Teams zufolge auch grammatische Phänomene, die besonders typisch für das gesprochene AmE sind, sowie im AmE gängige Diskurselemente genau analysieren. Dabei

²⁰¹ Nach Aussage der Redaktion sind Kinder ebenfalls in einer großen Zahl von Unterhaltungen vertreten.

²⁰² Ein Problem, das auch bei anderen Korpora eine Rolle spielt und bereits mehrfach erwähnt wurde, macht sich vielleicht besonders in bezug auf das *Longman Spoken American Corpus* bemerkbar: Da das Korpus statisch ist, sind einige seiner Daten vielleicht schon nach kürzerer Zeit nicht mehr auf dem neuesten Stand. Gerade die gesprochene Sprache und insbesondere Slang-Ausdrücke neigen zu schnellem Wandel (vgl. das OCEL, S. 940), so daß die Aussage, man könne mit Hilfe dieses Korpus die neuesten Slang-Ausdrücke ermitteln, nach wenigen Jahren wahrscheinlich schon nicht mehr zutrifft.

kämen oft neue Erkenntnisse zum Vorschein, wie z.B. im Falle der Wendung „I guess“, die zusätzlich zu den bekannten Verwendungsweisen auch dazu benutzt werden könne anzuzeigen, daß jemand von einem Ereignis aus zweiter Hand berichtet.

Die Anlage des *Longman Spoken American Corpus* weist große Ähnlichkeit mit dem ‚demographischen‘ Teil des *BNC Spoken Corpus* auf, dessen Aufbau als Muster gedient haben mag. Die Parallelität der Struktur beider Korpora hat den Vorteil, daß man die Frequenz bestimmter sprachlicher Phänomene im gesprochenen BrE und AmE gut vergleichen kann. Allerdings darf man dabei nicht vergessen, daß Korpora, die ca. 5 Mio. Wörter umfassen, hinsichtlich der Menge gesprochenen Sprachmaterials, das täglich von Vertretern der jeweiligen Varietät produziert wird, verschwindend klein sind, und die Ergebnisse des Vergleichs daher mit Vorsicht behandelt werden müssen.

Die wichtigsten Daten zum *Longman Spoken American Corpus* im Überblick:

<i>Longman Spoken American Corpus</i>	
Typ	<i>sample corpus</i> , Spezialkorpus (aus gesprochenem AmE bestehend) statisch, synchron
Größe	5 Mio. Wörter
Varietäten des Englischen	AmE
Alter der Quellen	1994/95
Art der Quellen	gesprochen (Tonbandaufnahmen der Gespräche von 100 amerikanischen Informanten aus 12 Regionen der kontinentalen USA)
Struktur	Auswahl der Informanten auf der Basis der demographischen Verhältnisse in den USA

Tab. 3.6 *Longman Spoken American Corpus*

3.2.1.1.6 Abschließende Bemerkungen zum *Longman Corpus Network*

Mit seinem *Longman Corpus Network* verfügt der *Longman* Verlag über eine bei der lexikographischen Arbeit sehr nützliche Reihe sich ergänzender Korpora. Im Zusammenhang mit dem Thema dieser Arbeit ist vor allem die Tatsache hervorzuheben, daß die Redaktion auf zwei eigenständige Korpora des AmE zurückgreifen kann, die nicht nur das geschriebene AmE, sondern ebenfalls, wenn auch in viel geringerem Maße, das gesprochene AmE berücksichtigen.

Diese beiden Korpora werden entscheidend dazu beitragen, Amerikanismen in den Wörterbuchprodukten des *Longman* Verlags präzise darzustellen. Korpora zum AmE sind in diesen Größenordnungen nicht nur in Verlagen, sondern auch in der linguistischen Forschung überhaupt bisher nicht häufig zu finden. Das *Longman Written American Corpus* ist anderthalbmal so groß wie das BNC und bietet allein durch seine Datenmenge interessante Forschungsmöglichkeiten, wenn auch nur das eine Teilkorpus, nämlich das Referenzkorpus mit 50 Mio. Wörtern, als strukturell ausgewogen bezeichnet werden kann. Das *Longman Spoken American Corpus* liefert eine bisher in diesem Umfang noch nicht gesammelte Menge von gesprochenem AmE einer repräsentativen Auswahl von Informanten. Allerdings deckt das *Longman* Team mit dem ‚demographischen‘ Korpus nur eine Gattung gesprochener Sprache (den in privaten Unterhaltungen vorkommenden Sprachgebrauch) ab, während andere Genres, wie z.B. Interviews, Debatten, Fernseh- und Radiosendungen, Vorträge o.ä., die in das BNC Eingang gefunden haben (vgl. den sog. *context-governed part*), unberücksichtigt bleiben.

Das *Longman Learner's Corpus* ist mit Sicherheit ein ausgezeichnetes Hilfsmittel, um die Lernerwörterbücher noch ‚lernerfreundlicher‘ zu gestalten, indem man auf Fehlerquellen, die man anhand des Korpus ermittelt hat, in den Wörterbüchern ausführlich eingeht. So kann das LLC u.a. auch dazu beitragen, daß den Lernern Unterschiede zwischen AmE und BrE Gebrauch verdeutlicht werden. Zur Ermittlung dieser Unterschiede zwischen den beiden Varietäten zogen amerikanische und britische Lexikographen vor allem die ‚muttersprachlichen‘ Korpora des *Longman* Verlags heran.²⁰³

Die begrüßenswerte Auffassung des Verlags, daß man aufgrund dessen, daß Englisch eine Weltsprache sei, auch andere Varietäten des Englischen außer dem BrE in den Wörterbüchern in angemessener Weise berücksichtigen müsse (vgl. Kap. 3.2.1.1.1), findet u.a. seinen Niederschlag in dem für britische Verhältnisse hohen Prozentsatz des AmE im LLELC (40%). Dieser große Anteil AmE Sprachdaten im Korpus wird sicherlich zusätzlich zu den zwei AmE Korpora zu einer besseren Erfassung, Markierung, Definition und Illustration von Amerikanismen in den Wörterbüchern beitragen. Vor allem im Zusammenhang mit dem LLELC (aber auch z.B. in bezug auf das *Longman Spoken American Corpus* und das Teilkorpus des Referenzkorpus im *Longman Written American Corpus*) muß man positiv hervorheben, daß sich das *Longman* Team sehr um die Objektivität der bei der Korpuserstellung angewandten Kriterien bemüht hat und bestrebt war, ein möglichst ‚repräsentatives‘ Korpus zu schaffen, weil sich die Redaktion der Wichtigkeit der ‚Repräsentativität‘ der Daten für die Verlässlichkeit der in den Wörterbüchern angegebenen Informationen bewußt war. Ebenso begrüßenswert ist die Tatsache, daß das

²⁰³ Persönliche Mitteilung von Steve Crowdy vom 25.04.01.

Longman Team schon während der Entstehungsphase des LLELC das unfertige Korpus testete, also eine Art Selbstkontrolle ausübte, um vorhandene Mängel abzumildern. Das ernsthafte Interesse des Verlags an der Erstellung ausgewogener Korpora spiegelt sich auch in dem mehrfach geäußerten Wunsch nach linguistischen Analysen des LLELC, die auf statistische Methoden zurückgreifen (und schon bereits von Biber an anderen Korpora durchgeführt wurden), wider. Trotzdem bleiben manche beim Korpusaufbau des LLELC getroffene Entscheidungen problematisch, wie z.B. der hohe Prozentsatz fiktionaler gegenüber non-fiktionalen Texten, der Einbezug einer großen Anzahl älterer Quellen sowie die deutlich stärkere Gewichtung des Mediums Buch gegenüber anderen Medien (Zeitungen, Zeitschriften, 'Ephemera' o.ä.).

Abschließend muß man sich zudem erneut vor Augen halten, daß vor allem das *Longman Written American Corpus* zum Zeitpunkt der Erstellung des LDOCE₃ von seiner heutigen Größe noch weit entfernt war, da es 1995 noch nicht einmal 5 % seiner jetzigen Größe umfasste. Aber auch das Lernerkorpus und wahrscheinlich ebenfalls das *Longman Spoken American Corpus* waren damals noch deutlich kleiner. Der Nutzen dieser drei Korpora konnte daher bei der dritten Auflage des LDOCE erst begrenzt zur Geltung kommen. Im Jahre 1998, als die zweite Auflage des LDEL₂ veröffentlicht wurde, konnten die Lexikographen dagegen auf eine AmE Datenbasis von 120 Mio. Wörtern sowie auf ein Lernerkorpus von 8 Mio. Wörtern rekurrieren. Leider scheinen die Verantwortlichen von diesem erweiterten Datenbestand keinen Gebrauch gemacht zu haben, denn den Recherchen der Verfasserin zufolge wurden nur die enzyklopädischen Einträge des Wörterbuchs für das LDEL₂ inhaltlich überarbeitet.

Die erweiterte Ausgabe des LDOCE₃ basiert laut Angaben auf dem Rückendeckel auf über 200 Millionen Wörtern aus dem *Longman Corpus Network* und dem BNC. Diese Auskunft verwundert etwas, da Steve Crowdys Informationen zur Größe der einzelnen Korpora im *Longman Corpus Network* zufolge das *Longman Corpus Network* und das BNC ohne die zum BNC neu hinzugefügten Daten insgesamt z.Zt. ca. 295 Mio. Wörter umfassen müßten, mit diesen dagegen schon 315 Mio. Wörter²⁰⁴.

3.2.1.2 Das *British National Corpus* (BNC)

Das bereits mehrfach erwähnte *British National Corpus* (BNC), ein Meilenstein in der britischen Korpuslinguistik, soll aufgrund seiner großen Bedeutung für eine Reihe der hier untersuchten Wörterbücher und wegen der von der Verfasserin dieser Arbeit durchgeführten eigenen BNC-

²⁰⁴ Persönliche Mitteilung von Steve Crowdy vom 25.04.01.

Recherchen nun genauer vorgestellt und beurteilt werden.²⁰⁵ Die Erstellung dieses *sample corpus*, das etwas mehr als 100 Mio. Wörter aus 4.124 Texten enthält, fand zwischen 1991 und 1994 statt. An dem Projekt waren mehrere Verlage beteiligt, nämlich *Oxford University Press*, *Longman* und *Chambers Harrap*, sowie die akademischen bzw. wissenschaftlichen Einrichtungen *Oxford University Computing Services*, das *Centre for Computer Research on the English Language* der Universität Lancaster und das *Research and Innovation Centre* der *British Library*. Das Projekt kam zustande, weil alle genannten Parteien sowie die britische Regierung ein großes Interesse an der Erstellung eines Korpus dieser Größenordnung hatten. So sahen sich z.B. die Verlage durch den Erfolg der COBUILD Wörterbücher, die auf dem *Birmingham Corpus* basierten, unter Druck gesetzt, selber Korpusressourcen zu schaffen, während die britische Regierung die Informationstechnologien in Großbritannien finanziell fördern wollte.²⁰⁶ Die Hälfte der Kosten für das Projekt wurde von den Verlagen getragen, die andere Hälfte von der britischen Regierung.²⁰⁷ Die große finanzielle Beteiligung der britischen Regierung ist nach Leech (1993:10) auch ein ausschlaggebender Faktor dafür gewesen, warum das BNC nur BrE enthalten sollte: Mit dem Geld wollte man ausschließlich die eigene Sprachgemeinschaft unterstützen. Wissenschaftlich wurde diese Entscheidung u.a. damit begründet, daß das BNC als Modell für „nationale Korpora“ anderer Länder dienen sollte.²⁰⁸ In Bezug auf das AmE ist man mit der (noch im Prozess befindlichen) Erstellung des ANC dieser Intention bereits gefolgt.

Das BNC sollte außer für die Lexikographie und die elektronische Sprachverarbeitung z.B. auch für die Sprachwissenschaft und den Sprachunterricht von Nutzen sein. Das Hauptanliegen der beteiligten Verlage bei der Erstellung des BNC faßt Allen (1995:9) folgendermaßen zusammen: „The main concern of the three dictionary houses has been to collect objective data for the re-evaluation of the core of the language, as distinct from the development of new language, which is more appropriately monitored by a focused reading programme [...]“. Schon aus diesem Grund - abgesehen von den aus anderen Disziplinen gestellten Anforderungen an das BNC - durfte das BNC kein *opportunistisches Korpus*²⁰⁹ sein, sondern wurde nach genau festgelegten Strukturprinzipien entwickelt. Rezipierte und produzierte Sprachäußerungen sollten

²⁰⁵ Die folgenden Informationen zum BNC stammen aus Burnard (1995:1ff.), Aston/ Burnard (1997:Preface, 1-44), Kennedy (1998:50ff.), Svartvik (1996:7f.), Quirk (1992:460ff.), Clear (1993:167ff.) und den Informationen auf den Webseiten von *Oxford University Press*: <http://info.ox.ac.uk/bnc/what/index.html> und http://www.oup.com/elt/global/products/oald/more_on_dicts/bnc/ (Stand Oktober 2004).

²⁰⁶ Vgl. Burnard (2001:2).

²⁰⁷ Vgl. Leech (1993:10). Genauere Auskünfte liefert z.B. das LDOCE₃ (iv) : „The project received funding from the UK Department of Trade and Industry and the Science and Engineering Research Council and was supported by additional grants from the British Academy and the British Library.“ Den Aussagen Burnards (2001:1) zufolge erhielten sogar auch die Verlage bei ihren Ausgaben zu 50 % finanzielle Unterstützung, und zwar durch das *Department of Trade and Industry*.

²⁰⁸ Ebd.

²⁰⁹ Vgl. zur Definition dieses Begriffs Kap. 3.1.6.2 dieser Arbeit.

gleichermaßen in das Korpus eingehen. Den für das Projekt verantwortlichen Linguisten und Lexikographen war damals jedoch bereits bewußt, daß das BNC niemals als im statistischen Sinne repräsentativ für das BrE werde gelten können, da dies eine klar definierte Grundgesamtheit, der man die Stichprobe entnimmt, voraussetzt. Es bestand allerdings das Ziel, eine geschichtete Stichprobe nach bestimmten Kriterien so zu definieren, daß das Korpus zumindest das bekannte Maß an Variabilität in der Sprache widerspiegeln würde, das z.B. durch verschiedene Modi der Sprachproduktion (geschrieben versus gesprochen), Textsorten, Sachgebiete u.ä. zum Ausdruck kommt.²¹⁰ Zum einen sollte das BNC als ‚Mikrokosmos des BrE‘ gelten können, zum anderen Varietätsstudien verschiedenster Art ermöglichen.

Im folgenden wird zunächst auf den Aufbau des schriftlichen Subkorpus eingegangen, danach auf die Struktur des *spoken subcorpus*. Das BNC besteht zu etwa 90% aus geschriebenen Quellen. Die Texte wurden anhand der Selektionsmerkmale *domain*, *time*, and *medium* ausgewählt. Jede dieser drei Kategorien wurde wiederum nach Subkategorien differenziert, für die jeweils bestimmte Zielprozentsätze festgelegt wurden. So sollten 75% der Samples aus non-fiktionalen Texten und 25% aus fiktonalen Texten stammen. In Bezug auf das Kriterium *Medium* wurde vorgesehen, daß 60% der Texte aus Büchern, 30% aus regelmäßig erscheinenden Zeitschriften und 10% aus verschiedenen anderen Quellen (veröffentlichte und unveröffentlichte Texte sowie zum mündlichen Vortrag vorgesehene Texte²¹¹) stammen sollten. Wie die folgenden Tabellen zeigen, weichen die Datenmengen im BNC etwas von den Vorgaben ab.²¹²

²¹⁰ Vgl. Kennedy (1998:53) und Burnard (2001:5).

²¹¹ Die Kategorie „Miscellaneous published“ enthält z.B. Broschüren, Bedienungsanleitungen und Werbung, die Kategorie „Miscellaneous unpublished“ Briefe, Protokolle, Essays, Berichte, firmeninterne Mitteilungen u.ä. Die Kategorie „written-to-be-spoken“ umfasst Fernsehdrehbücher, Regiebücher u.ä.

²¹² Gründe dafür sowie Informationen zu den Zielprozentsätzen der Sachgebiete aus non-fiktionalen Texten liefert Burnard (1995:11). Bezüglich der Kategorie *Medium* ist allerdings hinzuzufügen, daß ca. 60% der Wörter im *written subcorpus* des BNC aus Büchern, ca. 30% aus regelmäßig erscheinenden Zeitschriften und ca. 10% aus verschiedenen anderen Quellen stammen.

<i>Domain:</i>			<i>Medium:</i>		
	Texts	Percentage		Texts	Percentage
Imaginative	625	19.47	Book	1488	46.36
Arts	259	8.07	Periodical	1167	36.36
Belief and Thought	146	4.54	Misc. published	181	5.64
Commerce	284	8.85	Misc. unpublished	245	7.63
Leisure	374	11.65	To-be-spoken	49	1.52
Natural Science	144	4.48	Unclassified	79	2.46
Applied Science	364	11.34			
Social science	510	15.89			
World affairs	453	14.11			
Unclassified	50	1.55			

Time:

	Texts	Percentage
1960-1974	53	1.65
1975-1993	2596	80.89
Unclassified	560	17.45

(aus Burnard 1995: 11ff.)

Non-fiktionale Texte wurden erst ab dem Erscheinungsjahr 1975 berücksichtigt, fiktionale Texte allerdings schon ab 1960, da sie als „langlebiger“ eingestuft wurden.

In der Regel nahm man keine vollständigen Texte in das BNC auf, die Ausschnitte sollten nicht mehr als 45.000 Wörter umfassen.

Neben den genannten Auswahlkriterien schenkte man auch sog. *descriptive criteria* (z.B. *author sex, author domicile, target audience age, place of publication*) Aufmerksamkeit. Den Auskünften Burnards (1995:12) zufolge versuchte man während der Datensammelungsphase, wann immer möglich, auch hinsichtlich dieser Merkmale für eine relative Ausgewogenheit zu sorgen. Was die Wohnorte der Autoren betrifft, wird sich Burnard mit dieser Aussage aber wohl nur auf das Vereinigte Königreich bezogen haben, denn wenn der Wohnort eines britischen Autors oder einer britischen Autorin im englischsprachigen Ausland liegt, besteht die Möglichkeit, daß das von ihm gesprochene BrE durch die andere Varietät des Englischen beeinflusst ist. Wie die Informationen zu dem Merkmal *author domicile* zeigen (ebd. S.14f.), lebten z.B. sieben der Autoren zum Zeitpunkt der Erstellung des BNC in den Vereinigten Staaten und vier in Kanada, so daß eine Beeinflussung ihres individuellen Sprachprofils durch das AmE

bzw. CanE nicht auszuschließen ist. Da diese Autoren jedoch glücklicherweise mittels eines Codes identifizierbar sind, können sie bei der Ergebnisauswertung der eigenen BNC-Analysen der Verfasserin dieser Arbeit in besonderer Weise berücksichtigt werden.

Neben dem auf geschriebenen Quellen basierenden Subkorpus besteht das BNC zu etwa 10% aus gesprochenen Daten.²¹³ Für ein Korpus gesprochener Sprache gilt in besonderem Maße, daß es keine objektiven Maßstäbe gibt, nach denen die Grundgesamtheit definiert und ein Stichprobenauswahlverfahren erstellt werden könnte. Um sowohl dem Aspekt der Sprachproduktion als auch -rezeption gerecht zu werden, entwickelte man einen sog. *demographically sampled part* und einen sog. *context-governed part*. Beim ersteren benutzte man demographische Parameter, um eine Stichprobe der Grundgesamtheit von Sprechern des BrE im Vereinigten Königreich zu gewinnen. Mit Hilfe eines Zufallsstichprobenverfahrens wurden 124 Erwachsene ausgewählt, die in einem Zeitraum von zwei Tagen bis zu einer Woche ihre eigenen mündlichen Äußerungen sowie die ihrer Gesprächspartner aufnehmen sollten. Bei der Auswahl der Probanden wurde darauf geachtet, daß man Männer und Frauen sowie unterschiedliche Altersgruppen, Schichten und Regionen in ausgewogenem Maße einbezog. Zusätzlich zu diesen Aufnahmen verwandte man für den *demographically sampled part* auch Aufnahmen aus dem *University of Bergen COLT Teenager Language Project*, bei dem die gleichen Aufnahmemethoden verwandt worden waren wie beim BNC Projekt. Insgesamt wurden auf diese Weise ca. 4,2 Mio. Wörter spontaner Gesprächsäußerungen gewonnen.

Da der demographische Ansatz gewisse Genres der gesprochenen Sprache, bei denen es nur wenige Sprecher, aber viele Hörer gibt (vgl. z.B. Fernseh- oder Hörfunkinterviews oder Vorlesungen), kaum abdeckt, wurde in Ergänzung zu diesem Teil des Korpus der bereits erwähnte *context-governed part* entwickelt. Er enthält insgesamt ca. 6,1 Mio. Wörter. Ihm wurden vier kontextbezogene Kategorien zugrundegelegt: *educational and informative, business, public/ institutional, and leisure*, die jeweils in die Subkategorien *monologue* (40%) und *dialogue* (60%) unterteilt wurden. Für jede Subkategorie definierte man eine Reihe von Genres, wobei die Bandbreite jedoch nicht vollkommen festgelegt war, sondern die Hinzunahme zusätzlicher Genres während des Stichprobenverfahrens ermöglichte. Die folgende Tabelle gibt eine Übersicht über die vier Kategorien und die ihnen zugewiesenen Genres:

²¹³ Detaillierte Informationen zum *spoken subcorpus* finden sich in Crowdy (1993), Crowdy (1994) und Crowdy (1995).

Educational and informative	Business	Public/ or institutional	Leisure
lectures, talks, educational demonstrations News commentaries Classroom interaction	Company talks and interviews Trade union talks Sales demonstrations Business meetings Consultations	Political speeches Sermons Public/government talks Council meetings Religious meetings Parliamentary proceedings Legal proceedings	Speeches Sports commentaries Talks to clubs Broadcast chat shows and phone-ins Club meetings

aus Burnard (1995: 22f.)

Ungeachtet des unbestreitbar großen Nutzens des BNC gibt es auch einige Kritikpunkte an diesem Korpus, die teilweise sogar von den am Projekt beteiligten Wissenschaftlern selbst geäußert werden. So räumt Burnard (2001:5) ein, daß das Verhältnis zwischen geschriebenem und gesprochenem Sprachmaterial im BNC nicht der Vorstellung entspricht, daß die gesprochene Sprache im Sprachalltag eine mindestens ebenso große Rolle wie die geschriebene Sprache spielt. Ausschließlich ökonomische Gründe seien dafür ausschlaggebend gewesen, daß gesprochenes Sprachmaterial nur ein Zehntel des BNC ausmacht, denn die Gewinnung und Bearbeitung gesprochener Daten sei um ein Vielfaches teurer als die geschriebener Daten. Das Missverhältnis zwischen gesprochenen und geschriebenen Daten im BNC ist auch hinsichtlich des Themas dieser Arbeit bedauerlich, da man davon ausgehen kann, daß Amerikanismen (insbesondere solche der Umgangssprache oder des Slang) im gesprochenen BrE häufiger vorkommen als im geschriebenen. Allerdings sollte nicht unerwähnt bleiben, daß nach Angaben von Leech (1993:12) das BNC zur Zeit seiner Fertigstellung über die größte Ansammlung spontanen gesprochenen Englischs verfügte.

Den Informationen von Burnard (2001:5) zufolge waren praktische Überlegungen ebenfalls dafür verantwortlich, daß das Sprachmaterial aus Büchern und regelmäßig erscheinenden Zeitschriften insgesamt etwa 90% der Daten im *written subcorpus* ausmacht. Auch bei einem *sample corpus* wie dem BNC beeinflussten also ökonomische Faktoren die Korpuszusammensetzung, obwohl Burnard (ebd.) die große Menge an Büchern und Zeitschriften im *written subcorpus* schließlich mit der Argumentation verteidigt, daß diese Gattungen in Bezug auf die Rezeption von geschriebenem Sprachmaterial im Alltag eine sehr bedeutende

Rolle spielten. Die Auswahl und Gewichtung von Genres bei der Korpuserstellung muß zwangsläufig bis zu einem gewissen Grad subjektiv bleiben, da man weder die Anzahl aller Genres in einer Sprache noch ihre proportionale Bedeutung in dieser Sprache abschließend ermitteln kann (s.o.). Dessenungeachtet kritisiert Burnard (2001:6) im Rückblick auf das BNC-Projekt, daß vor allem einige der im *context-governed part* verwendeten Kategorien nicht präzise genug definiert worden seien und für viele von ihnen auch nicht in ausreichendem Maß Daten zur Verfügung ständen.

Für das Thema dieser Arbeit ist nicht zuletzt auch das Alter der in das BNC einbezogenen Quellen von Interesse. Da die überwiegende Mehrheit der Texte im Korpus älter als 1990 ist, drängt sich die Frage auf, inwieweit das Korpus bereits bei seiner Veröffentlichung im Jahr 1994 den aktuellen Sprachstand des BrE noch widerspiegeln konnte. Wenn es darum geht, dem in den letzten Jahren bzw. Jahrzehnten steigenden Gebrauch von Amerikanismen im BrE Rechnung zu tragen, und für Lexikographen in diesem Zusammenhang um die Entscheidung, ob und wie man bestimmte Amerikanismen in den Wörterbüchern markieren sollte, sind aktuelle Sprachdaten unverzichtbar. Zehn Jahre nach der Veröffentlichung des Korpus sehen nun auch viele Lexikographen das Alter der Quellen im BNC als Problem an, und greifen verstärkt auf andere Datenbasen zurück. Der *Longman* Verlag fügte jedoch laut Steve Crowdy (persönliche Mitteilung vom 25.04.01) dem BNC im Jahr 2001, allerdings nur für den Gebrauch im eigenen Hause, eine Datenmenge von 20 Mio. Wörtern aus den Jahren 2000-2001 hinzu. Rückblickend auf das BNC-Projekt bedauert es Burnard (2001:12), einer der Projektleiter, daß man das BNC damals nicht als *monitor corpus* konzipiert habe, um den im Laufe der Zeit stattfindenden Sprachwandel beobachten zu können. Für ein *monitor corpus* gibt es zudem auch keine Beschränkung der Größe des Korpus - das BNC ist nach heutigen Maßstäben nicht sehr groß. Allerdings weisen *monitor corpora* im Stil der BoE auch einige Nachteile auf (s. Kap. 3.1.6.2).

Für das Thema dieser Arbeit ist ebenfalls von besonderer Bedeutung, inwieweit die am Projekt beteiligten Lexikographen und Wissenschaftler ausschließen konnten, daß AmE Sprachdaten in das BNC Eingang gefunden haben. AmE Sprachdaten im BNC würden die aus diesem Korpus abgeleiteten Aussagen zum Gebrauch von Amerikanismen im BrE (wie sie von Wörterbuchverlagen und von der Verfasserin dieser Arbeit angestrebt werden) in Frage stellen. Den Angaben von Aston/ Burnard (1997:38) zufolge waren alle identifizierbaren Autoren und Sprecher Muttersprachler des BrE. Im *written subcorpus* gibt es jedoch auch Sprachdaten von anonym gebliebenen Autoren, deren Varietät des Englischen nicht festgestellt werden konnte. Für das *spoken corpus* gilt, daß sich unter den Gesprächspartnern der 124 Probanden des *demographically sampled part* sowie in Aufnahmesituationen für den *context-governed part* auch Sprecher anderer Varietäten des Englischen befanden, darunter auch solche des AmE. Mit

Hilfe des Codes *flang* („first language“) ließ sich herausfinden, daß in den *context-governed part* Sprachmaterial von 11 Muttersprachlern des AmE eingegangen ist (insg 12673 Wörter). Bezüglich des *demographically sampled part* ließen sich Sprachdaten im Umfang von 148 Wörtern einer Muttersprachlerin des AmE ermitteln. Da die Codes für diese Samples von AmE Muttersprachlern bekannt sind, lassen sie sich bei den BNC-Analysen in dieser Arbeit glücklicherweise herausfiltern. Wieviele nicht identifizierte AmE Sprachdaten im BNC enthalten sind, bleibt offen. Nach Auskunft Berglunds (1999:40) sind nur für 58% der Sprecher im *spoken subcorpus* Informationen zur Muttersprache erhältlich. Trotzdem wird die Menge weiterer nicht identifizierter AmE Daten gering sein, da alle BNC-Mitarbeiter über die Konzeption des BNC als Korpus BrE Sprachdaten informiert gewesen sein müßten.

Wie bereits in Kapitel 3.1.5 erwähnt, befinden sich in der ersten Fassung des BNC viele Fehler (Tagging-Fehler, mehrfache Auflistungen derselben Belege, Fehler in den demographischen Angaben, Druckfehler, Rechtschreibfehler, u.a.), von denen ein großer Teil in der *BNC World Edition* verbessert worden ist (außer den Druck- und Rechtschreibfehlern, da die Größe des Korpus das Korrekturlesen unmöglich macht). Die *BNC World Edition* (auch BNC₂ genannt) wurde im Dezember 1999 veröffentlicht.²¹⁴

Die wichtigsten Daten zum *British National Corpus* im Überblick:

<i>British National Corpus</i>	
Typ	<i>sample corpus</i>
Größe	etwas mehr als 100 Mio. Wörter (100.106.008 Wörter)
Varietäten des Englischen	BrE (bis auf wenige Ausnahmen)
Alter der Quellen	geschriebene Quellen: (Prozentangaben beziehen sich auf die Gesamtzahl laufender Wörter in diesem Teilkorpus.) 1960-1974: 2.26 % 1975-1993: 89.23 % Unclassified: 8.49 % gesprochene Quellen: überwiegende Mehrheit der Daten von Anfang der 90er Jahre bis 1994

²¹⁴ Weitere Informationen zu den Änderungen in der *BNC World Edition* finden sich bei Burnard (2001:10f.). Die bisherigen Ausführungen zum BNC sowie die eigenen Untersuchungen der Verfasserin dieser Arbeit am BNC beziehen sich auf die erste Ausgabe des Korpus.

<p>Art der Quellen</p>	<p>Anfang der 90er Jahre bis 1994</p> <p>geschrieben: ca. 90 % (Prozentangaben beziehen sich auf die Gesamtzahl laufender Wörter in diesem Teilkorpus.)</p> <p>ca. 59 % aus Büchern, ca. 31 % aus regelmäßig erscheinenden Zeitschriften und ca. 10% aus verschiedenen anderen Quellen (veröffentlichte und unveröffentlichte Texte sowie zum mündlichen Vortrag vorgesehene Texte)</p> <p>gesprochen: ca. 10 %</p> <p><i>demographically sampled part</i> (ca. 4 %): spontane Gesprächsäußerungen</p> <p><i>context-governed part</i> (ca. 6 %): Vorlesungen/ Referate, Aufnahmen von Schulunterricht, Live-Reportagen, Firmengespräche, Vorstellungsgespräche, Gespräche unter Gewerkschaftsvertretern, Verkaufsveranstaltungen, politische Reden, Predigten, Treffen von Clubs, Talkshows u.a.</p>
<p>Struktur</p>	<p><i>written subcorpus</i>:</p> <p>Auswahl der Texte nach den Selektionskriterien: Sachgebiet, Erscheinungsdatum, Medium</p> <p>Festlegung von Subkategorien und Zielprozentsätzen für die Subkategorien</p> <p><i>spoken subcorpus</i>:</p> <p><i>demographically sampled part</i>: mittels demographischer Parameter (Alter, Geschlecht, Gesellschaftsschicht, Herkunftsregion) Auswahl von Sprechern des BrE zum Zwecke von Gesprächsmitschnitten</p> <p><i>context-governed part</i>: Auswahl der Sprachdaten nach vier kontextbezogenen Kategorien (<i>educational</i></p>

	<i>and informative, business, public/institutional, leisure</i>), Festlegung von Prozentsätzen für Monologe und Dialoge
--	--------------------------------------------------------------------------------------------------------------------------

Tab. 3.7 *British National Corpus*

3.2.2 Wörterbuchproduktionen von Oxford University Press

Bei der Erstellung des COD₁₀ und OALD₆ standen den Herausgebern verschiedene Datenquellen zur Verfügung. Das BNC (s.o.) nimmt unter diesen einen wichtigen Platz ein, obwohl das darin enthaltene Material zum größten Teil deutlich mehr als zehn Jahre alt ist. So führt Frank Keenan, *Publishing Systems Manager (ELT Dictionaries)* aus: „The BNC remains our major source of information as it is the only truly large balanced corpus available. However we are aware that the material is now 10 years old and we take steps to keep in touch with more up-to-date usage.“²¹⁵ Neben dem BNC konnten die Lexikographen z.B. auf die Datenbanken der *Oxford Reading Programmes* zurückgreifen.

3.2.2.1 Die *Oxford Reading Programmes*²¹⁶

Im Gegensatz zu einem Korpus wie dem BNC werden in den Belegsammlungen der *UK and North American Reading Programmes* neue oder ungewöhnliche Wörter, Bedeutungen oder Wendungen (zusammen mit einem gewissen Maß an Kontext) festgehalten, die von freien Mitarbeitern des Verlags bei ihrer Lektüre aufgespürt werden. Die Sätze, in denen sich diese sprachlichen Phänomene befinden, werden sorgfältig getaggt und in die Datenbanken eingestellt.

Die Datenbanken dieser beiden *Reading Programmes* enthalten insgesamt ca. 83,6 Mio. Wörter aus 27.253 verschiedenen Quellen (Stand: April 2004). Die Zahl der Belege liegt bei ca. 2,6 Mio., wovon ca. 1,16 Mio. im Rahmen des *North American Reading Programme* und ca. 1,45 Mio. in dem des *UK Reading Programme* gesammelt wurden. Da das *UK Reading Programme* nicht nur BrE Texte, sondern auch solche aus anderen Varietäten außer dem AmE

²¹⁵ E-Mail vom 27.4.2004.

²¹⁶ Die Informationen dieses Unterkapitels gehen auf elektronische Mitteilungen von Angus Stevenson (*Project Manager, English Dictionaries and Thesauruses*) und Frank Keenan (*Publishing Systems Manager, ELT Dictionaries*) vom 27.4.2004 zurück sowie auf folgende Internetseiten: <http://www.oed.com/newsletters/2000-02/na-read.html>, <http://www.oed.com/newsletters/2000-02/uk-read.html> (Stand: jeweils April 2004), <http://www.oed.com/about/writing/evidence.html> und <http://www.oed.com/about/writing/resources.html> (Stand jeweils März 2004).

(Texte aus Australien, Neuseeland, Malaysia, der Karibik, Indien und Südafrika) berücksichtigt²¹⁷, ist die Gewichtung der BrE und AmE Belege fast ausgeglichen.

Über das *UK Reading Programme* erhält man im Internet die Auskunft, daß die von den freien Mitarbeitern ausgewählten Zitate vor allem aus Büchern, Fach- und anderen Zeitschriften und Zeitungen stammen, der Einbezug anderer Quellen aber zur Zeit vergrößert wird. So wurden z.B. im Jahr 1999 Skripts einer Reihe von britischen Fernsehprogrammen in deutlich größerem Maße berücksichtigt. Diese werden laut Eleanor Rands vor allem deswegen geschätzt, weil sie - im Gegensatz zu gedrucktem Material - wertvolle Quellen umgangssprachlichen Wortschatzes oder des Slang seien.²¹⁸ Insofern sind sie auch in bezug auf das Thema dieser Arbeit besonders interessant, da anzunehmen ist, daß die Zahl der im BrE verwendeten Amerikanismen in der Umgangssprache und im Slang höher ist als in der Schriftsprache. Dennoch sind diese Fernsehskripts selbstverständlich nicht mit spontanen Sprachäußerungen zu vergleichen.

Zum Start des *North American Reading Programme* führt Edmund Weiner (*Principal Philologist, Oxford English Dictionary*) aus: „We decided in 1989 to begin to collect quotations from North American sources in North America itself, instead of having the temerity to think that editors sitting in Oxford, England, speaking British English, could realistically organize this collection remotely.“²¹⁹ Zur Art der verwandten Quellen werden leider keine Auskünfte gegeben. Neben den Belegen aus dem *North American Reading Programme* spielen mittlerweile nach Aussage von Edmund Weiner auch andere Quellen hinsichtlich des AmE (und anderer Varietäten) eine wichtige Rolle, denn Datenbanken und Internetkorpora seien nun in großer Fülle für Lexikographen verfügbar (dazu weiter unten mehr).

Die beiden beschriebenen *Reading Programmes* werden in erster Linie dazu verwendet, neue Sprachentwicklungen zu verfolgen und in den Wörterbüchern, auch im COD und OALD, erfassen zu können.

²¹⁷ Nach Eleanor Rands (*Senior Editor, Oxford English Dictionary*) bezieht sich die Angabe *UK* in dem Namen *UK Reading Programme* in erster Linie auf das administrative Zentrum der Arbeit und weniger auf das geographische Reichweite (vgl. die Internetseite <http://www.oed.com/newsletters/2000-02/uk-read.html>, Stand: April 2004). Bezüglich des Australischen, Kanadischen und Südafrikanischen Englisch existieren auch separate *Reading Programmes* für das *Australian National Dictionary*, das *Canadian Oxford Dictionary* und das *Dictionary of South African English*, die aber vom Verlag für andere Wörterbücher ebenfalls genutzt werden können.

²¹⁸ Ebd.

²¹⁹ Dieses Zitat steht auf der o.g. Internetseite <http://www.oed.com/newsletters/2000-02/na-read.html>. Dort erfährt man auch, daß die redaktionelle Bearbeitung des nordamerikanischen Wortschatzes ab 2000 in Nordamerika selbst stattfinden soll.

3.2.2.2 Andere Quellen

Zwar waren das BNC und die *Oxford Reading Programmes* laut Angaben von Angus Stevenson die wichtigsten Belegquellen bei der Erstellung des COD₁₀, gelegentlich wurde aber auch ein Korpus mit AmE Sprachmaterial und ein Korpus mit historischem Sprachmaterial verwendet.²²⁰ Über das Korpus des AmE konnte die Verfasserin dieser Arbeit leider nicht viel in Erfahrung bringen, da nach Auskunft von Angus Stevenson die Person, die für das AmE Korpus verantwortlich war, nicht mehr bei *Oxford University Press* arbeitet. Zudem werde das Korpus heutzutage auch nicht mehr benutzt, weil der Verlag mittlerweile auf bessere Quellen Zugriff habe. Nach Ansicht von Angus Stevenson bestand das Korpus nur aus journalistischen Texten und enthielt etwas weniger als 50 Mio. Wörter.

Über die im ELT-Bereich verwandten Quellen gab Frank Keenan der Verfasserin dieser Arbeit Auskunft.²²¹ So hat der ELT-Bereich z.B. Zugriff auf Lernerkorpora. Das Sprachmaterial besteht zum einen aus schriftlichen Arbeiten, die bei Prüfungen einer Sprachschule entstanden sind. Diese Daten wurden eingegeben und getaggt, wodurch Abfragen von Sprachproduktionen bestimmter Lernergruppen (z.B. bezüglich eines Lernniveaus oder einer Nationalität) möglich sind. Dieses Sprachmaterial wurde zum anderen durch die veröffentlichten Teile des *International Corpus of Learner English (ICLE)* ergänzt.

Was das AmE angeht, so steht den Aussagen von Frank Keenan zufolge ein selbst entwickeltes Korpus, das geschriebene Texte enthält, sowie ein andernorts erworbenes Korpus, das gesprochenes Sprachmaterial umfasst, zur Verfügung (detailliertere Informationen zu diesen beiden Korpora wurden auch auf mehrfache Anfrage hin leider nicht gegeben). Außerdem habe man Zugriff auf eine Reihe von US- amerikanischen Fernseh- und Filmskripts. Da *Oxford University Press* Konsortiumsmitglied für das ANC ist, kann der Verlag überdies seit Oktober 2003 vom ersten Teilkorpus des ANC (10 Mio. Wörter) Gebrauch machen.²²² Gleichzeitig ist der Verlag ebenfalls Mitglied im *Linguistic Data Consortium* der Universität von Pennsylvania, einer Organisation, die u.a. auch viele Korpora des AmE öffentlich zugänglich macht und *Oxford University Press* daher weitere nützliche Daten liefert.

Oxford University Press verwendet zudem das Internet, um Korpusmaterial zu sammeln. So werden z.B. im Internet verfügbare Texte von Zeitungen und Zeitschriften dazu benutzt, neueste Sprachentwicklungen zu verfolgen. Dies geschieht, indem Lexikographen in regelmäßigen Abständen Daten dieser Zeitungen und Zeitschriften herunterladen und überprüfen, inwieweit Neologismen auftreten und ob sie sich im Sprachgebrauch halten. Der

²²⁰ E-Mail vom 27.04.2004. Zusätzlich wurden verschiedene Nachschlagewerke benutzt, um Sachinformationen zu erhalten, sowie Experten für verschiedene Sachgebiete konsultiert.

²²¹ E-Mail vom 27.04.2004. ELT ist eine Abkürzung für *English Language Teaching*.

Verlag benutzt aber auch andere Quellen. So heißt es auf einer Internetseite von *Oxford University Press*: „In the last few years there has been an explosion in the number of resources available to lexicographers on the Internet“.²²³ Unter den besonders oft eingesetzten großen Textdatenbanken aus dem Internet befindet sich u.a. JSTOR (enthält Faksimiles vieler Fachzeitschriften) und *Nexis* (ein Archiv für Zeitungen und Zeitschriften aus den letzten 30 Jahren).

Nach Informationen von Angus Stevenson²²⁴ ist der Verlag seit der Veröffentlichung des COD₁₀ im Jahr 1999 bestrebt, das BNC zur ergänzen. Allerdings soll diesem Korpus kein weiteres Material hinzugefügt werden. Stattdessen hat *Oxford University Press* die Erstellung eines neuen Korpus in Auftrag gegeben, das bedeutend größer als das BNC werden und Varietäten des Englischen aus der ganzen Welt umfassen soll. Während die Größe des BNC, das ein *sample corpus* ist, von vorneherein festlag, ist geplant, dem neuen Korpus ständig Material hinzuzufügen. Zur Struktur des neuen Korpus äußert sich Stevenson folgendermaßen: „It’s being compiled with a balanced selection of material from various sources, including OUP publications.“²²⁵ Detailliertere Auskünfte werden vom Verlag aus Wettbewerbsgründen nicht gegeben. Allerdings erhielt die Verfasserin dieser Arbeit die Information, daß das neue Korpus bereits bei der Vorbereitung des *Oxford Dictionary of English* (ODE₂)²²⁶, das 2003 erschien, sowie für die elfte Auflage des COD herangezogen wurde, die der Verlag im Juli 2004 veröffentlichte. Den Angaben im Internet zufolge²²⁷ heißt das neue Korpus *Oxford English Corpus* und enthält „hundreds of millions of words“. Es umfasst „all varieties of English from that used in formal, business, or scientific contexts to slang and informal usage. Texts are gathered from a range of sources including comics, newspapers, TV scripts, novels, the Web, and scholarly journals“.

3.2.2.3 Abschließende Bemerkungen zu den Datenquellen des Verlags

In den letzten Jahren ist die Anzahl der vom Verlag verwandten Datenquellen deutlich angestiegen. Neben den *Oxford Reading Programmes* und dem BNC spielen nun auch eine Reihe anderer Quellen empirischer Sprachdaten eine Rolle - darunter auch solche des AmE - , über deren Qualität jedoch in vielen Fällen aufgrund nicht ausreichender Informationen keine

²²² Vgl. zum ANC Kap. 3.1.6.1 dieser Arbeit.

²²³ Vgl. die Internetseite <http://www.oed.com/about/writing/resources.html>.

²²⁴ E-Mail vom 27.04.2004.

²²⁵ E-Mail vom 27.05.2004. OUP ist eine Abkürzung für *Oxford University Press*.

²²⁶ Hierbei handelt es sich um die zweite Auflage des *New Oxford Dictionary of English* von 1998.

²²⁷ Vgl. die Internetseiten <http://askoxford.com/worldofwords/wordfrom/concise11/?view=uk> und <http://www.askoxford.com/pressroom/archive/odelaunch/?view=uk> (Stand November 2004).

Aussage gemacht werden kann. Für das COD₁₀ und OALD₆ stellten das BNC und die *Oxford Reading Programme* allerdings immer noch die wichtigste Bezugsbasis dar, wie der bereits erwähnte Projektmanager und Co-Herausgeber des COD₁₁, Angus Stevenson, bestätigte²²⁸. Bei der Erstellung des COD₁₀ konnten die Lexikographen leider noch nicht auf einen großen AmE Datenbestand zurückgreifen, da außer der Datenbank des *North American Reading Programme* nur ein AmE Korpus von weniger als 50 Mio. Wörtern, das zudem nur journalistische Texte enthielt, zur Verfügung stand. Dieses Korpus war also noch nicht einmal halb so groß wie das BNC und zudem nicht so ausgewogen wie dieses. Das *North American Reading Programme* konnte zwar Informationen zu neuen Sprachentwicklungen im AmE liefern, aber den Mangel an zusätzlichen Korpusdaten für das AmE nicht wettmachen.

Von welchen AmE Datenquellen die Lexikographen bei der Erstellung des OALD₆ neben dem *North American Reading Programme* schon Gebrauch machen konnten, war leider nicht zu ermitteln. So wäre es interessant gewesen, zu erfahren, ob die zwei unter 3.2.3.2 erwähnten Korpora des AmE bereits zur Verfügung standen. Sollte das von Frank Keenan genannte geschriebene Korpus des AmE mit dem für das COD₁₀ verwandte identisch sein, hätte es bei der Arbeit am OALD₆ schon verwendet werden können. Auf das erste Teilkorpus des ANC von 10 Mio. Wörtern konnten die Lexikographen jedoch bei der Erstellung des OALD₆ mit Sicherheit noch nicht zurückgreifen, da dieses erst drei Jahre nach Erscheinen des OALD₆ den Konsortiumsmitgliedern zugänglich gemacht wurde. Ungewiß ist, inwieweit AmE Sprachdaten aus amerikanischen Fernseh- und Filmskripts, aus dem Datenbestand des Linguistic Data Consortiums oder aus dem Internet bei der Arbeit am OALD₆ bereits herangezogen wurden. Die nächste Auflage des Wörterbuchs wird jedenfalls von der erweiterten Datenbasis bezüglich des AmE profitieren. Noch vielversprechender ist allerdings das im Aufbau befindliche *Oxford English Corpus*, da dieses eine ‚ausgewogene‘ Auswahl von Material aus verschiedensten Quellen enthalten soll. Wenn das Korpus, wie angestrebt, schließlich viele hundertmillionen Wörter umfasst, die aus Varietäten des Englischen aus der ganzen Welt stammen, wird vermutlich auch der AmE Anteil daran groß sein.

3.2.3 Wörterbuchproduktionen von Cambridge University Press

Das *Cambridge Language Survey Corpus*, auf dem das CIDE₁ basiert, umfasste nach Informationen von Patrick Gillard (*Senior Commissioning Editor* bei *Cambridge University Press*)²²⁹ 102 Millionen Wörter.²³⁰ Da das Korpus in sehr kurzer Zeit entwickelt werden mußte -

²²⁸ E-Mail vom 01.10.2004.

²²⁹ Persönliche Mitteilung per E-Mail vom 10.03.2004. Da die Informationen zum *Cambridge Language Survey Corpus* sowohl im CIDE₁ selbst als auch in der Forschungsliteratur nicht sehr umfangreich waren, erbat die Verfasserin dieser Arbeit weitere Auskünfte vom Verlag selber, die auch bereitwillig gewährt wurden. So

für die Erstellung dieses vollkommen neuen Wörterbuchs waren nur vier Jahre vorgesehen - , wurde vor allem in der Anfangsphase auf öffentlich zugängliches gedrucktes Material zurückgegriffen, das dem BrE und AmE entstammte. Dies mag ein entscheidender Grund dafür gewesen sein, daß das Korpus schließlich zu 63% aus Zeitungstexten bestand. Die Verarbeitung von Zeitungsmaterial ist heutzutage schneller und kostengünstiger als es bei den meisten anderen Textsorten möglich ist, denn diese Sprachdaten liegen normalerweise schon in elektronischer Form vor. Die Sammlung von sog. 'Ephemera' ist dagegen viel zeitaufwendiger und kostspieliger. Material dieser Art fand sich dementsprechend im *Cambridge Language Survey Corpus* nur in sehr geringer Menge.²³¹ Jedoch war man nach Angaben von Baugh/Harley/Jellis (1996:39) trotzdem darum bemüht, eine möglichst große Bandbreite geschriebener Quellen einzubeziehen. Weitere 31% des Korpus bestanden daher aus fiktionalen und non-fiktionalen Texten, Zeitschriften und wissenschaftlichen Texten. Der Prozentsatz von Zeitungsmaterial ist jedoch, wie bereits erwähnt, demgegenüber mehr als doppelt so hoch. Obwohl Zeitungen im Alltagsleben vieler Menschen sicherlich eine große Rolle spielen, dürfte ihre Bedeutung im Korpus überbewertet sein. Es bleibt zu überprüfen, inwieweit sich dies auf die Erfassung und Markierung von Lemmata, insbesondere auf Amerikanismen, im CIDE₁ auswirkt.

Der Anteil gesprochener Daten war im *Cambridge Language Survey Corpus* dagegen leider sehr gering: er lag nur bei 6% (= 6,12 Mio. Wörtern).²³² Nach Informationen im CIDE₁ (S. xii) stammten die gesprochenen Daten aus „natural conversations and discussions, television and radio“. Den Angaben Patrick Gillards zufolge (persönliche Mitteilung vom 19.03.04) benutzte der Verlag vor allem öffentlich zugängliche Quellen (wie z.B. solche, die durch das *Linguistic Data Consortium* der Universität von Pennsylvania bereit gestellt werden²³³), begann aber parallel mit einem eigenen Programm zur Sammlung mündlicher Sprachdaten. Es ist zu vermuten, daß spontan gesprochenes Material (z.B. aus persönlichen Gesprächen in der Familie oder mit Freunden) im *Cambridge Language Survey Corpus* nur wenig vertreten war.²³⁴ Für den geringen Anteil des „spoken English“ insgesamt im Korpus mag auch wieder der hohe Zeit- und Kostenaufwand zur Beschaffung und Verarbeitung solchen Materials verantwortlich sein. Im

stützen sich die nachfolgenden Ausführungen zu dem Korpus hauptsächlich auf diese Quelle. Zusätzlich wurden aber auch Informationen zu dem Thema in CIDE₁ (S. viii und xii) und Baugh/Harley/Jellis (1996:39,55) herangezogen. Insgesamt liegen aber für dieses Korpus, das unter dem Namen *Cambridge International Corpus* weiterentwickelt wurde, weniger Auskünfte vor als beispielsweise für die BoE oder das *Longman Corpus Network*.

²³⁰ Das Lernerkorpus, von dem später noch die Rede sein wird, ist darin noch nicht enthalten. In den anderen Quellen zum CIDE₁ (s.o.) wird nur von 100 Millionen Wörtern gesprochen.

²³¹ Hierbei handelt es sich laut persönlicher Mitteilung von Patrick Gillard vom 19.03.2004 um „adverts, flyers, forms and mails“.

²³² Vgl. auch Baugh/Harley/Jellis (1996:39,55).

²³³ Weitere Informationen zum *Linguistic Data Consortium* s.o. in Kap. 3.2.3.2.

²³⁴ Zur Bedeutung dieses Materials vgl. das Kap. 3.2.1 zur BoE.

derzeitigen Korpus von *Cambridge University Press*, dem *Cambridge International Corpus*, macht der gesprochene Anteil auch nur ca. 5 % aus, doch sind dies bei einem Umfang von insgesamt 804 Mio. Wörtern immerhin schon 40 Mio. Wörter.²³⁵ Der geringe Prozentsatz des gesprochenen Englisch im *Cambridge Language Survey Corpus* wirkt sich auch negativ auf die Erfassung des gesprochenen Englisch im CIDE₁ aus, wie die Linguisten Coffey (1995:385) und Scholfield (1995:175) herausgefunden haben. Es könnte also sein, daß informelle und Slang-Ausdrücke des AmE nicht sehr umfassend im Wörterbuch vertreten sind - was nachzuprüfen bleibt.

Das für das Korpus herangezogene geschriebene und gesprochene Material stammte zu 95% aus der zweiten Hälfte der 80er Jahre und der ersten Hälfte der 90er Jahre. Somit konnte bei der Erstellung des CIDE₁, das 1995 erschien, auf eine zu diesem Zeitpunkt große aktuelle Menge von Sprachdaten zurückgegriffen werden. Es wurden aber auch einige ältere fiktionale Texte einbezogen, um älteren Wortbestand präsentieren zu können und den Lernern somit die Lektüre häufig gelesener älterer Texte zu erleichtern. Die ältesten berücksichtigten Texte gingen ungefähr bis auf das Jahr 1850 zurück.²³⁶

Am wichtigsten für die vorliegende Arbeit ist jedoch der Anteil des AmE am Korpus: Er betrug ca. 46 Mio. Wörter, machte also ca. 45% des Korpus aus. Ungefähr 56 Mio. Wörter (ca. 55%) stammten demgegenüber aus dem BrE. Dem AmE Sprachmaterial wurde somit fast die gleiche Stellung eingeräumt wie dem BrE. Der prozentuale Anteil des AmE im *Cambridge Language Survey Corpus* liegt sogar über dem im *Longman/Lancaster English Language Corpus* und ist für ein gemeinenglisches, in Großbritannien erstelltes Korpus damals wie heute bemerkenswert (auch wenn die heutigen Korpora teilweise um ein Vielfaches größer sind).

Nach Aussage von Patrick Gillard²³⁷ wurden auch für die Gewinnung AmE Sprachdaten alle oder zumindest fast alle der oben erwähnten Textsorten einbezogen. Bei dem gesprochenen Sprachmaterial überwog seiner Auskunft zufolge sogar der AmE Anteil.

Andere Varietäten außer dem BrE und AmE waren so gut wie gar nicht im *Cambridge Language Survey Corpus* vertreten. Dies gilt auch für das Australische Englisch, mit dessen gründlicher Behandlung im Wörterbuch sowohl im CIDE₁ selbst als auch in Verlagskatalogen geworben wurde.

²³⁵ Während den Aussagen von Baugh/Harley/Jellis (1996:55) zufolge allerdings anfangs geplant war, im aktuellen Korpus das geschriebene und gesprochene Englisch gleichermaßen zu berücksichtigen, gibt Patrick Gillard an (persönliche Mitteilung vom 19.03.04), daß z. Zt. nur ein Prozentsatz von 5 - 10% mündlicher Daten im *Cambridge International Corpus* angestrebt wird. Bei der großen Datenmenge von momentan 804 Mio. Wörtern im Korpus (Stand November 2004) würde schon ein geringfügig höherer Prozentsatz gesprochener Sprachdaten deutlich mehr Kosten bedeuten, die der Verlag offensichtlich scheut.

²³⁶ Diese Informationen und die folgenden stammen von Patrick Gillard (E-Mail vom 12.03.04).

²³⁷ E-Mail vom 19.03.04.

Ergänzend zum *Cambridge Language Survey Corpus* wurde bei der Erstellung des CIDE₁ von einem „Lernerkorpus“ Gebrauch gemacht, das 1995 über 5 Mio. Wörter verfügte. Es diente (wie auch das *Longman Learner's Corpus*) dazu, aufzudecken, welche sprachlichen Bereiche Lernern vor allem Schwierigkeiten bereiten, damit diesen im Wörterbuch besondere Beachtung geschenkt werden konnte.

Die wichtigsten Daten zum *Cambridge Language Survey Corpus* im Überblick:

<i>Cambridge Language Survey Corpus</i>	
Typ	nach Angaben von Patrick Gillard eher ein <i>sample corpus</i> als ein <i>monitor corpus</i> ²³⁸
Größe	102 Mio Wörter
Varietäten des Englischen	BrE: ca. 55 % (= ca. 56 Mio. Wörter) AmE: ca. 45 % (= ca. 46 Mio. Wörter) Andere: keine Angaben möglich
Alter der Quellen	95 % des Korpusmaterials: aus der zweiten Hälfte der 80er Jahre und der ersten Hälfte der 90er Jahre
Art der Quellen	Fiktionale Texte teilweise älter (ältestes Erscheinungsdatum ca. 1850) geschrieben: ca. 94% (ca. 95,9 Mio. Wörter) (fiktionale und non-fiktionale Texte, unter letzteren auch wissenschaftliche Texte, Zeitungen, Zeitschriften, - kaum 'Ephemera') gesprochen: ca. 6 % (ca. 6,1 Mio. Wörter) („natural conversations and discussions, television and radio“)
Struktur	Es ist unklar, inwieweit Richtlinien für den Korpusaufbau vorhanden waren.

Tab. 3.8 *Cambridge Language Survey Corpus*

²³⁸ E-Mail von Patrick Gillard vom 21.05.2004. Das Korpus ist zwar statisch, erfüllt aber ein anderes wichtiges Kriterium eines *sample corpus* nicht zufriedenstellend: eine möglichst ausgewogene Auswahl an Textsorten im Korpus widerzuspiegeln, denen jeweils eine bestimmte Anzahl von Textproben mit einer festgelegten Größe zugeordnet wird.

Wie bereits angedeutet, wurde das *Cambridge Language Survey Corpus* weiterentwickelt. Dies ist zwar für das im Rahmen dieser Arbeit behandelte Wörterbuch *Cambridge International Dictionary of English* (1. Aufl.) nicht mehr relevant, aber für weitere Wörterbuchproduktionen des Verlags. Ist die Datenbasis im aktuellen Korpus bezüglich des AmE erweitert worden, z.B. hinsichtlich des gesprochenen Englisch? Inwieweit können weitere Auflagen des CIDE₁ von einer insgesamt verbesserten Datenbasis profitieren? Um diese Fragen beantworten zu können, soll auch auf das heute vom Verlag verwendete Korpus *Cambridge International Corpus of English* eingegangen werden. Es umfasst laut Informationen auf der diesbezüglichen Internetseite über 700 Millionen Wörter²³⁹. Beim Addieren der Größenangaben für die einzelnen Subkorpora kommt man sogar auf eine Gesamtgröße des Korpus von 804 Millionen Wörter.²⁴⁰ Nach Auskunft von Patrick Gillard sollte man sich auf die zuletzt genannte Zahl verlassen, da das Korpus jedes Jahr vergrößert werde und die Informationen im Internet trotz der Bemühungen des Verlags nicht immer auf dem neusten Stand seien. Das *Cambridge International Corpus of English* ist also nach den verfügbaren Informationen unter den hier behandelten Korpora zur Zeit das größte.

Als Quellen für das *Cambridge International Corpus* dienen z.B. Zeitungen, Zeitschriften, Bestseller, non-fiktionale Bücher zu einer großen Bandbreite von Themen, 'Ephemera' (z.B. „junk mail“), Webseiten, Radio- und Fernsehsendungen und Aufnahmen von Alltagsgesprächen. Weitergehende Informationen zu den Quellen selbst oder zu der Menge der entnommenen Daten aus den einzelnen Genres sind für Außenstehende aus Wettbewerbsgründen nicht oder nur eingeschränkt verfügbar. Die Verfasserin dieser Arbeit erhielt von Patrick Gillard jedoch folgende Auskunft: „I can safely say that the majority of most of the written corpora [im Sinne von Subkorpora - meine Hinzufügung] would come from news media sources but that a significant amount also comes from areas such as fiction, non-fiction and academic“ (E-Mail vom 21.05.2004). Diese Aussage läßt vermuten, daß das Datenmaterial des geschriebenen Teils des Korpus ähnlich wie im Fall des *Cambridge Language Survey Corpus* in hohem Maße aus Zeitungstexten und Webseiten stammt, was die Frage nach der Ausgewogenheit des Korpus aufwirft. Die vom Verlag eingesetzte Software kann jedoch glücklicherweise zumindest anzeigen, ob ein Wort besonders häufig oder ausschließlich in bestimmten Textsorten vorkommt, woraus der Lexikograph Rückschlüsse auf den Verbreitungsgrad des jeweiligen Wortes ziehen und eventuell entsprechende Markierungen vergeben kann.

²³⁹ Vgl. die Webseite <http://uk.cambridge.org/elt/corpus/cic.htm> (Stand: Nov. 2004). Die folgenden Ausführungen basieren auf dieser Webseite sowie auf den Darlegungen von Patrick Gillard.

²⁴⁰ Darunter fällt auch das Lernerkorpus mit 27 Mio. Wörter. Informationen zum *Cambridge Learner Corpus* bietet die o.g. Webseite.

Einige wenige für das Korpus herangezogene Texte gehen bis auf die Mitte des 19. Jahrhunderts zurück, doch sind die meisten Quellen nach 1980 entstanden. Da auch Daten bis einschließlich 2004 im Korpus enthalten sind, ist die Aktualität des Sprachmaterials gewährleistet. Die fortlaufende Sammlung von neuen Daten garantiert, daß sprachliche Trends beobachtet werden können.

Von den 804 Mio. Wörtern im Korpus sind laut Angaben im Internet (Stand: Nov. 2004) ca. 518 Mio. dem BrE zuzurechnen, von denen ca. 500 Mio. aus geschriebenem und ca. 18 Mio. gesprochenem Sprachmaterial entnommen sind. Unter den letzteren Daten befinden sich auch solche aus dem *CANCODE Corpus*, das 5 Mio. Wörter spontan gesprochener Sprache aus den Jahren 1995 - 2000 umfasst und 1 Mio. Wörter aus dem *Cambridge and Nottingham Spoken Business English Corpus*. Zu dem geschriebenen BrE Korpusmaterial erfährt man auf der o.g. Internetseite, daß 20 Mio. aus „Written British academic English“ und 30 Mio. aus „Written British business English“ stammen. Dem AmE gehören ca. 259 Mio. Wörter an, davon 237 Mio. aus geschriebenem und 22 Mio. aus gesprochenem AmE. Zur Gewinnung der gesprochenen AmE Daten wurden viele verschiedene Quellen in Anspruch genommen, darunter das *Callhome Corpus* und andere Korpora, die vom *Linguistic Data Consortium* zur Verfügung gestellt werden, das *Cambridge/Cornell Corpus of Spoken North American English*, Daten von Fernsehagenturen sowie gesprochenes Material, das von *Cambridge University Press* selber aufgenommen und transkribiert wurde.²⁴¹ Da die Beschaffung eigenen gesprochenen Sprachmaterials mühsam und teurer ist, behilft sich der Verlag, indem er auf verschiedenste „fremde“ Subkorpora zurückgreift. Untersucht werden müßte, inwieweit diese Daten eine gewisse Ausgewogenheit hinsichtlich unterschiedlicher Genres (z.B. Alltagsgespräche im privaten Bereich versus Kommunikationssituationen im Berufsleben) innerhalb des gesprochenen Englisch aufweisen oder eher beliebig zusammengetragen wurden. Zudem wäre eine Vergleichbarkeit der Zusammensetzung des gesprochenen AmE Sprachmaterials mit dem des BrE für komparative Studien wünschenswert (dies trifft natürlich auch auf den schriftlichen Teil zu), um beispielsweise zutreffendere varietätenspezifische Markierungen vergeben zu können. Zumindest bezüglich des *Cambridge/Cornell Corpus of Spoken North American English* wurde erfreulicherweise darauf geachtet, daß dieses Korpus der Grundstruktur des *CANCODE Corpus* (s.o.) folgt, welches sich zudem durch eine Bandbreite verschiedener gesprochener Genres auszeichnet.²⁴² Hinsichtlich des geschriebenen Korpusmaterials scheint man auch auf eine Vergleichbarkeit BrE und AmE Subkorpora Wert zu legen, denn für beide Varietäten gibt es

²⁴¹ E-Mail von Patrick Gillard vom 21.05.04.

²⁴² Vgl. die Informationen auf der Webseite <http://www.bol.ucla.edu/~ht37/Oxford.pdf> (Stand: Mai 2004). Leider ist unklar, ob das AmE Korpus schon die angestrebten 5 Mio. Wörter umfaßt.

die Subkorpora „Written academic English“ und „Written business English“.²⁴³ Während das Subkorpus „Business English“ für beide Varietäten gleich groß ist, variiert der Umfang zwischen dem Subkorpus „Written American academic English“ (7 Mio.) und dem Subkorpus „Written British academic English“ (20 Mio) deutlich. Inwieweit die Struktur des nicht-spezialisierten geschriebenen BrE Subkorpus im Umfang von 450 Mio. Wörtern mit dem des AmE derselben Gattung (Größe: 200 Mio. Wörter) vergleichbar ist, ist unklar. Dieses BrE Subkorpus umfasst jedenfalls deutlich mehr Daten. Insgesamt gesehen liegt der Anteil des AmE am Gesamtkorpus bei ca. 32%, der des BrE dagegen bei ca. 64%. Im *Cambridge Language Survey Corpus* war das Verhältnis zwischen den Varietäten noch ausgewogener: Aus dem BrE stammten ca. 55 % der Daten, aus dem AmE ca. 45 %. Allerdings war dieses Korpus auch bedeutend kleiner: Während es nur ca. 46 Mio. Wörter aus dem AmE enthielt, verfügt das *Cambridge International Corpus* nun über mehr als die fünffache Menge an AmE Sprachdaten (ca. 259 Mio. Wörter). Eine möglichst große Ausgewogenheit zwischen dem AmE und BrE Anteil am *Cambridge International Corpus* herzustellen, ist dem Verlag aber trotzdem immer noch wichtig, wie Patrick Gillard verdeutlicht: „We try for as close a balance as possible between British English and American English, as both of these are important to learners all around the world.“ (E-Mail vom 19.03.04). Im Bereich des gesprochenen Englisch ist dies bereits geglückt, hier überwiegt zur Zeit sogar die Menge der Daten aus dem AmE. Der Umfang des gesprochenen AmE Sprachmaterials hat sich im Vergleich zum *Cambridge Language Survey Corpus* um ein Vielfaches gesteigert und liegt jetzt, wie bereits erwähnt, bei immerhin 22 Mio Wörtern. Der prozentuale Anteil des gesprochenen Englisch am *Cambridge International Corpus* ist zwar insgesamt sogar niedriger als im Vorgängerkorpus und spiegelt somit in keiner Weise die Bedeutung dieser Gattung für die Sprachteilnehmer wider, doch ist die absolute Zahl der Daten auf Grund der enormen Vergrößerung der Datenmenge im *Cambridge International Corpus* von ca. 6,1 Mio. Wörter auf ca. 40 Mio. erhöht worden. Die empirische Datenbasis hat sich also - zumindest was den Umfang des Sprachmaterials angeht - insgesamt deutlich verbessert, wovon auch das AmE sehr profitiert. Man kann daher annehmen, daß die Darstellung des AmE in Wörterbuchproduktionen, die auf dem *Cambridge International Corpus* basieren, bezüglich der Erfassung und Markierung dieser Varietät noch umfassender und präziser ausfällt als zuvor. Leider ist es im Rahmen der vorliegenden Arbeit nicht mehr möglich, dies z.B. anhand der neuen Auflage des CIDE, des *Cambridge Advanced Learner's Dictionary*, zu untersuchen.

Andere Varietäten als das AmE und BrE, nämlich Australisches und Singapurisches Englisch, sind nach Angaben des Verlags im *Cambridge International Corpus* geringer

²⁴³ Diese Subkorpora wurden laut Patrick Gillard (E-Mail vom 19.03.04) für spezifische Benutzergruppen zusammengestellt.

vertreten. Wie eine Mitarbeiterin mitteilte, spielen sie für die Veröffentlichungen des Verlags nur eine untergeordnete Rolle.²⁴⁴

Die wichtigsten Daten zum *Cambridge International Corpus* im Überblick (Stand Nov. 2004):

<i>Cambridge International Corpus</i>	
Typ	nach Angaben von Patrick Gillard ein <i>monitor corpus</i> ²⁴⁵
Größe	ca. 804 Mio Wörter (Stand: Nov. 2004)
Varietäten des Englischen	BrE: ca. 64 % (= ca. 518 Mio. Wörter) AmE: ca. 32 % (= ca. 259 Mio. Wörter) geringe Mengen von Sprachdaten aus dem Australischen u. Singapurischen Englisch (genaue Prozentzahlen nicht vorhanden) Andere, nicht muttersprachliche Varietäten (Lernerkorpus): ca. 3 % (= ca. 27 Mio. Wörter)
Alter der Quellen	Mitte des 19. Jahrhunderts - 2004,
Art der Quellen	Mehrheit der Texte aber nach 1980 entstanden geschrieben: ca. 95 % (764 Mio. Wörter) (Zeitungen, Zeitschriften, Bestseller, non-fiktionale Bücher zu einer großen Bandbreite von Themen, 'Ephemera' [z.B. „junk mail“], Webseiten u.a.) gesprochen: ca. 5 % (40 Mio. Wörter) (Radio- und Fernsehsendungen, Aufnahmen von Alltagsgesprächen)
Struktur	Vergrößerung des Korpus jedes Jahr wird angestrebt; möglichst große Ausgewogenheit zwischen den Datenmengen für das BrE und AmE ist erwünscht;

²⁴⁴ E-Mail von Ann Fiddes, *Corpus Controller* bei *Cambridge University Press*, vom 16.10.2002.

²⁴⁵ E-Mail vom 19.3.2004. Es ist allerdings unklar, ob Patrick Gillard den Begriff *monitor corpus* im Sinne von Sinclair (s.o. Kap. 3.1.6.2) verwendet, d.h. ob z. B. ältere Daten im Korpus durch neuere ersetzt werden.

	gesprochenes Material soll 5% bis 10% der Gesamtmenge der Daten ausmachen
--	---------------------------------------------------------------------------

Tab. 3.9 *Cambridge International Corpus*

3.2.4 Wörterbuchproduktionen des Cassell Verlags

Zur Erstellung des *Cassell Concise Dictionary* in der Ausgabe von 1997 und 1998 wurde kein Korpus herangezogen, wie eine Verlagsmitarbeiterin bestätigte.²⁴⁶ Der Verlag scheint bei der Wörterbucharbeit auf keinerlei empirische Daten zurückgreifen zu können: „We are a small reference department with limited resources. Dictionaries are produced from a central database which undergoes a continuous editorial process [...] We do not use corpora.“²⁴⁷ Auch für die Erfassung von Neologismen scheint es kein spezifisches Programm (vergleichbar z.B. mit dem *Chambers Wordtrack* Programm) zu geben, denn nach Informationen von Rebecca Skipwith werden Ergänzungen im Bereich der Neologismen bisher noch nicht systematisch durchgeführt, obwohl dies geplant sei (Stand: Oktober 1999). Für den Einbezug des AmE in die Wörterbuchproduktionen des Verlags ist nur ein amerikanischer Redakteur verantwortlich.²⁴⁸

Über die bereits erwähnte zentrale Datenbank erhält man nicht viele Auskünfte. Im Verlagsprojekt von 1997 wird versucht, diese Datenbank als ein hoch einsatzfähiges und vielseitiges Arbeitsmittel darzustellen:

At the heart of the new dictionary programme lies an immensely powerful, custom-built database that offers lexicographers complete control over the millions of pieces of information stored in the system. It also gives Cassell the flexibility to produce a whole range of new dictionaries and to shape them very precisely to customers' requirements.

Über die Art und Quellen der der Datenbank zugrundeliegenden Informationen erhielt die Verfasserin dieser Arbeit vom Verlag keine Auskünfte. Der bereits erwähnte Redakteur R. Milbank äußerte sich nur zu den Funktionen des verwendeten Tagging-Programms: „The database is tagged to generate several sizes of dictionary: pocket, compact, concise and complete“ (E-Mail vom 23.7.2004).

²⁴⁶ Brief von Rebecca Skipwith, *Database Assistant*, vom 5. Oktober 1999.

²⁴⁷ Ebd. Diese Aussage wird von Richard Milbank, Redakteur bei *Cassell*, bestätigt (E-Mail vom 23.7.2004).

²⁴⁸ E-Mail von Richard Milbank vom 23.7.2004.

3.2.5 Wörterbuchproduktionen von Chambers Harrap²⁴⁹

Bei der Erstellung der „revised edition“ des im Rahmen dieser Arbeit behandelten *Chambers 21st Century Dictionary* (CH21₂) standen den Lexikographen das BNC und das *Chambers Wordtrack* Programm als empirische Datenbasen zur Verfügung.²⁵⁰ Das BNC wurde allerdings nach Informationen einer Mitarbeiterin des Verlags in erster Linie nur für die Auswahl von Anwendungsbeispielen herangezogen. Somit wären nur BrE Anwendungsbeispiele im CH21₂ vorhanden, keine aus dem AmE, da das BNC fast ausschließlich aus BrE Sprachdaten besteht. Das *Chambers Dictionary*, das 1998 in der achten Auflage erschienen war, spielte bei der Arbeit am CH21₂ wahrscheinlich aber eine sehr viel größere Rolle als das BNC, wie die folgende Aussage der o.g. Mitarbeiterin verdeutlicht: „The revised edition of the *Chambers 21st Century Dictionary* was cut down from the *Chambers Dictionary*“. Da nach Angaben der Redaktion bei der „Überarbeitung“ des CH21₂ Hunderte von Neologismen dem Wörterbuch hinzugefügt wurden, wird das *Chambers Wordtrack* Programm als Datenbasis auch von großer Bedeutung gewesen sein.

3.2.5.1 Das *Chambers Wordtrack* Programm

Auf dem Buchumschlag des CH21₂ heißt es: „Chambers Wordtrack is the Chambers new-word monitoring programme, collecting around 500 new words and meanings every month“. Nach Auskünften des Verlags verfügt *Chambers* über einen Stab von freiberuflichen Mitarbeitern, die angewiesen werden, bestimmte Quellen auf Neologismen hin zu überprüfen. Außer Neologismen werden zudem auch Wörter, Wendungen und Bedeutungen gesucht, die nicht im *Chambers Dictionary* (CHD) verzeichnet sind. Unter den Quellen befinden sich laut Verlagsinformationen neben Zeitungen, Zeitschriften und Romanen auch beispielsweise Verpackungsbeilagen sowie Radio- und Filmaufnahmen. Ob die von den freien Mitarbeitern gefundenen Wörter, Wendungen und Bedeutungen in ein Wörterbuch aufgenommen werden, hängt den Auskünften des Verlags zufolge von vielen Faktoren ab: So ist im Falle eines Wortes z.B. entscheidend, wie viele Belege für das Wort vorliegen, in welchen Arten von Quellen und in welchem Zeitraum das Wort benutzt wird, und welche Personen das Wort verwenden. Zusätzlich wird der Gebrauch des Wortes z.B. im BNC, in der Zeitung *The Independent* (auf CD-Rom) und im *World Wide Web* überprüft. Bei der Frage der Aufnahme eines *items* spielt selbstverständlich

²⁴⁹ Während *Chambers* seit der Fusion mit *Harrap* ausschließlich monolinguale Wörterbücher und Nachschlagewerke für die englische Sprache erstellt, hat sich *Harrap* auf bilinguale Wörterbücher und Nachschlagewerke zu Fremdsprachen spezialisiert.

²⁵⁰ Diese (und andere) Informationen erhielt die Verfasserin der vorliegenden Arbeit während ihrer Übersetzungs- und redaktionellen Tätigkeit für *Chambers Harrap* in den Jahren 1999 -2000 sowie von einer leitenden Redakteurin des *Chambers Harrap* Verlags, Mary O'Neill, die zu dem Thema im Juni 2004 befragt wurde (E-Mail vom 10.6.2004).

die Größe des jeweiligen Wörterbuchs eine Rolle, d.h. in eine Neuauflage des CH21₂ werden deutlich weniger durch das *Chambers Wordtrack* Programm identifizierte Wörter, Wendungen und Bedeutungen einbezogen als in eine Neuauflage des CHD.

Für das Thema dieser Arbeit ist jedoch in erster Linie interessant, inwieweit das AmE im *Chambers Wordtrack* Programm berücksichtigt wird. Als die erste Auflage des CH21₂ überarbeitet wurde, war kein amerikanischer Mitarbeiter bzw. keine amerikanische Mitarbeiterin in das Programm involviert, obwohl es z.B. eine entsprechende Person in Australien gab. Diese Situation scheint sich bis ins Jahr 2004 nicht verändert zu haben, wie die Informationen der leitenden Redakteurin Mary O'Neill verdeutlichen.²⁵¹ Amerikanische Publikationen, z.B. wissenschaftliche Zeitschriften und Romane, werden zwar in das *Chambers Wordtrack* Programm einbezogen, doch werden diese von britischen Mitarbeitern bzw. Mitarbeiterinnen gelesen. Es ist daher eher zweifelhaft, ob das *Chambers Wordtrack* Programm zu sprachlichen Neuerungen im AmE in ausreichender Weise Informationen liefern kann, denn zum einen sind die britischen Mitarbeiter bzw. Mitarbeiterinnen mit dem AmE vermutlich nicht vertraut genug, um diese Neuerungen zu erkennen oder adäquat beschreiben zu können. Zum anderen sieht es nicht so aus, als ob es eine breite Auswahl von AmE Quellen gäbe, in denen nach Neologismen im AmE geforscht würde.

3.2.5.2 Andere Datenquellen des Verlags

Neben dem BNC und *Chambers Wordtrack* Programm greifen die Verlagsmitarbeiter, wie bereits erwähnt, auch auf die Zeitung *The Independent* (auf CD-Rom) und das *World Wide Web* zurück. Besonders das Internet und solche Korpora, die auf dem Internet basieren, haben wegen ihrer großen Aktualität bei *Chambers Harrap* in den letzten Jahren an Bedeutung gewonnen. Das Internet ist nach Angaben der Verlagsmitarbeiter zudem eine gute Quelle für AmE Sprachdaten und wird gerne zur Untersuchung des AmE herangezogen. Allerdings sind sich die Redakteure auch der Nachteile des Internets bewußt. So schreibt Mary O'Neill: „[...] care has to be taken when analysing the results. The language can often tend towards informal language and structures, and searches can produce a number of repetitions and false results.“

Neben den bereits erwähnten Quellen verwenden die Redakteure für die Erstellung von Wörterbüchern auch eine hausinterne Datenbank, in der Neologismen (mit dem dazugehörigen Kontext) verzeichnet sind, auf die sie selbst gestoßen sind.

²⁵¹ E-Mail vom 10.06.2004.

3.2.5.3 Abschließende Bemerkungen zu den Datenquellen des Verlags

Leider hat der Verlag kaum Zugriff auf verlässliche AmE Sprachdaten, geschweige denn auf eine umfassende Datenbasis auf der Grundlage eines Korpus des AmE. Das BNC enthält, wie bereits mehrfach erwähnt, kein AmE Sprachmaterial. Die Zeitung *The Independent* kann ebenfalls nur BrE Sprachdaten liefern, da es sich um eine britische Zeitung handelt. AmE Sprachdaten aus dem Internet sind jedoch mit Vorsicht zu behandeln, wie oben bereits dargelegt wurde. Das *Chambers Wordtrack* Programm schließlich zielt nur auf Neologismen, und sein Nutzen hinsichtlich des AmE wird deswegen beschränkt sein, weil es ausschließlich britische Redakteure sind, die (in einer zudem wahrscheinlich eingeschränkten Auswahl von Quellen) nach AmE Neologismen suchen. Eine Belegsammlung, wie sie im Rahmen des *Chambers Wordtrack* Programm erstellt wird, kann zudem nur als zusätzliche Datenquelle dienen und kein Ersatz für mangelnden Korpusdaten für das AmE und andere Varietäten außer dem BrE darstellen. *Chambers Harrap* ist auch leider nicht unter den Konsortiumsmitgliedern für das ANC zu finden (Stand Mai 2005), was insofern verwundert, als der Verlag bereits Konsortiumsmitglied für das BNC war und fast alle wichtigen konkurrierenden britischen Verlagshäuser im Konsortium für das ANC vertreten sind (so z.B. *Oxford University Press*, *Cambridge University Press*, *Pearson Education* und *Macmillan Publishers*). Die Verwendung des ANC wäre auch für *Chambers Harrap* sehr vorteilhaft, da das ANC dieselben Genres wie das BNC einbezieht und daher in Größe und Struktur dem BNC entsprechen wird.

3.2.6 Wörterbuchproduktionen des Penguin Verlags

Wie im Kapitel 3.1.4 bereits angesprochen, verwendet der *Penguin* Verlag offensichtlich keine Korpora zur Erstellung von Wörterbüchern. Das einzige empirische Datenmaterial scheint eine Belegsammlung zu sein. Sie wird überdies wohl nur dazu genutzt, Beispiele für die Verwendung vieler in dem Wörterbuch verzeichneter Stichwörter zu liefern. Dennoch werben die Herausgeber ausdrücklich mit den aus literarischen Werken entnommenen Belegen für das NPED₁: „examples drawn from the works of major writers, from Shakespeare to Jane Austen to P G Wodehouse to Salman Rushdie“. Dies erinnert sehr an die Auswahl der Quellen, die schon für Wörterbücher wie z.B. das *Dictionary of the English Language* (1755) eine Rolle spielten (vgl. Kap. 3.1.4). Unter den für die Erstellung des NPED₁ berücksichtigten Werken sind auch solche amerikanischer Schriftsteller, z.B. von Philip Roth. Allerdings sind im NPED₁ auch sehr viele Beispiele ohne Quellenangabe anzutreffen, die daher wahrscheinlich erdacht sind, obwohl dies an keiner Stelle im Wörterbuch explizit erwähnt wird. Zudem zeigt eine flüchtige Stichprobe, daß vereinzelt auch z.B. Zitate aus Zeitungen, Zeitschriften oder Parlamentsberichten zur Illustration von Wortbedeutungen verwendet werden, obwohl auf dem

Rückendeckel und in der Einleitung des NPED₁ nur auf literarische Quellen Bezug genommen wird. Die Zusammensetzung der Belegsammlung und die Kriterien für die Aufnahme von Belegen sind also nicht klar ersichtlich.

3.3 Abschließende Beurteilung der den Verlagshäusern für die untersuchten Wörterbücher zugrundeliegenden Belegquellen

	Harper Collins	Longman	Oxford University Press	Cambridge University Press	Cassell	Chambers Harrap	Penguin
Stand	Juni 2004	Februar 2004	April 2004	Nov. 2004	Juli 2004	Juni 2004	Juni 2004
Werden Korpora ²⁵² verwendet? Welche Korpora werden verwendet?	ja BoE (524 Mio.)	ja <i>Longman Corpus Network</i> ²⁵³ (ca. 300 Mio.) und <i>BNC</i> (ca. 100 Mio.) und ANC (ca. 10 Mio)	ja <i>Oxford English Corpus</i> (Größe ?) und <i>BNC</i> (ca. 100 Mio.) und ANC (ca. 10 Mio) u.a.	ja <i>Cambridge International Corpus of English</i> (ca. 804 Mio.) und ANC (ca. 10 Mio.)	nein	ja <i>BNC</i> (ca. 100 Mio.)	nein
Werden Belegsammlungen verwendet?	nein	ja	Ja	nein	nein	ja	ja
Werden andere Quellen verwendet? (Internet, Zeitungen auf CD-ROM u.ä.)	nicht bekannt	nicht bekannt	Ja	nicht bekannt	nicht bekannt	ja	nicht bekannt

Tab. 3.10 Überblick über die Belegquellen der untersuchten Verlage (soweit bekannt)

²⁵² Die Größenangaben für die Korpora beziehen sich auf die Anzahl der enthaltenen Wörter.

²⁵³ Das *Longman Corpus Network* wird im folgenden der Einfachheit halber mit LCN abgekürzt.

	Harper Collins	Longman	Oxford University Press	Cambridge University Press	Chambers Harrap
Stand	2002	2001	2000	1995	1999
Größe der Korpora	BoE (ca. 448,5 Mio.)	1) LCN (ca. 95 Mio.) 2) BNC (ca. 100 Mio.+ 20 Mio. von Longman)	1) BNC (ca. 100 Mio.) 2) für COD ₁₀ (und OALD ₆ ?): AmE Korpus (weniger als 50 Mio.); weitere Korpora für OALD ₆ ?	<i>Cambridge Language Survey Corpus</i> ²⁵⁴ (ca. 102 Mio.) (und Lernerkorpus, ca. 5 Mio.)	BNC (ca. 100 Mio.)
Korpustyp	<i>monitor corpus</i>	1) LCN: <i>sample corpora</i> u. <i>dynamische corpora</i> 2) BNC: <i>sample corpus</i>	1) BNC: <i>sample corpus</i> 2) AmE Korpus: Spezialkorpus, statisch	CLSC: eher <i>sample corpus</i> als <i>monitor corpus</i>	<i>sample corpus</i>
Berücksichtigung des AmE	ca. 76,4 Mio. (= ca. 17%)	1) LCN: ca. 67 Mio. (= ca. 71 %) 2) BNC: fast nur BrE	1) BNC: fast nur BrE 2) AmE: weniger als 50 Mio.	CLSC: ca. 46 Mio. Wörter (= 45 %)	fast nur BrE
Alter der Quellen	Mehrheit der Texte ab 1990 (einige schon ab 1980)	1) LCN: ab 1900 2) BNC: 1960-2001	1) BNC: 1960-1994 2) AmE Korpus: ?	CLSC: bis auf Ausnahmen: 2. Hälfte der 80er - 1. Hälfte der 90er Jahre	1960-1994
Art der Quellen	geschrieben: ca. 86 % (= ca. 385,6 Mio., davon 52,2 Mio. AmE) gesprochen: ca. 14 % (= ca. 62,9 Mio., davon 24,2 Mio. AmE)	1) LCN: geschrieben: ca. 94,7 %, (= ca. 90 Mio., davon 67 Mio. AmE); gesprochen: ca. 5,3 % (= ca. 5 Mio., nur aus AmE bestehend) 2) BNC: ? ²⁵⁵	1) BNC: geschrieben: ca. 90 % (= ca. 90 Mio.), gesprochen: ca. 10 % (= ca. 10 Mio.) 2) AmE Korpus: nur geschrieben	CLSC: geschrieben: ca. 94 % (= 95,9 Mio. Wörter, Anteil des AmE ?) gesprochen: ca. 6 % (= 6,1 Mio., Anteil des AmE höher als des BrE)	geschrieben: ca. 90 % (= ca. 90 Mio.), gesprochen: ca. 10 % (= ca. 10 Mio.)

Tab. 3.11 Übersicht über die etwa zur Zeit der Erstellung der untersuchten Wörterbuchausgaben verwandten Korpora (soweit bekannt)

²⁵⁴ Das *Cambridge Language Survey Corpus* wird im folgenden der Einfachheit halber mit CLSC abgekürzt.

²⁵⁵ Prozentangaben sind hier nicht möglich, da die Zusammensetzung des neu hinzugefügten Teils von 20 Mio. Wörter aus den Jahren 2000/2001 nicht bekannt ist (s. Kap. 3.2.2.2).

Wie die beiden Übersichtstabellen verdeutlichen, haben sich vor allem die Korpusressourcen einiger der in dieser Arbeit behandelten Verlage in den letzten Jahren immens vergrößert. Da die untersuchten Wörterbuchausgaben aus den Jahren 1995 bis 2001 stammen, waren die Verlage zum Zeitpunkt der Veröffentlichung ihrer Produkte unterschiedlich weit in der Entwicklung ihrer Belegquellen, vor allem der Korpusressourcen. Um diesem Umstand Rechnung zu tragen, soll am Ende dieses Kapitels auch kurz auf das Belegmaterial eingegangen werden, auf das die Verlage zur Zeit (Stand 2004) zurückgreifen. Bei der Beurteilung der Belegquellen darf allerdings nicht vergessen werden, daß die (leider „kurzlebigen“) Informationen, auf denen diese Beurteilung basiert, zum größten Teil von den Verlagen selbst stammen. In der Annahme, daß die zugrundeliegenden Auskünfte zutreffend und vollständig sind, soll nun zunächst vergleichend auf die Belegquellen eingegangen werden, die den Verlagen zur Zeit der Veröffentlichung der analysierten Wörterbuchaufgaben zur Verfügung standen.

Der einzige Verlag unter den insgesamt sieben behandelten, der sich anscheinend keinerlei empirischer Daten bei der Erstellung seines hier untersuchten Wörterbuchprodukts, des CACD, bediente, ist der Cassell Verlag. Da im CACD keine Angaben darüber gemacht werden, auf welche Datenbasis die Herausgeber für die Produktion des Wörterbuchs zurückgegriffen haben, ist anzunehmen, daß sie sich der Introspektion und vielleicht ebenfalls der Auswertung anderer Wörterbücher zur Gewinnung lexikographischer Daten bedient haben. Diese Methoden zur Datengewinnung weisen jedoch gegenüber der Verwendung von Korpora erhebliche Nachteile auf (s. Kap. 3.2.4).

Auch die an der Arbeit am NPED₁ beteiligten Lexikographen werden - wie bereits erwähnt - neben der Belegsammlung, die allem Anschein nach in erster Linie nur als Quelle für viele der im Wörterbuch verzeichneten Anwendungsbeispiele benutzt wurde, wahrscheinlich diese Methoden angewandt haben, um lexikographische Daten zu erhalten. Auf Grund der in Kapitel 3.2.4 formulierten Einschätzungen zur Introspektion und der Auswertung anderer Wörterbücher ist eher zu bezweifeln, daß die im CACD und NPED₁ zur Verfügung gestellten Informationen beispielsweise in Bezug auf die Erfassung und Markierung von Amerikanismen genauso verlässlich sind wie diejenigen der Wörterbücher, die auf Korpora basieren (s. Kap. 7 dieser Arbeit).

Dem Verlag *Chambers Harrap* stand bei der Erstellung des CH21₂ zwar ein Korpus, das BNC, zur Verfügung, doch enthält dieses - wie bereits mehrfach erwähnt - fast nur BrE Sprachmaterial. AmE Korpusdaten konnte *Chambers Harrap* nicht verwenden; andere verwandte Quellen, wie z.B. das *Chambers Wordtrack* Programm oder das Internet, sind kein adäquater Ersatz. Daher ist zu vermuten, daß auch im CH21₂ die Qualität der Informationen zu Amerikanismen nicht so gut ist wie in den behandelten Wörterbuchprodukten von *Harper*

Collins, Oxford University Press, Cambridge University Press und *Longman*, die u.a. auf AmE Korpusdaten basieren.

Zur Zeit der Veröffentlichung des COBUILD₃ und CCD₅ im Jahr 2001 verfügte *Harper Collins* zwar erst über ein etwas kleineres Korpus (zwischen 418 und 448,5 Mio. Wörtern) als in Tabelle 3.1 angegeben, doch bestand der Anteil AmE Sprachdaten wohl schon in diesem Umfang. Auf den ersten Blick scheint *Harper Collins* daher bei der Publikation der o.g. zwei Wörterbuchausgaben im Besitz einer größeren Menge an AmE Korpusdaten gewesen zu sein als die anderen drei Verlage bei der Veröffentlichung ihrer Konkurrenzprodukte. Allerdings spielt auch die Zusammensetzung des AmE Korpusmaterials eine wichtige Rolle für die Nutzung. So hätte *Longman* nämlich - zumindest bei der Publikation der erweiterten Ausgabe des LDOCE₃ im Jahr 2001 - auch auf ein ca. 100 Mio. Wörter großes „newsdata“ Subkorpus des *Longman Written American Corpus* zurückgreifen können, doch wurde dieses anscheinend darum nicht verwendet, weil es strukturell nicht ausgewogen ist. Rechnet man dieses Subkorpus jedoch zu den AmE Korpusressourcen des Verlags hinzu, verfügte *Longman* mit 167 Mio. Wörtern über deutlich mehr AmE Korpusdaten als die anderen Verlage. Läßt man das besagte Subkorpus außer Acht, so war zwar *Harper Collins* - wie bereits erwähnt - im Besitz der größten Sammlung von AmE Korpusdaten, doch war die Zusammensetzung der AmE Subkorpora von *Longman* weniger ‚opportunistisch‘. *Cambridge University Press* konnte bei der Veröffentlichung des CIDE₁ im Jahr 1995 auf ca. 46 Mio. AmE Wörter im CLSC rekurren. Damit verfügte der Verlag zur Zeit der Herausgabe der o.g. Wörterbuchprodukte über weniger AmE Sprachmaterial als *Longman* oder *Harper Collins*. Die Zusammensetzung des AmE Sprachmaterials im CLSC scheint zudem auch eher dem Prinzip der leichten Verfügbarkeit der Quellen unterworfen gewesen zu sein als dies bei *Longman* der Fall war. *Oxford University Press* benutzte zwar zur Erstellung des COD₁₀ (und OALD₆?) ein AmE Korpus von weniger als 50 Mio. Wörtern, doch das enthielt ausschließlich journalistische Texte. So verfügte der Verlag zumindest bei der Arbeit am COD₁₀ über eine qualitativ weniger gute Korpusdatenbasis bezüglich des AmE als *Cambridge University Press*, *Longman* oder *Harper Collins*. Zudem stellten das BNC und die *Oxford Reading Programmes* nach Verlagsangaben immer noch die wichtigsten Bezugspunkte für das COD₁₀ und OALD₆ dar. Das *North American Reading Programme*, das zu den *Oxford Reading Programmes* gehört, konnte und kann AmE Korpusmaterial jedoch aus verschiedenen Gründen, die bereits in Kap. 3.2.3.3. dargelegt wurden, nicht ersetzen.

Betrachtet man den Einbezug des gesprochenen AmE, so standen *Harper Collins* zur Zeit der Veröffentlichung des COBUILD₃ und CCD₅ im Jahr 2001 deutlich mehr gesprochene AmE Daten (ca. 24 Mio. Wörter) zur Verfügung als *Cambridge University Press*, *Longman* oder

Oxford University Press bei der Publikation der hier behandelten Wörterbuchausgaben. Allerdings war die Zusammensetzung dieser gesprochenen AmE Daten recht einseitig (ca. 22 Mio. Wörter stammten aus dem *American Radio* Subkorpus). Aufnahmen von Alltagsgesprächen waren anscheinend kaum vorhanden, so daß es möglich ist, daß Ausdrücke aus der informellen Sprache oder dem Slang dieser Varietät nicht in ausreichendem Maße in der BoE repräsentiert sind, was wiederum Auswirkungen auf die Qualität der diesbezüglichen Wörterbuchinformationen hätte. *Longman* konnte zur Zeit der Veröffentlichung des LDEL₂ und der erweiterten Ausgabe des LDOCE₃ (sowie wohl bei der Publikation von LDOCE₃ 1995) nur auf ein ca. 5 Mio. Wörter großes *Spoken American Corpus* rekurren, das Sprachmaterial aus gesprochener Alltagssprache enthält und im Bemühen um ‚Repräsentativität‘ nach demographischen Prinzipien entwickelt wurde. Nach Angaben der Herausgeber war dieses Korpus z.B. ein wichtiges Hilfsmittel bei der Darstellung von AmE Slang-Ausdrücken in den Wörterbüchern, doch finden sich andere Genres der gesprochenen Sprache, die z.B. im *context-governed part* des BNC vorhanden sind, im *Longman Corpus Network* für das AmE nicht. *Cambridge University Press* verfügte bei der Veröffentlichung des CIDE₁ im Jahr 1995 vermutlich über noch weniger gesprochenes AmE Datenmaterial als *Longman* bei den o.g. Wörterbuchprodukten. Da aber zu vermuten ist, daß spontan gesprochenes Material (z.B. aus persönlichen Gesprächen in der Familie oder mit Freunden) nur wenig im *Cambridge Language Survey Corpus* vertreten war, stellt sich hier wieder die Frage, inwieweit informeller Wortschatz oder Slang- Ausdrücke des AmE im CIDE₁ angemessen dargestellt werden konnten. Diese Problematik betrifft jedoch *Oxford University Press* noch viel mehr, da dieser Verlag - zumindest bei der Herausgabe von COD₁₀, wahrscheinlich auch bei der von OALD₆ - auf keinerlei gesprochene AmE Korpusdaten Zugriff hatte. Abschließend läßt sich also festhalten, daß von den in dieser Arbeit untersuchten sieben Verlagen zur Zeit der Veröffentlichung der behandelten Wörterbücher nur drei über gesprochene AmE Korpusdaten verfügten (*Harper Collins*, *Longman* und *Cambridge University Press*). Bezüglich der beiden letzteren Verlage hätte man sich zudem eine deutlich höhere Menge an gesprochenen AmE Korpusdaten gewünscht und für alle drei Verlage eine größere Vielfalt der Genres des gesprochenen AmE z.B. nach dem Muster des BNC.

Neben den Aspekten des Umfangs und der Zusammensetzung des vorhandenen geschriebenen und gesprochenen Korpusmaterials aus dem AmE, die vor allem für die Erfassung von Amerikanismen in den Wörterbüchern von Interesse sind, spielt auch der gesamte Korpusbestand des jeweiligen Verlags eine Rolle. Denn um Amerikanismen in den Wörterbüchern als solche markieren zu können, braucht man die Gewissheit, daß sie in anderen Varietäten des Englischen - oder zumindest im BrE - weniger häufig oder gar nicht verwendet

werden. Zum Vergleich der Frequenz sprachlicher Phänomene in verschiedenen Varietäten eignen sich, wie bereits in Kap. 3.1.6.1 dargelegt, am besten varietätenspezifische (Sub-) Korpora, die etwa gleich groß und in gleicher Weise aufgebaut sind.

Leider konnte keiner der behandelten Verlage zur Zeit der Veröffentlichung der analysierten Wörterbuchprodukte mit einem Korpusbestand aufwarten, der diesem Modell vollends entspricht. *Longman* hatte sich zwar bei der Entwicklung des *Longman Spoken American Corpus* in Größe und Aufbau nach den Prinzipien des ‚demographischen‘ Teils des *BNC Spoken Corpus* gerichtet, aber das 50 Mio. Wörter umfassende Subkorpus des *Longman Written American Corpus* war, obwohl es nach dem Muster des *BNC Written Corpus* entstand, sehr viel kleiner als dieses. Bei *Harper Collins* gab es zum Zeitpunkt der Veröffentlichung des COBUILD₃ und CCD₅ nur wenige AmE und BrE Subkorpora, die nicht nur bezüglich des Inhalts, sondern auch der Größe sehr gut vergleichbar gewesen wären. Über den Aufbau des CLSC von *Cambridge University Press* liegen keine detaillierten Informationen vor, aber der Anteil der AmE Korpusdaten war in jedem Fall niedriger als der des BrE Korpusmaterials. *Oxford University Press* konnte gar keine annähernd vergleichbaren varietätenspezifischen (Sub-) Korpora vorweisen. Für andere Varietäten außer dem AmE und BrE sah die Korpuslage bei allen Verlagen noch schlechter aus.

Die jüngsten Korpusentwicklungen in einigen Verlagshäusern sind jedoch vielversprechend. Vor allem bei *Cambridge University Press* und *Oxford University Press* lassen sich große Fortschritte feststellen. *Cambridge University Press* besitzt zur Zeit (Stand November 2004) mit dem *Cambridge International Corpus* (804 Mio. Wörter ohne die ersten 10 Mio. des ANC) nicht nur das unter den hier behandelten Verlagen größte Korpus, sondern bemüht sich augenscheinlich auch um die Erstellung vergleichbarer BrE und AmE Subkorpora.²⁵⁶ Insgesamt wird zudem eine möglichst große Ausgewogenheit zwischen dem AmE und BrE Anteil am *Cambridge International Corpus* angestrebt. Zwar gehören bisher nur ca. 32% der Korpusdaten dem AmE an, aber dies sind in absoluten Zahlen immerhin ca. 259 Mio. Wörter. Im Bereich des gesprochenen Englisch überwiegt zur Zeit sogar die Menge der Daten aus dem AmE (22 Mio. Wörter) vor dem BrE (18 Mio. Wörter). Die empirische Datenbasis hat sich also auch hinsichtlich des AmE - zumindest was den Umfang des Sprachmaterials angeht - enorm verbessert. Wie ausgewogen das *Cambridge International Corpus* insgesamt gestaltet ist, läßt sich aufgrund mangelnder Informationen leider nicht abschließend beurteilen. Da aber dem Korpus beständig neue Daten hinzugefügt werden, finden dort auf jeden Fall neuere Sprachentwicklungen Niederschlag.

²⁵⁶ Vgl. z.B. die „Business English“ Subkorpora für die beiden Varietäten und das *Cambridge/Cornell Corpus of Spoken North American English* und das *CANCODE Corpus*.

Oxford University Press arbeitet seit einigen Jahren nun auch an einem eigenen umfangreichen Korpus, dem sog. *Oxford English Corpus*, das derzeit schon viele hundertmillionen Wörter umfassen soll. Da leider keine genauen Zahlen vorliegen, ist ungewiß, ob dieses Korpus schon die Größe der BoE oder des *Cambridge International Corpus* erreicht hat. Zur Struktur des Korpus stehen ebenfalls keine detaillierten Informationen zur Verfügung, aber nach Angaben der Verlagsmitarbeiter enthält das Korpus eine ausgewogene Auswahl von Sprachmaterial aus verschiedensten Quellen und Varietäten des Englischen. Im Gegensatz zu *Cambridge University Press* legt der Verlag somit auch Wert auf den Einbezug anderer Varietäten außer dem BrE und AmE. Die Aktualität des *Oxford English Corpus* soll dadurch gewährleistet werden, daß dem neuen Korpus beständig neues Material hinzugefügt wird.

Harper Collins verfügt mittlerweile über ein noch größeres Korpus (524 Mio. Wörter, Stand Juni 2004), dessen derzeitige Zusammensetzung leider nicht genau bekannt ist. Geplant war jedoch, eine größere Vergleichbarkeit zwischen BrE und AmE Subkorpora in Größe und Struktur zu schaffen.

Auskünfte über neuere Korpusentwicklungen bei *Longman* liegen leider nicht vor. Die Informationen im Internet zur Größe der Subkorpora im *LCN* haben sich bis Nov. 2004 nicht verändert. Entweder werden die Auskünfte im Internet nicht auf den neuesten Stand gebracht, oder die dynamischen Korpora im *LCN* sind tatsächlich nicht vergrößert worden. Letzteres würde jedoch verwundern, da *Longman* ebensowenig wie die anderen Verlage auf aktuelle Korpusdaten des BrE und AmE (und anderer Varietäten) verzichten kann.

Bei *Cassell*, *Chambers Harrap* und *Penguin* scheint es bis heute (Stand Juni/ Juli 2004) keine Bemühungen um die Entwicklung von hausinternen Korpora zu geben. Trifft dies zu, hinken diese drei Verlage der lexikographischen Forschung deutlich hinterher.

4. Die ausgewählten Wörterbücher

4.1 Bedeutung der in den Wörterbüchern außerhalb des Lexikon-Teils gegebenen Informationen zur Berücksichtigung des AmE

Bevor eine Überprüfung einer begrenzten Auswahl von Amerikanismen in den Einträgen der verschiedenen Wörterbücher stattfindet, ist es lohnenswert, sich detailliert mit der Konzeption und der Zielsetzung der einzelnen Werke auseinanderzusetzen. Auf diese Weise läßt sich feststellen, unter welchen Voraussetzungen die Erfassung und Markierung von Amerikanismen in den untersuchten Wörterbüchern vonstatten ging. Außerdem muß man, um bei der Beurteilung von Wörterbüchern fair sein zu können, deren Zielsetzung kennen.²⁵⁷ Im folgenden wird daher auf solche Auskünfte in den einführenden Kapiteln zu den einzelnen Wörterbüchern und auf andere, außerhalb der eigentlichen Einträge befindlichen Informationen eingegangen, die für den Untersuchungsgegenstand dieser Arbeit von Bedeutung sind.

Beispielsweise sind Angaben darüber, ob **amerikanische Lexikographen oder Redakteure** an den Wörterbuchausgaben beteiligt waren, von Interesse. Man kann annehmen, daß bei einer Mitarbeit amerikanischer Lexikographen der Einbezug des AmE wahrscheinlich größer und die Darstellung der Amerikanismen vielleicht präziser ist als bei Wörterbuchausgaben, an denen nur britische Lexikographen mitgewirkt haben.

Informationen über die Anzahl der in das jeweilige Wörterbuch einbezogenen **Lemmata** für den Untersuchungsgegenstand dieser Arbeit sind insofern interessant, als eine höhere Anzahl gleichzeitig eine umfassendere Aufnahme von Amerikanismen bedeuten kann. So sind z.B. *taffy* oder *graft off* im LDOCE₃, das nach Angaben der Herausgeber ca. 24.000 Stichwörter und Wendungen mehr als das LDOCE₂ enthalten soll, verzeichnet, während diese Lemmata in das LDOCE₂ nicht einbezogen wurden.²⁵⁸ Auskünfte zu den Auswahlkriterien für die in den Wörterbüchern einbezogenen Lemmata sind in diesem Zusammenhang von großer Wichtigkeit.

Auf die von den Verlagshäusern der hier untersuchten Wörterbücher benutzten **Belegquellen** (vor allem Korpora) wurde schon in Kapitel 3 ausführlich eingegangen. Als Datenbasis für die Wörterbuchproduktionen kommt ihnen eine besondere Rolle zu.

Erklärungen der Herausgeber der Wörterbücher zu dem **label**, den sie jeweils für die Markierung des AmE verwenden, sind ebenfalls von Bedeutung. Im OCEL (S. 41f.) wird darauf hingewiesen, daß sich die Ambiguität und Vagheit der Begriffe *British English* und *American English*, die auf verschiedene Weise von verschiedenen Personen für verschiedene Zwecke

²⁵⁷ Anhand der stichprobenartigen Kontrolle der von mir ausgewählten Amerikanismen in den Wörterbüchern läßt sich feststellen, ob manche der in den einleitenden Kapiteln der Wörterbücher gemachten Angaben hinsichtlich des Einbezugs des AmE bestätigt werden können.

verwendet werden, auch in den Markierungen der Wörterbücher niederschlägt. Dem OCEL zufolge kann ein Wörterbuchbenutzer bei der Etikettierung *BrE* davon ausgehen, daß das so Gekennzeichnete in Großbritannien gebräuchlicher ist als in den USA, aber er könne nicht sicher sein, ob dieses Wort, diese Wendung etc. nur auf Großbritannien beschränkt ist oder z.B. auch in Australien oder Neuseeland vorkommt.²⁵⁹ Die internationale Verwendung der Varietätenbezeichnungen AmE und BrE sei unzureichend bestimmt.

Auch die **Darstellung lexikalischer** und **grammatikalischer Unterschiede** zwischen dem AmE und BrE in den einleitenden Kapiteln eines Wörterbuchs sowie Informationen darüber, wie diese Unterschiede in den Einträgen gekennzeichnet werden, sind von Wichtigkeit. Um so ausführlicher und detaillierter diese Wiedergabe lexikalischer und grammatikalischer Unterschiede in der Einleitung oder in den zusätzlichen Informationen eines Wörterbuchs ist, desto höher ist die Wahrscheinlichkeit, daß einer Vielzahl von Unterschieden in der Lexik und Grammatik zwischen dem AmE und BrE auch in den Einträgen Rechnung getragen wird.

Auskünfte solcher Art zu den einzelnen Wörterbüchern werden nach diesen Aspekten geordnet in Kapitel 4.3. wiedergegeben. Die jeweils ältere Auflage eines Wörterbuchs wird im folgenden immer dann herangezogen, wenn sie in bestimmten, für den Untersuchungsgegenstand wichtigen Aspekten von der neuen abweicht.

Informationen, die die Darstellung aller Stichwörter in den jeweiligen Wörterbüchern - und somit auch die der Amerikanismen, aber nicht nur dieser - betreffen, wie z.B. Auskünfte über die Definitionspraxis oder die Art und Zahl der angefügten Beispiele, werden nur in dem Umfang dargelegt, wie sie für den Untersuchungsgegenstand der vorliegenden Arbeit interessant erscheinen.

In Kap. 4.2. werden jedoch zunächst die gattungsspezifischen Unterschiede zwischen monolingualen Wörterbüchern für fortgeschrittene Lerner und Wörterbüchern für Muttersprachler der *Concise*-Sparte dargelegt.

²⁵⁸ Vgl. auch Norri (1996:3).

²⁵⁹ Laut OCEL wissen die Lexikographen, die die Markierung benutzen, das selber oft nicht. Hier kann nur ein angemessener Einbezug anderer Varietäten außer dem BrE und AmE in die den jeweiligen Wörterbüchern zugrundeliegenden Korpora für größere Sicherheit in der Markierungspraxis sorgen.

4.2 Gattungsspezifische Unterschiede zwischen monolingualen Wörterbüchern für fortgeschrittene nicht-englischsprachige Lerner und Wörterbüchern für Muttersprachler der *Concise*-Sparte

Die Unterschiede zwischen diesen beiden Wörterbuchtypen ergeben sich aus den divergierenden Anforderungen, die Lerner und Muttersprachler an ein Wörterbuch des Englischen stellen.²⁶⁰ Aus einer Reihe von Studien zur Verwendung von Wörterbüchern des Englischen durch Muttersprachler geht hervor, daß für diese Benutzergruppe das Nachschlagen von Bedeutungen am wichtigsten ist.²⁶¹ Dieses Ergebnis zeigt, daß Muttersprachler Wörterbücher des Englischen häufiger zur Dekodierung als zur Enkodierung von Informationen verwenden. Der Einbezug einer möglichst großen Zahl von Stichwörtern bzw. Bedeutungen ist somit das Ziel von Lexikographen, die Wörterbücher für Muttersprachler erstellen.²⁶² Um dieses Ziel zu erreichen, wird den Informationen zu den einzelnen Stichwörtern oder Bedeutungen gewöhnlich weniger Platz eingeräumt als in einem Lernerwörterbuch.

Die Bedürfnisse der Lerner sind demgegenüber anders: Lerner brauchen ein monolinguales Wörterbuch, das ihnen außer bei der Dekodierung vor allem bei der Enkodierung Hilfestellung leistet.²⁶³ Daraus folgt, daß in einem Lernerwörterbuch die Zahl der aufgenommenen Stichwörter viel beschränkter sein muß als in einem Wörterbuch für Muttersprachler: Wenig frequente oder sehr fachsprachliche Wörter, auf die der Lerner wahrscheinlich nicht stoßen wird, werden beispielsweise in ein Lernerwörterbuch nicht aufgenommen, um Platz für solche Informationen zu schaffen, die für die Sprachproduktion hilfreich sind. Ein Lernerwörterbuch sollte nach Rundell (1988:127) „a multi-faceted language-learning resource“ sein.²⁶⁴ Neben der Bedeutung spielen daher viele andere Informationen in einem Lernerwörterbuch eine Rolle, z.B. Angaben zur Aussprache: Um ein Wort in der mündlichen Kommunikation verwenden zu können, muß der Lerner wissen, wie man es ausspricht. So enthalten alle der hier untersuchten

²⁶⁰ Vgl. auch Herbst (1990:1379).

²⁶¹ Vgl. z.B. Quirk (1973) und Barnhart (1962). Siehe auch Rundell (1988:128). Daneben sind Muttersprachler vor allem an Informationen zur Orthographie und zu Synonymen interessiert. Vgl. dazu auch Whitcut (1989:90). Wörterbücher für Muttersprachler unterscheiden sich von Lernerwörterbüchern auch durch den Einbezug etymologischer Informationen, die in Lernerwörterbüchern keine große Rolle spielen, vgl. auch Herbst (1990:1379) und Heuberger (2000:2). Das OALD₆ gibt allerdings bei einigen ausgewählten Wörtern und Ausdrücken Einblicke in die Wortgeschichte mittels sog. *origin notes*.

²⁶² Vgl. auch Rundell (1988:128) und Pätzold (1994:50).

²⁶³ Vgl. Svensen (1993:13f.), Herbst (1990:1379), Pätzold (1994:50), Ilson (1988:9) und Ilson (1990b:1975). So konstatiert Pätzold (1994:51) mit Recht, daß monolinguale Wörterbücher für Muttersprachler unzureichend für die Sprachproduktion von Nicht-Muttersprachlern sind, weil sie für die Enkodierung nicht genügend Unterstützung bieten.

²⁶⁴ Svensen (1993:22) beschreibt den Unterschied zwischen monolingualen Wörterbüchern für Lerner und für Muttersprachler folgendermaßen: „The student does not need a description of the language so much as an effective tool for communication.“

Lernerwörterbücher nicht nur BrE Ausspracheangaben zu jedem Stichwort²⁶⁵, sondern berücksichtigen auch die AmE Aussprache. Viele der von den Verlagen anvisierten Lerner lernen nämlich AmE und benötigen diese Ausspracheangaben. Bei den britischen monolingualen Wörterbüchern der *Concise*-Sparte kommt es jedoch, wie im Falle des COD₁₀, vor, daß Ausspracheangaben nur für schwer auszusprechende Wörter angeführt werden²⁶⁶, und dazu ausschließlich in BrE. (Britische monolinguale Wörterbücher des Englischen werden in den USA in der Regel nicht verkauft.)

Außerdem werden in Lernerwörterbüchern in der Regel detailliertere grammatikalische Informationen gegeben als in Wörterbüchern für Muttersprachler²⁶⁷, z.B. zur Valenz von Verben, da dies für die Sprachproduktion ebenfalls wichtig ist.²⁶⁸

In Lernerwörterbüchern wird überdies auch Kollokationen mehr Aufmerksamkeit geschenkt. Rundell (1988:135) führt dazu aus: “In most situations, word choice is quite significantly constrained [...] and for the learner who is attempting to use the dictionary as an aid to encoding, this information is absolutely vital.”²⁶⁹

Für Lerner sind Markierungen zudem von größerer Bedeutung als für Muttersprachler.²⁷⁰ Während Muttersprachler in der Regel um Wortverwendungsgrenzen intuitiv wissen, müssen Lerner diese eigens lernen. Markierungen können den Lerner vor situationsunangemessener Sprachverwendung warnen (und somit beispielsweise vor peinlichen Fauxpas bewahren). Vor allem diatopische Markierungen sind jedoch auch für Muttersprachler von großem Interesse, denn oft ist diesen nicht bewußt, daß z.B. eine Wortform oder Bedeutung auf die von ihnen gesprochene Varietät beschränkt ist, und man in einer anderen Varietät eine andere Wortform oder Bedeutung bevorzugt. Niebaum (1989:662) äußert daher zurecht: „Einschlägige Verständnisschwierigkeiten oder zumindest Einschätzungsunsicherheiten ergeben sich hier nicht nur für den Ausländer, sondern nicht selten auch für den Muttersprachler.“ So hat z.B. das Substantiv *tramp* neben der gemeinenglischen Bedeutung „an itinerant homeless person who lives by begging or doing casual work“ (COD₁₀) auch die vor allem AmE Bedeutung „a promiscuous woman“ (COD₁₀), die mancher Muttersprachler des BrE vielleicht nicht kennt, wodurch im Gespräch mit Sprechern des AmE eventuell Mißverständnisse auftreten könnten. Für BrE Muttersprachler sind Markierungen von Amerikanismen sowie Querverweise zu AmE Synonymen sicherlich für die transatlantische Kommunikation sehr hilfreich (ebenso wie die

²⁶⁵ Ausnahmen betreffen z.B. Komposita oder *phrasal verbs*, deren Aussprache man anhand der Wörter, aus denen sie sich zusammensetzen, erschließen kann.

²⁶⁶ Die Herausgeber setzen voraus, daß Muttersprachler Alltagswörter richtig aussprechen können.

²⁶⁷ Vgl. auch Rundell (1998:317), Heuberger (2000:60), Herbst (1996:328f.) und Whitcut (1985:76).

²⁶⁸ Vgl. z.B. den Eintrag zu *cop out* im OALD₆ und im COD₁₀.

²⁶⁹ Vgl. hierzu z.B. den Eintrag zu *baggage* im OALD₆ und im COD₁₀.

Kennzeichnung der auf das BrE beschränkten Wortformen, Wendungen, Bedeutungen u.ä., um den regional begrenzten Geltungsraum dieser Britizismen aufzuzeigen). Die Verlage der hier untersuchten Wörterbücher der *Concise*-Sparte scheinen nicht immer dieser Ansicht zu sein, denn nach Auffassung einer Lexikographin von *Chambers Harrap* beispielsweise ist es in Wörterbüchern für Muttersprachler weniger wichtig, Amerikanismen zu markieren, da Muttersprachler „a feel for the language“ hätten.²⁷¹ Hier werden jedoch die angesprochenen möglichen transatlantischen Verständigungsschwierigkeiten außer acht gelassen.

Anwendungsbeispiele haben in Lernerwörterbüchern ebenfalls einen viel höheren Stellenwert als in monolingualen Wörterbüchern für Muttersprachler.²⁷² Sie tragen nicht nur zu einem besseren Verständnis des jeweiligen Stichworts bei, sondern dienen auch als Modell, an dem sich der Lerner bei der eigenen Sprachproduktion orientieren kann. In Wörterbüchern für Muttersprachler finden sich dagegen in der Regel kaum Anwendungsbeispiele; sie scheinen zudem oft willkürlich ausgewählt. Muttersprachler brauchen natürlich nicht die gleiche Hilfestellung wie Lerner, daher sind die in monolingualen Wörterbüchern für Muttersprachler aufgeführten Anwendungsbeispiele auch nicht vereinfacht (wie dies in Lernerwörterbüchern manchmal der Fall ist).

Sog. *Usage Notes* haben ebenfalls das Ziel, Lernern Unterstützung bei der Sprachproduktion bieten.²⁷³ Sie sollen dem Lerner z.B. helfen, seine Sprechintentionen angemessen zu formulieren, die richtige Wortwahl zu treffen, seinen Wortschatz zu erweitern oder grammatikalisch richtige Sätze zu bilden. Außerdem werden auch *Usage Notes* angeboten, die kulturelle Informationen enthalten. In Wörterbüchern für Muttersprachler sind zwar auch *Usage Notes* enthalten, aber in diesen geht es in erster Linie um strittigen Sprachgebrauch im Bereich der Grammatik, Rechtschreibung oder Aussprache. Es finden sich jedoch sowohl in den *Usage Notes* der Lernerwörterbücher als auch in denen der Wörterbücher für Muttersprachler Hinweise auf Divergenzen zwischen BrE und AmE Sprachgebrauch, wobei diese in Lernerwörterbüchern generell eingehender erörtert werden. *Usage Notes* bieten sich besonders dazu an, komplexere Unterschiede zwischen den beiden Varietäten zu behandeln. Diese Ansicht teilt auch Whitcut (1985:79f.), die ausführt: „Varieties of English, or at least some broad British-American distinctions, are interesting to both natives and advanced learners; and the subject can be discussed more delicately in *Usage Notes* than in the simple nationality labels [...]“

²⁷⁰ Vgl. auch Norri (1996:1), Norri (2000:75) und Hausmann (1989:649). Bei Gorbahn/ Hausmann (1989:52) heißt es sogar: „Knowing about marked forms, social varieties, is of the greatest importance for non-native users.“

²⁷¹ Diese Meinungsäußerung stammt vom August 1999.

²⁷² Vgl. auch Pätzold (1994:51), Kirkpatrick (1985:11) und Rundell (1998:316).

²⁷³ Vgl. auch Whitcut (1985:78).

Andere typische Merkmale von Lernerwörterbüchern, die die Sprachproduktion und/ oder Sprachrezeption für die Lerner erleichtern sollen, wie z.B. Illustrationen oder Appendizes, kommen in den im Rahmen dieser Arbeit untersuchten Wörterbüchern der *Concise*-Sparte nur vereinzelt vor.²⁷⁴

Eines der wichtigsten Charakteristika von Lernerwörterbüchern ist ihr Definitionsstil, der sich von dem in Wörterbüchern für Muttersprachler durch die Einfachheit und Vorkommenshäufigkeit des verwendeten Definitionswortschatzes deutlich unterscheidet.²⁷⁵ Die Verwendung eines Definitionswortschatzes kann in manchen Fällen auch dazu führen, daß Bedeutungserklärungen länger ausfallen als in einem Wörterbuch für Muttersprachler, weil der Rückgriff auf eine einfachere Wortwahl mehr Platz benötigt.²⁷⁶ Definitionen in Wörterbüchern für Muttersprachler des Englischen waren dagegen vor allem in der Vergangenheit durch einen sehr knappen Definitionsstil gekennzeichnet, der dem o.g. Ziel, möglichst viele Stichwörter in das jeweilige Wörterbuch aufzunehmen, deutlich untergeordnet war, auch wenn die Verständlichkeit der Bedeutungserklärungen dadurch erschwert wurde. Diesen Definitionsstil, der sich unter den hier untersuchten Wörterbüchern der *Concise*-Sparte am deutlichsten im CACD widerspiegelt, charakterisiert Rundell (1999:43) mit folgenden Worten: “elliptical structures, lists of (supposed) synonyms, genus words used in archaic or non-central meanings [...], abstract formulae [...] and a use of parentheses [...]”

In Bezug auf das Thema dieser Arbeit und die hier analysierten Wörterbücher kann man zusammenfassend feststellen, daß von den Wörterbüchern für Muttersprachler der *Concise*-Sparte aufgrund ihrer oben beschriebenen Konzeption ein größerer Einbezug von Amerikanismen als in den Lernerwörterbüchern zu erwarten wäre. Die Lernerwörterbücher wiederum werden, durch ihre Zielsetzung bedingt, detailliertere Informationen zu den verzeichneten Amerikanismen liefern können, da diese für die Sprachproduktion der Lerner wichtig sind. Unter den Käufern der Lernerwörterbücher sind nämlich auch viele Lerner des AmE, weshalb es sogar schließlich fraglich ist, ob die Zahl der aufgenommenen Amerikanismen in den Lernerwörterbüchern entscheidend geringer ist als in solchen der *Concise*-Sparte. Dieser Frage wird in Kapitel 7 nachgegangen.

²⁷⁴ Im CACD gibt es z.B. einen Anhang zu Rechtschreib- und Wortbildungsregeln, im CH21₂ kleine Schwarz-Weiß-Illustrationen zu den Sternzeichen.

²⁷⁵ Vgl. auch Heuberger (2000:2,5).

²⁷⁶ Vgl. z.B. die Definition des Amerikanismus *taffy* in den hier untersuchten Lernerwörterbüchern und in den Wörterbüchern der *Concise*-Sparte: In den letzteren wird in der Bedeutungserklärung das Wort *toffee* verwendet, das sich nicht im Definitionswortschatz der Lernerwörterbücher befindet und in diesen daher nicht benutzt wird, in manchen Wörterbüchern der *Concise*-Sparte aber eine kürzere Definition des Wortes ermöglicht. Vgl. auch Kirkpatrick (1985:10) und Heuberger (2000:5).

4.3 Analyse der für die Berücksichtigung des AmE wichtigen Informationen außerhalb des Lexikon-Teils

4.3.1 Allgemeine Informationen

Zunächst sollen einige allgemeine Informationen zu den benutzten Wörterbuchausgaben, die zwischen 1995 - 2001 veröffentlicht wurden, angeführt werden. Neuauflagen der untersuchten Wörterbücher, die ab 2003 erschienen, konnten in dieser Arbeit leider nur in geringem Maße oder überhaupt nicht mehr berücksichtigt werden.

a) Oxford Advanced Learner's Dictionary of Current English (OALD₆)

Die sechste Ausgabe des *Oxford Advanced Learner's Dictionary* erschien 2000 als "grundlegende Neubearbeitung"²⁷⁷. Schon der Titel des Wörterbuchs macht deutlich, daß die Adressatengruppe fortgeschrittene Englischlerner sind.²⁷⁸

b) Longman Dictionary of Contemporary English (LDOCE₃)

Die dritte Ausgabe des *Longman Dictionary of Contemporary English* (als "Völlige Neuentwicklung"²⁷⁹ bezeichnet) wurde 1995 publiziert. Als Zielgruppe des Wörterbuchs werden Englisch-Lehrende und Englisch-Lernende mit fortgeschrittenen Sprachkenntnissen genannt ("at intermediate to advanced level")²⁸⁰.

Ab Juli 2001 wurde das LDOCE₃ zusammen mit einem 64-seitigen Anhang (*New Words supplement*) verkauft, der aktuelle Neuwörter, neue Bedeutungen und Wendungen enthält.²⁸¹ Seit 2003 ist die vierte Auflage des LDOCE auf dem Markt.²⁸²

²⁷⁷ Vgl. den Rückendeckel der deutschen Lizenzausgabe. Das OALD₆ unterscheidet sich in Inhalt und Struktur tatsächlich deutlich von seinem Vorgänger, wie in dieser Arbeit dargelegt wird.

²⁷⁸ Das Lernniveau, für welches das OALD₆ geeignet sein soll, reicht gemäß der Informationen auf der Internetseite des Verlags (<http://www.oup.com/elt/global/isbn/6710/>; Stand: November 2003) von „upper-intermediate“ bis „advanced“.

²⁷⁹ Vgl. den Buchdeckel und S. iii der Lizenzausgabe der Langenscheidt-Longman GmbH, München.

²⁸⁰ Vgl. die Einleitung, S. xi. In der 2. Auflage werden "advanced students and teachers of English" (*General Introduction*) als Adressatenkreis des Wörterbuchs angegeben. Das LDOCE₃ senkt also das Niveau der erforderlichen Sprachkenntnisse etwas und zielt somit auf eine größere Lernergruppe ab.

²⁸¹ Außerdem wurden alle Stichwörter sowie Kollokationen, idiomatische Wendungen und wichtige Zusatzinformationen in Blau gedruckt und somit farblich abgesetzt, um dem Benutzer das Auffinden von Einträgen zu erleichtern. Als kostenloses Extra erhielt der Käufer zudem die zum Wörterbuch gehörende CD-ROM. Dies schien zumindest für die Lizenzausgabe der Langenscheidt-Longman GmbH, München, zuzutreffen (vgl. die Langenscheidt-Longman Internetseite http://www.langlong.de/neues/wb_lgdce_d.htm), in der schon ab Januar 2000 die LDOCE CD-ROM beigelegt wurde.

Der harte Wettbewerb auf dem Wörterbuchmarkt, gerade im ELT - Bereich, mag dafür verantwortlich gewesen sein, daß diese verkaufsfördernden Neuerungen von *Longman* angeboten wurden, bevor die 4. Auflage zur Veröffentlichung bereitstand. Verlage wie *Oxford University Press* und *Harper Collins* brachten nämlich, wie bereits erwähnt, schon 2000 bzw. 2001 Neuauflagen ihrer Konkurrenzprodukte zum LDOCE₃ heraus.

²⁸² Informationen zu den Neuerungen im LDOCE₄ erhält man beispielsweise unter <http://www.longman.com/ldoce/about.html>.

c) Longman Dictionary of English Language and Culture (LDEL_{C2})

Die zweite Auflage des *Longman Dictionary of English Language and Culture* wurde 1998 auf den Markt gebracht. Während es im Vorwort auf S. V heißt, daß dieses Wörterbuch für Lehrer und fortgeschrittene Lerner gedacht sei²⁸³, ist die Zielgruppe auf einer der Internetseiten des *Longman* Verlags weiter gefaßt: „upper intermediate - Proficiency“.²⁸⁴

d) Collins COBUILD English Dictionary for Advanced Learners (COBUILD₃)

Die dritte Auflage des *Collins COBUILD English Dictionary* (als vollständige Neubearbeitung bezeichnet²⁸⁵) wurde mit leicht verändertem Titel 2001 von *Harper Collins* veröffentlicht.²⁸⁶ Wie der neue Titel bereits erkennen läßt, ist das Wörterbuch für fortgeschrittene Lerner gedacht.²⁸⁷

Im September 2003 erschien - nach nur zwei Jahren - die vierte Auflage des Wörterbuchs unter dem Titel *Collins COBUILD Advanced Learner's English Dictionary*.

e) Cambridge International Dictionary of English (CIDE₁)

Die erste Auflage des *Cambridge International Dictionary of English* erschien 1995.²⁸⁸ Durch die auf der Buchvorderseite befindliche Aufschrift „For Advanced Learners“ gibt sich das Wörterbuch sofort als Lernerwörterbuch zu erkennen, und auf dem Rückendeckel heißt es: "It is aimed at learners and users of English as a Foreign Language from intermediate level upwards". Somit soll das CIDE₁ auch schon für Lerner geeignet sein, die man als fortgeschrittene Anfänger bezeichnen kann.

Im Jahr 2003 wurde die zweite Auflage des Wörterbuchs unter dem Titel *Cambridge Advanced Learner's Dictionary* veröffentlicht: Wie schon der *Harper Collins Verlag* (s.o.) wollte nun wohl auch *Cambridge University Press* den Lernerbezug im Titel ihres Wörterbuchs hervorheben. Zudem ist der Gebrauch der Formulierung *Advanced Learner's* möglicherweise

²⁸³ Auch für die erste Auflage wurde diese Angabe gemacht.

²⁸⁴ Vgl. die Informationen unter http://www.longman.com/dictionaries/which_dict/ldelc.html (Stand Februar 2004). Dies hat höchstwahrscheinlich etwas mit verkaufsstrategischen Überlegungen zu tun.

²⁸⁵ Vgl. die Informationen des *Klett* Verlags auf der folgenden Webseite: <http://www.pons.de/isbn/517905-X/> (Stand Mai 2003).

²⁸⁶ Jede der bisher erschienenen Auflagen des Wörterbuchs trägt einen etwas anderen Titel als sein Vorgänger, so hieß die erste Auflage z.B. *Collins COBUILD English Language Dictionary*, die zweite dagegen *Collins COBUILD English Dictionary* (s.o.). Damit wollen die Herausgeber vermutlich verdeutlichen, daß es sich tatsächlich jeweils um eine völlig überarbeitete Ausgabe des Wörterbuchs handelt.

²⁸⁷ Auf der im Internet befindlichen Übersichtsseite des Verlags zu seinen Wörterbüchern erhält man allerdings die Information, daß das COBUILD₃ auch schon für die Sprachstufe geeignet sei, die im Englischen mit „intermediate“ beschrieben wird.

²⁸⁸ Dies war gleichzeitig das erste von dem Verlag *Cambridge University Press* herausgegebene monolinguale Lernerwörterbuch (vgl. das Vorwort von CIDE₁).

auch darauf zurückzuführen, daß das *Oxford Advanced Learner's Dictionary* viele Jahre das einzige Wörterbuch dieser Sparte und sein Titel im Bereich des ELT überall bekannt war. Diese Bekanntheit des Titels wollten sich *Harper Collins* und *Cambridge University Press* wahrscheinlich für ihre Produkte ebenfalls zunutze machen. Im übrigen soll die Titeländerung des CIDE vermutlich auch unterstreichen, daß es sich bei der zweiten Auflage um eine „vollständige Neubearbeitung“²⁸⁹ handelt. Allerdings wird nun der internationale Bezug nicht mehr so direkt ersichtlich wie in der ersten Auflage (*Cambridge International English Dictionary*), der den Herausgebern in bezug auf das CIDE₁ sehr wichtig erschien.²⁹⁰

Das Sprachniveau, auf das die zweite Auflage abzielt, liegt etwas höher als bei CIDE₁, nämlich bei „advanced to proficiency“²⁹¹.

f) Cassell Concise Dictionary (CACD)

Eine neue Auflage des *Cassell Concise Dictionary* erschien 1998, ein Jahr, nachdem eine grundlegende Neubearbeitung des Wörterbuchs stattgefunden hatte. Das CACD wurde nach Informationen des Verlags erstmals 1891 veröffentlicht, allerdings unter dem Namen *Cassell's English Dictionary*. 1989 erschien das Wörterbuch dann als *Cassell's Concise English Dictionary* und erst ab 1997 unter dem heutigen Titel *Cassell Concise Dictionary*. Der Verlag konnte der Verfasserin daher nicht sagen, als welche Auflage die 1998er Ausgabe des Wörterbuchs zu gelten hat. Nach Ansicht von Richard Milbank, Redakteur bei *Cassell*, könnte man sie sogar als 23. Auflage bezeichnen, wenn man das Jahr 1891 als Ausgangsdatum nimmt. Der Auskunft von R. Milbank zufolge ist das CACD das älteste Wörterbuch der *Concise*-Sparte, älter noch als das *Concise Oxford Dictionary*.²⁹² Da aber unklar ist, wieviel das Wörterbuch von 1891 mit dem von 1998 gemein hat, wird im folgenden auf eine Angabe zur Auflage verzichtet.

Mittlerweile wird das Wörterbuch nicht mehr aufgelegt. Die Anfrage der Verfasserin dieser Arbeit nach dem Grund hierfür beantwortete der Verlag leider nicht.²⁹³ Vielleicht hat sich das Wörterbuch aufgrund der großen Konkurrenz nicht gut genug verkauft. Das einzige Standardwörterbuch des Englischen, das sich noch im Verlagsprogramm befindet, ist das *Cassell's English Dictionary*, das früher unter dem Titel *Cassell Compact Dictionary* verkauft wurde.

²⁸⁹ Vgl. den Rückumschlag der Lizenzausgabe des Klettverlags und die Internetseite <http://www.pons.de/produkte/katalog/suche/suche.php?tmp=suche&isbn=3-12-517993-9> (Stand: Februar 2004).

²⁹⁰ Vgl. Procter (1996:4).

²⁹¹ Vgl. die Internetseite <http://publishing.cambridge.org/ge/elt/dictionaries/158668/>, Stand Februar 2004.

²⁹² E-Mail vom 23.7.2004.

²⁹³ Ebd.

Nach Angaben von Richard Milbank, Redakteur bei *Cassell*, wurde das CACD als „general desk reference [book] and family purchase“ konzipiert.²⁹⁴

g) Concise Oxford Dictionary (COD₁₀)

Die zehnte Auflage des *Concise Oxford Dictionary* erschien im Juni 1999. Laut Herausgeber ist das Wörterbuch für den Gebrauch beim Lernen, bei der Arbeit und zu Hause vorgesehen.²⁹⁵

Im Juli 2001 wurde eine sog. „revised tenth edition“ des COD veröffentlicht. Die Änderungen, die diese Auflage gegenüber der Ausgabe von 1999 aufweist, bestehen in der Aufnahme von 200 neuen Einträgen und in einem Anhangsteil, in dem sich z.B. Kapitel zu den Ländern der Erde, verschiedenen Alphabeten, Gewichten und Maßen sowie andere wissenswerte Informationen befinden²⁹⁶ (u.a. wird auch ein „Guide to good English“ angeboten). Bei der „Überarbeitung“ des COD₁₀ handelt es sich daher eigentlich um eine Ergänzung der bisherigen Auflage und nicht um eine inhaltliche Veränderung bereits existierender Wörterbucheinträge. Eine kurze Stichprobe bestätigte diesen Sachverhalt. Die Verfasserin dieser Arbeit verglich den Text der Einträge für fünf Amerikanismen (*baggage, unlisted, junior college, package* und *mailman*) in der Ausgabe von 1999 mit dem der Auflage von 2001 und stellte fest, daß er bis ins kleinste Detail identisch war. Insofern kann man die Etikettierung „completely revised“ auf der Buchvorderseite der Auflage von 2001 als irreführend bezeichnen.²⁹⁷ Für die im Rahmen dieser Arbeit durchgeführte Untersuchung bedeutet dieser Umstand, daß bei den Analysen der Wörterbucheinträge des COD₁₀ noch auf die Auflage von 1999 zurückgegriffen wird, da sich unter den 200 neuen Einträgen der Ausgabe von 2001 keiner der in der Dissertation behandelten Amerikanismen befindet.

Ab Mai 2002 wurde die „revised tenth edition“ unter dem Titel *Concise Oxford English Dictionary* verkauft, um daran zu erinnern, daß von *Oxford University Press* ebenfalls das renommierte *Oxford English Dictionary* herausgegeben wird, das bei der Erstellung des COD₁₀ herangezogen werden konnte.²⁹⁸ Im Juli 2004 erschien die elfte Auflage des Wörterbuchs.

h) Collins Concise Dictionary (CCD₅)

²⁹⁴ Ebd.

²⁹⁵ Vgl. die Internetseite <http://www.oup.co.uk/isbn/0-19-860572-2?view=ask> (Stand: März 2004).

²⁹⁶ Informationen dieser Art waren teilweise auch schon in den älteren Auflagen des COD vorhanden.

²⁹⁷ Auch die Titellei der Ausgabe von 2001 ist - abgesehen von einem zusätzlichen kurzen Vorwort zu dieser „überarbeiteten“ Auflage - mit der von 1999 deckungsgleich.

²⁹⁸ Vgl. die Internetseite <http://www.oup.co.uk/isbn/0-19-860572-2?view=ask> (Stand: März 2004) und auch Görlach (1990c:312).

Die fünfte Auflage des *Collins Concise Dictionary* wurde 2001 veröffentlicht, nur zwei Jahre nach dem Erscheinen der vierten Auflage. Nach Ansicht der Herausgeber ist das Wörterbuch ein Nachschlagewerk „for home, office, and study“.²⁹⁹

i) Chambers 21st Century Dictionary (CH21₂)

Eine überarbeitete Auflage des *Chambers 21st Century Dictionary* wurde 1999 veröffentlicht. Das Wörterbuch war 1996 neu auf den Markt gekommen und sollte ursprünglich das *Chambers Concise Dictionary* ersetzen, das 1991 in der dritten Auflage erschienen war.³⁰⁰ Das *Chambers Concise Dictionary* verkaufte sich aber auch nach der Veröffentlichung des CH21₂ noch so gut, daß es aus der Produktpalette des Verlags nicht herausgenommen wurde. Im Juni 2004 wurde die vierte Auflage des *Chambers Concise Dictionary* veröffentlicht, dessen Wörterbuchtext laut Angaben einer leitenden Redakteurin bei *Chambers Harrap* auf dem CH21₂ basiert, wobei der Wörterbuchtext jedoch überarbeitet und auf den neusten Stand gebracht wurde. Diese vierte Auflage des *Chambers Concise Dictionary*, die ihrerseits nun das CH21₂ ersetzen soll, konnte leider nicht mehr in die vorliegende Arbeit einbezogen werden.

Den Aussagen der Herausgeber zufolge eignet sich das CH21₂ sehr gut als Informationsquelle „for family and general reference“ (Vorwort, S. vi).

j) The New Penguin English Dictionary (NPED₁)

Das Wörterbuch *The New Penguin English Dictionary*, bei dem es sich um eine Neuerscheinung handelt, wurde im Jahr 2000 veröffentlicht.

Seit Juli 2003 ist die zweite Auflage unter dem Titel *The Penguin English Dictionary* auf dem Markt. Sie konnte leider nicht mehr in dieser Arbeit berücksichtigt werden.

Die unterschiedlichen Erscheinungsdaten der hier untersuchten Wörterbücher könnten zur Folge haben, daß neuere Ausgaben aufgrund des Einbezugs aktuellerer Sprachentwicklungen gegenüber älteren im Rahmen der in dieser Arbeit durchgeführten Analysen im Vorteil sind. So stammt das CIDE₁ aus dem Jahr 1995 und ist damit die älteste hier berücksichtigte Auflage, während das COBUILD₃ und CCD₅ erst 2001 veröffentlicht wurden. Auf die Frage, ob die unterschiedlichen Erscheinungsdaten Auswirkungen auf die Untersuchungsergebnisse haben, soll in Kapitel 7 eingegangen werden.

²⁹⁹ Vgl. den Verlagskatalog von 2001.

³⁰⁰ Diese (und andere) Informationen erhielt die Verfasserin der vorliegenden Arbeit während ihrer Übersetzungs- und redaktionellen Tätigkeit für *Chambers Harrap* in den Jahren 1999 -2000 sowie in einer E-Mail einer leitenden Redakteurin von *Chambers Harrap* vom 10.6.2004.

4.3.2 Beteiligung amerikanischer Lexikographen

Wenn man die zehn Wörterbücher in Hinblick auf den Einbezug amerikanischer Lexikographen oder Redakteure miteinander vergleicht, fällt auf, daß an der Erstellung der Lernerwörterbücher insgesamt deutlich mehr Experten für das AmE beteiligt waren als es bei den Wörterbüchern der *Concise*-Sparte der Fall gewesen ist. So waren in die Arbeit am CCD₅ beispielweise überhaupt keine amerikanischen Lexikographen oder Berater eingebunden. Für Mitarbeiter von *Chambers Harrap* bestand laut Angaben des Verlags die Möglichkeit, bei Fragen zum AmE den Rat eines *American Consultant* namens John Wright einzuholen, es ist aber offen, inwieweit dies bei der Erstellung des CH21₂ tatsächlich genutzt wurde. Im NPED₁ wird ebenfalls nur auf einen Berater Bezug genommen, und zwar auf Malcolm Bradbury, *Professor of American Studies, University of East Anglia* (vgl. S. iv). Es ist allerdings zu vermuten, daß dieser Amerika-Experte mehr Informationen landeskundlicher, geographischer oder historischer Art gegeben hat als sprachliche Auskünfte. Der im COD₁₀ erwähnte Berater bzgl. der Behandlung des AmE könnte dagegen eine größere Rolle für die Erfassung und Markierung von Amerikanismen in diesem Wörterbuch gespielt haben. Zumindest im COD₉ wird nämlich die Bedeutung der „consultants for North American usage“ (S. vii) für den Einbezug des AmE besonders hervorgehoben: „Another important feature of this edition has been the enlisting of special consultants for North American usage, which has enabled us to improve our coverage of this area and to apply geographical labels more accurately“³⁰¹ (Vorwort, S. vii). Was das CACD betrifft, so war nach Informationen von Richard Milbank, Mitarbeiter bei Cassell, ein amerikanischer Redakteur beauftragt worden, den Einbezug des AmE im Wörterbuch zu überprüfen³⁰².

Unter den Lernerwörterbüchern weist das COBUILD₃ den geringsten Einbezug von Experten für das AmE auf, denn bei diesen handelt es sich nur um eine Phonetikerin und einen „US Consultant“ für das AmE. Da das Redaktionsteam des COBUILD₃ angibt, in die neue Auflage u.a. mit Hilfe von Experten eine deutlich größere Menge AmE Sprachmaterials aufgenommen zu haben³⁰³, hätte man erwartet, daß mehr amerikanische Linguisten oder Lexikographen im COBUILD₃ als im COBUILD₂ angeführt werden, was nicht der Fall ist. Allerdings scheint der *US Consultant* John Wright für die Erstellung des COBUILD₃ eine wichtigere Rolle gespielt zu haben als die in den Danksagungen des COBUILD₂ angegebenen fünf Personen, die hinsichtlich der Darstellung des AmE im COBUILD₂ Rat und Unterstützung gegeben haben, denn er wird als einer der Mitarbeiter am Wörterbuch aufgelistet. In die Erstellung des CIDE₁ waren immerhin ein amerikanischer Redakteur, ein Phonetiker des AmE

³⁰¹ Außer einem Berater für das AmE wird auf S. v auch ein *special consultant* für das Kanadische Englisch angegeben. Dieser wird auch im COD₁₀ genannt.

³⁰² E-Mail vom 23.7.2004.

und zwei Experten für US-bezogene Sachthemen involviert. Den Angaben Paul Procters (1996:4) zufolge wurde jeder Eintrag im CIDE₁ von dem o.g. amerikanischen Verlagslektor dahingehend überprüft, ob der Inhalt aus der Sicht eines Muttersprachlers des amerikanischen Englischs korrekt sei.³⁰⁴ Die meisten amerikanischen Lexikographen und Redakteure waren jedoch an der Erstellung des LDOCE₃ beteiligt, nämlich insgesamt vier. Der Einbezug dieser Experten für das AmE ist laut Aussage der Redaktion auch mit dafür verantwortlich, daß im LDOCE₃ in großem Maße sowohl das BrE als auch das AmE erfaßt ist (S. xxi). Leider geht aus den Informationen des Anhangs nicht hervor, inwieweit amerikanische Lexikographen an der Erstellung des 2001 hinzugefügten *New Words supplement* mitgearbeitet haben. Unklar ist auch, wieviele amerikanische Lexikographen bei der zweiten Auflage des LDELIC mitgewirkt haben, für die erste Auflage waren es jedenfalls genauso viele wie im Falle des LDOCE₃, nämlich vier. Doch auch das OALD₆ steht mit drei Redakteuren für das AmE nicht weit dahinter zurück. Erfreulich ist, daß bei der Erstellung des OALD₆ und LDOCE₃ mehr Experten für das AmE direkt involviert waren als in den vorhergehenden Auflagen, da dies für einen größeren Stellenwert des AmE in den neueren Auflagen spricht. Im Vorwort zum OALD₅ wird nämlich nur auf eine Dame Bezug genommen, die für die Erfassung des AmE im Wörterbuch zuständig war. Was das LDOCE₂ betrifft, kann man den Danksagungen zu Beginn des Wörterbuchs entnehmen, daß die Redaktion zwar Berater für die Darstellung des AmE in der 2. Auflage hinzugezogen hatte, aber keine amerikanischen Lexikographen direkt an der Arbeit am Wörterbuch beteiligt waren.³⁰⁵

4.3.3 Lemmata

4.3.3.1 Vergleich der Größe der untersuchten Wörterbücher

Bei dem Versuch, den Umfang verschiedener Wörterbücher zu vergleichen, stößt man auf eine Reihe von Problemen. So stellt sich zunächst die Frage, wie man den Umfang eines Wörterbuchs faßt. Ilson (1988:9) bemerkt diesbezüglich folgendes: „the size of a dictionary is a function of two variables: the number of items entered (its „macrostructure“), and the amount of information given about them (its „microstructure“). In Hinblick auf das Thema dieser Arbeit läßt sich festhalten, daß nicht nur die breite Aufnahme von Amerikanismen in ein Wörterbuch, sondern auch ihre eingehende Behandlung (d.h. Informationen zur Bedeutung, zum Gebrauch, zu

³⁰³ Vgl. S. xi. Hierauf wird im weiteren noch eingegangen.

³⁰⁴ Für das AusE wurde diese Arbeit von einem australischen Verlagslektor durchgeführt.

³⁰⁵ Dagegen wird im LDOCE₂ nicht nur - wie im LDOCE₃ - erwähnt, daß zahlreiche Lehrer und Lerner des Englischen mit nützlichen Ratschlägen bei der Wörterbucharbeit behilflich gewesen seien, sondern auch, daß die Lexikographen des LDOCE₂ eng mit amerikanischen und britischen Linguisten zusammengearbeitet hätten, vgl. F7.

grammatikalischen Besonderheiten u.a.) vor allem für Lerner, aber auch für Muttersprachler des BrE, von großem Interesse sein muß. Die Verlage der hier untersuchten Wörterbücher beziehen sich jedoch in ihren Größenangaben auf die Anzahl der in dem jeweiligen Wörterbuch enthaltenen *references*, *words and phrases*, *definitions*, *meanings* und/ oder *entries*. Die Herausgeber der deutschen Lizenzausgaben machen z.B. Aussagen zur Menge der in die Wörterbücher einbezogenen „Wörter und Wendungen“ oder gehen, wie im Falle des OALD₆, auf die Anzahl der Definitionen und Anwendungsbeispiele ein. Unterschiedliche Meßmethoden sowie die Vagheit der meisten der o.g. Begriffe machen einen Vergleich zwischen den verschiedenen Wörterbüchern sehr schwierig.³⁰⁶ Da den Verlagen sehr viel daran liegt, Konkurrenzprodukte nicht zuletzt durch die Größe des eigenen Wörterbuchs zu übertrumpfen, scheinen die jeweiligen Herausgeber nicht sehr darauf bedacht zu sein, ihr Zählverfahren präzise offen zu legen. Erschwerend kommt hinzu, daß die Makrostruktur der untersuchten Wörterbücher sehr unterschiedlich ist: Folgt man Svensén (1993:35) und definiert *entry* als „words which begin a dictionary entry and are printed in bold or semi-bold type“, so würde z.B. ein Wörterbuch wie das COBUILD₃ bei einem Zugrundelegen dieses Zählkriteriums sehr schlecht abschneiden, denn seine Makrostruktur wurde nach dem Prinzip "one word, one entry" gestaltet³⁰⁷. Aufgrund der dargestellten Schwierigkeiten, die Größe verschiedener Wörterbücher zu vergleichen, ist man in der Forschungsliteratur dazu übergegangen, nach alternativen Meßmethoden zu suchen.³⁰⁸ Heuberger (2000:115f.) hat beispielsweise in mehreren Wörterbüchern die Zahl der Definitionen zwischen bestimmten Stichwörtern, die nach dem Zufallsverfahren ausgesucht wurden, ermittelt. Aber auch dieses Verfahren ist keineswegs unproblematisch, denn eine Definition kann z.B. lediglich aus der Angabe eines Synonyms bestehen oder 30 oder mehr Wörter enthalten. Zu klären ist auch, wie man mit Worterläuterungen innerhalb von Beispielen umgeht, die beispielsweise im CIDE₁ häufig vorkommen.³⁰⁹

Wie bereits in Kap. 4.2 dargelegt, kann man trotz der geschilderten Problematik bei dem Versuch, den Umfang verschiedener Wörterbücher zu vergleichen, davon ausgehen, daß die

³⁰⁶ So bezieht sich der Verlag *Chambers Harrap* beispielsweise bei dem Gebrauch des Wortes *references* auf „definitions plus undefined run-on words“ (Mary O’Neills E-Mail vom 10.06.2004), während *Harper Collins* Stichwörter, Flexionsformen, Wendungen und Querverweise in Fettdruck darunter faßt (Cormac McKeowns E-Mail vom 23.06.2004). Zu den Schwierigkeiten bei dem Versuch, den Umfang verschiedener Wörterbücher zu vergleichen, äußern sich auch Norri (1996:3) und Whitcut (1989:89). Whitcut (1989:89) stellt u.a. diesbezüglich fest: „no outside body has yet managed to set up an objective standard“.

³⁰⁷ Vgl. S. xii der Einleitung des COBUILD₃.

³⁰⁸ Vgl. z.B. Piotrowski (1997:292).

³⁰⁹ Siehe auch Bogaards (1996:281f.), der sich über weitere Probleme dieser Methode äußert. Dennoch soll das Ergebnis von Heuberger (ebd.) Untersuchung, die sich u.a. auf CIDE₁, COBUILD₂, OALD₅ und LDOCE₃ (ohne Anhang) bezieht, und mit dem Resultat einer ähnlichen Zählung von Bogaards (1996:282) übereinstimmt,

Wörterbücher der *Concise*-Sparte über eine deutlich höhere Zahl von Stichwörtern verfügen als die Lernerwörterbücher. Dies liegt daran, daß sie auf viele der in den Lernerwörterbüchern bereitgestellten Informationen ganz oder weitestgehend verzichten (s.o. Kap. 4.2). Im CACD sind nach Auffassung der Verfasserin dieser Arbeit wahrscheinlich die meisten Stichwörter behandelt worden, da das CACD im Gegensatz zu den anderen hier untersuchten Wörterbüchern der *Concise*-Sparte noch sehr stark dem nestalphabetischen Prinzip folgt, d.h. etymologisch zusammengehörende Stichwörter in einem „Nest“ aneinanderreihet, wodurch Platz gespart wird und besonders viele Stichwörter insgesamt aufgenommen werden können. Ob dies gleichzeitig bedeutet, daß das CACD auch mehr Amerikanismen enthält als die anderen hier analysierten Wörterbücher, wird sich anhand der im Rahmen dieser Arbeit durchgeführten - wenn auch begrenzten - Untersuchung zeigen. Den Herausgebern des CACD war der Umfang des Wörterbuchs zumindest sehr wichtig, denn auf der Vorderseite des Schutzumschlags befindet sich als allererstes die Information, daß das Wörterbuch mehr Wörter und Wendungen enthalte als jedes andere „Concise Dictionary“ (zur Zeit der Veröffentlichung des CACD). Diese Auskunft wird auch an anderer Stelle auf dem Schutzumschlag gegeben und in den Verlagsbroschüren zum CACD mehrfach betont.³¹⁰

Auskünfte der Verlage zum Umfang ihrer Wörterbücher:

OALD ₆	„80,000 references“ ³¹¹ „160.000 Definitionen und Anwendungsbeispiele“ ³¹²
LDOCE ₃	„über 82.000 Stichwörter und Wendungen“ ³¹³
LDEL ₂	„80,000 words and phrases and 15,000 cultural and encyclopedic entries“ ³¹⁴
COBUILD ₃	„over 110,000 references“ ³¹⁵
CIDE ₁	„100.000 Stichwörter und Wendungen“ ³¹⁶

kurz wiedergegeben werden: Unter diesen Wörterbüchern in der jeweiligen Auflage von 1995 schneidet das LDOCE₃ am besten ab, verfügt also über die größte Anzahl von Definitionen.

³¹⁰ Im Verlagskatalog von 1998 heißt es auch, das CACD enthalte mehr Definitionen als jedes andere Wörterbuch der *Concise*-Sparte. Im Verlagskatalog von 1997 versuchen die Herausgeber sogar, ihre Behauptungen bezüglich des Wörterbuchumfangs anhand ausgewählter Einträge zu belegen, indem sie die Vorzüge des CACD vor allem gegenüber dem COD herausstellen.

³¹¹ Vgl. die Auskünfte auf der Webseite des Verlags <http://www.oup.com/elt/global/isbn/6710/> (Stand: November 2003).

³¹² Vgl. den Rückendeckel der dt. Lizenzausgabe.

³¹³ Vgl. den Rückendeckel der dt. Lizenzausgabe von 2001. In dieser Zahl sind die über 1.500 Neologismen des neuen Anhangs wohl enthalten. Bei der Erstveröffentlichung des LDOCE₃ im Jahre 1995 hieß es dagegen, etwas vage, daß das Wörterbuch „über 80.000 Stichwörter und Wendungen“ enthalte.

³¹⁴ Vgl. die Internetseite http://www.longman.com/dictionaries/which_dict/ldelc.html (Stand Februar 2004).

³¹⁵ Vgl. die Informationen auf dem Umschlag des Wörterbuchs.

CACD	„well over 250,000 entries and definitions“ ³¹⁷
COD ₁₀	„240,000 words, phrases, and definitions“ „240,000 meanings“ „240,000 entries and definitions“ ³¹⁸
CCD ₅	„125,000 references“ und „16,000 encyclopaedic entries“ ³¹⁹
CH21 ₂	„140,000 definitions“ ³²⁰
NPED ₁	keine Angaben erhältlich

4.3.3.2 Auskünfte der Verlage zum Einbezug des AmE und anderer Varietäten des Englischen

Vor allem die Herausgeber des OALD₆, LDOCE₃, COBUILD₃ und LDEL₂ unterstreichen, daß in ihren Wörterbüchern neben dem BrE auch das AmE umfassend dargestellt werde. So wird auf dem Rückendeckel der deutschen Lizenzausgabe des OALD₆ die „besondere Berücksichtigung auch des amerikanischen Englisch“ als Neuerung im OALD₆ hervorgehoben. Daraus läßt sich schließen, daß in der Konzeption der vorgehenden Auflage dem BrE noch Priorität gegenüber dem AmE eingeräumt wurde und der Einbezug des AmE im OALD₅ daher nicht so umfassend wie im OALD₆ ist. Dementsprechend heißt es kurz nach der Veröffentlichung der neuen Auflage auf den Internetseiten des Verlags: „There is thorough coverage of American English words and expressions. New usage notes help learners to become more aware of the differences between British and American English.“³²¹ Im November 2003 wurde auf den überarbeiteten Webseiten des Verlags zum OALD₆ zudem darauf hingewiesen, daß das in diese Auflage eingegangene BrE und AmE aktuell sei und sowohl schriftliches als auch gesprochenes Sprachmaterial umfasse.³²² Des weiteren wurde nach Auskunft der Herausgeber auch die Darstellung der Aussprache des AmE im OALD₆ verbessert.³²³

Zum Umfang des Einbezugs von Amerikanismen in das LDOCE₃ (in der Fassung von 1995) erhält man, wie bereits erwähnt, die Information, daß in großem Maße sowohl das BrE als

³¹⁶ Vgl. den Rückendeckel der dt. Lizenzausgabe.

³¹⁷ Vgl. die Informationen auf dem Umschlag des Wörterbuchs.

³¹⁸ Die Informationen zum Umfang des Wörterbuchs sind leider nicht eindeutig. Die erste Angabe befindet sich auf dem Rückumschlag des Wörterbuchs, die zwei anderen wurden im Internet veröffentlicht, und zwar im Juni 2000 (vgl. die Internetseite <http://www.oup.co.uk/dictionaries/english/list/>) und im Oktober 2001 (vgl. die Internetseite <http://www.oup.co.uk/isbn/0-19-860470-X?view=ask>).

³¹⁹ E-Mail von Cheryl Bellingham, *Admin. Assistant* bei *Harper Collins Dictionaries*, vom 16.06.2004.

³²⁰ E-Mail von Mary O’Neill, leitende Redakteurin bei *Chambers Harrap*, vom 10.06.2004.

³²¹ Vgl. <http://www1.oup.co.uk/OxfordAdvancedLearner’sDictionarySixthEdition> (Stand: Juni 2000).

³²² Vgl. <http://www.oup.com/elt/global/isbn/6710/>. Vgl. auch Kap. 3.2.3 zu den für das OALD₆ herangezogenen Belegquellen.

³²³ Vgl. das Vorwort, S. vi.

auch das AmE erfaßt sei.³²⁴ Besonders interessant ist jedoch die dem Anhang von 2001 vorangestellte Einleitung, in der u.a. auch auf die Bedeutung des AmE und seinen Einfluß auf andere Varietäten des Englischen eingegangen wird. So heißt es z.B. auf S. 1: „A strong trend is for words from American English to spread to other varieties of English“. Und auf S. 3 wird diesbezüglich noch weiter ausgeführt:

The importance of American English is growing all the time [...]. Words that are not new in the US, but which have become widely accepted and used throughout the world, are included as new words in the New Words supplement to the *Longman Dictionary of Contemporary English*. [...] Of course, words that are new in American English are also included, as the *Longman Dictionary of Contemporary English* includes wide coverage of American and other varieties of English, as well as British English. [...] The adoption of words from American English, coming particularly from the worlds of business, youth and black culture, pop music, computers, and the Internet, is a significant trend in British English, which is reflected in the New Words supplement.

Im Vorwort zu der 1995er Ausgabe des LDOCE wurde noch nicht so ausführlich auf die große Bedeutung, die dem AmE im Vergleich zu anderen Varietäten des Englischen zukommt, eingegangen. Das könnte zum einen daran liegen, daß sich der Einfluß des AmE auf das BrE (und andere Varietäten) in den letzten Jahren noch einmal merklich erhöht hat³²⁵ oder daß sich die Wörterbuchmacher der Tatsache bewußt sind, daß der Verkaufswert des LDOCE₃ auf dem ELT - Wörterbuchmarkt einen größeren Einbezug des AmE für sinnvoll erscheinen läßt, da beispielsweise ELT-Lerner in Asien stark am AmE orientiert sind (s. unten).

Der Einleitung zur *New Words Section* zufolge finden sich unter den verzeichneten neuen Wörtern also auch viele, die im AmE schon längere Zeit verwendet wurden, jetzt aber mittlerweile auch im Gemeinenglischen gebräuchlich sind (und daher nicht mehr als Amerikanismen gekennzeichnet werden, vgl. S. 3)³²⁶. Dem Leser stellt sich die Frage, warum diese Wörter, auch wenn sie ehemals noch Amerikanismen waren, nicht schon 1995 in das LDOCE₃ aufgenommen wurden. Denn in der Einleitung zum Neologismen - Anhang wird ebenfalls hervorgehoben, daß sich das Wörterbuch durch eine breite Aufnahme sowohl des AmE und anderer Varietäten des Englischen als auch des BrE auszeichnet (ebd.). Zudem waren laut Herausgeber viele dieser ehemaligen Amerikanismen im AmE sehr geläufig (vgl. S. 1), können also daher nicht aus dem Grunde vernachlässigt worden sein, daß ihr Gebrauch nur schwer nachzuweisen war.

³²⁴ Vgl. Kap. 4.3.2.

³²⁵ Dies wird vielleicht in der schon oben zitierten Aussage angedeutet: „The importance of American English is growing all the time [...]“ (*New Words Supplement*, S. 3), die an dieser Stelle auch begründet wird.

³²⁶ Diese Wörter sind daher eigentlich nur in Bezug auf andere Varietäten des Englischen außer dem AmE als „neue Wörter“ zu bezeichnen.

Die Herausgeber des COBUILD₃ äußern sich folgendermaßen zur Berücksichtigung des AmE in diesem Wörterbuch:

For this edition, the coverage of American English has been greatly extended, and with the advice of experts together with evidence from our extensive US corpora, we have made a large number of additions that feature meanings and usages that are characteristic of American English, to ensure a more comprehensive coverage of that variety. (S.xi)³²⁷

Indirekt wird hier zum Ausdruck gebracht, daß die Behandlung des AmE im COBUILD₂ noch nicht sehr umfassend war. Wenn man die Aussagen der Herausgeber des COBUILD₂ zum Einbezug des AmE betrachtet, wird deutlich, daß das AmE bei der Erstellung dieser Auflage noch einen wesentlich geringeren Stellenwert hatte. So heißt es auf S. xx: „In general, this dictionary focuses on British English, but when there is sufficient evidence in the Bank of English, differences in British and American usage are indicated.“³²⁸

Daher findet sich auf dem Rückendeckel der *Klett*-Lizenzausgabe des COBUILD₂ auch nur die Aussage, daß das AmE berücksichtigt wurde, während auf dem Rückenumschlag des COBUILD₃ damit geworben wird, daß das AmE umfassend in das Wörterbuch einbezogen worden sei, alle Angaben zum AmE aktualisiert worden seien und die Unterschiede zwischen dem BrE und dem AmE in den Einträgen deutlich hervorgehoben würden.³²⁹ Das Augenmerk der Herausgeber des COBUILD₃ liegt ihren Aussagen zufolge gleichermaßen auf dem BrE und dem AmE: „This dictionary focuses on both British and American English using evidence from the Bank of English.“ (S. xx). Begründet wird die vorrangige Behandlung des BrE und AmE gegenüber anderen Varietäten damit, daß die meisten Bücher, Zeitungen, Radio- und Fernsehprogramme sowie die Mehrheit der für den internationalen Gebrauch erstellten Lehrmaterialien in Großbritannien und den USA produziert würden. Für *HarperCollins* standen bei der Entscheidung, das AmE im COBUILD₃ aufzuwerten, laut persönlicher Mitteilung des Mitarbeiters Simon Bell (Oktober 2000) eigentlich aber vor allem ökonomische Interessen im Vordergrund. Der asiatische Markt, auf dem das Wörterbuch verkauft werde, sei nämlich auf das AmE hin orientiert.

Die Herausgeber des LDEL₂ heben wie bereits schon in bezug auf das LDEL₁ hervor, daß das Wörterbuch das BrE und AmE gleichermaßen abdecke.³³⁰ Auf dem Rückendeckel des LDEL₂ wird das Wörterbuch sogar mit folgenden Worten angepriesen: „full coverage of

³²⁷ Vgl. auch S. xiii der Einleitung.

³²⁸ Vgl. auch S. ix.

³²⁹ Vgl. auch die Informationen auf der Webseite des *HarperCollins*-Verlags zum COBUILD₃ (<http://titania.cobuild.collins.co.uk/catalogue/cob3flash.html>, Stand: Mai 2003). Dort wird die neue Auflage mit folgenden Worten angepriesen: „extensive coverage of American English“, „completely revised to give comprehensive and consistent coverage of American English usage“.

³³⁰ Vgl. die oben genannte Internetseite.

British and American English“³³¹. In der neuen Auflage wird zudem betont, daß man sich sehr darum bemüht habe, die amerikanische Kultur im gleichen Maße wie die britische zu berücksichtigen³³²:

Every attempt has been made to ensure that American culture as well as British culture has been covered, not only by using our extensive written and spoken American corpora that make up part of the Longman Corpus Network, but also by having American lexicographers working on the book. (s. v)

Die Untersuchungen von Kokawa/ Yamada (1998:344ff.) belegen, daß bereits im LDEL_{C1} nicht nur in Bezug auf die sog. *Features*, *Cultural Notes* oder Illustrationen, sondern auch hinsichtlich der angegebenen biographischen und geographischen Namen sowie Institutionen eine große Ausgewogenheit zwischen Einträgen zu Großbritannien und den USA bestand.

Während andere Varietäten außer dem BrE und AmE im OALD₆, LDOCE₃, LDEL_{C2} und COBUILD₃ keine große Rolle zu spielen scheinen, stellen die Herausgeber des CIDE₁ heraus, daß in dieses Wörterbuch auch das Australische Englisch in großem Maße einbezogen worden sei:

CIDE is an international dictionary, providing coverage of British, American and Australian English, including spelling, pronunciation, grammar patterns and usage of all varieties. Furthermore, the dictionary's cultural content is, as far as possible, international, reflecting the fact that English is often used today as the only common tongue between groups of speakers of other languages.

Procter (1996:4)

Auf dem Rückendeckel wird unter dem Stichwort *International* (das auch Teil des Namens des CIDE₁ ist) ebenfalls die gründliche Behandlung der sprachlichen Unterschiede zwischen dem BrE, AmE und AusE betont.³³³ Die Menge der in das Wörterbuch aufgenommenen Amerikanismen wird zwar in den einleitenden Kapiteln von CIDE₁ nicht genauer spezifiziert, aber eine kleine Stichprobe von Stephen Coffey (1996:385) deutet darauf hin, daß der Einbezug des AmE in das Wörterbuch relativ umfassend ist. Görlach (1995:147f.) hält die Aufnahme des BrE und AmE im CIDE₁ ebenfalls für ausgewogen und den Gebrauch der Britizismen und Amerikanismen als adäquat markiert.³³⁴ Allerdings ist er der Auffassung, daß das AusE im

³³¹ Diese Behauptung ist selbstverständlich sehr fragwürdig, da besonders ein Lernerwörterbuch auf Grund seiner Zielgruppe bei der Aufnahme von Stichwörtern immer eine Auswahl treffen muß.

³³² Die zahlreichen Informationen zur amerikanischen Kultur auf den sog. *Feature Pages*, den ganzseitigen Farbillustrationen und im Anhang beweisen, daß es den Herausgebern mit diesem Vorhaben ernst war.

³³³ Vgl. auch die Internetseite <http://uk.cambridge.org/elt/catalogue/0521482364/default.htm> (Stand:Oktober 2001), auf der es heißt: „Full coverage of International English is given including British, American and Australian varieties.“ Zur Formulierung „full coverage“ s.o. In den Produktinformationen für die neue Auflage wird der Einbezug verschiedener Varietäten des Englischen in das Wörterbuch allerdings nicht mehr so sehr in den Vordergrund gerückt.

³³⁴ Wie bzw. in welchem Umfang er die Erfassung und Markierung von Britizismen und Amerikanismen im CIDE₁ überprüft hat, bleibt allerdings offen. Auch Scholfield (1995:173) vertritt den Standpunkt, daß sich die Erfassung des BrE und AmE im Wörterbuch die Waage halte.

Wörterbuch weniger vertreten sei als die anderen beiden Varietäten. Varietäten wie z.B. das Kanadische, Südafrikanische, Indische und Neuseeländische Englisch sind seiner Ansicht nach gänzlich unberücksichtigt geblieben.

Tatsächlich bestand das *Cambridge Language Survey Corpus*, auf dem das CIDE₁ basiert, fast ausschließlich aus BrE und AmE Sprachdaten, wobei das AmE Sprachmaterial immerhin ca. 45 % (= ca. 46 Mio. Wörter) des Korpus ausmachte.³³⁵ Hinsichtlich des Einbezugs des AusE in das Wörterbuch wird aufgrund der mangelnden AusE Sprachdaten im Korpus dem vom Verlag eingesetzten australischen Lektor eine Schlüsselrolle zugekommen sein.³³⁶

Die Herausgeber der Wörterbücher der *Concise*-Sparte geben entweder überhaupt keine Informationen zum Einbezug des AmE oder anderer Varietäten des Englischen (vgl. das CH21₂ und NPED₁) oder nicht sehr ausführliche (vgl. das CACD, COD₁₀, CCD₅). Das CACD ist das einzige Wörterbuch unter den zuletzt genannten, dessen Herausgeber, wenn auch nur in geringem Maße, speziell auf die Darstellung des AmE im Wörterbuch eingehen. So erfährt man im Verlagskatalog von 1997 zum Einbezug des AmE in das Wörterbuch, daß AmE Schreibweisen und AmE Wortschatz aufgeführt seien. Auf dem Schutzumschlag des Wörterbuchs selbst und im Verlagskatalog von 1998 findet sich auch der Hinweis, daß Amerikanismen umfassend markiert worden seien. Da der Verlag - soweit bekannt ist - nicht auf empirische Sprachdaten des AmE zurückgreifen konnte, aber zumindest einen amerikanischen Redakteur beauftragt hatte, den Einbezug des AmE im Wörterbuch zu überprüfen, bleibt nachzukontrollieren, wie gut die Qualität der im Wörterbuch verzeichneten Informationen zum AmE ist. Andere Varietäten außer dem BrE und AmE, so z.B. das Australische, Neuseeländische und Südafrikanische Englisch, sind den Auskünften im Verlagskatalog von 1998 zufolge ebenfalls in das Wörterbuch eingegangen.

Die Herausgeber des COD₁₀ betonen in der zehnten Auflage nicht mehr so sehr den Einbezug des „North American English“ wie im COD₉ (s.o.), sondern die Berücksichtigung des „World English“, worunter sie das Nordamerikanische, Südafrikanische, Westindische, Australische, Neuseeländische, Kanadische, Schottische, Irische und Indische Englisch fassen. Die Herausgeber des CCD₅ geben ebenfalls an, daß ihr Wörterbuch „a truly international survey of English as a world language“ sei.

Zusammenfassend läßt sich also feststellen, daß die Redaktionen der Lernerwörterbücher in den Verlagsinformationen sehr viel stärker auf die Rolle des AmE in ihren Wörterbuchprodukten eingehen als die der *Concise*-Sparte, wogegen zumindest von einigen

³³⁵ Vgl. Kapitel 3.2.2.3.

³³⁶ Vgl. Kapitel 4.3.2.

Herausgebern der letzteren Gattung der Einbezug möglichst vieler Varietäten des Englischen in ihre Wörterbücher hervorgehoben wird.

4.3.4 Definitionen und Beispiele

Wie bereits in Kap. 4.2 erläutert, gehört die Verwendung eines Definitionswortschatzes, mit dessen Hilfe die in ein Wörterbuch einbezogenen Lemmata erklärt werden, zu den Merkmalen von Lernerwörterbüchern. Unter den untersuchten fünf Lernerwörterbüchern macht nur das COBUILD₃ nicht von einem festgelegten Definitionsvokabular Gebrauch, den Angaben der Herausgeber zufolge sind jedoch, wenn möglich, die für die Bedeutungserläuterungen verwendeten Wörter einfacher als die zu erklärenden. Aus dieser Vorgehensweise entstehe ein "natural defining vocabulary with most words in our definitions being amongst the 2,500 commonest words of English" (S. xviii). Inwieweit die Wortwahl in den Definitionen auf ihre Einfachheit hin tatsächlich abschließend überprüft wurde, bleibt offen.

Eine Besonderheit des COBUILD₃ ist, daß sich die Herausgeber des Wörterbuchs bei der Formulierung der Definitionen ganzer Sätze bedienen. Dieses Verfahren, mit dem das COBUILD Team Neuland betrat, da es der traditionellen Wörterbuchpraxis entgegensteht, ist für den Lerner in vielerlei Hinsicht vorteilhaft.³³⁷ Denn durch diese Art der Bedeutungserklärung kann der Wörterbuchbenutzer viele Informationen zur Gebrauchweise eines Wortes erhalten.³³⁸ Die Verfasser des Wörterbuchs bevorzugen nicht zuletzt auch deswegen diese Art der Definitionen, weil sie der Ansicht sind, daß die Wörterbuchbenutzer Erklärungen in vollständigen Sätzen sehr viel einfacher lesen und verstehen können. Der Stil der Definitionen soll dem von Lehrern oder Freunden ähneln, die ihrem Gegenüber Wörter erklären. Ein Nachteil dieser Definitionspraxis ist allerdings, daß die Erklärungen zwangsläufig mehr Platz einnehmen als solche in anderen Wörterbüchern, und dadurch z.B. insgesamt weniger Stichwörter aufgenommen werden können (darunter auch weniger Amerikanismen). Außerdem besteht die Gefahr, daß manche Definitionen etwas langatmig ausfallen, dadurch einen komplizierten Satzbau erhalten und Schlüsselwörter in den Erklärungen nicht mehr ausreichend zur Geltung kommen, wie es z.B. bei dem Eintrag zu *clothes horse* oder *pocket* (Bedeutung 4) der Fall ist.³³⁹

Unter den anderen vier Lernerwörterbüchern ist der Definitionswortschatz, der bei der Erstellung des OALD₆ verwendet wurde, mit 3000 Wörtern der größte. Die Herausgeber des LDOCE₃ und LDEL₂ griffen dagegen auf ein kleineres Definitionsvokabular, das 2000 Wörter

³³⁷ Vgl. auch die positiven Einschätzungen zum Definitionsstil des Wörterbuchs von Yorkey (1997:180), Walter (1996:357) und Herbst (1996:326).

³³⁸ Hierauf soll nicht näher eingegangen werden. Vgl. die Einleitung zum COBUILD₃, S. xiii und S. xviii f.

³³⁹ Vgl. auch Gorbahn/ Hausmann (1989:47f.) Scholfield (1999:28), Rundell (1998:333f.) und Bogaards (1996:292f.). Die beiden letzteren sind der Ansicht, daß nur durch eine größere Anzahl von Forschungsarbeiten der pädagogische Nutzen unterschiedlicher Definitionsstile ermittelt werden kann.

umfassende *Longman Defining Vocabulary*, zurück. Der für das CIDE₁ benutzte Definitionswortschatz enthielt laut Angaben der Redaktion sogar etwas weniger als 2000 Wörter. Der Vorteil des etwas größeren Definitionswortschatzes im OALD₆ könnte darin liegen, daß Bedeutungserklärungen präziser und auch evtl. knapper ausfallen, was auch der Behandlung der hier untersuchten Amerikanismen zu Gute käme.

Im COD₁₀ wird zwar genausowenig wie in den anderen hier untersuchten Wörterbüchern für Muttersprachler ein begrenzter Definitionswortschatz verwendet, aber die Herausgeber bemühen sich deutlich um einen noch klareren und zugänglicheren Definitionsstil als in den vorhergehenden Auflagen.³⁴⁰ So wurden die Bedeutungsangaben nicht nur in vielen Fällen in der Wortwahl vereinfacht und insgesamt gestrafft, sondern man machte sich auch das für das *New Oxford Dictionary of English* entwickelte Definitionsprinzip zu eigen, nach dem jeder Eintrag aus einer oder mehreren Kernbedeutung(en) („core sense(s)“) besteht. Diesen Kernbedeutungen folgen in vielen Fällen Subbedeutungen („subsenses“), die den Kernbedeutungen untergeordnet sind. Die genannten Änderungen bewirkten, daß ein Großteil der Einträge für die neue Auflage vollkommen neu geschrieben werden mußten. Die übersichtlichere Struktur der Einträge führte bei der Suche mancher der hier analysierten Amerikanismen (wie z.B. *fall*) zu einem schnelleren Ergebnis, aber das beschriebene Definitionsprinzip kann auch zu fragwürdigen Zuordnungen führen: So findet man z.B. in der neuen Auflage die AmE Bedeutung von *block* („a group of buildings bounded by four streets“ bzw. „the length of one side of such a block“) als Subbedeutung unter der BrE Kernbedeutung („a large single building subdivided into separate flats or offices“).³⁴¹

Bezüglich der Definitionen im NPED₁ heißt es auf dem Schutzumschlag des Wörterbuchs: „Clear, detailed definitions, with particular emphasis on difficult concepts and meanings“. Tatsächlich sind manche Definitionen der hier untersuchten Amerikanismen präziser und klarer als in anderen Wörterbüchern der *Concise*-Sparte, so z.B. bei *block*, *dry goods* und *graft*.³⁴²

Da den Herausgebern des CACD vor allem daran gelegen war, möglichst viele Stichwörter ins Wörterbuch aufzunehmen, sind z.B. Komposita, Verben mit Präpositionen und mehrgliedrige Wendungen zusammen mit ihren Definitionen in dem Eintrag zu dem jeweiligen Lemma eingebettet. Durch diese Vorgehensweise wird Platz gespart, andererseits dauert es länger, z.B. die hier untersuchten Amerikanismen *junior college*, *dry goods*, *cop out* und *right off* mit den jeweiligen Bedeutungserklärungen im Wörterbuch zu finden.

³⁴⁰ Vgl. zum Definitionsstil der Wörterbücher des Englischen für Muttersprachler auch das Kapitel 4.1.

³⁴¹ Vgl. auch Kapitel 7.2.1.

³⁴² Vgl. das Kapitel 7.2.

Was die Zahl der Anwendungsbeispiele betrifft, liegt das COBUILD₃, verglichen mit den anderen hier untersuchten Wörterbüchern, an der Spitze, wenn man den Angaben der Herausgeber vertrauen darf.³⁴³ So enthält das COBUILD₃ 105.000 Anwendungsbeispiele, während es im CIDE₁ etwas weniger, nämlich 100.000, und im OALD₆ und LDOCE₃ nur 90.000 bzw. 63.000³⁴⁴ sind. Über die Zahl der im LDELC₂ enthaltenen Anwendungsbeispiele erhält man leider keine präzise Auskunft. Die hier analysierten Wörterbücher der *Concise*-Sparte machen mit Ausnahme des NPED₁ ebenfalls keine Angaben zu der Menge der einbezogenen Anwendungsbeispiele. Das ist auch nicht verwunderlich, da in den Wörterbüchern dieser Gattung kaum Anwendungsbeispiele zu finden sind (vgl. Kap. 4.2). Umso erstaunter ist man, daß die Herausgeber des NPED₁ damit werben, daß sie eine Fülle von Anwendungsbeispielen in das NPED₁ aufgenommen hätten (genaue Zahlenangaben werden aber nicht zur Verfügung gestellt): „It contains thousands of illustrative examples, many of them drawn from the works of major literary figures, that serve not only to show how individual words work in context, but also to give a sense of the richness and variety of English.“³⁴⁵ Tatsächlich finden sich auch bei einigen der in dieser Arbeit behandelten Amerikanismen Beispiele, so bei *block*, *can*, *chipper*, *protest sthg* und *a half hour*. Während jedoch bei *chipper* die Quelle angegeben wird, bleibt die Herkunft der anderen Beispiele offen. Die Länge der verwandten Beispiele variiert sehr.³⁴⁶

4.3.5 Darstellung der Unterschiede zwischen dem Amerikanischen und Britischen Englisch in den einleitenden Kapiteln der Wörterbücher

a) Oxford Advanced Learner's Dictionary of Current English (OALD₆)

Den Unterschieden zwischen dem BrE und AmE ist in den einleitenden Kapiteln des OALD₆ genausowenig wie in der vorhergehenden Auflage ein eigenes Kapitel gewidmet. In dem Kapitel *Key to dictionary entries* (S. vii ff.) wird jedoch darauf hingewiesen, daß Divergenzen in der Aussprache³⁴⁷ und Orthographie zwischen dem BrE und AmE angegeben werden. Im Gegensatz zu dem entsprechenden Kapitel im OALD₅ wird jedoch auf lexikalische Unterschiede zwischen den Varietäten und ihre Darstellung im Wörterbuch nicht eingegangen, und die Behandlung orthographischer Varianten des AmE wird nur kurz am Ende der allgemeinen Ausführungen zu

³⁴³ Vgl. die Informationen auf der Umschlagrückseite des Wörterbuchs und auf der Webseite des Verlags (<http://titania.cobuild.collins.co.uk/catalogue/cob3flash.html>, Stand Mai 2003).

³⁴⁴ Über die Zahl der Anwendungsbeispiele im LDOCE₃ und dem 2001 veröffentlichten Anhang erhält man keine direkte Auskunft. Da aber das LDOCE₄ mit seinen 88.000 Anwendungsbeispielen 25.000 mehr als bisher enthalten soll (vgl. die Informationen auf der Webseite <http://www.longman.com/ldoce/members/ruth.html>, Stand Februar 2004), müßte das LDOCE₃ in der Ausgabe von 2001 63.000 Anwendungsbeispiele umfaßt haben.

³⁴⁵ Vgl. die Innenseite des Schutzzumschlags.

³⁴⁶ Vgl. das Kapitel 7.2. und 7.3.

³⁴⁷ Siehe auch Appendix 7 des OALD₆.

alternativen orthographischen Formen erwähnt und nicht, wie noch im OALD₅, anhand eines Beispiels veranschaulicht.³⁴⁸ Laut Informationen der Herausgeber erfolgt nach einem Stichwort, das eine orthographische Variante des AmE darstellt, ohne Angabe einer Definition ein Querverweis auf den „Haupteintrag“, bei dem es sich um das BrE oder gemeinenglische Pendant handelt.³⁴⁹ Da über die Darstellung der lexikalischen Unterschiede zwischen dem BrE und AmE in den einleitenden Kapiteln des OALD₆ keine Auskunft gegeben wird, bleibt es dem Wörterbuchbenutzer überlassen, herauszufinden, daß in vielen Fällen in dem Eintrag für das AmE Synonym eines BrE Lexems ebenfalls keine Bedeutungserklärung vorhanden ist, sondern auf das BrE Lemma verwiesen wird. Die Behandlung von AmE Synonymen ist jedoch im OALD₆ nicht immer gleich: Während z.B. *mailman* eine eigene Definition erhält, werden im Falle von *fall* (= autumn), *baggage* (= luggage) und *clipping* lediglich die jeweiligen BrE Synonyme angeführt. Allerdings finden sich z.B. unter *fall* bzw. *baggage* zumindest Beispiele zum Gebrauch, während bei *clipping* kein Anwendungsbeispiel gegeben wird.³⁵⁰

Im Gegensatz zum OALD₆, wo die Auskünfte hinsichtlich der Divergenzen zwischen dem AmE und BrE und ihrer Darstellung im Wörterbuch in den einleitenden Kapiteln also sehr knapp ausfallen, enthält das OALD₅, wie bereits angedeutet, diesbezüglich deutlich mehr Informationen. So wird z.B. in dem Kapitel *Using your dictionary* (S. ixf.) unter dem Abschnitt *Alternative forms and synonyms* zusätzlich noch einmal auf orthographische und lexikalische Unterschiede zwischen dem AmE und BrE eingegangen, die durch Beispiele veranschaulicht werden.³⁵¹

Unterschieden zwischen dem AmE und BrE auf den Ebenen der Grammatik sowie der Wendungen und Kollokationen wird weder in den einleitenden Kapiteln des OALD₆ noch des OALD₅ Aufmerksamkeit geschenkt.

Als Markierungen für das BrE und AmE verwenden die Lexikographen nicht mehr, wie noch in der 5. Auflage, *Brit* bzw. *US*, sondern *BrE* und *AmE*. Der Grund für diese Änderung mag darin liegen, daß diese Markierungen präziser sind: In den vorher verwendeten kam nämlich der Bezug zur Sprache (Britisches Englisch bzw. Amerikanisches Englisch) gar nicht zum Ausdruck. Unter der Überschrift *Labels* (S. i) wird näher erläutert, was mit den Etikettierungen *AmE* und

³⁴⁸ Vgl. S. vii.

³⁴⁹ So wird z.B. bei dem Stichwort *center*, das als *AmE* markiert ist, ohne weitere Bedeutungserklärung auf das unmarkierte *centre* verwiesen, das erst im „Haupteintrag“ mit *BrE* gekennzeichnet wird. Letzteres ist jedoch schon eine Verbesserung gegenüber dem OALD₅, denn dort wurden nirgendwo britische Schreibweisen als solche hervorgehoben. Bezüglich der Aussprachemarkierungen hat sich allerdings auch im OALD₆ nichts geändert, denn nach wie vor werden nur AmE Aussprachevarianten markiert, nicht BrE.

³⁵⁰ Weitere Informationen dazu sind in Kapitel 7.2 nachzulesen. Auch im OALD₅ war die Darstellung AmE Synonyme nicht immer gleich.

³⁵¹ In der 4. Auflage des Wörterbuchs beschreiben die Herausgeber in dem Kapitel *Using the Dictionary - A Detailed Guide to the Entries* (S. 1545ff.) sogar noch ausführlicher die Unterschiede zwischen dem BrE und AmE.

BrE gemeint ist. Zu *AmE* heißt es: "**AmE** describes expressions, spellings and pronunciations used in American English and not in British English, for example *bleachers*, *blindside*, *blooper*."³⁵² Die Markierung *BrE* wird so erklärt: "**BrE** describes expressions used in British English and not in American English, for example *jumble sale*, *agony aunt*, *chinwag*." Die Verfasser setzen die beiden Varietäten in den Definitionen der Markierungen zueinander in Beziehung und schließen explizit nur aus, daß z.B. ein im AmE auftretendes Wort auch im BrE vorkommt und umgekehrt. Sie präzisieren aber nicht, in welchem geographischem Raum AmE und BrE gesprochen wird und lassen somit offen, ob ein z.B. als *AmE* gekennzeichnetes Wort nicht auch anderswo, z.B. in Kanada, vorkommt. In den Definitionen wird zudem leider nicht erwähnt, daß auch varietätenspezifische Bedeutungen oder grammatikalische Besonderheiten des AmE oder BrE eine entsprechende Markierung erhalten. Überdies wird in der Erläuterung der Etikettierung *BrE* nicht angegeben, daß, wie bereits angesprochen, auch orthographische Varianten des BrE als solche gekennzeichnet werden.

Vergebens sucht man im übrigen in den einleitenden Kapiteln nach Hinweisen auf die Markierungen *especially AmE* bzw. *especially BrE*, die im Lexikon-Teil oft Verwendung finden.

b) Longman Dictionary of Contemporary English (LDOCE₃)

Im Kapitel 7 des *Guide to the Dictionary* wird auf die Unterschiede zwischen dem BrE und AmE in Aussprache, Orthographie, Lexik, Grammatik und Phraseologie eingegangen. Die Unterschiede werden nach Angaben der Redaktion folgendermaßen gekennzeichnet:

Words, senses of words, grammatical constructions, phrases, and collocations, that only occur in British English are marked *BrE*, and those that exist only in American English are marked *AmE*. Those that are much more frequent in one form of English than in the other are marked *especially BrE* or *especially AmE*.³⁵³

Diese Erläuterung der Etikettierungen *BrE* und *AmE* definiert die eine Markierung auf der Folie der anderen, sagt jedoch noch nichts über den geographischen Geltungsraum der beiden Varietäten aus. Bei der Erklärung von Abkürzungen und Markierungen auf der ersten Seite des Wörterbuchs unter *Labels* heißt es hierzu: "1 Words which are used only or mainly in one region or country are marked: *AmE* American English [...] *BrE* British English..." *AmE* kann dieser Formulierung nach also heißen, daß ein Wort hauptsächlich, aber nicht ausschließlich, in den USA verwendet wird. Diese Ausdrucksweise hat natürlich den Vorteil, daß die

³⁵² Die Beispiele hätten ein wenig ausgewogener sein können: Orthographische Varianten im AmE spiegeln sie z.B. nämlich nicht wider.

³⁵³ LDOCE₃, S.xxi. Unterschiede in der Orthographie werden ebenfalls mit BrE oder AmE kenntlich gemacht (aber nicht mit *especially BrE* oder *especially AmE*).

Lexikographen sich nicht festlegen müssen, ob z.B. ein in den USA gebräuchliches Wort nicht auch in Australien benutzt wird.

In der 2. Auflage heißt es im Abschnitt *Short Forms and Labels* auf der Innenseite des Rückendeckels nur: *Labels used in the dictionary: AmE American English BrE British English*. In der 2. Auflage schließt die Markierung *AmE* nach Angaben der Herausgeber in der Regel auch das Kanadische Englisch mit ein und *BrE* steht für gewöhnlich ebenfalls für das Englisch Irlands, Australiens, Neuseelands, Indiens, Pakistans, Afrikas und der Karibik³⁵⁴. Diese Aussage wird im LDOCE₃ nicht gemacht³⁵⁵. Die Markierungen *AustrE*, *CanE*, *CarE*, *IndE*, *PakE*, *IrE*, *NZE*, *SAfrE* und *ScotE* werden zwar auch schon in der 2. Auflage aufgeführt, aber in den Einträgen nach Angaben der Redaktion nur benutzt, wenn ein Wort ausschließlich auf eine dieser Regionen beschränkt sein sollte. Es ist anzunehmen, daß in der 3. Auflage genauso vorgegangen wird, ohne daß die Herausgeber allerdings explizit darauf hinweisen. Markierungen außer *AmE* und *BrE* sind auch im LDOCE₃ selten. Der Hauptakzent des Wörterbuchs liegt deutlich auf dem Britischen und Amerikanischen Englisch (s.o.). Die Etikettierung *especially AmE* oder *especially BrE* wird in der 2. Auflage bei der Erläuterung der Markierungen nicht erwähnt, aber in den Einträgen benutzt, wie sich bei der Überprüfung der von mir ausgewählten Amerikanismen zeigte.

Im Kapitel 7.1 *Pronunciation and spelling differences* wird dem Wörterbuchbenutzer vermittelt, daß unterschiedliche Ausspracheformen des BrE und AmE durch zwei senkrechte Linien voneinander getrennt dargestellt werden (links befindet sich die britische Aussprache, rechts die amerikanische). Außerdem erfährt der Lerner, wie Unterschiede in der Orthographie verdeutlicht werden, wenn sowohl das AmE als auch das BrE eine varietätenspezifische Schreibweise bevorzugen oder wenn es eine gemeinenglische Form gibt, zu der in einer der beiden Varietäten eine Alternativform besteht.

Im Kapitel 7.2 *Word meanings - British and American differences* des *Guide to the Dictionary* wird dargelegt, wie lexikalische Unterschiede zwischen dem BrE und dem AmE im Wörterbuch kenntlich gemacht werden. Im Falle von zwei Synonymen, die jeweils auf eine Varietät beschränkt seien, werde nach der Definition des BrE Wortes auf das AmE Synonym hingewiesen und umgekehrt (so z.B. bei *lift BrE* und *elevator AmE*).³⁵⁶ Wenn ein BrE Wort oder

³⁵⁴ Vgl. F45.

³⁵⁵ Wahrscheinlich deswegen, weil sie viel zu vereinfachend ist und sachlich wohl auch oft nicht zutrifft. Vertreter der erwähnten nationalen Varietäten des Englischen heben zudem seit einiger Zeit die Eigenständigkeit ihrer Varietäten hervor.

³⁵⁶ Im LDOCE₂ wird direkt hinter dem BrE Lemma das AmE Synonym angegeben. Während das LDOCE₂ die Definition meistens nur bei dem Eintrag des BrE Wort angibt und nach dem AmE Stichwort auf das BrE Lemma verweist, findet sich im LDOCE₃ in der Regel auch nach dem AmE Wort eine Bedeutungserklärung (vgl. z.B. *market garden* und *truck farm* im LDOCE₂ und LDOCE₃). Dies zeigt, daß man im LDOCE₃ um eine gleichwertige Behandlung des AmE bemüht ist.

ein AmE Wort kein genaues Äquivalent in der anderen Varietät habe oder eines, das kaum benutzt werde, habe man dieses nur mit *BrE* oder *AmE* markiert. Als Beispiele werden *airing cupboard* (als *BrE* markiert) und *airhead* (als *slang especially AmE* gekennzeichnet) angegeben. Wenn eine bestimmte Bedeutung eines Wortes nur in einer der beiden Varietäten existiere, wie beispielsweise bei *homely*, werde diese ebenfalls entsprechend als *AmE* oder *BrE* gekennzeichnet.

Die Darstellungen der lexikalischen Unterschiede zwischen dem BrE und AmE unterscheiden sich in den Einführungen zum Gebrauch des Wörterbuchs in allen drei Ausgaben des LDOCE erheblich voneinander. In der 1. Ausgabe werden sechs Gruppen lexikalischer Unterschiede genannt³⁵⁷, von denen nur zwei in die 2. Auflage übernommen werden: Die eine Gruppe bezieht sich auf den Fall, daß zwei Synonyme existieren, die jeweils auf eine Varietät beschränkt sind; die andere Gruppe umfaßt auch zwei Synonyme, wobei aber das eine Synonym gemeinenglisch und das andere auf eine Varietät beschränkt ist. In Algeos Kategorisierung (1989a:227-240) würde die zuerst genannte Gruppe der Klasse 6 (*equivalent terms=interdialectal synonymy*) und die zweite Gruppe der Klasse 7 (*overlapping terms=one-sided dialect synonymy*) entsprechen. Die letztere Gruppe wird im LDOCE₃ nicht mehr ausdrücklich im *Guide to the Dictionary* erwähnt, obwohl sie in den Einträgen vorkommt und dort in der gleichen Weise wie in der 1. und 2. Auflage markiert wird.³⁵⁸ Zu der oben zuerst genannten Gruppe (*equivalent terms=interdialectal synonymy* in Algeos Klassifizierung) treten zwei weitere, die in keiner der vorherigen beiden Auflagen erwähnt werden, obwohl sie in den Einträgen dieser Auflagen entsprechend markiert sind. Davon umschließt die eine die Klassen 2, 3 und 4 der Kategorisierung Algeos (*referential gap, lexical gap, cultural gap*) und die andere die Klassen 10, 11, 12 und 13 (*interdialectal polysemy, one-sided dialectal polysemy, interdialectal polysemy plus a common sense, interdialectal diversity*).

Im Kapitel 7.3 *Differences in grammar* werden grammatikalische Unterschiede zwischen dem BrE und AmE angeführt. Es wird erwähnt, daß einige Wörter nur im BrE oder AmE eine bestimmte Präposition oder Konstruktion nach sich ziehen. Als Beispiel für grammatikalische Unterschiede zwischen den Varietäten wird genannt, daß Substantive wie *government* oder *class* nur im BrE ein Verb im Plural erfordern und dementsprechend gekennzeichnet sind. Die grammatikalischen Divergenzen werden in diesem Abschnitt nur kurz und eher oberflächlich behandelt; vieles, wie z.B. Unterschiede im Gebrauch des Konjunktivs, der Modalverben, des

³⁵⁷ Vgl. S. xvi. Siehe auch Wächters Beurteilung dieser sechs Typen des Gebrauchs (1980:148f.).

³⁵⁸ Vgl. die Erläuterungen zur Überprüfung von *fall* und *autumn* im Wörterbuch in Kap. 7.2.6 der Arbeit.

Artikels oder der *tags*³⁵⁹ bleibt unerwähnt. Zu überprüfen ist, ob diesen Punkten in den konkreten Einträgen mehr Beachtung geschenkt wurde. Man muß auch zur Kenntnis nehmen, daß es im LDOCE₂ noch keinen Abschnitt zu den grammatikalischen Verschiedenheiten zwischen dem BrE und AmE gab, und man nur beiläufig, bei der Auflistung der *Grammar codes*³⁶⁰, auf die Tatsache verwiesen wurde, daß bestimmte Substantive wie *committee* nur im BrE ein Verb im Plural nach sich ziehen können oder z.B. in den *Language Notes: Modals* oder *Articles* über grammatikalische Unterschiede zwischen dem AmE und BrE erfuhr.

Im Kapitel 7.4 *Differences in phrases and collocations* wird darauf hingewiesen, daß sich das BrE und AmE durch eine Vielzahl varietätenspezifischer idiomatischer Wendungen und Kollokationen unterscheiden. Wendungen und Kollokationen, die nur im BrE oder AmE vorkommen, würden gekennzeichnet, so z.B. *jack sb around* phr v (als *AmE slang* markiert) und *jack sth in* phr v (als *BrE informal* bezeichnet). In der 2. Auflage des LDOCE wurde den Unterschieden zwischen dem BrE und AmE auf der Ebene idiomatischer Wendungen und Kollokationen noch kein eigener Abschnitt gewidmet. Nur das Beispiel *a new lease of life (BrE)/on life(AmE)* unter der Überschrift *Where labels can be shown in dictionary entries*³⁶¹ und die Beispiele der *Language Notes: Collocations* oder *Make and Do* machen Wörterbuchbenutzer darauf aufmerksam, daß es solche Unterschiede gibt.

c) Longman Dictionary of English Language and Culture (LDEL_{C2})

Die Beschreibungen der Unterschiede zwischen dem BrE und AmE sowie Informationen über deren Darstellung im Wörterbuch und zur Verwendung der Markierungen AmE und BrE sind in der neuen Auflage deutlich gekürzt worden. Während der Lerner in der ersten Auflage unter 4.1. *Labels* erfährt, was im Wörterbuch unter AmE und BrE gefaßt wird („American English usually includes the English of both the US and Canada. British English is the variety of English spoken in Britain, and usually also includes the English of Ireland, Australia and New Zealand, India and Pakistan, Africa and the Caribbean“, F24³⁶²), findet man im LDEL_{C2} keinerlei Erläuterung mehr zu den Markierungen AmE und BrE. Unklar ist also, ob die Herausgeber den beiden Markierungen immer noch stillschweigend die Definitionen aus der ersten Auflage zugrunde

³⁵⁹ Vgl. z.B. Benson/ Benson/ Ilson (1986), Algeo (1988), Matsuda (1990), Trudgill/ Hannah (1985), das OCEL (1992) und Gramley & Pätzold (1992) hinsichtlich einer ausführlichen Behandlung der grammatikalischen Unterschiede zwischen dem BrE und AmE.

³⁶⁰ Vgl. F 41.

³⁶¹ Vgl. F 46.

³⁶² Zwar gibt es auch in der ersten Auflage die Markierungen *AustrE*, *CanE*, *CarE*, *IndE & PakE*, *IrE*, *NZE*, *SAfrE* und *ScotE*, doch werden diese nur dann verwendet, wenn ein Wort, eine Bedeutung o.ä. vollkommen auf eine dieser Varietäten beschränkt ist. Die erste Auflage stützt sich in ihrer Definitionspraxis auf das LDOCE₂, vgl. auch das Kapitel 4.3.7.2.

legen oder die Etikettierungen AmE und BrE anders verstanden wissen wollen.³⁶³ Aufgrund der geringen Datenmenge, die es bezüglich anderer Varietäten außer dem BrE und AmE im *Longman Corpus Network* gibt, können sicherlich kaum stichfeste Aussagen über Gebrauchsgewohnheiten in diesen Varietäten gemacht werden.³⁶⁴

Die Markierungen *especially AmE* bzw. *especially BrE* werden leider weder in der ersten noch in der zweiten Auflage des Wörterbuchs erwähnt oder erläutert, obwohl diese sehr häufig in den Einträgen verwendet werden.

Während das LDEL_{C1} im Kapitel 4.2. (*Word Variation*) einen Abschnitt zu *British and American word differences* aufführt, erhält der Lerner in der neuen Auflage keinerlei Erklärungen zu den lexikalischen Unterschieden zwischen den beiden Varietäten und ihrer Darstellung im Wörterbuch. Bezüglich der Erfassung der Unterschiede im Wörterbuch selber hat sich aber de facto nichts geändert. Im Falle zweier Synonyme, die jeweils auf das BrE und AmE beschränkt sind, gilt auch in der zweiten Auflage, daß nach dem AmE Stichwort keine eigenständige Worterklärung, sondern der Querverweis auf das BrE Synonym und dessen Definition erfolgt. An dieser Stelle kommt eine größere Gewichtung des BrE gegenüber dem AmE zum Ausdruck, die auch für die Darstellung der AmE Aussprache im Wörterbuch gilt: Im LDEL_{C2} wird nämlich nicht, wie im LDOCE₃ oder OALD₆, die vom BrE abweichende AmE Aussprache eines Lexems in den Einträgen im Lexikon-Teil separat angeführt, wenn es sich um „systematische“ Divergenzen zwischen den Varietäten handelt (wie z.B. bei der Aussprache des /r/, das im BrE abhängig von der Lautumgebung artikuliert wird, im *Network* AmE jedoch durchgängig).

Orthographische Unterschiede zwischen den beiden Varietäten sowie solche in der Aussprache werden dagegen auch im *Guide to the Dictionary* der zweiten Auflage behandelt, und zwar in Kapitel 4.4 bzw. 5.3 und 5.5. Während jedoch der Abschnitt zu den orthographischen Unterschieden aus dem LDEL_{C1} in voller Länge übernommen wurde, hat man bei der Beschreibung der Ausspracheunterschiede den Text unter 5.5. um die veranschaulichenden Beispiele gekürzt. Die Entscheidung der Herausgeber, das in der ersten Auflage 20 Seiten umfassende Kapitel *Guide to the Dictionary* zu straffen, war sicherlich richtig, aber dadurch sind leider auch eine Reihe wissenswerter Informationen zu den BrE-AmE Unterschieden sowie zu den Markierungen AmE und BrE selber weggefallen.

³⁶³ Vgl. zu den im LDEL_{C1} und LDOCE₂ angegebenen Definitionen der Markierungen BrE und AmE meine Ausführungen unter b) in diesem Kapitel.

³⁶⁴ Dieses Problem betrifft sehr viele der hier untersuchten Korpora.

Unterschiede zwischen den Varietäten hinsichtlich des grammatischen Verhaltens oder der Kollokationen und Wendungen werden im LDELC₂ genauso wenig wie im LDELC₁ in einem gesonderten Abschnitt in der Einführung des Wörterbuchs behandelt.

d) Collins COBUILD English Dictionary for Advanced Learners (COBUILD₃)

Wie schon in den ersten beiden Auflagen ist auch im COBUILD₃ den Unterschieden zwischen dem BrE und AmE kein eigenes Kapitel in der Einleitung gewidmet. Einem im *Guide to the Dictionary Entries* abgedruckten Beispiel zu der Darstellung flektierter Formen eines Lexems in den Einträgen kann man entnehmen, daß Schreibweisen, die im AmE üblich sind und von denen des BrE abweichen, angeführt werden.³⁶⁵ Im Kapitel *Pronunciation* erfährt man, daß in der Regel nur die britische Aussprache eines Wortes angezeigt wird, da die meisten Ausspracheunterschiede zwischen dem AmE und BrE systematischer Natur seien. Wenn aber die übliche AmE Aussprache von der im BrE gängigen in erheblicher Weise differiere, werde eine separate Lautschriftangabe für das AmE gemacht.³⁶⁶ An dieser Stelle wird erneut die in den älteren Auflagen bereits ersichtliche Konzentration der Herausgeber auf das BrE deutlich, denn andere Lernerwörterbücher wie das OALD₆ oder das LDOCE₃ geben auch systematische Ausspracheunterschiede zwischen den beiden Varietäten in den Einträgen wieder (s.o.).

Nur anhand der Einträge selber kann man ersehen, daß auch Unterschiede in den Betonungsmustern zwischen dem AmE und BrE dargestellt werden.

In der Einleitung des Kapitels *Style and Usage* wird zunächst darauf hingewiesen, daß sich Markierungen jeder Art in der Regel am Ende der Definitionen befinden und dort in Großbuchstaben und eckige Klammern gesetzt sind. Somit hat sich die Markierungspraxis im Vergleich zu COBUILD₂ geändert, denn in der vorgehenden Auflage wurden die durchgehend unabgekürzten Markierungen ohne typographische Hervorhebung in den ausformulierten Definitionstext integriert. Vor allem diatopische Markierungen sind im COBUILD₂ an unterschiedlichen Stellen in den Einträgen anzutreffen (d.h. am Anfang und am Ende, je nachdem, ob ein Wort vollständig oder nur teilweise auf eine Varietät beschränkt ist). Eine einheitliche Positionierung von Markierungen in den Einträgen, wie sie im COBUILD₃ durchgeführt wurde, ist sicherlich benutzerfreundlicher, denn der Lerner weiß nun, an welcher Stelle eine Markierung zu erwarten ist. Ebenso ist auch die typographische Hervorhebung der Markierungen zu begrüßen, obwohl diese noch deutlicher hätte ausfallen können, damit dem

³⁶⁵ In demselben Kapitel wird ebenfalls darauf aufmerksam gemacht, daß auch Querverweise auf varietätenspezifische Synonyme eines Wortes im Wörterbuch angegeben werden und sich Informationen zu der Verwendung von diatopischen Markierungen unter *Style and Usage* befinden.

³⁶⁶ Die im COBUILD₃ befindlichen Ausspracheinformationen für das BrE und AmE basieren auf den Ausspracheformen des RP (*Received Pronunciation*) bzw. GenAm (General American).

Lerner beispielsweise die Beschränktheit eines Wortes auf eine Varietät noch besser in die Augen springt. Die Abkürzung der Markierungen *British English* und *American English* in BRIT und AM und der Verzicht auf die Einbettung dieser Markierungen in ausformulierte Halbsätze (z.B. „used mainly in American English“) dürfte den durchschnittlichen Wörterbuchbenutzer wohl nicht vor größere Probleme stellen, außerdem hilft diese Praxis, Platz für andere Informationen oder sogar für mehr Stichwörter im Wörterbuch zu schaffen. Fragwürdig ist allerdings die Angabe der Markierungen nach den Definitionen, denn manche Lerner werden vielleicht nur den Beginn einer Bedeutungserklärung lesen, um sich in kürzester Zeit die Bedeutung eines Lexems zu vergegenwärtigen.

In dem Abschnitt *Geographical labels* desselben Kapitels werden die Markierungen BRIT und AM erläutert und in ihrem Gebrauch anhand der auf der rechten Seite abgedruckten Bedeutungserklärungen zum Lexem *boot*, in denen auf lexikalische Unterschiede zwischen dem AmE und BrE eingegangen wird, veranschaulicht.³⁶⁷ BRIT wird folgendermaßen definiert: "used mainly by speakers and writers in Britain, and in other places where British English is used or taught". Zu AM heißt es: "used mainly by speakers and writers in the USA, and in other places where American English is used or taught".³⁶⁸ In diesen Definitionen werden die beiden Varietäten nicht jeweils in Beziehung zueinander gesetzt, sondern es wird auf den geographischen Geltungsraum der jeweiligen Varietät eingegangen. Die Erläuterungen der Markierungen BRIT und AM lassen allerdings absichtlich offen, welche Orte bzw. Länder - abgesehen von Großbritannien und den USA - gemeint sind, in denen BrE und AmE verwendet oder gelehrt wird. Verwunderlich ist in diesem Zusammenhang, daß in den Einträgen auch die Markierung [mainly AM] bzw. [mainly BRIT] auftaucht, denn - abgesehen davon, daß auf diese Markierung in dem Kapitel *Style and Usage* mit keinem Wort verwiesen wird - erscheint sie auch darum nicht sinnvoll, weil es bereits in den oben angeführten Definitionen von AM und BRIT heißt: „used mainly [meine Hervorhebung] by speakers and writers in ...“. Die Wörterbuchmacher haben das Wort „mainly“ sicherlich ganz bewußt in die beiden Definitionen eingefügt, da sie sich so nicht festlegen müssen, ob ein varietätenspezifisches Lexem nicht vielleicht auch (wenn auch weniger häufig) in anderen Varietäten verwendet wird. Die Markierungen [mainly AM] bzw. [mainly BRIT] sind dann aber als tautologisch anzusehen.

³⁶⁷ Ein eigener Abschnitt zu den lexikalischen Unterschieden zwischen dem BrE und AmE findet sich im COBUILD₃ aber nicht.

³⁶⁸ In der ersten Auflage werden unter dem Kapitel *Usage* die geographischen und stilistischen Markierungen noch nicht getrennt aufgeführt und neben den zur Veranschaulichung beigefügten Beispielen erklärt. Stattdessen wird in einem einzigen Abschnitt allgemein auf das Auftreten von Markierungen im Wörterbuch eingegangen, wonach in ungeordneter Reihenfolge nacheinander einmal ein Beispiel für stilistische Markierungen, dann wieder ein Beispiel für die Kennzeichnung varietätenspezifischen Gebrauchs gegeben wird.

In dem Abschnitt *Geographical labels* erfährt der Wörterbuchbenutzer auch, daß am Ende des Eintrags zu einem varietätenspezifischem Lexem des BrE oder AmE das jeweilige Pendant in der anderen Varietät - soweit vorhanden - angegeben wird. Dieser Passus findet sich im COBUILD₂ noch nicht. In der Praxis sind aber auch dort schon Querverweise auf Synonyme vorhanden, die in der jeweils anderen Varietät gängig sind. Im COBUILD₃ werden diese jedoch neuerdings ebenfalls typographisch hervorgehoben, und zwar in einer Weise, daß die Unterschiede zwischen dem BrE und AmE sehr gut ins Auge fallen³⁶⁹. Überraschend ist allerdings, daß im Lexikon-Teil die Formulierungen zu den Synonymangaben variieren: So heißt es am Ende des Eintrags zu BrE *bookshop*: „in AM, use bookstore“, aber nach der Erläuterung der BrE Bedeutung von *bookstall* wird angeführt: „in AM, usually [meine Hervorhebung] use newsstand“. Es bleibt unklar, warum der Wortlaut sich hier unterscheidet, vor allem, da beide Britizismen mit BRIT gekennzeichnet sind.

Markierungen für Wörter oder Wortbedeutungen, die ausschließlich oder überwiegend in anderen Varietäten als dem BrE und AmE gebräuchlich sind, finden sich nicht (auch nicht für das Australische Englisch, dessen Anteil an der BoE im Jahr 2000 8.28% betrug).

e) Cambridge International Dictionary of English (CIDE₁)

Eine ausführliche Darstellung der Unterschiede zwischen dem BrE und AmE erfolgt nicht in den einleitenden Kapiteln des CIDE₁, sondern in den *Language Portraits*, vor allem in dem *Language Portrait 'Varieties of English'* (S. 1608), das diese Unterschiede zum Thema hat. Dort werden zunächst die Markierungen *Am*, *Aus* und *Br* erläutert: *Am* steht für "North American English", *Aus* bedeutet "Australian English" und *Br* ist die Abkürzung für "British English".³⁷⁰ Diese Markierungen beziehen sich nach Auskunft der Redaktion auf "the standard type of English used by speakers from the country mentioned."³⁷¹ Im CIDE₁ werden also - im Gegensatz zu anderen Wörterbüchern - nicht nur das AmE und BrE als die zwei Hauptvarietäten des Englischen gegenübergestellt, sondern das AusE wird als ebenso wichtig betrachtet und von diesen beiden Varietäten unabhängig behandelt. Ein Wort, das als (*Am*) markiert ist, wird nach Auskunft der Redaktion so gut wie nie im BrE oder im AusE verwendet.³⁷² Wörter, Wendungen etc., die in den drei Varietäten verwendet werden, seien nicht gekennzeichnet worden.³⁷³ Das

³⁶⁹ Dies wird durch die Umrandung des Querverweises, den Fettdruck des in der jeweils anderen Varietät gebräuchlichen Synonyms und ein graphisches Symbol (ein schwarz unterlegtes Häkchen zu Beginn dieser Informationen) erreicht.

³⁷⁰ Vgl. S. 1608.

³⁷¹ Ebd.

³⁷² Herausgeber anderer Wörterbücher legen sich dagegen nicht fest, ob ein Wort, das AmE gekennzeichnet ist, im AusE verwendet wird oder nicht.

³⁷³ An anderer Stelle (S.790) heißt es, daß unmarkierte Wörter und orthographische Schreibweisen in allen Varietäten gebräuchlich sind.

Nordamerikanische Englisch und das Britische Englisch umfassen der Redaktion zufolge weitere "varieties", darunter das Kanadische, Irische und Schottische Englisch sowie regionalen Sprachgebrauch innerhalb der USA und Großbritanniens.³⁷⁴ Für diese Varietäten sowie für den regionalen Sprachgebrauch innerhalb der beiden Staaten werden aber zusätzlich noch eigene Markierungen angegeben. Aus dieser Tatsache läßt sich schließen, daß z.B. die Markierung *Am* vermutlich in der Regel den kanadischen Sprachgebrauch einschließt und die Markierung *Canadian English* nur verwendet wird, wenn der kanadische Sprachgebrauch vom AmE abweicht. In dem *Language Portrait 'Labels in the Dictionary'* (S. 790) wird unter dem Abschnitt *Labels referring to different varieties of English* verdeutlicht, daß z.B. ein als *Am* markiertes Wort vor allem, aber nicht nur, in den USA verwendet werde. Dort heißt es nämlich: "The following labels are applied to words or phrases which are used especially by people in or from particular countries."³⁷⁵ An dieser Stelle werden jedoch das Kanadische, Irische und Schottische Englisch nicht dem AmE bzw. BrE zugeordnet, sondern als besondere Varietäten aufgelistet.³⁷⁶

In dem Abschnitt *Differences in words and phrases* des *Language Portraits 'Varieties of English'* wird auf lexikalische Unterschiede zwischen dem AmE, BrE und AusE eingegangen. Die Herausgeber führen zunächst aus, daß in manchen Bereichen der Sprache, z.B. in solchen, die das Transportwesen, Nahrungsmittel oder Kleidung betreffen, die lexikalischen Unterschiede zwischen den Varietäten besonders groß seien. AmE Wörter werden nach Ansicht der Verfasser häufig in das BrE übernommen, während Sprecher des AmE seltener neue Wörter aus dem BrE verwendeten. Danach gehen die Herausgeber zunächst auf den Fall ein, daß z.B. ein Wort oder eine Wendung weder im BrE noch im AusE vorkommt, sondern nur im AmE auftritt und mit *Am* gekennzeichnet wird (*panhandle, math*). Das an dieser Stelle angeführte Beispiel *panhandle* würde in Algeos Kategorisierung lexikalischer Unterschiede in die Klasse 3 "lexical gap" eingeordnet werden, während *math* und (*Br and Aus*) *maths* der Klasse 6 "equivalent terms=interdialectal synonymy" zugehören müßten.³⁷⁷ Im folgenden Unterpunkt werden weitere drei Arten lexikalischer Variationen zwischen dem AmE einerseits und dem BrE bzw. AusE andererseits genannt, die unterschiedlich im Wörterbuch gekennzeichnet werden: 1. Eine im BrE und AusE geläufige Wortform wird im AmE selten verwendet; im AmE tritt für gewöhnlich eine andere synonyme Wortform auf, die manchmal auch im BrE und AusE benutzt wird (Beispiel:

³⁷⁴ Man kann sich denken, daß das Irische und Schottische Englisch unter das BrE fallen sollen und das Kanadische Englisch eher unter das AmE, doch wird dies von den Herausgebern nicht ausdrücklich festgehalten.

³⁷⁵ Vgl. S.790. Hier wird nicht spezifiziert, daß z.B. ein mit *Am* markiertes Wort den Herausgebern zufolge nicht im BrE oder AusE vorkommt. Es wird allerdings auf das *Language Portrait Varieties of English* verwiesen.

³⁷⁶ Die Markierung *Am* wird mit "American English" wiedergegeben statt mit "North American English" und die Erklärung zu der Markierung *regional* weicht auch von der oben genannten Formulierung ab. Vgl. auch die Auflistung der Markierungen unter *Usage labels in the dictionary* auf der ersten Seite des Wörterbuchs.

*biscuit/ (esp Am) cookie*³⁷⁸). Dieser Fall würde in Algeos Kategorisierung ebenfalls in die Klasse 6 gehören, wobei *biscuit* eine "fully limited"³⁷⁹ Form darstellt, während *cookie* eine "partially limited"³⁸⁰ Form ist. 2. Eine im BrE und AusE gebräuchliche Wortform wird auch bisweilen im AmE verwendet, doch im AmE ist eine andere synonyme Wortform üblich, die manchmal im BrE und AusE benutzt wird (Beispiel: *flat/ (Am usually) apartment*). Diese Art der lexikalischen Beziehungen würde in Algeos Kategorisierung ebenfalls in die Klasse 6 fallen, wobei beide Wortformen als "partially limited" zu bezeichnen wären. 3. Das AmE verwendet zusätzlich zu einer gemeinenglischen Wortform eine synonyme Wortform, die im BrE und AusE selten ist (Beispiel: *car/ (Am also) auto*). Dieser Fall würde in Algeos Kategorisierung der Klasse 7 "overlapping terms=one-sided dialect synonymy" entsprechen. Auf Unterschiede zwischen dem AmE und dem BrE in den Bedeutungen mancher Wörter wird in der Darstellung der lexikalischen Unterschiede dieses *Language Portraits* nicht eingegangen.³⁸¹

Eine Liste, die jeweils auf das BrE und AmE beschränkte Wörter, nach Sachgebieten geordnet, in der Art der sog. *Parallel Lists*³⁸² gegenüberstellt, findet sich in dem *Language Portrait 'Australian English'* (S. 81).³⁸³ Auf die Kennzeichnung lexikalischer Unterschiede (und solche in der Orthographie) zwischen dem BrE, AmE und AusE in den Einträgen wird auch in dem Kapitel *How to use the dictionary* (S.x f.) hingewiesen.

Drei andere Abschnitte des *Language Portrait 'Varieties of English'* beschäftigen sich jeweils mit den Unterschieden in der Orthographie, der Betonung und Aussprache zwischen dem AmE und BrE, auf die hier nicht näher eingegangen werden soll, da sie nicht Untersuchungsgegenstand der Arbeit sind.³⁸⁴

Dem Abschnitt *Differences in grammar* soll an dieser Stelle jedoch Aufmerksamkeit geschenkt werden. Nach der Feststellung, daß zwischen dem AmE, AusE und BrE nicht viele grammatikalische Unterschiede bestehen, listen die Herausgeber einige der von ihnen als wichtig erachteten Distinktionen zwischen dem AmE einerseits und dem BrE bzw. AusE andererseits

³⁷⁷ Vgl. für diese und die folgenden Ausführungen Algeo (1989a:227-240).

³⁷⁸ Die Beispiele vorwiegend im BrE und AusE verwandter Synonyme werden in diesem Abschnitt merkwürdigerweise nicht gekennzeichnet.

³⁷⁹ Vgl. Algeo (1989a:231f.).

³⁸⁰ Ebd.

³⁸¹ Vgl. aber S. xi des CIDE₁.

³⁸² Vgl. dazu Algeo (1989a:221 und 231).

³⁸³ Dort werden auch die "Paare" BrE *exdirectory number/ AmE unlisted number* und BrE *luggage van/ AmE baggage car* angeführt, auf die im Rahmen der Überprüfung der von mir ausgewählten Amerikanismen noch einmal eingegangen werden soll.

³⁸⁴ Unterschiede in der Orthographie werden auch in dem *Language Portrait 'Spelling Differences between American and British English'* (S. 40) dargestellt, solche in der Betonung werden ebenfalls in dem *Language Portrait 'Stress in English pronunciation'* (S. 1439) behandelt. Auf Unterschiede in der Aussprache und ihre Darstellung im Wörterbuch wird auch in den *Language Portraits 'Pronunciation'* (S. 1133), *'Pronunciation of words with -ough-'* (S. 999) und *'Pronunciation of 'r' in British and American English'* (S. 1165) sowie in dem Kapitel *Phonetic Symbols* (S.1774) und unter *How to use the dictionary* hingewiesen.

auf. Es werden folgende Unterschiede genannt, die die Redaktion jeweils mit einigen Beispielen veranschaulicht: 1. Das AmE verfüge in bezug auf manche unregelmäßigen Verben auf eine zusätzliche Form des Präteritums oder Partizip Perfekts, die im BrE nicht verwendet werde. 2. Das AmE könne in vielen Situationen auch das *Simple past* verwenden, in denen das BrE die Form des *Present perfect* benutze. Unter den angefügten Beispielen sind zwei, die die Verwendung des Präteritums im AmE nach den Adverbien *already* und *yet* illustrieren und mit (*Am also*) gekennzeichnet werden. Dieses grammatikalische Verhalten des AmE gehört zu den Amerikanismen, auf die in Kapitel 5 und 7 dieser Arbeit genauer eingegangen werden soll. 3. Das AmE und BrE bevorzugen manchmal unterschiedliche Präpositionen. Unter den angegebenen Beispielen befindet sich auch die Verwendung von AmE *after* statt des gemeinenglischen *past* bei der Angabe von Uhrzeiten, die mit (*Am also*) gekennzeichnet wird. Dieser Amerikanismus wird ebenfalls in den Kapiteln 5 und 7 dieser Arbeit detaillierter behandelt. 4. Im BrE können bestimmte Gruppenbezeichnungen, wie z.B. *government*, *public* oder *company*, von einer Verbform im Singular oder Plural gefolgt werden, während das AmE nur die Singularform des Verbs verwende. Auf diesen Unterschied wird auch schon in dem Kapitel *Grammar labels in the dictionary* bei der Erklärung der Markierung [+sing/pl v] hingewiesen.

Die Darstellung der Unterschiede zwischen dem BrE und AmE ist im CIDE₁ also sehr viel ausführlicher als in den meisten anderen untersuchten Wörterbüchern. Es bleibt zu überprüfen, ob die hier angeführten lexikalischen und grammatikalischen Unterschiede auch in den Wörterbucheinträgen entsprechend gekennzeichnet sind.

In der neuen Auflage des CIDE₁ werden die Unterschiede zwischen den Varietäten nicht mehr so eingehend erläutert - zwar gibt es eine sog. *Study Section*³⁸⁵ zu den Varietäten des Englischen, aber keine Informationsseiten mehr zu 'Labels in the Dictionary' oder 'Australian English'. Die Auskünfte in der *Study Section* 'Varieties of English' sind zudem z.B. hinsichtlich der Darstellung der lexikalischen Unterschiede deutlich gekürzt worden.

f) Cassell Concise Dictionary (CACD)

Unterschiede zwischen dem BrE und AmE werden in den einleitenden Kapiteln des Wörterbuchs nicht explizit behandelt. Bezüglich der Darstellung der Unterschiede zwischen den beiden Varietäten erfährt man in dem Kapitel *How to use The Cassell Concise Dictionary* unter der Überschrift *Spellings* (S. viii) nur, daß nordamerikanische Rechtschreibvarianten dort, wo es angebracht erscheint, nach den empfohlenen BrE Schreibweisen angegeben werden. Die BrE

³⁸⁵ Statt von *Language Portraits* wird nun von *Study Sections* gesprochen.

Schreibweisen sind jedoch im Gegensatz zu denen des AmE unmarkiert. AmE Aussprachevarianten werden, wie bereits erwähnt, nur in einigen wenigen *Usage Notes* aufgegriffen, während die BrE Aussprache (in Form der *Received Pronunciation*) nach fast jedem Stichwort im Lexikonteil ohne Markierung aufgeführt ist.

Lexikalische Unterschiede zwischen dem BrE und AmE und ihre Darstellung im Wörterbuch werden in den einleitenden Kapiteln des Wörterbuchs genausowenig angesprochen wie Divergenzen in grammatikalischen Konstruktionen sowie in Wendungen und Kollokationen

Wie schon im Zusammenhang mit anderen Wörterbüchern der *Concise*-Sparte erwähnt, wird auch im CACD für ein AmE Synonym eines BrE Lexems normalerweise keine Bedeutungserklärung angeführt, sondern nur das BrE Lemma als Entsprechung angegeben.

Die Zentrierung der Herausgeber auf das BrE wird besonders dadurch deutlich, daß nicht nur BrE Ausspracheformen und Schreibweisen unmarkiert sind, sondern auch BrE Wortschatz. So findet sich auch in den einleitenden Kapiteln des Wörterbuchs keine Angabe bezüglich einer das BrE betreffenden Markierung. Für die Kennzeichnung der Amerikanismen wird die Markierung *N Am.* benutzt, unter der wahrscheinlich auch das Kanadische Englisch fällt. Die im CACD verwandten diatopischen Markierungen sind leider nur im Abkürzungsverzeichnis unter vielen anderen Abkürzungen zu finden, werden also nirgendwo separat aufgeführt und erläutert.

Vergeblich sucht man in den einleitenden Kapiteln auch nach Hinweisen auf die Markierungen *chiefly N Am.* und *esp. N Am.*, die im Lexikon-Teil beide manchmal verwendet werden (wobei der gleichzeitige Gebrauch dieser beiden Markierungen wohl auf eine inkonsistente Verfahrensweise zurückzuführen sein dürfte).

g) Concise Oxford Dictionary (COD₁₀)

Auf die Unterschiede zwischen dem BrE und AmE und ihre Darstellung im Wörterbuch wird - wie schon in der vorhergehenden Auflage - nicht in einem gesonderten Kapitel eingegangen.

In dem Kapitel *Spelling and Inflection* (S. xii) heißt es jedoch, daß die AmE Standardschreibweise hinter der BrE Standardform angegeben wird und zusätzlich auch einen Querverweis auf die BrE Standardschreibweise erhält, wenn ihre alphabetische Position mehr als vier Einträge vom „Haupteintrag“ entfernt ist. Während die AmE Standardform aber als solche gekennzeichnet ist, ist die des BrE unmarkiert. Unter *Hyphenation* (S.xiii) werden Unterschiede zwischen dem AmE und BrE in der Schreibung von Komposita beleuchtet, wobei schließlich darauf hingewiesen wird, daß nur die Standardform des BrE in das Wörterbuch eingegangen ist („to save space and avoid confusion“). Die Zentrierung der Herausgeber auf das BrE wird auch bei den Ausspracheangaben deutlich, denn in dem Kapitel „Pronunciations“ erfährt der

Wörterbuchbenutzer, daß für das AmE keine Informationen zur Aussprache bereitgestellt werden.³⁸⁶

Auf lexikalische Unterschiede zwischen dem BrE und AmE und ihre Darstellung im Wörterbuch wird genausowenig eingegangen wie auf Divergenzen in grammatikalischen Konstruktionen sowie in Wendungen und Kollokationen.³⁸⁷

Bezüglich der Darstellung der lexikalischen Unterschiede zwischen dem BrE und AmE ist festzuhalten, daß für ein AmE Synonym eines BrE Lexems normalerweise keine Bedeutungserklärung angeführt, sondern nur das BrE Lemma als Entsprechung angegeben wird.

Als Markierungen für das BrE und AmE verwenden die Lexikographen ‚Brit.‘ bzw. ‚US‘.³⁸⁸ Unter *Geographical Labels* (S. x) wird näher erläutert, was mit diesen Etikettierungen gemeint ist. Zu ‚Brit.‘ heißt es: „The geographical label ‚Brit.‘ implies that the use is found typically in standard British English but is not found in standard American English, though it may be found in other varieties such as Australian or South African English.“ Die Markierung ‚US‘ wird folgendermaßen erklärt: „The label ‚US‘, on the other hand, implies that the use is typically US and is not standard in British English, though it may be found elsewhere.“ Wie schon für das OALD₆ gilt auch für das COD₁₀, daß die Verfasser die beiden Varietäten in den Definitionen der Markierungen zueinander in Beziehung setzen und explizit nur ausschließen, daß z.B. ein mit ‚US‘ gekennzeichnetes Wort auch in der Standardform des BrE vorkommt und umgekehrt. Sie präzisieren aber nicht, in welchem geographischem Raum AmE und BrE gesprochen wird und lassen somit offen, ob ein z.B. als ‚US‘ gekennzeichnetes Wort nicht auch in anderen Ländern der Welt, z.B. in Asien, vorkommt.³⁸⁹

Unverständlicherweise wird unter dem Abschnitt *Geographical Labels* auf die Markierung ‚N. Amer.‘ nicht Bezug genommen, obwohl sie im Zusammenhang mit den in der Arbeit untersuchten Amerikanismen viel häufiger verwendet wird als die Markierung ‚US‘. Mit ‚N. Amer.‘ soll mit großer Wahrscheinlichkeit sowohl das AmE als auch das Kanadische Englisch bezeichnet werden. Man müßte bei der Verwendung dieser Markierung also davon ausgehen können, daß sich die Lexikographen sicher sind, daß das jeweilige Wort, die jeweilige Bedeutung o.ä. in beiden Varietäten vorkommt.³⁹⁰

³⁸⁶ Im COD₁₀ werden allerdings im Gegensatz zu der vorhergehenden Auflage Auskünfte zur Aussprache generell nur bei solchen Wörtern gegeben, deren Aussprache schwierig ist.

³⁸⁷ Im COD₉ erhält man allerdings außer bei den Erklärungen zu den Markierungen *Brit.* und *US* gar keine Informationen zu Unterschieden zwischen dem BrE und AmE und ihrer Darstellung im Wörterbuch.

³⁸⁸ Im OALD₆, das bekanntermaßen auch von *Oxford University Press* veröffentlicht wurde, hatte man sich dagegen erstmals für die Markierungen *AmE* und *BrE* entschieden, die den Bezug zur Sprache erkennbarer machen, s.o.

³⁸⁹ Im COD₉ wird in ähnlicher Weise verfahren (vgl. S. xix).

³⁹⁰ Dabei haben die beiden *consultants* für das Nordamerikanische Englisch (vgl. Kap. 4.3.2) zweifellos eine Schlüsselrolle gespielt.

Vergebens sucht man im übrigen in den einleitenden Kapiteln nach Hinweisen auf die Markierungen ‚chiefly US‘ bzw. ‚chiefly N. Amer.‘ oder ‚chiefly Brit.‘, die im Lexikon-Teil oft Verwendung finden.

Im Gegensatz zur vorhergehenden Auflage sind alle Markierungen nicht mehr kursiv, sondern in einer kleineren Schriftgröße und einer anderen Schriftart als der Definitionstext gedruckt. Dadurch wollte man wahrscheinlich Platz sparen. Da die Markierungen in der neuen Darstellungsweise jedoch immer noch deutlich ins Auge fallen, ist an diesem Vorgehen nichts auszusetzen.

h) Collins Concise Dictionary (CCD₅)

Auf die Unterschiede zwischen dem BrE und AmE wird in den einleitenden Kapiteln des Wörterbuchs kaum eingegangen.³⁹¹ Unter *Guide to the Use of the Dictionary* (S. xiii) wird darauf hingewiesen, daß AmE Schreibweisen, die von solchen des BrE abweichen, angeführt werden (wobei die BrE Schreibweisen unmarkiert sind, vgl. z.B. den Eintrag zu *centre*).

Bei den Ausspracheangaben wird auf die Angabe AmE Aussprachevarianten verzichtet, der *Wörterbuchbenutzer* erhält nur Informationen zu der Aussprache, wie sie „in educated British English speech“ (S. xiii) üblich ist.

Lexikalische Unterschiede zwischen dem BrE und AmE und ihre Darstellung im Wörterbuch werden genausowenig angesprochen wie Divergenzen in grammatikalischen Konstruktionen sowie in Wendungen und Kollokationen.

Hinsichtlich der Darstellung der lexikalischen Unterschiede zwischen dem BrE und AmE gilt auch für das CCD₅, daß für ein AmE Synonym eines BrE Lexems normalerweise keine Bedeutungserklärung angeführt, sondern nur das BrE Lemma als Entsprechung angegeben wird. Das Hauptinteresse der Herausgeber gilt also dem BrE.

Als Markierungen für das BrE und AmE verwenden die Lexikographen ‚Brit.‘ bzw. ‚US‘, wie man dem Lexikonteil entnehmen kann. Die im Wörterbuch benutzten diatopischen Markierungen werden jedoch leider an keiner Stelle separat aufgelistet oder definiert (während diastratische, diaphasische und diaevaluative Markierungen unter dem Abschnitt „Usage labels“ erläutert werden). Nur in der Liste der Abkürzungen (S. xix) stößt man z.B. auf ‚Brit.‘ („Britain; British“) und ‚US‘ („United States“). Es ist fast unnötig zu sagen, daß der Wörterbuchbenutzer somit auch nur aus den Einträgen selbst ersehen kann, daß z.B. eine Markierung wie *chiefly US* im Lexikonteil vorkommt.

³⁹¹ Dasselbe gilt auch für die vorgehende Auflage.

i) Chambers 21st Century Dictionary (CH21₂)

In den einleitenden Kapiteln des Wörterbuchs finden Unterschiede zwischen dem BrE und AmE keinerlei Erwähnung. Dem Vorwort kann man nur entnehmen, daß ausschließlich die BrE Aussprache der Stichwörter angegeben wird.

Die genaue Analyse der Einträge im Lexikonteil bringt allerdings zu Tage, daß divergierende AmE Schreibweisen der Stichwörter angeführt werden (vgl. z.B. den Eintrag zu *colour* oder *theatre*), wobei die BrE Schreibweisen jedoch nicht markiert sind.

Auch hinsichtlich der Behandlung der lexikalischen Unterschiede zwischen dem BrE und AmE wird deutlich, daß das Hauptaugenmerk der Herausgeber auf dem BrE liegt, denn für ein AmE Synonym eines BrE Lexems wird in der Regel keine Bedeutungserklärung angeführt, sondern nur das BrE Lemma als Entsprechung angegeben.

Die von den Lexikographen verwendeten diatopischen Markierungen werden in den einleitenden Kapiteln des Wörterbuchs leider genausowenig wie die diaphasischen separat aufgelistet und definiert. In der Liste der Abkürzungen (S. x) stößt man zwar z.B. auf ‚Brit.‘ (‚Britain; British‘) und ‚US‘ (‚United States‘), aber nur anhand der Einträge im Lexikonteil läßt sich ersehen, daß für das BrE die Markierung ‚Brit.‘ und für das AmE ‚US‘ bzw. *N Amer* benutzt wird. Vor allem, weil im Zusammenhang mit den Amerikanismen einmal ‚US‘ und ein anderes Mal *N Amer* vorkommt (vgl. z.B. *Britisher* oder *dumb* und *chipper* oder *chutzpah*), hätte man sich eine Erläuterung zu den Verwendungsweisen der Markierungen gewünscht. Den Angaben einer Mitarbeiterin des Verlags nach bedeutet die Markierung *N Amer* jedoch nicht - wie man annehmen könnte - daß das so Gekennzeichnete sowohl in den USA als auch in Kanada gebräuchlich ist. Vielmehr scheint es sich hier um eine inkonsistente Verwendung der Markierungen *N Amer* und *US* zu handeln: Denn es war wohl eigentlich vorgesehen, mit *N Amer* ein AmE Lexem, mit *US* dagegen eine AmE Schreibweise oder Bedeutung zu kennzeichnen (obwohl dieses Vorgehen für einen Außenstehenden nicht sehr einsichtig ist). Wie die bereits erwähnte Mitarbeiterin zugab, ist die genannte Unterscheidung jedoch während des redaktionellen Prozesses etwas in Vergessenheit geraten. Das läßt sich auch anhand einiger im Rahmen dieser Arbeit untersuchten Amerikanismen nachweisen (vgl. z.B. *Britisher*, *dumb* und *to table*). Inkonsistent ist leider auch die Art und Weise, wie die Markierung *N Amer* bzw. *US* modifiziert wird: Neben *chiefly* wird nämlich auch *esp* verwendet, um anzuzeigen, daß ein AmE Lexem o.ä. nicht nur auf das AmE beschränkt ist (vgl. z.B. *mailman*, das mit *esp N Amer* gekennzeichnet wird, und *unlisted*, das die Markierung *chiefly N Amer* erhalten hat).

j) The New Penguin English Dictionary (NPED₁)

Auf die Unterschiede zwischen dem BrE und AmE und ihre Darstellung im Wörterbuch wird nicht in einem gesonderten Kapitel eingegangen. In dem Kapitel *Alternative spellings* (S. viii) des *Guide to the Dictionary* findet sich jedoch der Hinweis, daß AmE Schreibweisen hinter der BrE Standardform angegeben und mit *NAmer* markiert werden (während die BrE Schreibweisen unmarkiert sind). Da der Begriff *North American* sich aber - übereinstimmend mit dem entsprechenden Wörterbucheintrag im NPED₁ - sowohl auf die USA als auch Kanada bezieht, hätte in diesem Zusammenhang eigentlich von Schreibweisen des Amerikanischen und Kanadischen Englisch die Rede sein müssen. Augenscheinlich sind sich die Herausgeber oft nicht ganz sicher, für welchen spezifischen geographischen Raum eine Schreibweise gilt. So heißt es auch in demselben Kapitel zur Markierung *NAmer*:

The distinction implied here is not absolute, since usage varies considerably within North America. Spellings labelled as American are sometimes also found in British English and in other varieties of English; conversely British spellings are also sometimes found in American English and in other varieties. Information of this kind indicates what is typical rather than what is invariable. (S. viii)

Dieser Hinweis ist selbstverständlich berechtigt und sehr zu begrüßen. Er sollte auch in anderen Wörterbüchern zu finden sein. Allerdings kann der Gebrauch von varietätenspezifischen Korpora, die in Größe und Struktur vergleichbar sind, bei dem Problem des geographischen Geltungsraums eines Wortes, seiner Schreibweise, seiner Aussprache u.ä. eine große Hilfe sein, wie in Kap. 3.1.6 ausgeführt wird. Da der *Penguin* - Verlag - soweit bekannt - bei seinen Wörterbuchproduktionen bisher jedoch über keinerlei Korpusbasis verfügt und bei der Erstellung des NPED₁ auch keine muttersprachlichen Lexikographen anderer Varietäten außer dem BrE herangezogen wurden, wird die Vergabe von diatopischen Markierungen im Wörterbuch besonders schwierig gewesen sein (s. auch unten).

In dem Kapitel *Pronunciation* (S. ix) wird zwar festgestellt, daß die Aussprachekonventionen auf den Britischen Inseln und im gesamten englischsprachigen Gebiet variieren können, aber gleichzeitig wird konstatiert, daß im NPED₁ nur die Ausspracheformen der *Received Pronunciation* dargestellt werden. AmE Ausspracheformen werden somit nicht im Wörterbuch berücksichtigt.

Auf lexikalische Unterschiede zwischen dem BrE und AmE wird in den einleitenden Kapiteln des NPED₁ nicht speziell eingegangen. In dem Kapitel *Regional labels* (S.ix) wird nur allgemein festgestellt, daß Wörter und Bedeutungen, die hauptsächlich in bestimmten regionalen Varietäten des Englischen verwendet werden, entsprechend gekennzeichnet sind. Divergenzen zwischen den Varietäten in grammatikalischen Konstruktionen sowie in Wendungen und Kollokationen finden keinerlei Erwähnung.

Als Markierung für das BrE verwenden die Lexikographen *Brit.* Amerikanismen werden wie bereits erwähnt - mit *NAmer* gekennzeichnet, worunter wohl auch das Kanadische Englisch gefaßt wird. Leider wird nicht näher erläutert, wie diese Etikettierungen definiert sind. Die Herausgeber weisen nur noch einmal darauf hin, daß die Unterschiede zwischen den Varietäten des Englischen nicht immer eindeutig sind:

Again distinctions are not always clear-cut. For example, many uses identified as North American or as British are also used in other parts of the English-speaking world, such as Australia and New Zealand, and uses identified as Australian may also have filtered into North American use and even into occasional British use. (S. ix)

Aus diesen Ausführungen läßt sich zumindest entnehmen, daß die Herausgeber ausschließen, daß ein mit *NAmer* gekennzeichnetes Wort auch in der Standardform des BrE vorkommt und umgekehrt.

Anhand des Eintrags zu *bagman*, der zur Veranschaulichung der Verwendung von Markierungen im NPED₁ auf derselben Seite abgedruckt ist, läßt sich erkennen, daß im Lexikonteil auch *chiefly NAmer* oder *chiefly Brit* benutzt wird. Mit diesen Markierungen soll vermutlich zum Ausdruck gebracht werden, daß z.B. ein so gekennzeichnetes Wort auch in der jeweils anderen Varietät - natürlich deutlich weniger häufig - vorkommt. Die Markierung *chiefly NAmer* ist mehrmals in den Einträgen der analysierten Amerikanismen zu finden.

5. Begründung der Auswahl der untersuchten Amerikanismen

Aufgrund des begrenzten Rahmens dieser Arbeit beschränkt sich die stichprobenartige Überprüfung der Erfassung und Markierung von Amerikanismen in den Einträgen der zehn ausgewählten Wörterbücher auf 30 lexikalische und 5 grammatikalische Amerikanismen, wobei die Aufnahme und Kennzeichnung von vorwiegend oder ausschließlich im AmE auftretender grammatikalischer Variation in der Regel jeweils anhand mehrerer Lemmata überprüft wird.

Die Konzentration auf lexikalische Amerikanismen läßt sich zum einen dadurch begründen, daß die Divergenzen zwischen dem BrE und AmE in der Lexik weitaus zahlreicher sind als beispielsweise solche in der Orthographie oder Grammatik.

Der Fülle lexikalischer Amerikanismen muß in einem britischen Wörterbuch aber auch aufgrund des starken Einflusses des AmE auf andere Varietäten Beachtung geschenkt werden³⁹². So bezeichnet Algeo (2001:xxii) den Wortschatz des AmE als „the most innovative and influential aspect of American English“. Ein großer Teil dieser Wortformen und Wortbedeutungen ist nicht nur auf die USA beschränkt, sondern stellt in anderen englischsprachigen Ländern, wie z.B. Kanada oder Australien, oft die gültige Norm dar, und wird ebenfalls zunehmend - wenn auch noch weniger häufig als in den USA - in Großbritannien verwendet. In manchen Fällen muß sogar entschieden werden, ob eine Wortform oder Wortbedeutung überhaupt noch als Amerikanismus anzusehen ist, oder ob sie schon dem Gemeinenglischen zugerechnet werden muß.

Zudem besteht im Bereich der Lexik die wahrscheinlich größte Variationsbreite bezüglich der Unterschiede zwischen dem BrE und AmE, während z.B. Divergenzen in der Aussprache oder Orthographie in großem Maße Regelmäßigkeiten aufweisen. Abweichungen im Wortgebrauch zwischen dem AmE und BrE können daher öfter zu Verständigungsschwierigkeiten führen als in anderen Bereichen. So heißt es bei Trudgill/ Hannah (1985:45): "Vocabulary differences [...] are very numerous and are capable of causing varying degrees of comprehension problems". Das Zweiwortlexem *first floor* ist ein gutes Beispiel dafür, wie Verständnisprobleme zwischen Sprechern des BrE und AmE entstehen können: Im BrE hat dieses Zweiwortlexem für gewöhnlich die Bedeutung 'erster Stock', während es im AmE in der Regel für 'Erdgeschoß' steht. Für 'erster Stock' wird im AmE normalerweise *second floor* verwendet, während im BrE für 'Erdgeschoß' gewöhnlich *ground floor* gebraucht wird.³⁹³ Die präzise Wiedergabe der lexikalischen Unterschiede in den Lernerwörterbüchern ist nicht zuletzt also auch deshalb wichtig, damit Lerner des Englischen die Bedeutungen solcher Lexeme in der

³⁹² Vgl. auch die Einleitung.

jeweiligen Varietät richtig verstehen und diese Lexeme selber ebenfalls richtig im BrE und AmE verwenden. In den Wörterbüchern für Muttersprachler sollten diese Divergenzen ebenfalls verzeichnet sein, um die Kommunikation zwischen den Sprechern des BrE und AmE zu erleichtern.

Die Überprüfung der Darstellung lexikalischer Unterschiede zwischen dem BrE und AmE in den Wörterbüchern ist daher auch aus dem folgenden Grund von Interesse: Die präzise Wiedergabe dieser Unterschiede ist nämlich schwieriger als z.B. die orthographischer oder phonologischer Distinktionen, da die beiden Varietäten gerade im Bereich der Lexik durch das Zusammenspiel einer teilweisen, überlappenden, gegenseitigen Beeinflussung miteinander verbunden sind, und nur gelegentlich direkte Entsprechungen bestimmter varietätenspezifischer Formen auftreten.³⁹⁴ Die Komplexität der lexikalischen Unterschiede zwischen dem BrE und AmE spiegelt sich in einer Vielzahl von Kategorisierungsversuchen wider, unter denen sich die Typologie John Algeos (1989a) in seinem Aufsatz "British-American lexical differences: A typology of interdialectal variation" durch ihre Differenziertheit besonders auszeichnet.³⁹⁵ Von den in den zehn Wörterbüchern überprüften lexikalischen Amerikanismen stammen 14 aus dieser Typologie, da man anhand dieser sehr gut feststellen kann, ob die Wörterbuchherausgeber nur AmE Synonyme BrE Lexeme einbezogen haben (wie sie z.B. in den sog. *Parallel Lists*³⁹⁶ vorkommen) oder auch Wortformen und Wortbedeutungen in die Wörterbücher korrekt aufgenommen und entsprechend markiert wurden, die zum BrE in komplexerer Beziehung stehen.³⁹⁷ Der Kategorisierung von Algeo wurden somit folgende Lexeme entnommen³⁹⁸:

- *block*
- *junior college*
- *unlisted*
- *package*
- *mailman*
- *fall*
- *Britisher*
- *robin*
- *chancellor*

³⁹³ Vgl. die *usage note* zu *floor* im LDOCE₃, S. 537.

³⁹⁴ Vgl. Algeo (1989a:221).

³⁹⁵ Görlach bezeichnet Algeos Typologie als "insightful discussion of types of equivalences" (1990a:239).

³⁹⁶ Vgl. zu dem Begriff *Parallel List* Algeo (1989a:221).

³⁹⁷ Algeo (1989a:234) weist z.B. darauf hin, daß in fünf britischen Wörterbüchern die von ihm erläuterte AmE Bedeutung von *chancellor* nicht richtig dargestellt worden sei. Auf welche Wörterbücher er sich bezieht, wird leider nicht erwähnt.

³⁹⁸ Algeos Angaben zu diesen Wörtern werden in Kap. 7 wiedergegeben.

- *to table*
- *staff*
- *baggage*
- *taffy*
- *granola*

Drei weitere in dieser Arbeit behandelte Lexeme stammen aus Algeos Aufsatz „Types of English Heteronyms“ (1997), in dem der Autor eine differenzierte Typologie verschiedener Arten von Heteronymie zwischen dem BrE und AmE vorlegt. Auch hier schien es lohnenswert, nachzuprüfen, inwieweit einige der angeführten Amerikanismen in den zehn Wörterbüchern korrekt erfaßt und markiert worden sind. Bei den ausgewählten drei Lexemen handelt es sich um:

- *outhouse*
- *undershirt*
- *hockey*.

Weitere sieben Amerikanismen wurden dem Aufsatz von Matsuda (1990) „The American Impact on British English“ entnommen. Die Untersuchung dieser Amerikanismen in den hier behandelten Wörterbüchern erschien deswegen besonders interessant, weil Matsuda ebenso wie die Verfasserin der vorliegenden Arbeit britische Muttersprachler des Englischen zu ihrem Gebrauch bestimmter Amerikanismen befragt hat, um dadurch beurteilen zu können, ob die betreffenden Amerikanismen in den von ihm analysierten Wörterbüchern sowie bei Benson/ Benson/ Ilson (1986) richtig markiert worden sind. Im Zusammenhang mit den oben erwähnten sieben Amerikanismen weist Matsuda (1990) aufgrund seiner empirischen Studien darauf hin, daß die Markierung der Bedeutungen dieser sechs Lexeme sowie der idiomatischen Wendung *right off* bei Benson/ Benson/ Ilson (1986) (und in verschiedenen Wörterbüchern) nicht zutreffend sei. In einem der sieben Fälle ist er sogar der Ansicht, daß das Lexem schon dem Gemeinenglischen zuzuordnen sei, also gar kein Amerikanismus mehr ist. Da die empirische Untersuchungen, die die sieben Amerikanismen betreffen, zum einen bereits länger zurückliegen (sie stammen aus den 80er Jahren) und zum anderen nur mit einer begrenzten Zahl von Universitätsangehörigen durchgeführt wurden³⁹⁹, bot sich eine neuerliche Untersuchung der Amerikanismen mit Hilfe der im Rahmen dieser Arbeit erstellten Fragebögen an. Dem Aufsatz von Matsuda wurden entnommen⁴⁰⁰:

³⁹⁹ Vgl. Matsuda (1990:80). Er gibt an, bei drei Gelegenheiten Befragungen durchgeführt zu haben, in denen jeweils der Gebrauch eines Teils der von ihm in seinem Aufsatz behandelten Lexeme und Wendungen überprüft wurde. An den Befragungen waren einmal 42 Personen, einmal 49 und einmal 50 Personen beteiligt.

⁴⁰⁰ Matsudas Angaben zu diesen Lexemen und der Wendung, die sich problemlos auch in Algeos Kategorisierung einordnen lassen würden, werden detailliert in Kap. 7 erörtert.

- *right off*
- *dry goods*
- *clipping*
- *can* n.
- *graft* n.
- *billion*
- *dumb*

Die übrigen sechs lexikalischen Amerikanismen stammen aus dem Artikel „Regional Labels in some British and American Dictionaries“ von Norri (1996). Norri vergleicht den Gebrauch von diatopischen Markierungen in sieben britischen und drei amerikanischen Wörterbüchern. Er führt somit ähnliche Untersuchungen durch wie die Verfasserin dieser Arbeit, wobei das Interesse von Norri jedoch nicht nur der Darstellung von Amerikanismen in britischen Wörterbüchern gilt und er auch nicht auf eigene Befragungen von Muttersprachlern zurückgreifen kann. Dessenungeachtet sind die Ergebnisse von Norris Analysen in Bezug auf die Behandlung einer Reihe von Amerikanismen in britischen Wörterbüchern des Englischen aufschlußreich, da sie zum Teil große Divergenzen in der Vergabe von Markierungen nicht zuletzt auch in solchen Wörterbüchern aufzeigen, die in dieser Arbeit behandelt werden (z.B. OALD, COD, LDELIC, CHD und COBUILD). Allerdings sind die von Norri benutzten Auflagen älter als die in der vorliegenden Arbeit verwandten, weswegen es interessant erschien, zu überprüfen, ob die Markierungen der sechs ausgewählten Amerikanismen in den neueren Auflagen überarbeitet worden sind. Von diesen sechs lexikalischen Amerikanismen lassen sich zudem mindestens drei dem Slang zuordnen. Der Einbezug von AmE Slang in die Untersuchungen ist deswegen besonders sinnvoll, weil Slang-Ausdrücke als für das AmE charakteristisch erachtet werden können.⁴⁰¹ Aus Norris (1996) Aufsatz wurden diese Lexeme entnommen:

- *chipper*
- *cop out*
- *sashay*
- *patsy*
- *nerd*
- *chutzpah*

⁴⁰¹ Vgl. Algeo (2001:xxiii) und Lighter (2001:251).

Die in dieser Arbeit behandelten Lexeme werden zum größten Teil auch in der ausführlichen Liste lexikalischer Unterschiede zwischen dem AmE und dem BrE (bzw. Gemeinenglischen) von Benson/ Benson/ Ilson (1986:102-152) aufgeführt.⁴⁰²

Grammatikalische Unterschiede zwischen dem BrE und dem AmE sind weitaus geringer als lexikalische und führen auch nicht so sehr zu Verständnisproblemen⁴⁰³. Dennoch ist auch die Aufnahme grammatikalischer Amerikanismen in die Lernerwörterbücher wichtig, damit Lerner des Englischen vorwiegend oder ausschließlich im AmE auftretende grammatikalische Phänomene verstehen und auch selber im AmE verwenden können.

Es wurden grammatikalische Amerikanismen ausgewählt, die in der Forschungsliteratur oft erwähnt werden. Die von mir ausgesuchten Amerikanismen lassen sich nicht nur bei Benson/ Benson/ Ilson (1986) finden, sondern werden sehr häufig auch in anderen Darstellungen der grammatikalischen Unterschiede zwischen dem BrE und AmE genannt, beispielsweise bei Algeo (1988), Gramley und Pätzold (1992), Trudgill/ Hannah (1985), Butters (2001), im OCEL und in *The Right Word at the Right Time*. Folgende grammatikalische Amerikanismen wurden in den Wörterbüchern überprüft:

1. Der häufigere Gebrauch des Konjunktiv Präsens im AmE nach Verben und Adjektiven des Forderns, Drängens oder Vorschlagens, so z.B. nach
 - 1a) *important*
 - 1b) *to insist*
 - 1c) *to demand*

2. Die Verwendung anderer Präpositionen als im BrE bzw. die Auslassung von Präpositionen, z.B. in folgenden Fällen:
 - 2a) *to protest sth*
 - 2b) *a new lease on life*
 - 2c) *to cater to*
 - 2d) *to work on weekends/ to work weekends*
 - 2e) *after* (in Zeitangaben, neben *past*)
 - 2f) *in* (in Ausdrücken, die eine Zeitspanne wiedergeben, statt *for*)

3. Die häufigere Verwendung von *have* als Vollverb in der Bedeutung 'besitzen, haben' im AmE in Fragesätzen und Negationssätzen, die mit *do* gebildet werden

⁴⁰² Görlach nennt diese Auflistung lexikalischer Unterschiede zwischen dem BrE und AmE bei Benson/ Benson/ Ilson "the best list of AmE vs. BrE 'pairs' available to date"(1990:239).

⁴⁰³ Vgl. z.B. Algeo (1988:2).

-
4. Der häufige Gebrauch des *past tense* im AmE im Zusammenhang mit bestimmten Adverbien, z.B. mit
- 4a) *yet*
 - 4b) *already*
 - 4c) *just*
5. Der sich vom BrE manchmal unterscheidende Gebrauch der Artikel, z.B. in folgenden Fällen:
- 5a) *to go to the hospital*
 - 5b) *in the future* (in der Bedeutung 'von jetzt an')
 - 5c) *a half hour*

6. Die Fragebogenuntersuchung

6.1 Einleitung

Nachdem offensichtlich geworden war, wie sehr die in den zehn untersuchten Wörterbüchern vorhandenen Auskünfte zu den in dieser Arbeit analysierten Amerikanismen stellenweise divergieren, stellte sich die Frage, welche Möglichkeiten bestehen, die Informationen der Lexikographen zu überprüfen. Eigene empirische Untersuchungen schienen sich am besten dazu anzubieten, die Angaben in den zehn Wörterbüchern beurteilen zu können, denn die in anderen Wörterbüchern, Grammatiken oder in der Forschungsliteratur bereitgestellten sekundären Auskünfte sind als alleinige Informationsquelle zu diesem Zwecke nicht verlässlich genug. Die Entscheidung für die Durchführung eigener empirischer Forschungen schloß jedoch die Erstellung einer Belegsammlung aus, da Belegsammlungen, abgesehen von dem mit ihnen verbundenen großen Zeit- und Materialaufwand, im Vergleich zu Korpora erhebliche Nachteile aufweisen (vgl. Kapitel 3.1.4). Der Aufbau eines eigenen ausgewogenen Korpus in einer angemessenen Größe mußte selbstredend von vornherein ausgeschlossen werden. Es war allerdings möglich, das BNC zur Überprüfung der in dieser Arbeit untersuchten Lexeme heranzuziehen. Da das BNC, wie bereits in Kap. 3.2.2.2 dargelegt, jedoch (fast) nur BrE Sprachdaten enthält, gaben die BNC-Recherchen vor allem Aufschluß über die Gebrauchshäufigkeit der analysierten Amerikanismen im BrE und somit Hinweise auf die Validität der Markierungen in den Wörterbüchern. Informationen zur Verwendung der ausgewählten Amerikanismen im AmE konnten dagegen auf diese Weise nicht gewonnen werden. Zudem stellte auch das Alter der für das BNC verwandten Quellen sowie der relativ geringe Prozentsatz mündlicher Sprachdaten ein Problem dar (vgl. Kap. 3.2.2.2). Eine Informantenbefragung schien somit ebenfalls sinnvoll zu sein, um dadurch weitere wichtige Auskünfte zu den untersuchten Amerikanismen zu gewinnen. Nach Retti (1999:1)

bietet sich das Instrument der Informantenbefragung als Instanz zu Verifizierungen und Korrektur dessen, was in Wörterbüchern zu finden ist, an, stellt es doch einen Rückgriff auf die Sprachkompetenz der Sprachteilnehmer an sich dar und ermöglicht so, Quellenmaterial von den Trägern der Sprachgemeinschaft den mehrfach vermittelten Informationen der Wörterbücher gegenüberzustellen.

Schriftliche Befragungen werfen jedoch auch eine Reihe von Problemen auf: Da die Informanten die Fragebögen nicht im Beisein des Fragestellers ausfüllen, kann man nicht sicher sein, ob die Fragen in jedem Falle richtig begriffen und der Frageintention gemäß beantwortet werden. Man kann davon ausgehen, daß die Gewährspersonen auftretende Verständnisprobleme nicht in jedem

Falle zu erkennen geben.⁴⁰⁴ Zudem kann der Leiter der Fragebogenerhebung selber keine Rückfragen an die Informanten stellen, wenn er Schwierigkeiten damit hat, den Gedankengang in den Antworten nachzuvollziehen, oder wenn er die Ausführungen der Gewährspersonen aufgrund unleserlicher Schrift kaum entziffern kann. Weiterhin bleibt dem Fragesteller verborgen, wieviel Zeit sich die Informanten für die Beantwortung des Fragebogens genommen haben, wie sorgfältig sie also die Aufgabenstellungen studiert haben und auf diese eingegangen sind. Überdies ist nicht auszuschließen, daß im Verlauf der Bearbeitung des Fragebogens Ermüdungserscheinungen und (bei sprachlichen Untersuchungen) Veränderungen des Sprachgefühls auftreten.⁴⁰⁵

6.2 Informationen zur durchgeführten Befragung

6.2.1 Die Fragebogenkonzeption

Um bei divergierenden Informationen zu den Amerikanismen in den zehn Wörterbüchern Auskünfte von *native speakers* zum Gebrauch der untersuchten Wörter, Wortbedeutungen, Wendungen und grammatikalischen Konstruktionen zu erhalten, entwarf die Verfasserin dieser Arbeit zwei unterschiedliche Fragebögen: einen für Sprecher des BrE (Fragebogen B) und einen für Sprecher des AmE (Fragebogen A).⁴⁰⁶ Die Anfertigung eines Fragebogens für Sprecher des AmE war notwendig, um herauszufinden, ob und wie einige der als Amerikanismen erachteten Wörter, Wortbedeutungen, Wendungen und grammatikalischen Konstruktionen von den Sprechern des AmE tatsächlich verwendet werden. Wie sich nämlich z.B. bei *Britisher* herausstellte, scheinen einige Wörter, die als AmE angesehen werden, auch im AmE kaum noch gebräuchlich zu sein, müßten also zusätzlich noch mit einem *label* wie z.B. *old-fashioned* gekennzeichnet werden.

Während der Teil, der das grammatikalische Verhalten betrifft, in den Fragebögen A und B jeweils gleich ist⁴⁰⁷, unterscheidet sich der auf die Lexik bezogene erste Teil in vielen Punkten, da einige Fragestellungen nur in bezug auf eine Varietät wichtig erschienen. Die in den Fragebögen abgedruckten Fragen wurden möglichst genau auf die divergierenden Informationen in den Wörterbüchern abgestimmt. In bezug auf den Fragebogen für Sprecher des BrE bemühte sich die Verfasserin dieser Arbeit darum, hinsichtlich der Reihenfolge der abgefragten *items* zwischen Amerikanismen, die von vielen britischen Muttersprachlern eindeutig als

⁴⁰⁴ Vgl. dazu auch Zöfgen (1994:42 f.)

⁴⁰⁵ Vgl. Bergenholtz/ Mugdan (1990:1617).

⁴⁰⁶ Zwei Muster der Fragebögen befinden sich im Anhang B.

⁴⁰⁷ Im Fragebogen A beginnt der grammatikalische Teil mit Frage 25, im Fragebogen B mit Frage 24.

Amerikanismen identifiziert werden und solchen, bei denen dies nicht der Fall ist, abzuwechseln (zu diesem Zweck wurden dementsprechende Auskünfte von *native speakers* eingeholt). So sollte vermieden werden, daß die britischen Informanten sofort bemerkten, worum es in ihrer Befragung ging, nämlich die Verwendung von Amerikanismen im BrE. Diese Vorgehensweise sollte verhindern, daß viele der britischen Informanten aufgrund ihrer Abneigung gegen Amerikanismen angaben, die betreffenden Wörter, Wendungen und grammatikalische Phänomene nie zu verwenden oder seltener zu benutzen, als es vielleicht tatsächlich der Fall ist. Die eigenen Beobachtungen der Verfasserin dieser Arbeit bestätigen, daß die Einstellung der Informanten gegenüber Amerikanismen einen Einfluß auf die Fragebogenergebnisse haben kann. Deswegen wurde zu Beginn der für die britischen Informanten konzipierten Fragebögen auch nur darüber informiert, daß das Ziel der Untersuchung sei, den tatsächlichen Gebrauch bestimmter sprachlicher Phänomene im BrE mit den Angaben in Wörterbüchern zu vergleichen.

Da Klassifikationsmerkmale wie Herkunft, Alter, Geschlecht oder Bildungsgrad Ergebnisse von Befragungen nachweislich beeinflussen können, wurden alle Gewährsleute gebeten, vor der Beantwortung der Fragen diesbezügliche Informationen zu ihrer Person zu geben. Eine möglichst große Streuung innerhalb der genannten Parameter erschien wünschenswert, um ein möglichst ausgewogenes Bild der Verwendung von Amerikanismen in Großbritannien bzw. des Gebrauchs bestimmter sprachlicher Phänomene des AmE in den USA zu erhalten. Diese Auffassung soll kurz anhand des Parameters ‚regionale Varietät‘ begründet werden. Die Verteilung der Fragebögen an Informanten verschiedener regionaler Sprachräume in Großbritannien ist u.a. deshalb sinnvoll, weil selbst viele gebildete Sprecher durchaus manchmal auf lokal gebräuchliche Ausdrücke zurückgreifen (vgl. das CED, S. xxv). Im Süd-Westen ist z.B. dialektal noch das Wort *fall* für *autumn* üblich, das im Standardenglischen in Großbritannien nicht mehr gebräuchlich ist. Zudem ist anzunehmen, daß der Süd-Osten mit der Metropole London noch stärker dem AmE Einfluß ausgesetzt ist als andere Regionen des Landes.

Die Kontaktpersonen in Großbritannien und in den USA wurden daher gebeten, bei der Weitergabe der Fragebögen an britische bzw. US-amerikanische Testpersonen ihrer Wahl auf die Streuung innerhalb der o.g. vier Parameter zu achten.

Um den Gewährspersonen die Mitteilung von Informationen zu ihrem Bildungsgrad und zu ihrer regionalen Varietät zu erleichtern, wurden bestimmte Vorgaben gemacht, die nur noch angekreuzt werden mußten. Der Karte zu den regionalen Varietäten in Großbritannien liegt die von Peter Trudgill (1990:33) zugrunde, der sich über Jahrzehnte hinweg mit den regionalen Varietäten des Englischen auf den Britischen Inseln und auf der ganzen Welt beschäftigt hat.⁴⁰⁸

⁴⁰⁸ Die Verfasserin greift auf die Karte Trudgills zu den traditionellen Dialektgebieten (*Traditional Dialect Areas*) zurück, da es viel mehr regionale Variation im Wortschatz der traditionellen Dialekträume gibt als in den sog.

Die Karte zu den regionalen Varietäten des AmE basiert auf derjenigen, die bei Craig Carver (1987:248) abgedruckt ist. Carvers sprachräumliche Gliederung des AmE fußt vornehmlich auf den umfangreichen Feldforschungen für das renommierte *Dictionary of American Regional English (DARE)*.

Die von Matsuda (1990) verwandte Tabellenform, mit der er die Ergebnisse seiner empirischen Untersuchungen darstellt, wurde zunächst von der Verfasserin dieser Arbeit weiterentwickelt und dann in den Fragebögen häufig eingesetzt. Die Tabellenform schien oftmals die zweckmäßigste und gleichzeitig die differenzierte Methode zu sein, um die Verwendung bestimmter Amerikanismen festzustellen.

Die Fragen, Frageformulierungen und der Aufbau der Fragebögen wurden vor der eigentlichen Erhebung durch jeweils zwei Muttersprachler des BrE und AmE getestet. Ihre Anregungen wurden in der Fragebogenkonzeption berücksichtigt. Zudem hatte die Verfasserin dieser Arbeit bereits 1996 eine ähnliche Studie zum Gebrauch von Amerikanismen im BrE durchgeführt und konnte von den in dieser Befragung gewonnenen Erfahrungen profitieren.

6.2.2 Die Durchführung der schriftlichen Befragung

Von Juli bis September 1999 wurden insgesamt 230 Fragebögen für Sprecher des BrE an Kontaktpersonen in Großbritannien weitergegeben bzw. in geringem Maße von der Verfasserin dieser Arbeit in Großbritannien selbst verteilt. Von diesen 230 Exemplaren wurden 176 ausgefüllt zurückgeschickt oder zurückgegeben. Leider konnten aber nur 156 gewertet werden, wofür es verschiedene Gründe gab: Bei 10 Informanten mußte angenommen werden, daß ihr Sprachgebrauch durch das AmE beeinflusst ist, da sie z.B. angegeben hatten, eine Zeitlang in den USA gelebt zu haben (bzw. dort sogar noch zu leben), mit einem US-Amerikaner verheiratet zu sein, für eine US-amerikanische Firma gearbeitet oder eine Ausbildung an einer US-amerikanischen Hochschule genossen zu haben. Weitere sechs Personen erwähnten, einen längeren Aufenthalt in Kanada bzw. in Australien verlebt zu haben, wodurch ihr Sprachgebrauch ebenfalls beeinflusst sein könnte. Unter den zurückgesandten Fragebögen waren zudem trotz

Modern Dialects: Die Unterteilung Englands in traditionelle u. moderne regionale Varietäten des Englischen (englisch *Traditional Dialects* u. *Modern Dialects*) beruht auf Untersuchungen im Bereich der Aussprache und Grammatik; regionale Variationen bezüglich des Wortschatzes gibt es zwar auch in den Gebieten der *Modern Dialects*, viel häufiger lassen sie sich jedoch auf die traditionellen Dialektgebiete beziehen.

Einige der britischen Informanten waren allerdings mit der vorgegebenen sprachräumlichen Gliederung nicht zufrieden (wahrscheinlich hätten sie die Karte zu den *Modern Dialects* vorgezogen). Ihre Kritik richtete sich zum einen gegen die Lage bestimmter Isoglossen (ein Informant bemängelte z.B., dass *Manchester English* mehr Gemeinsamkeiten mit dem Nordenglischen habe als mit der Varietät des Englischen, die in den Midlands gesprochen wird), zum anderen aber auch gegen die wenig differenzierten Untergliederungen der Sprachräume. Eine noch detailliertere Vorgabe von regionalen Varietäten innerhalb Großbritanniens wäre allerdings nicht

gegenteiliger Anweisungen an die Kontaktpersonen auch zwei von Gewährsleuten nicht-britischer Nationalität⁴⁰⁹, die ebenso wenig gewertet wurden wie die von jeweils einer Gewährsperson deutsch-britischer bzw. libanesisch-britischer Nationalität. Bei den letzteren beiden Informanten konnte nicht verifiziert werden, ob sie schon lange genug in Großbritannien lebten, um dem Zweck der Fragebogenuntersuchung dienlich sein zu können. Die insgesamt zwanzig Fragebögen, die nicht gewertet werden konnten, wurden lediglich bei solchen Antworten berücksichtigt, die Meinungsäußerungen betrafen.

Über die Sozialdaten der 156 britischen Informanten, deren Angaben gewertet wurden, können folgende Auskünfte gegeben werden:

Briten: 156 Gewährspersonen

Alter:

Statistics	0-19	20-29	30-39	40-49	50-59	60-69	>70
Total	9	23	46	28	32	12	5

Geschlecht:

Statistics	male	female
Total	61	95

Bildungsgrad:

Statistics	school to 14-16	school to 18	university/college	postgraduate
Total	33	14	59	50

Regionale Varietät:

Statistics	English English						Welsh English	Scottish English	Northern Irish English	Other ⁴¹⁰
	Northern	Western Central	Eastern Central	Western	Eastern	Urban Area of London				
Total	8	24	17	25	21	12	4	23	2	20

sinnvoll gewesen, da der Rahmen der Untersuchung nicht groß genug war, um für jeden der kleinsten Sprachräume eine hinreichende Anzahl von Befragten aufbieten zu können.

⁴⁰⁹ Bei diesen Informanten handelt es sich um eine Irin und eine Australierin.

⁴¹⁰ In die Kategorie ‚Other‘ fallen solche Informanten, die sich keiner bestimmten regionalen Varietät zuordnen ließen.

Von November 1999 bis März 2000 wurden ebenfalls insgesamt 230 Fragebögen an US-amerikanische Kontaktpersonen bzw. Kontaktpersonen in den USA weitergegeben. Die Rücklaufquote war jedoch nicht ganz so hoch wie im Falle der Fragebögen für die Sprecher des BrE. Von den 230 Fragebögen wurden nur 144 an die Verfasserin der vorliegenden Arbeit zurückgesandt oder zurückgegeben. Ferner konnten leider nur 138 Fragebögen gewertet werden. Bei einem Fragebogen war beispielsweise nur das Deckblatt ausgefüllt worden⁴¹¹. Fünf andere Fragebögen konnten nicht in die Auswertung einbezogen werden, weil die Informanten zum Ausdruck gebracht hatten, daß ihr Sprachgebrauch vom BrE beeinflusst sei, denn sie hatten sich entweder eine Weile in Großbritannien aufgehalten oder waren mit Sprechern des BrE in anderen Ländern außer Großbritannien oder den USA in engen Kontakt getreten. Die Fragebögen, die nicht gewertet werden konnten, wurden lediglich bei solchen Antworten berücksichtigt, die Meinungsäußerungen betrafen.

Über die Sozialdaten der 138 US-amerikanischen Informanten, deren Angaben gewertet wurden, können folgende Auskünfte gegeben werden:

US-Amerikaner: 138 Gewährspersonen

Alter:

Statistics	0-19	20-29	30-39	40-49	50-59	60-69	>70
Total	3	21	26	33	34	14	7

Geschlecht:

Statistics	male	female
Total	57	81

Bildungsgrad:

Statistics	school to 16	school to 17/18	college/university	postgraduate
Total	4	20	68	46

Regionale Varietät:

Statistics	Upper North	New England	Lower North	Upper South	Lower South	West	Other ⁴¹²
Total	43	9	31	2	15	29	9

⁴¹¹ Die Angaben zur Person auf diesem Fragebogen gleichen denen auf einem anderen Fragebogen, so daß anzunehmen ist, daß ein Informant versehentlich die erste Seite eines weiteren Fragebogensvordrucks ein zweites Mal ausgefüllt hat.

⁴¹² In die Kategorie ‚Other‘ fallen wiederum solche Informanten, die sich keiner bestimmten regionalen Varietät zuordnen ließen.

6.2.3 Anmerkungen zur Fragebogenerhebung

Leider ergab sich trotz der o.g. Anweisungen der Verfasserin der vorliegenden Arbeit, daß die Fragebögen vorwiegend an Akademiker weitergegeben wurden. Dies gilt für die US-amerikanischen Informanten noch stärker als für die britischen Gewährsleute. Der Grund dafür wird sein, daß die Kontaktpersonen in den USA und Großbritannien selber zum größten Teil dem akademischen Umfeld angehörten und sich die Weitergabe von Fragebögen an Kollegen, Studenten oder Freunde aus dem Universitätsmilieu für sie daher besonders anbot.

Wie die Verfasserin dieser Arbeit zudem jedoch bei der Verteilung von Fragebögen selbst feststellte, schien für Informanten mit geringer formaler Bildung das Ausfüllen des Fragebogens schwierig zu sein, so daß die Bereitschaft, an der Fragebogenuntersuchung teilzunehmen, bei Personen aus unteren Schichten geringer war.⁴¹³ Hierbei mögen verschiedene Faktoren eine Rolle gespielt haben: Zum einen empfanden die weniger gebildeten Informanten vielleicht eine gewisse Scheu vor dem auf den ersten Blick eventuell kompliziert wirkenden Fragebogen. Zum anderen waren sie vermutlich - auch durch ihren Beruf bedingt - nicht daran gewöhnt, längere Texte zu lesen und selber zu bestimmten Fragen schriftlich Stellung zu nehmen. Hiermit steht in Einklang, daß Nicht-Akademiker oder solche Personen, die es nicht gewohnt sind, viel zu schreiben, kaum Kommentare zu den angekreuzten Antworten abgaben. In manchen Fällen hätten zusätzliche Erläuterungen größere Klarheit darüber geschaffen, warum bei einer bestimmten Frage so und nicht anders geantwortet wurde, aber es fiel vor allem den weniger gebildeten Informanten wahrscheinlich schwer, ihre Gedanken zu formulieren. Vielleicht bestand aber auch bei dem einen oder anderen die Angst, daß sich in die Ausführungen Rechtschreibfehler einschleichen würden.

Eine homogene Verteilung der Gewährspersonen über die Parameter Alter, Geschlecht und regionale Varietät konnte ebenfalls weder für die Gruppe der britischen noch für die der US-amerikanischen Informanten erreicht werden; diese war aber auch kaum zu erwarten. Was das Merkmal ‚regionale Varietät‘ in bezug auf die britischen Gewährsleute angeht, wäre allerdings eine größere Zahl von Sprechern des Nordirischen und Walisischen Englisch sowie des *Northern English English* wünschenswert gewesen. Unter den US-amerikanischen Informanten hätte die Zahl der Sprecher aus den Regionen *Upper South* und *New England* größer sein können. Regionale Variationen innerhalb Großbritanniens und den USA konnten im Zusammenhang mit den Fragebögen und ihren Ergebnissen leider nicht so differenziert betrachtet werden, wie es eine Untersuchung größerer Art notwendig machen würde.

⁴¹³ Dies wurde z.B. ersichtlich, als die Verfasserin der vorliegenden Arbeit Angestellte eines Fish & Chips Shops und einer Sandwich Bar in Schottland bat, Fragebögen auszufüllen.

Leider konnte die Verfasserin der vorliegenden Arbeit nicht feststellen, ob einige Informanten nicht doch aufgrund bestimmter Faktoren in ihrem Sprachgebrauch durch die jeweils andere Varietät (AmE bzw. BrE) beeinflusst waren, obwohl sie dazu keine Angaben gemacht hatten. Um zu vermeiden, daß sich die Informanten von Anfang an über den Schwerpunkt der Untersuchung bewußt waren, wurden auf der ersten Seite des Fragebogens nur ganz allgemeine Erkundigungen nach möglichen sprachlichen Einflüssen eingeholt. Auf die diesbezügliche Frage „Are you conscious of any modification of your speech through influences like moving house, marriage or popular culture?“ antworteten jedoch einige Gewährsleute nur mit „ja“, ohne weitere Auskünfte zu geben. Man hätte vielleicht noch expliziter darauf hinweisen sollen, daß die Informanten, die der Ansicht sind, daß sich ihr Sprachgebrauch verändert hat, eine Erklärung dazu abgeben mögen. Allerdings ließ der Platz, der zur Beantwortung der Frage zur Verfügung stand, eigentlich darauf schließen, daß weitere Ausführungen erwünscht waren. Bei den 35 Fragebögen für Sprecher des BrE, die als letzte weggeschickt wurden, sowie bei allen Fragebögen für Sprecher des AmE wurde nach der o.g. Frage noch verdeutlichend hinzugefügt: „If so, please specify.“

Es muß außerdem festgehalten werden, daß unter den britischen Informanten auch eine Reihe von Lexikographen waren, die durch ihre Arbeit bedingt wahrscheinlich über einen größeren Wortschatz verfügen als viele andere Menschen und daher vielleicht eher mit einigen der untersuchten Amerikanismen vertraut waren als andere Personen.

Doch nicht nur über die Auswahl der Gewährsleute, sondern auch über die Art der Fragestellung in den Fragebögen sollte abschließend reflektiert werden. Diese verlangt von den Informanten eine metasprachliche Betrachtung des eigenen Sprachgebrauchs. Dabei stellt sich die Frage, inwieweit sich die Gewährspersonen darüber bewußt sind, welche Wörter, Wendungen und grammatikalische Strukturen sie tatsächlich verwenden. Bei der Auswertung der Befragung mußte daher berücksichtigt werden, daß es sich bei den Antworten der Informanten um Urteile zum Sprachgebrauch und nicht um tatsächliches Sprachverhalten handelt. So gaben die Gewährsleute eventuell in den Fragebögen an, daß sie z.B. gewisse Wörter oder grammatikalische Konstruktionen nicht benutzen, weil sie diese ablehnen, obwohl sie sie im tatsächlichen Sprachgebrauch vielleicht dennoch verwenden. Da viele der britischen Informanten den Einfluß des AmE auf das BrE nicht gutheißen (vgl. die Beantwortung der letzten Frage des Fragebogens⁴¹⁴) besteht die Möglichkeit, daß diese Befragten anführten, die Amerikanismen, die von ihnen deutlich als solche erkannt wurden, nicht oder nur selten zu benutzen, obwohl sie sie in der Sprachwirklichkeit vielleicht doch bzw. häufiger als angegeben

⁴¹⁴ Die Antworten sind in Anhang C abgedruckt.

verwenden. Ähnliches gilt für solche Lexeme, die als Slang angesehen werden: Besonders die Informanten, die für sich beanspruchten, ‚gutes Englisch‘ zu sprechen, brachten zum Ausdruck, diese Wörter und Wendungen nicht zu benutzen. Offen bleibt, ob sie sie nicht tatsächlich doch einmal im entsprechenden Kontext gebrauchen. Insofern ergänzten sich die BNC-Recherchen und die Fragebogenuntersuchung gut: Das BNC spiegelt tatsächliches Sprachverhalten wider, das aber aus bereits veraltenden Quellen stammt, während die Fragebogenuntersuchung jüngere Sprachdaten liefert, die sich jedoch nur auf die Angaben der Informanten zu ihrem Sprachgebrauch beziehen.

Was die Beantwortung der Fragebögen durch manche britische Informanten angeht, schien es der Verfasserin dieser Arbeit manchmal, als ob die Gewährleute die Definitionen einiger analysierter Lexeme nicht genau genug gelesen hatten. Dieser Umstand könnte dazu geführt haben, daß die Befragten bei solchen Wörtern, deren Bedeutung im AmE und BrE unterschiedlich ist, angaben, das Wort in der AmE Bedeutung zu verwenden, da sie nur das Wort, aber nicht seine im AmE divergierende Definition zur Kenntnis genommen hatten.

Abschließend bleibt zu erwähnen, daß die durchgeführte Befragung britischer und US-amerikanischer Sprecher des Englischen keinen Anspruch auf Repräsentativität erheben kann, da nicht vollkommen gewährleistet ist, daß die Stichproben die zugrundeliegenden Grundgesamtheiten widerspiegeln. Dennoch lassen sich anhand der Ergebnisse der Fragebogenerhebung Annahmen formulieren, die im Zusammenspiel mit den Resultaten der BNC-Recherchen dazu beitragen, die Qualität der Wörterbuchinformationen zu den analysierten Amerikanismen beurteilen zu können.